# WORLD HEALTH ORGANIZATION

**Meeting on Health System Performance Measurement**
**New Orleans, USA, 08 January 2001**

**GPE/EQC/HSPM/00.3**

# REPORT ON WHO MEETING OF EXPERTS

# ON THE MEASUREMENT OF EFFICIENCY OF HEALTH SYSTEMS

**GLOBAL PROGRAMME ON EVIDENCE FOR HEALTH POLICY**
**WORLD HEALTH ORGANIZATION**
**GENEVA, SWITZERLAND**

**Introduction:** This report is a summary of major conclusions and recommendations of a meeting of experts on efficiency measurement organized by WHO and held in New Orleans, USA, on Monday, 8 January, 2001. To maximize the potential participation of internationally renowned experts, the meeting was added at the end of the annual meetings of the American Economic Association. In addition to 3 WHO staff, the participants included major world experts on the measurement of economic efficiency. None had been involved in the preparation of the World Health Report 2000 and the associated performance (efficiency) rankings. A list of participants and their affiliations, as well as details of the agenda, can be found in the Annex. **The rapporteur was Dr. K. Kalirajan** from the Australian National University who has published and consulted widely in this area.

**Objectives and Agenda:** There were two objectives of the meeting. The first was to obtain the opinions of a group of recognized international experts on the approach taken by WHO to measure the efficiency of health systems. The second was to obtain their advice and suggestions on ways this work could develop in the future.

The meeting began with an outline of recent developments in the estimation of economic efficiency, particularly those using frontier production technology, by Dr Kalirajan. It was followed by a presentation of the work WHO has undertaken to measure and explain the performance of health systems, both for the World Health Report 2000 and subsequently. Following this, three of the world's top experts (Professors William Greene, Subal Kumbhakar, and C.A. Knox Lovell – alphabetical order) responded formally to WHO's work taking into account the recent developments discussed earlier. Subsequently, Marijn Verhoeven provided an outline of the work the IMF had undertaken in this area and its current plans. The rest of the agenda involved an open discussion by all participants of the WHO approach and ways that it could be developed in the future.

**Main Conclusions and Recommendations:** This section summarizes the conclusions and discussion for which there was general agreement.

1) **On Frontier Production Functions**

a) *Appropriateness*: there was general agreement that the frontier approach was an appropriate technique for measuring the efficiency of health systems. It had not been used widely before in the health sector, apart from assessing the efficiency of hospitals, because economists have been uncertain how to measure outputs of the health sector and what the appropriate inputs to the production process are. Not only was it felt to be technically appropriate, but the participants were excited about the potential of the technique to address important practical policy questions which would make a difference to the lives of many people.

b) *Stochastic versus Deterministic Frontiers*: There was general consensus that, given the nature of the data, stochastic frontier methods like those used by the WHO were preferable to deterministic frontier methods for assessing health system efficiency. This was because deterministic frontiers attribute all of the deviation from the frontier to inefficiency, whereas stochastic frontiers allow for the possibility that deviations from the frontier may also be due to random unobserved factors and measurement problems.

c) *Estimating Efficiency for all Countries Together or for Sub-Groups of Countries.* WHO had estimated efficiency for all countries together. The assumption was that the technologies available to all countries to improve health were the same and that the main limiting factor to their use was the availability of resources. Moreover, the appropriateness of the technologies does not differ by setting – the technology which provides the greatest possible benefit at the individual level in the treatment of cancer would be considered the best in all settings . If these arguments are accepted, the efficiency scores for all countries could be estimated together and there is no need to estimate them for sub-groups of countries separately. This question was discussed extensively. Similar assumptions are unusual in the wider economic literature where the appropriateness of technologies differs across settings. It was agreed that the biological similarities between people mean that the assumption is appropriate to health and that resource availability is the major inhibitor of use. However, it was also suggested that there was value in estimating frontiers separately for selected subgroups of countries to check if the rankings of countries within the subgroups was consistent with the full ranking from the combined analysis.

d) *Fixed-Effects Model*: WHO estimated efficiency using a fixed-effects model, based on panel data from 1993-1997. This was based on the strong recommendation of some of the published literature for estimation of frontiers using panel data. Take the function,

$$Y_{it} = \alpha + X'_{it}\beta + v_{it} - u_i , \qquad \text{(i)}$$

where $X_{it}$ is a vector of inputs and $v_{it}$ is the error term with mean zero. The term $u_i \geq 0$ is a random variable representing country-specific technical inefficiency. For the fixed effects model, this can be rewritten as:

$$Y_{it} = \alpha_i + X'_{it}\beta + v_{it} . \qquad \text{(ii)}$$

where the new intercept $\alpha_i = (\alpha - u_i)$ is country-specific. The frontier intercept is represented by $\alpha$, and the $u_i$'s are the country-specific inefficiencies. In order to ensure that all the estimated $u_i$'s are positive, the country with the maximum $\alpha_i$ is assumed to be the reference and is deemed fully efficient. Mathematically,

$$\hat{\alpha} = \max(\hat{\alpha}_i) \qquad \text{(iii)}$$

and

$$\hat{u}_i = \hat{\alpha} - \hat{\alpha}_i . \qquad \text{(iv)}$$

This normalisation ensures non-negative $u_i$'s. Technical efficiency is defined as:

$$TE_i = \frac{E(Y_{it} \mid u_i, X_{it})}{E(Y_{it} \mid u_i = 0, X_{it})} . \qquad \text{(v)}$$

In order to allow for the fact that health outcomes would not be zero in the absence of any factor of production, unlike in other sectors, WHO modified this equation by subtracting out the predicted minimum level of $Y_{it}$ (denoted by $M_{it}$) from the numerator and denominator. Overall efficiency, or $E_i$, is now:

$$E_i = \frac{E(Y_{it} \mid u_i, X_{it}) - M_{it}}{E(Y_{it} \mid u_i = 0, X_{it}) - M_{it}} . \qquad \text{(vi)}$$

This formulation is easy to estimate using standard statistical packages.

The experts pointed out that one possible problem with the fixed effects approach is that the country-specific fixed effect might also include the influence of unmeasured determinants and not just efficiency. If there were missing explanatory variables, the form could overestimate the inefficiencies. On the other hand, if explanatory variables were included that were highly correlated with those already in the equation, the approach might well underestimate inefficiencies. (The question of which explanatory variables to include is discussed below.)

The meeting also suggested that a number of recent variations could be explored in the future. Professor Greene described how variable coefficient models have been developed which would allow greater flexibility in specifying the efficiency in a production function approach. One application would allow the error component associated with efficiency to be estimated as a random variable. An alternative form of the general approach described by Professor Green is the random coefficients model of the frontier itself which has been developed by Dr Kalirajan.

e) *Functional Form:* It was suggested that there should be a formal test of whether regularity conditions for the translog form used by the WHO are satisfied. Also, a test of the functional form across sub-samples would be useful. This would entail a test to see if the estimated parameters of the translog function were the same for different sub-samples of countries. Another possibility was to use "grade of membership" type models to subdivide the sample into sub-groups for this testing – these models allow for endogenous determination of sample sub-groups.

**2) The Choice of Inputs**

a) *Factors of Production*: The choice of what variables should be included as inputs to the production process was discussed. WHO had made a distinction, based on the literature, between variables that were truly factors of production, and those that might explain observed efficiencies. It had used health expenditure per capita as the summary indicator of health system inputs and the average years of schooling of the adult population as the indicator of non-health system inputs to the production of health system outcomes. The group agreed that this distinction was appropriate. Only variables which are direct factors of production (such as labour and capital inputs in traditional economics) should theoretically go into the estimation of efficiency of the production process. Variables that might explain observed differences in efficiency should not be used as factors of production but efficiency itself should be modelled as a function of those variables and kept separate from the factors of production. (How this should be done is described later).

Given the availability of data across a large number of countries, the experts felt that health expenditure per capita was an appropriate way to summarize health sector inputs, while average years of schooling was also appropriate for non health system inputs. Education could be considered as a direct factor of production in the sense of available knowledge, or as a proxy for other inputs such as housing and nutrition where data are not yet available for all countries and which would be highly correlated with it. Certainly, it would be preferable to measure these other determinants directly but it is not clear if they would be able to be included if they are already very highly correlated with the existing factors of production. As said earlier, this would lead to econometric problems with estimation.

b) *Income per Capita:* WHO explained that there is an extensive literature suggesting that major health improvements this century were highly correlated with improvements in income per capita. This had led some people to the conclusion that increases in income per capita should be included as a factor of production. An extreme form of this argument is that income per capita is the only determinant of health levels and that the health system has no impact on health. This can be shown to be incorrect in many ways, including the fact that the Preston curves – which plot income per capita against health outcomes in a given year - have moved up over the century, implying at least that there must be other factors influencing the health improvements.

These questions were discussed in relation to the need to keep factors of production separate from possible correlates with efficiency. It was agreed that money by itself did not produce health. It is the things that income purchases, such as food and housing, that produces health. So it is not theoretically correct to include income as a factor of production. As said above, if data were available, perhaps food (or nutritional intake) and housing could be used as factors of production, but these data are not available for most countries. The group felt that the appropriate way to introduce income, and a number of other variables such as geography, political institutions etc., was as explanators of efficiency and not as factors of production.

c) *Incorporating Factors Correlated with Efficiency*. There are two ways of doing this. WHO used a two stage process described in the literature, in which efficiency scores were estimated from the production functions first. Subsequently the estimated efficiencies were regressed on possible determinants. The participants recommended that the explanation of efficiency could be done better using the same maximum likelihood estimation process as used for the production function. In it, the component of the error term considered to be inefficiency is made a function of the possible correlates. This would be better econometrically than the method used by WHO and it is in this way that income per capita could possibly be introduced. The new variable coefficients technology could also be used at this stage. None of the experts felt certain whether this would make a difference to the results obtained by WHO.

**3) Uncertainty Analysis**

The approach to uncertainty analysis used in the WHO report was considered novel and did not have theoretical problems. However, it was noted that that information on the higher moments of the estimated outcome data (e.g., not just the mean and standard errors) had not been used and

could probably be incorporated into the analysis. It was not possible to say at the meeting how this could be done as considerable work would be required to sort out the technical details of doing this.

**4) Small Area Analysis**

The experts were asked if they saw any theoretical problems of including sub-national units in the estimation of the frontier. The rationale stated by WHO was that national efficiency is probably an average of different levels of efficiency achieved by sub-national units such as states or provinces. If data were available at this level, it would allow a better estimation of the "true" frontier. The experts did not see econometric or theoretical problems with this, but simply noted that data quality at the sub-national level was often of lower quality than at the national level, and sometimes definitions used for variables differed across these units.

**Other Discussion.** This section summarizes the discussion for which there was no general agreement or conclusion.

1) *MIMIC models:* There was some discussion about whether the estimated efficiency scores could be incorporated into multiple-indicator-multiple-causes (MIMIC) type models along with other possible factors believed to be indicators of efficiency – such as vaccine coverage rates, access, etc.. A MIMIC-type model would treat efficiency to be an unobserved latent variable. One formulation of such a framework would model this latent efficiency variable to be a function of observable exogenous determinants – as well as model the efficiency to be a determinant of several observable effects (or indicators) of efficiency. There was no general agreement. However, it was agreed that it would be useful to test if some of these other variables were correlated with WHO efficiency scores.

2) *Preference Weights*: To construct the overall attainment index, WHO used fixed weights for all countries for the 5 indicators of system attainment. These weights were estimated from the responses given by a sample of people from many different countries, so represent the weights that people felt **should** be guiding policy. There was some discussion about whether these weights would vary across settings and cultures, though it was agreed this was an empirical question. WHO's best evidence at the moment was that they did not, but the large sample surveys now underway would provide updated information.

   The experts pointed out that it was also possible to determine the weights that apparently were being used to guide policy in each country using data envelopment analysis (DEA) on the composite index. It is, obviously, not clear if policy makers really had those weights, or had other preferences but were unable to achieve them. But in any case, it would be possible to see if these implicit weights differed to those used by WHO. If so, it would be possible to use the weights implied for each country to recalculate attainment, and recalculate the efficiency scores. These could be called "benefit of the doubt" efficiency scores – that is, the efficiency given the weights which appeared to be driving the country's policy rather than the weights that people feel should be guiding policy.

   As a technical nicety, constrained DEA would be preferable to unconstrained DEA to do this, partly because it would be much easier for programmes to solve the algorithm. The other reason is that there we have prior knowledge at least of the range into which the weights must fall. For example, it is not possible that the weight for health level is zero. In fact, it is unlikely that the weight for any specific component is zero. A way of setting the constraints would be to use surveys for eliciting preferences (as done by WHO) – but to constrain the weights in the revealed preference analysis to be plus or minus x standard deviations from the mean of the surveys.

   If the two sets of efficiency scores differ, it is not clear which set is the most appropriate. This decision depends on judgement about whether people's expressed preferences should guide policy, or whether the apparent preferences of policy makers should be the yardstick. But it would be technically possible for WHO to provide the two types of efficiency estimates.

3) *Minimum Frontier*: The experts were not very excited about the current method for estimating the minimum but could not suggest an alternative given the data available. There was some discussion about whether the minimum really added anything to the analysis, and a suggestion that the minimum adjustment could be made to the dependent and independent variables before doing the efficiency estimation rather than after as in the WHO approach.

**ANNEX**

AGENDA FOR WHO MEETING OF EXPERTS HELD IN NEW ORLEANS
8 JANUARY, 2001

| | |
|---|---|
| 9:00-9:30 | **K. Kalirajan (rapporteur)**<br>Australian National University<br>Overview of Recent Developments for Measuring Efficiency |
| 9:30-10:30 | **C.J.L. Murray, D.B. Evans, A. Tandon**<br>World Health Organization<br>WHO's Measurement of Efficiency of Health Systems. |
| 10:30-11:00 | Coffee Break |
| 11:00-11.45 | **S. Kumbhaker**, University of Texas<br>**W. Greene**, New York University<br>**C.A. Knox Lovell**, University of Georgia<br>Lead Discussants: Measuring Health System Efficiency |
| 11:45-12:30 | General Discussion and Reactions |
| 12:30-2:00 | Lunch Break |
| 2:00-2:45 | **M. Verhoeven**, International Monetary Fund<br>Other Applications of Frontier Production Functions to Health |
| 2:45-3:30 | General Discussion |
| 3:30-4:00 | Coffee Break |
| 4:00-5:00 | Future Work on Efficiency of Health Systems. |

**LIST OF PARTICIPANTS**

1. Prof. William Greene
   Department of Economics
   Stern School of Business
   New York University
   New York, NY 10012
   USA

2. Prof. Subal Kumbhakar
   Department of Economics
   University of Texas
   Austin, TX 78712
   USA

3. Prof. C.A. Knox Lovell
   Department of Economics
   University of Georgia
   Athens, GA 30602
   USA

4. Prof. Kaliappa Kalirajan
   Division of Economics
   Research School of Pacific and Asian Studies
   Australian National University
   Canberra, ACT 0200

AUSTRALIA

5. Marijn Verhoeven
International Monetary Fund
Washington, DC 20431
USA

6. Prof. Paul Wilson
Department of Economics
University of Texas
Austin, TX 78712
USA

7. Prof. Christopher Tong
Department of Economics
Hong Kong Baptist University
Hong Kong, SAR
CHINA

8. Prof. Philip Grossman
Department of Economics
St Cloud State University
St Cloud, MN 56301
USA

9. Prof. Jaume Puig-Junoy*
Research Centre for Health and Economics
Pompeu Fabra University
Barcelona
SPAIN

10. Dr. Christopher Murray
Evidence and Information for Policy
World Health Organization
Geneva
SWITZERLAND

11. Dr David Evans
Global Programme on Evidence for Health Policy
World Health Organization
Geneva
SWITZERLAND

12. Dr Ajay Tandon
Global Programme on Evidence for Health Policy
World Health Organization
Geneva
SWITZERLAND

* Accepted the invitation, but could not attend at the last minute.