The Evolution of WHO’s Approach to Health System Performance Assessment

David B. Evans
Director, Global Programme on Evidence for Health Policy (GPE)
WHO is Committed to Regular Health System Performance Assessment for 191 Member States

Executive Board Resolution: “Takes note with satisfaction the measures proposed by the Director-General to help Member States contribute to the WHO assessment of their health system performance regularly, namely….

- to compile a report on the performance of Member States’ health systems every 2 years”

- After consultation and peer review
World Health Report 2000

Proposed:

• a framework to use consistently over time and across Member States
• major feature: outcome-focused

This requires:

• defining a parsimonious set of intrinsic goals, indicators developed and tested
• links between performance and system functions
Framework

• Definition of the Health System: Consists of all *health actions*

• A *health action* is any activity whose primary intent is to improve health
FRAMEWORK: Defining the Health System

- Personal medical services
- Non-personal health services
- Intersectoral action
- Other factors

COVERAGE

HEALTH
Goals

• Defined a limited set of *intrinsic goals* to which the system contributes

• *Intrinsic goals*: a. It is possible to raise the level of attainment of the goal while holding all other intrinsic goals constant

b. Raising the level of attainment of an intrinsic goal is always desirable holding attainment on other goals constant

• Desirable objectives not meeting these - *instrumental goals* e.g. coverage
Main Social Goals to Which Health Systems Contribute

<table>
<thead>
<tr>
<th></th>
<th>LEVEL</th>
<th>DISTRIBUTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Financial</td>
<td></td>
<td>✔</td>
</tr>
</tbody>
</table>

Quality          | Equity

Efficiency
4 Key functions influencing performance:

- Financing
- Service provision
- Resource generation
- Stewardship

• **Goal attainment** estimated for the 191 Member States - for the 3 intrinsic goals individually, and for a composite indicator

• **Efficiency** estimated - health system contribution to attainment compared to maximum possible contribution for observed resources

• Reviewed existing evidence on the link between functions and outcomes
Evolution of HSPA - Process

- WHR aroused considerable interest
- Many Member States requested collaboration in development of HSPA or application to national or sub-national levels
- A consultative process established - 6 regional consultation, 8 technical consultations (>300 experts)
- small Advisory Group
- scientific Peer Review Group
Reflections on Evolution of Performance Assessment

• Many components of work in response to debate and the consultations - inputs, outcomes, functions, data collection, translation into policy

• Talk will focus on 2 issues critical to ensuring the policy-relevance of performance assessment

1. Reliability, validity and comparability of measurement - illustrate with health and responsiveness

2. Value of measuring final outcome then disaggregating the components of goal attainment - health inequality
Purposes of Performance Assessment

1. Help Member States monitor and evaluate their own performance

2. Build the evidence base on the relationship between the structure, organisation and content of health systems and performance
Measurement

- Reliability, validity and comparability all important

- **Comparability**
  a. Across time in a given setting;
  b. Across populations in a given setting;
  c. Across settings
WHO Multi-Country Survey Study 2000-2001

Summary:
- HH 15
- BFTF 27
- Postal 28
- CATI 2
- Total 72

in 62 Countries

N = 185,000
# Health as a Multidimensional Construct

## List of ICF Domains used in Surveys

<table>
<thead>
<tr>
<th>Health Domains</th>
<th>Health Related Domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Vision</td>
<td>• <strong>Self-care</strong>: Including eating</td>
</tr>
<tr>
<td>• Hearing</td>
<td>• <strong>Usual activities</strong>: household activities; work or school activities</td>
</tr>
<tr>
<td>• Speech</td>
<td>• <strong>Social functioning</strong>: interpersonal relations</td>
</tr>
<tr>
<td>• Digestion</td>
<td>• <strong>Participation</strong>: societal participation including discrimination/stigma</td>
</tr>
<tr>
<td>• Bodily excretion</td>
<td></td>
</tr>
<tr>
<td>• Fertility</td>
<td></td>
</tr>
<tr>
<td>• Sexual activity</td>
<td></td>
</tr>
<tr>
<td>• Skin &amp; disfigurement</td>
<td></td>
</tr>
<tr>
<td>• Breathing</td>
<td></td>
</tr>
<tr>
<td>• Pain</td>
<td></td>
</tr>
<tr>
<td>• Affect</td>
<td></td>
</tr>
<tr>
<td>• Sleep</td>
<td></td>
</tr>
<tr>
<td>• Energy / vitality</td>
<td></td>
</tr>
<tr>
<td>• Cognition</td>
<td></td>
</tr>
<tr>
<td>• Communication</td>
<td></td>
</tr>
<tr>
<td>• Mobility and Dexterity</td>
<td></td>
</tr>
</tbody>
</table>

"Evidence and Information for Policy"
Cross Population Comparability

• People within a country and across countries have different norms and expectations.

• Point at which groups of people shift their rating of an experience from bad to moderate differ.
Response Category Cut-Point Shift

A

B

C

Latent mobility scale

N = None, Mi = Mild, Mo = Moderate, S = Severe, E = Extreme
Cross-Population Comparability: Health

- Same level of health ➞ Same measurement result
  - irrespective of culture, age, sex, income, expectations ...
- Strategies to obtain cross-population comparability
  - ensure similar content of questions
  - incorporate vignettes that give fixed level of health stimulus
  - measure certain domains that avail clinical tests to fix level of health
    - vision test
    - mobility tests
    - memory and concentration tests
- Novel statistical strategies to adjust for response category cut-point shift
  - Hierarchical Ordered Probit (HOPIT)
  - Compound HOPIT (CHOPIT)
Self-Report: How much difficulty did you have seeing and recognizing a person you know from across the road (i.e., from a distance of about 20 meters)?

Measured Test: Snellen’s eye examination.
HOPIT Model with Vignettes

Mean Categorical Response and Mean Adjusted $Y^*$ vs Age
Results for Colombia - Mobility
Average Level of Health (Mobility vs. Pain)
66 surveys, age-standardized, 2000-2001
Average Level of Health (Affect vs. Cognition)
66 surveys, age standardized, 2000-2001
Average Level of Health (Self-Care vs. Usual Activities)
66 surveys, age standardized, 2000-2001
Health State as a function of Six Core Domains
Population based Valuation Studies conducted in 15 Countries

Domain Levels

H.S. Value

Valuation Function

Multi-Method

Ordinal Domain Descriptions

Valuation Responses: VAS, SG, etc.

- China
- Colombia
- Egypt
- Georgia
- India
- Indonesia
- Iran
- Lebanon
- Mexico
- Myanmar
- Nigeria
- Singapore
- Slovakia
- Syria
- Turkey
Health State as a function of Six Core Domains
Population based Valuation Studies
VAS responses in 9 countries (N=28,000)
Life expectancy - HALE versus LE at birth, 191 countries, 2000
Satisfaction vs. Responsiveness

• Satisfaction focuses on whether people are satisfied with their interaction with the health system

• Responsiveness measures what actually happens when individuals interact with health system in key domains
Responsiveness Instrument Revised for 2000-2001 Household Survey Study

Respect of Persons
• dignity
• autonomy
• confidentiality
• communication

Client Orientation
• prompt attention
• access to social support networks
• quality of basic amenities
• choice
Autonomy Vignettes and mean Taus for 50 Countries

Evidence and Information for Policy
Differences in Domain Ranks
Responsiveness Country Mean Score (outpatient)
Quality of Basic Amenities vs Dignity
Levels of Prompt Attention (Inpatient), WHO 2000-2001 Survey
World Standardised Population Distribution

Denmark
Switzerland
Iceland
Finland (b)
USA
Ireland
Great Britain
Finland (p)
Australia
Canada (t)
Greece
Czech Republic
Hungary
Austria
Mexico
Slovakia
Luxembourg
Netherlands (b)
Turkey
Italy
Germany
Sweden
Spain
Belgium
Portugal
France (b)

* Finland (p) for Postal, (b) for BFTB
Evidence and Information for Policy

Levels of Prompt Attention (Outpatient), WHO 2000-2001 Survey
World Standardised Population Distribution

* Finland (p) for Postal, (b) for BFTB
Performance Assessment

1. Measurement only useful to decision-makers with valid, reliable and comparable indicators - considerable investment required

2. Disaggregating goal attainment by component provides policy relevant information - illustrate with health inequality
Health Inequality - WHR2000

- Measured child survival inequality in all countries with complete birth history data
- Beta-binomial model developed and applied
- 56 countries with detailed micro data
Causes of Child Survival Inequality

- In countries with DHS, analyzed the contribution of income inequality, education inequality and access to the health system as contributors to child survival inequality.

- Results show major factor explaining health inequality is access to the health system.
Decomposition of child survival inequality

Evidence and Information for Policy
## Health state inequality
(WHO Multi-country survey study 2000-2001)

Countries ranked by inequality in health states

<table>
<thead>
<tr>
<th>10 most unequal</th>
<th>10 most equal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Portugal 0.247</td>
<td>Denmark 0.119</td>
</tr>
<tr>
<td>Romania 0.237</td>
<td>United Arab Emirates 0.119</td>
</tr>
<tr>
<td>Russia 0.210</td>
<td>France 0.116</td>
</tr>
<tr>
<td>Netherlands 0.205</td>
<td>Oman 0.116</td>
</tr>
<tr>
<td>Georgia 0.205</td>
<td>Canada 0.115</td>
</tr>
<tr>
<td>Morocco 0.193</td>
<td>Mexico 0.115</td>
</tr>
<tr>
<td>Bulgaria 0.192</td>
<td>China 0.110</td>
</tr>
<tr>
<td>Estonia 0.190</td>
<td>Chile 0.096</td>
</tr>
<tr>
<td>Slovakia 0.188</td>
<td>Indonesia 0.094</td>
</tr>
<tr>
<td>India 0.181</td>
<td>Nigeria 0.056</td>
</tr>
</tbody>
</table>
Summary

• WHO Committed to Measuring & Reporting HSPA for all 191 Member States
• Methods have benefited from consultation and are being peer reviewed
• To be useful to policy-makers:
  1. measurement needs to be reliable, valid and comparable
  2. Have the ability to explore all possible determinants of goal attainment - disaggregate aggregates