Measuring health outcomes: from concepts to practice

Professor Paul Kind
Centre for Health Economics,
University of York, York England
Measuring health outcomes: from concepts to practice

Professor Paul Kind
Centre for Health Economics, University of York, York England
Objectives

• To establish the central concepts that underpin the measurement of health outcomes in clinical AND economic evaluation

• To review the historical development of the “science” of health status measurement

• To describe practical methods for the standardised measurement of health status (health-related quality of life)
Measuring health outcomes

*not specific for economics*

• We are **all** concerned with observing and measuring the effects of health care
  – professional health care workers
  – industry (pharmaceutical and devices)
  – providers (public and private sector)
  – politicians
  – patients / (tax)payers
Measuring health status

(a) $[\text{FEV1}]_{t0} - [\text{FEV1}]_{t1} \quad \Delta \text{FEV1}$

(b) $[\text{health}]_{t0} - [\text{health}]_{t1} \quad \Delta \text{health status}$

there is a calibrated test procedure for (a)

what do we use for (b)?
Outcome measurement addresses fundamental questions

- Does health care have an(y) effect on patient (health status)?
- Can we distinguish the positive and negative effects?
- Can those effects be quantified?
- Can effects be aggregated to yield a net measure of outcome?
Measuring health outcomes

• Health outcomes can be defined as changes in health status that result from the provision (or withholding of health care)
• Involves identifying and quantifying the consequences of health care
• As health care may impact on quantity of life and / or quality of life, health outcome measurement involves quantifying both
Outcome measures

- Biological markers
- Symptom counts
- Employment status
- Days lost through illness
- Length of hospital stay
- Functional status
- Patient-reported outcomes (PRO)

Survival rates
Time to disease progression
Life-years gained/lost
Number of symptom-free patients
Readmission rates
Economic evaluation

Choice: Treatment A or B?

Does the extra benefits (outcomes) justify the extra cost?
Comparing costs and benefits

<table>
<thead>
<tr>
<th>COSTS</th>
<th>BENEFITS</th>
</tr>
</thead>
<tbody>
<tr>
<td>lower</td>
<td>lower</td>
</tr>
<tr>
<td></td>
<td>same</td>
</tr>
<tr>
<td></td>
<td>higher</td>
</tr>
<tr>
<td>same</td>
<td>lower</td>
</tr>
<tr>
<td></td>
<td>same</td>
</tr>
<tr>
<td></td>
<td>higher</td>
</tr>
<tr>
<td>higher</td>
<td>lower</td>
</tr>
<tr>
<td></td>
<td>same</td>
</tr>
<tr>
<td></td>
<td>higher</td>
</tr>
</tbody>
</table>
Combining data on quality and quantity of life

Scenario A
5 years with QoL of 0.8 = 4 QALYs

Scenario B
9 years with QoL of 0.6 = 5.4 QALYs
Economic evaluation

Type of evaluation?

Cost effectiveness

Outcome measure

Cost-utility

- Health status
  - ‘Natural’ units
    - BP
    - Weight
    - etc

- EQ-5D
- HUI
- 15-D
- ......
Traditional indicators

*Employment status*

- influence of ill-health on desire to work
- age of patient / life-cycle
- regional / cyclical changes in economy
Total symptom score (TSS)

*a condition-specific measure*

- 4 symptoms
  - sneezing
  - rhinorrhea
  - itchy nose / throat
  - itchy / watery eyes

- each symptom rated on a 5-point scale where
  0 = absent ➔ 4 = very severe
Hypothetical profiles in SAR

Is A is “better” or “worse” than B, and if so, then by how much?
Hypothetical profiles in SAR

which is “better” B or C, and by how much?

<table>
<thead>
<tr>
<th>Severity Rating</th>
<th>Profile B</th>
<th>Profile C</th>
</tr>
</thead>
<tbody>
<tr>
<td>sneezing</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>itchy nose / throat</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>itchy / watery eye</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>rhinorrhea</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
TSS revisited

• underlying assumptions
  – all symptoms count equally
  – moderate (2) = 2 * mild
  – moderate (2) = 0.5 * very severe

• these ‘weights’ contain no information about the relative importance attached to these symptoms by patients or clinicians
Classifying health outcomes

*The Florence Nightingale method*

- Relieved
- Unrelieved
- Dead
- Alive
- Dead
Hospital mortality rates

• Crude mortality rates average around 3% for all patient admissions
• We can be certain about the outcome for around 3 in every 100 patient admissions
• We do not know whether the remaining 97 patients are ‘relieved’ or ‘unrelieved’
Mortality-based indicators

implications of a 5-year survival rate

• An individual who lives for 5 years and 1 month is a "success"

• An individual who lives for 4 years and 11 months is a "failure"

• 5 years 1 month of poor quality of life is "better" than 4 years 11 month with good quality of life
Quality of life

• Not a new concept
• A broadly recognised construct that is usually not formally defined and that is widely misused
• Not the sole property of any one discipline
• Used in many different ways
Health-related quality of life (HrQoL)

- **Health-related** quality of life
- The methodological foundations are generally agreed, but developers disagree on details
- These differences cannot be resolved on the basis of theory alone
- There are no 'gold' standards in HrQoL
Strategic approaches

• DIRECT
  – measure HrQoL by asking patients to describe and assess their own health status and to (somehow) place a value on it

INDIRECT
  record relevant information from a patient that allows a 3rd party / observer to categorise the health status of that patient in terms of a pre-specified descriptive classification system
  assign a corresponding value to it
Direct HrQoL measurement

Direct valuation of patient’s own health status and/or its change over time
Indirect HrQoL measurement

Indirect measurement via classification of health status
Designing / constructing measures

*Methodological components*

- Two principal elements
  - *description*
    establishing a nominal descriptive system that defines health may be expressed in terms of key domains of interest
  - *valuation*
    determining a set of weights associated with elements of that descriptive system
Description

• Forms the basis of all higher level measurement operations
  – identity (same / different)
  – classification
  – tally / counting
• Exclusions do "count"
  – elements missing from the descriptive system have an arbitrary zero weight
Defining a descriptive system

Source

• Whose descriptions?
  – people in poor health?
  – health professionals?
  – healthy taxpayers?
  – health planners / managers?
  – older / younger people?
  – health economists?
Health outcomes

• How we choose to measure health outcomes expresses a judgement about what WE think matters and what does not
• Hence the developing importance over the past 40 years of standardised measures of health status ("quality of life") measures
A short history of (HrQoL) time

North America

- Sullivan
  - 1966

- QWB
  - Bush et al
  - 1970

- HUI
  - Torrance
  - 1971

- SIP
  - Bergner et al
  - 1972

- SF-**
  - RAND

- Williams et al

- Rosser et al

- Grogono & Woodgate

- NHP
  - Hunt et al
  - 1980

- 15-D
  - Sintonen
  - 1982

- EQ-5D
  - EuroQoL Group
  - 1987

- WHOQOL
  - 1990

UK / Europe

- AQLQ
  - Richardson
  - 1990
Grading of angina

New York Heart Association

• Grade I
  • ordinary physical activity does not cause undue fatigue, palpitation or anginal pain
• Grade II
  • comfortable at rest. Ordinary physical activity results in fatigue, palpitation, dyspnoea or anginal pain
• Grade III
  • comfortable at rest. Less than ordinary physical activity causes fatigue, palpitation, dyspnoea or anginal pain
• Grade IV
  • inability to carry on physical activity without discomfort. Symptoms of cardiovascular insufficiency or the anginal syndrome may be present even at rest
### Karnofsky Performance Scale

<table>
<thead>
<tr>
<th>Description</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>100</td>
</tr>
<tr>
<td>Normal activity ; minor signs / symptoms</td>
<td>90</td>
</tr>
<tr>
<td>Subnormal activity ; some signs / symptoms</td>
<td>80</td>
</tr>
<tr>
<td>Unable to work or to continue normal activities</td>
<td>70</td>
</tr>
<tr>
<td>Requires occasional assistance</td>
<td>60</td>
</tr>
<tr>
<td>Requires considerable assistance and frequent care</td>
<td>50</td>
</tr>
<tr>
<td>Disabled ; requires special care</td>
<td>40</td>
</tr>
<tr>
<td>Severely disabled ; hospitalised</td>
<td>30</td>
</tr>
<tr>
<td>Very sick ; hospitalised with active support treatment</td>
<td>20</td>
</tr>
<tr>
<td>Moribund</td>
<td>10</td>
</tr>
<tr>
<td>Dead</td>
<td>0</td>
</tr>
</tbody>
</table>
### PHYSICAL WELL-BEING

<table>
<thead>
<tr>
<th>GP</th>
<th>Description</th>
<th>Not at all</th>
<th>A little bit</th>
<th>Somewhat</th>
<th>Quite a bit</th>
<th>Very much</th>
</tr>
</thead>
<tbody>
<tr>
<td>GP1</td>
<td>I have a lack of energy</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>GP2</td>
<td>I have nausea</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>GP3</td>
<td>Because of my physical condition, I have trouble meeting the needs of my family</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>GP4</td>
<td>I have pain</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>GP5</td>
<td>I am bothered by side effects of treatment</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>GP6</td>
<td>I feel ill</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>GP7</td>
<td>I am forced to spend time in bed</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>
SF-36 / RAND MOS-36

- 8 domains
  - physical functioning (10)
  - role limitations due to physical problems (4)
  - social functioning (2)
  - bodily pain (2)
  - mental health (5)
  - role limitations due to emotional problems (3)
  - energy / vitality (4)
  - general health perceptions (5)
  - plus 1 question on change in health status 12/12
Original (RAND) SF-36 coding

• General health (item 1 : 5)
  – in general would you say your health is excellent / very good / good / fair / poor
  – original response 1-5 recoded to value of excellent / very good / good / fair / poor
  
  100  75  50  25  0
SF-36 scoring system - Pain

- Pain (item 7)
  - none    6
  - very mild 5
  - mild     4
  - moderate 3
  - severe   2
  - very severe 1

- Pain (item 8)
  - not at all  5
  - a little bit 4
  - moderately 3
  - quite a bit 2
  - extremely 1

Pain score ==> 
\[
\frac{[(\text{item 7} + \text{item 8}) - 2]}{9} \times 100
\]
Medical Outcomes Study

*chronic disease profiles*

![Graph showing chronic disease profiles across different health domains](image-url)
SF-36 health profile

_all dimensions : aged 60 - 64_

Principal generic measures

PROFILE MEASURES

• Sickness Impact Profile (SIP)
• Nottingham Health Profile (NHP)
• SF-community
  • RAND MOS SF-36
  • SF-20 / 12 / 8 / 2
• WHOQOL

INDEX MEASURES

• QWB
• Rosser-Kind Index
• 15-D
• HUI cluster
• HUI II and III
• EQ-5D
• AQLQ
• YHL
• SF-6D
Value and health

- The value of health underpins and guides all aspects of healthcare decision-making
  - Delivery of health care to individual patients
  - Formulating health policies / programmes for Society
  - Developing new health technologies

It is central to the assessment of both effectiveness and cost-effectiveness
It provides a common language that links patients, clinicians, health economists and regulatory decision-makers