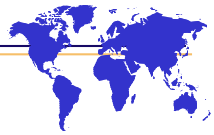
International Health Economics Association (iHEA)
6th World Congress, Copenhagen/Denmark, July 9, 2007

Lost in Translation?



Over-Reliance on QALYs
May Lead to Neglect of Relevant Evidence

Michael Schlander

Institute for Innovation & Valuation in Health Care (INNOVAL[HC])
University of Applied Economic Sciences Ludwigshafen and University of Heidelberg

IV

THE ISSUE

Introductory Remarks

"Economic Evaluation in Health Care:

Is It Really Useful or Are We Just Kidding Ourselves?"[1]

"Let's face it: most health economists have an **interest** in the continued growth of the subdiscipline."

Obstacles may be
"(i) the short-term nature of the decision making process;
(ii) problems in interpreting studies;
(iii) lack of timeliness in study results;
and (iv) importance of other factors in decision making."[1]

[1]**Michael Drummond (2004)**
Australian Economic Review 37 (1): 3-11

IV

THE ISSUE

Introductory Remarks

…'and what about the quality (objectivity, reliability, validity) of health economic evaluations as part of Health Technology Assessments?

"Let's face it: most health economists have an **interest** in the continued growth of the subdiscipline."

Obstacles may be
"(i) the short-term nature of the decision making process;
(ii) problems in interpreting studies;
(iii) lack of timeliness in study results;
and (iv) importance of other factors in decision making."[1]

[1]**Michael Drummond (2004)**
Australian Economic Review 37 (1): 3-11

IV

The Role of Quality-Adjusted Life Years (QALYs)

- **"Technical efficiency"**[1]
- **"Allocative efficiency"**[1]
  - Need some universal and comprehensive measure of benefit (enabling comparisons across individuals / across patient groups)
  - Objectives of health care (effectiveness criterion)**?**
    - **The extrawelfarist proposition**: maximization of (population) health (quantity and [health-related] quality of life)
  - Health-Adjusted Life Years (HALYs)
    - QALYs as the commonly used variant of HALYs
    - Conventional QALY aggregation rules (maximand proposition as basis)
  - **The attractiveness of QALYs** is largely derived from their (conceptualized) characteristics as a cardinal measure of health(-related) outcomes.

[1]While cost-minimization analysis is congruent with the concept of "technical efficiency", the distinction between cost-effectiveness and cost-utility analysis does not simply reflect the difference between technical and allocative efficiency

Some Issues with Quality-Adjusted Life Years (QALYs)

Despite an impressive research agenda
on preference-based measures of health, **there remain**:

- **Methodological Issues**[1]
  - "Cardinal **utilities**" based on Standard Gamble (Neumann-Morgenstern EUT)[2]
    - … consistency with[3] Time Trade-Off, Rating Scales, Person Trade-Off ?
    - … consistency with[3] index instruments: HUI3, EQ-5D, SF-36, AQoL, …?
    - … assumptions (constant proportional trade-off, additive separability[1] …)?
- **Normative Issues**[1]
  - Whose preferences should count from which perspective (*ex ante* / *ex post*)[4]?
  - Aggregation assumptions and derived decision rules[4]?
- **A Common Defense**[1]
  - "high face validity" (intuitively appealing), easy to explain
  - "good enough", "no better alternative", a "pragmatic" workable approach

[1]non-exhaustive lists; [2]cf. G.W. Torrance (1976) [3]and in-between; [4]conflicting empirical data

Some Issues with Quality-Adjusted Life Years (QALYs)

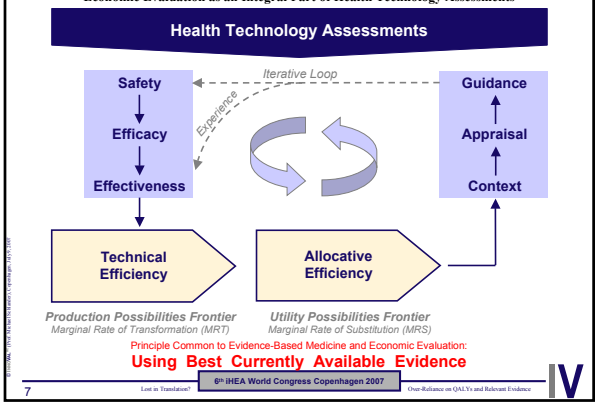
Despite an impressive research agenda
on preference-based measures of health, **there remain**:

- **Methodological Issues**[1]
  - "Cardinal **utilities**" based on Standard Gamble (Neumann-Morgenstern EUT)[2]
    - … consistency with[3] Time Trade-Off, Rating Scales, Person Trade-Off ?
    - … consistency with[3] index instruments: HUI3, EQ-5D, SF-36, AQoL, …?
    - … assumptions (constant proportional trade-off, additive separability[1] …)?
- **Normative Issues**[1]
  - Whose preferences should count from which perspective (*ex ante* / *ex post*)[4]?
  - Aggregation assumptions and derived decision rules[4]?
- **A Common Defense**[1]
  - "high face validity" (intuitively appealing), easy to explain
  - "good enough", "no better alternative", a "pragmatic" **workable** approach

[1]non-exhaustive lists; [2]cf. G.W. Torrance (1976) [3]and in-between; [4]conflicting empirical data

CONTEXT

Economic Evaluation as an Integral Part of Health Technology Assessments

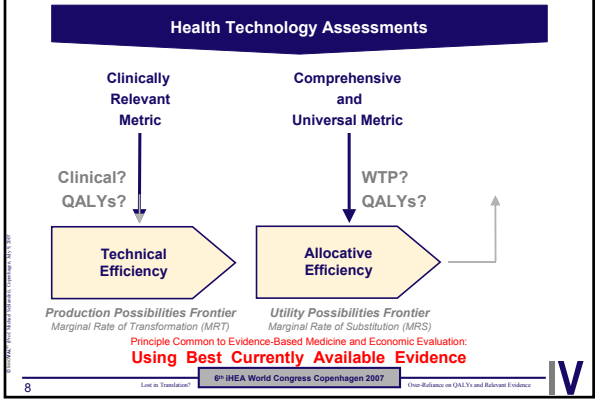


Principle Common to Evidence-Based Medicine and Economic Evaluation:

Using Best Currently Available Evidence

CONTEXT

Economic Evaluation as an Integral Part of Health Technology Assessments

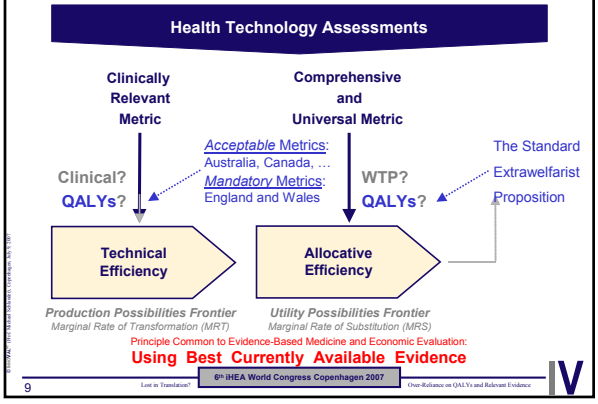


Principle Common to Evidence-Based Medicine and Economic Evaluation:

Using Best Currently Available Evidence

CONTEXT

Economic Evaluation as an Integral Part of Health Technology Assessments



Principle Common to Evidence-Based Medicine and Economic Evaluation:

Using Best Currently Available Evidence

CONTEXT

Using Best Currently Available Evidence

Economic Evaluation Objectives

- **"Technical Efficiency"**
  - Discriminate Between Alternative Interventions
    - with Same Objectives
    - for Same Patient (Group)s
  - Can Be Achieved Using
    - Cost Minimization Analysis (however, rarely applicable)
    - Cost Effectiveness Analysis (usually by way of approximation)
- **"Allocative Efficiency"**
  - Capture (Individual /"Social"?) Preferences
  - Need a Universally Applicable Metric of Benefit
  - Major Current Contenders:
    - Willingness-to-Pay (Cost Benefit Analysis)
    - QALY (Cost-per-QALY Gained; Cost Utility Analysis)
  - **In Order to Meet Empirical** ("Real World") **Stakeholders' Expectations, Both Will Have** to (a) Incorporate or (b) Be Extended to **Reflect Concerns for Fairness**[1]

[1]cf. Dolan et al. (2005); Schlander (2005), Richardson and McKie (2005, 2007)

QALYs

Using Best Currently Available Evidence

Using QALYs as a Universal Measure of Benefit

- **Some Potential Problems**
  - Patients with **behavioral / mental health problems** may not be the best judges of their impairment.
  - (Health-related) quality of life in **children** may be difficult to quantify because of (a) rapid developmental changes, (b) different cognitive abilities of children at various ages, (c) the role of parents as proxy-raters, and (d) its impact on parental utility[1].
- **National Institute for Health and Clinical Excellence (NICE)**
  - **NICE Technology Appraisal No. 98**[2] Treatment Strategies for Attention-Deficit/Hyperactivity Disorder (ADHD) in children and adolescents (England and Wales)
- **Cost-Effectiveness Analysis in Severe Mental Illness**
  - **Hallucination focused Integrative Treatment Program (HIT)**[3] in patients with schizophrenia (The Netherlands)

[1]cf. Griebsch et al. (2005); [2]King et al. (2004, 2006); NICE (2006); Schlander (2007); [3]Staun et al. (2003, 2007)

RELIANCE ON QALYs

NICE Standard: The Reference Case[1]

Problem definition	Scope from NICE
Comparator	Routine therapies in NHS
Evidence on outcomes	Systematic review
Economic evaluation	Cost-effectiveness analysis
Perspective on outcomes	All health effects on individuals
Perspective on costs	National Health Service
Discount rate	3.5% p.a. on costs and health effects
Addressing uncertainty	Probabilistic sensitivity analysis
Measure of health benefits	Quality adjusted life-years
Source of preference data	Representative sample of the public
Health state valuation method	Choice-based method - e.g. SG or TTO
Description of health states for calculating QALYs	Using a standardized and validated generic instrument
Equity position	Each additional QALY has equal value

[1]NICE (2004)

QALYs

Using Best Currently Available Evidence

NICE Technology Appraisal No. 98 (ADHD)

- Findings presented here are part of a more comprehensive qualitative study …
- **Technology Assessment of three molecular entities available as short- and long-acting formulations**
  - **Clinical effectiveness review** based on symptom normalization
  - **Cost-effectiveness analysis** (model) based on response rates, primarily based on **CGI-I sub-scores** (inter-preted as proxies for HRQoL), secondarily including responders based on symptom normalization
  - **Unable to differentiate products …**

QALYs

Over-restrictive use of evidence due to over-reliance on QALYs as a "universal and comprehensive" measure of effectiveness?

NICE Technology Appraisal No. 98 (ADHD)
Shrinkage of Evidence Base[1]

2,908 RCTs — **Literature search**

Filter 1 — **Clinical effectiveness review**
Evidence-based medicine ("EBM")

64 (+1) RCTs — "Efficacy"

Filter 2 — **Calculation of utilities ("QALYs")**
Addition of real-world data?

"Effectiveness" / "Cost-Effectiveness"
Real-world studies (prospective)?
Database analyses (retrospective)?
Economic models (cost-effectiveness analyses)?

05 (+1) RCTs
(3-8w duration)

[1]King et al. (2004, 2006); NICE (2006); Schlander (2007).

NICE TECHNOLOGY ASSESSMENT NO. 98 RELYING ON QALYs

5+1 RCTs
(3-8 weeks treatment duration)

"Very much improved"
"Much Improved"
"Minimally improved"
"No change"
"Minimally worse"
"Much worse"
"Very much worse"

CGI-I subscores
(secondary endpoint; one item only, 7-point scale for improvement "*over baseline*")

"Responder"

X

"Non-Responder"

Psychometric properties of CGI-I scores?

Secondary model extensions pooling heterogeneous response criteria

n=142 ADHD Patients
(of these: male, 87%; combined subtype, 89%; coexisting ODD, 38%)

0.837
standard error 0.039

0.773
standard error 0.039

Utility weights
(derived from parent proxy ratings based on EQ-5D)

"Withdrawal Rates"

Double-counting of nonresponders as a potential source of bias

=>

Mixed treatment comparison model

36+1 Treatment sequences
(12 months)

QALYs
(diff. ext. to the 3rd or 4th decimal place; assumed to capture compliance)

Over-restrictive use of evidence due to over-reliance on QALYs as a "universal and comprehensive" measure of effectiveness?

NICE Technology Assessment No. 98 (ADHD)[1]

- **Unable to differentiate between products on grounds of effectiveness**
  - relying on response rates based on CGI-I sub-score ratings for primary analysis (which were used to compute QALYs); secondary extensions adding heterogeneous outcome measures
- **NICE Assessment in contrast to *consistent* findings from**
  - One RCT using "pragmatic design" suggesting differences
  - Two RCTs reporting relevant head-to-head comparison
  - Two meta-analyses (endpoint: symptom normalization, effect sizes) based on phase III RCTs revealing differences
  - Two cost-effectiveness models indicative of differences (one including a meta-analysis of effectiveness data)
  - Scottish Medicines Consortium (SMC)
  - Australian PBAC

[1]Schlander (2007)

QALYs

Over-restrictive use of evidence due to over-reliance on QALYs as a "universal and comprehensive" measure of effectiveness?

Hallucination focused Integrative Treatment (HIT)[1]

- **Dennis Stant et al. (Groningen, NL):**
- Data of a previously conducted economic evaluation assessing the cost-effectiveness of the HIT intervention in patients with schizophrenia were used to compare
  - analyses based on the primary health outcome (PANSS);
  - results based on various other health outcomes assessed during the study;
  - cost-per-QALY analyses calculated using the EQ-5D.
- **No relevant differences between groups were found on the single primary health outcome initially included.**
- **In contrast, three out of four additional assessed health outcomes revealed significant and relevant differences.**
- **QALY results did not show differences between groups.**

[1]Stant et al. (2007)

QALYs

Over-restrictive use of evidence due to over-reliance on QALYs as a "universal and comprehensive" measure of effectiveness?

Conclusions

- Alan Williams: **"What more could anyone ask for?"**[1]
  - NICE has been acclaimed for representing "the closest anyone has yet come to fulfilling the economist's dream of how priority-setting in health care should be conducted."[1]
  - However; "it is not uncommon for an-economist's-dream-come-true to be seen as a nightmare by everyone else."[1]
- **There is reason for exercising caution concerning the generalizability of the QALY approach.**
  - Standard decision rules (derived on the QALY maximization assumption) have been shown to be "empirically flawed"[2].
  - Standardized (QALY-based) analytic approaches may fail to adequately address specific clinical decision problems.
  - It seems conceivable that the "**feasibility argument**" in favor of cost-per-QALY analyses may be **overstated**.[3]

[1]Williams (2004); [2]Dolan et al. (2005); cf. Schlander (2005) [3]relating to both technical and allocative efficiency