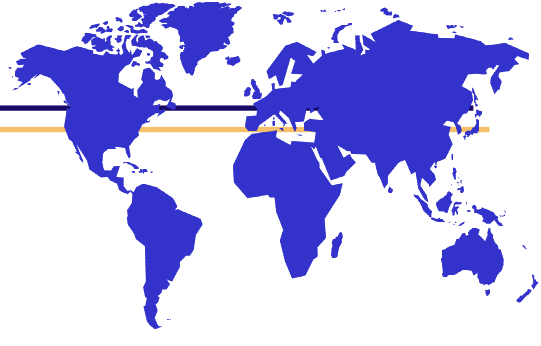
International Society for Pharmacoeconomics & Outcomes Research (ISPOR)
13th Annual International Meeting – Toronto, ON, Canada – May 03-07, 2008

Issues Panel "*QALYs Gone Wild?*"



Over-Reliance on QALYs

May Contribute to the Neglect of Relevant Evidence

Michael Schlander

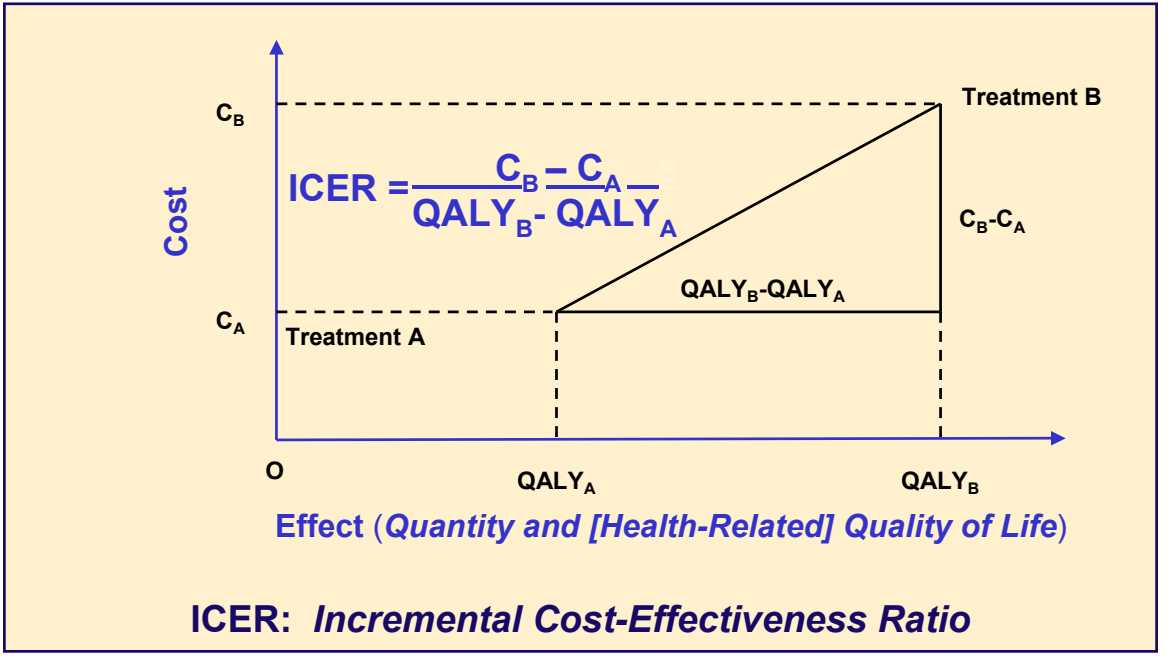
Institute for Innovation & Valuation in Health Care (INNOVALHC)
University of Applied Economic Sciences Ludwigshafen and University of Heidelberg

"How to Interpret Cost per Unit of Outcome?"

Need a "Common Currency"

Capturing Impact of Interventions on Morbidity and Mortality
(viz., on Quality and Quantity of Life)

CUA: Incremental Analysis



ICER: *Incremental Cost-Effectiveness Ratio*

Quality-Adjusted Life Years (QALYs)

Quality and Quantity of Life as Outcomes

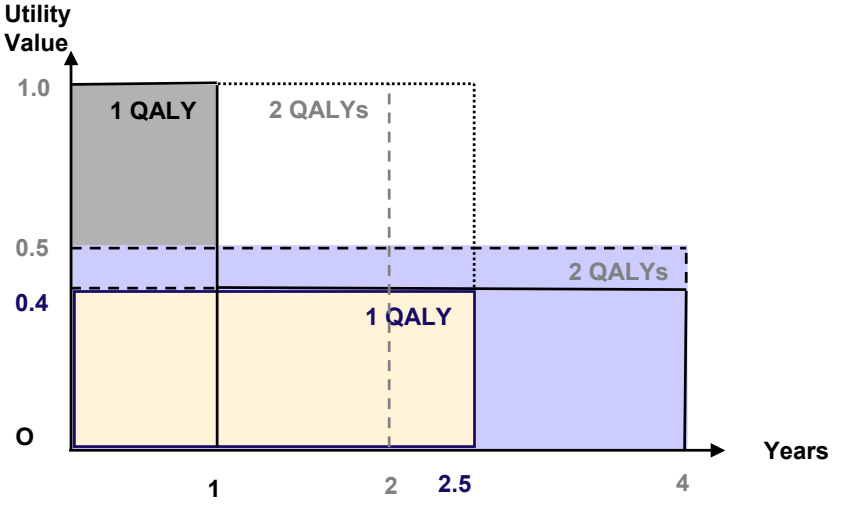
- **Basic idea underlying the QALY**
  - Combination of (health-related) quality of life and length of life into **one comprehensive and universal measure**
  - Intended to facilitate comparisons between different kinds of treatments and diagnoses
  - Should be measured on a cardinal scale to enable computations[1]
- **The concept of the QALY**
  - If the health state "blind" gives a quality weight (utility index) of 0.4, then one year as blind gives 0.4 QALYs ...
  - ... or 1 year in full health gives the same number of QALYs (1) as 2.5 years as blind

[1]According to expected utility theory (EUT), this can be achieved using standard gamble (SG) experiments.

Quality-Adjusted Life Years (QALYs)

Quality and Quantity of Life as Outcome

QALY: Quantity and Quality of Life = AUC



Quality-Adjusted Life Years (QALYs)

Some Utilities for Health States[1]

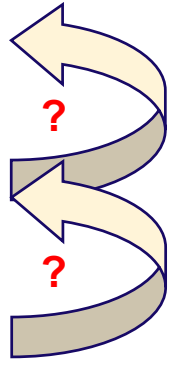
Health State	Utility
Full health (reference state)	1.00
Myocardial infarction, acute (TTO)	0.87
HIV infection, asymptomatic (TTO)	0.87
Hospital dialysis (TTO)	0.56
Liver cirrhosis, decompensated (SG and TTO)	0.54
Being blind or deaf or dumb (TTO)	0.39
Dead (reference state)	0.00
Confined to bed with severe pain	< 0

[1]Data from: G.W. Torrance (1987); T.O. Tengs (2000)

Quality-Adjusted Life Years (QALYs)

Measurement methods to generate quality weights

Decomposed Measurement	Holistic Measurement
Use a MAU[1] Instrument	**1. Health State Vignette**
1. Creation of a Validated Generic Index Instrument	**2. Scaling of Scenario:**
¬ Descriptive system	¬ **Standard Gamble (SG)**
¬ Scaling of instrument: development of a scoring system	**>**
2. Application of Instrument	¬ **Time Trade-Off (TTO)**
¬ Matching health states	**>**
¬ Reading utility scores	¬ **Visual Analogue Scales (VAS) [?]**



[1]MAU, multi-attribute utility theory

Deconstructing Health-Adjusted Life Years (HALYs)

Some Dimensions of Choice[1]

- Scaling instrument (SG, TTO, PTO, …)
- Time horizon (life time, episode, one year, …)
- Personal versus social (community) perspective
- Ex ante versus ex post perspective
- Respondent (patient, public, insured population, expert)
- Social values (age weights, severity, etc.)

Result:

- Large number of potential options
- Justification for selected option(s) ?

[1]From J. Richardson (2002); cf. also J. Richardson (1994)

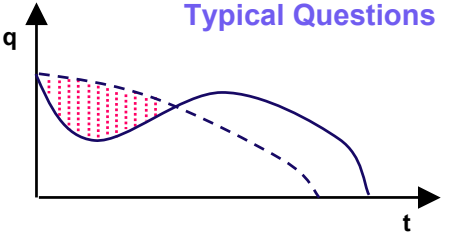
Quality-Adjusted Life Years (QALYs) as a measure of (health-related) outcomes[1]

Three Distinct Ways How to Use QALYs

Same intervention for **Same indication** (same patient group)

"Does the Utility Gain Outweigh the Disutility of Treatment?"
e.g., cancer chemotherapy

Typical Questions



Different interventions for **Same indication** (same patient groups)

"How Can We Integrate a Variety of Clinical Outcomes in one Summary Measure?"
Alternative: disaggregated (cost-consequence) analysis

Different interventions for **Different indications** (different patient groups)

"How Can We Determine the Most Efficient Allocation of Scarce Health Care Resources across a Wide Range of Competing Interventions?"
"Efficiency" usually defined in terms of QALY maximization

QALYs as a utility measure of health-related consequences

[1]This is *not* a comprehensive list. For example, QALYs may also be used in descriptive (non-comparative) economic analyses.

The logic of cost-effectiveness

QALY League Tables[1]

Ranking Interventions by Their Cost-Effectiveness

Example	Cost/QALY
¬ GP advice to stop smoking	220 £
¬ Antihypertensive treatment to prevent stroke (age 45-64 years)	940 £
¬ Hip replacement	1,180 £
¬ Kidney transplant	4,710 £
¬ Hospital hemodialysis	21,970 £
¬ Neurosurgical intervention for malignant intracranial tumors	107,780 £

[1]Data from: A. Maynard (1991); data for United Kingdom (in 1990 £)

HAS NICE GOT IT RIGHT?

"What More Could Anyone Ask For?"

NICE is **"the closest anyone has yet come to fulfilling the economist's dream of how priority-setting in health care should be conducted."**

Alan Williams (1927 – 2005)

… "[NICE] is transparent, evidence-based, seeks to balance efficiency with equity, and uses a **cost-per-QALY benchmark** as the focus for its decision-making. *What more could anyone ask for*?"

HAS NICE GOT IT RIGHT?

"What More Could Anyone Ask For?"

NICE is **"the closest anyone has yet come to fulfilling the economist's dream of how priority-setting in health care should be conducted."**

Alan Williams (1927 – 2005)

However:

"Experience has taught me that it is not uncommon for an-economist's-dream-come-true to be seen as a nightmare by everyone else."

Some Issues with Quality-Adjusted Life Years (QALYs)

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

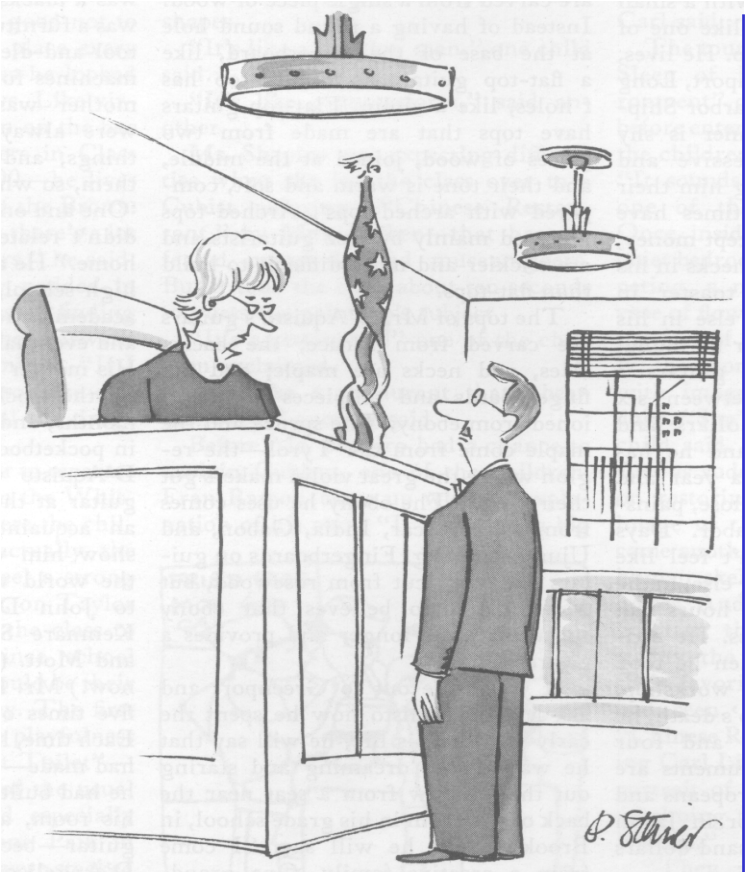
- **Methodological Issues**[1]
  - "Cardinal **utilities**" based on Standard Gamble (Neumann-Morgenstern EUT)[2]
    - … consistency with[3] Time Trade-Off, Rating Scales, Person Trade-Off ?
    - … consistency with[3] index instruments: HUI3, EQ-5D, SF-36, AQoL, …?
    - … assumptions (constant proportional 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 Cost-Effectiveness Benchmarks

- **No scientific basis**
- **Some international "de facto" benchmarks:**
  - **New Zealand** (PHARMAC):
    NZ-$ 20,000 / QALY[1]
  - **Australia** (PBAC):
    AUS-$ 42,000 / LYG to AUS-$ 76,000 / LYG[2]
  - **England and Wales** (NICE):
    £ 20,000 – £ 30,000 / QALY
  - **United States** (MCOs):
    US-$ 50,000 – US-$ 100,000 / QALY[3]
  - **Canada** (proposed "grades of recommendation"):
    CAN-$ 20,000 – CAN-$ 100,000 / QALY[4]

[1]C. Pritchard (2002); QALY: "quality-adjusted life year"; [2]George et al. (2001); LYG: "life year gained"
[3]D.M. Cutler, M. McClellan (2001); [4]A. Laupacis et al. (1992)



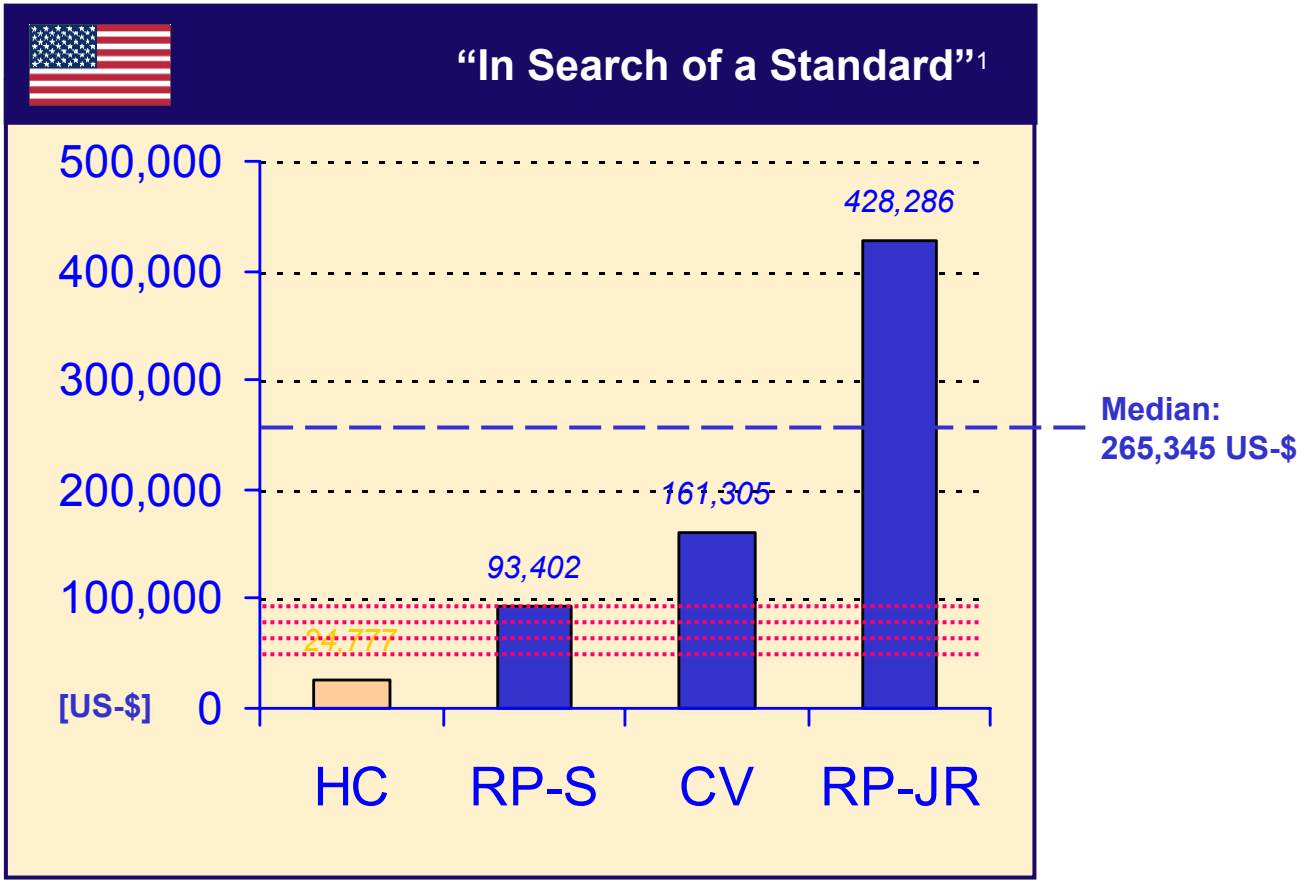
"I sentence you to five years in prison. For your information, that's fifteen thousand billable hours."

© THE NEW YORKER (1990)

Not *so* new:

The evaluation of human life time in economic / monetary terms

"Gaining a QALY may be worth more than analysts generally assume."[1]



[1]R.A. Hirth et al. (2000)

THE LOGIC OF COST-EFFECTIVENESS

A promise and a premise

"A QALY
is a QALY
is a QALY
–
regardless of
who gains and who
loses it."[1]

"The principal **objective** of the *National Health Service* **ought** to be **to maximize the aggregate improvement** in the health status of the whole community."[2]

[2]**Anthony J. Culyer (1997)**

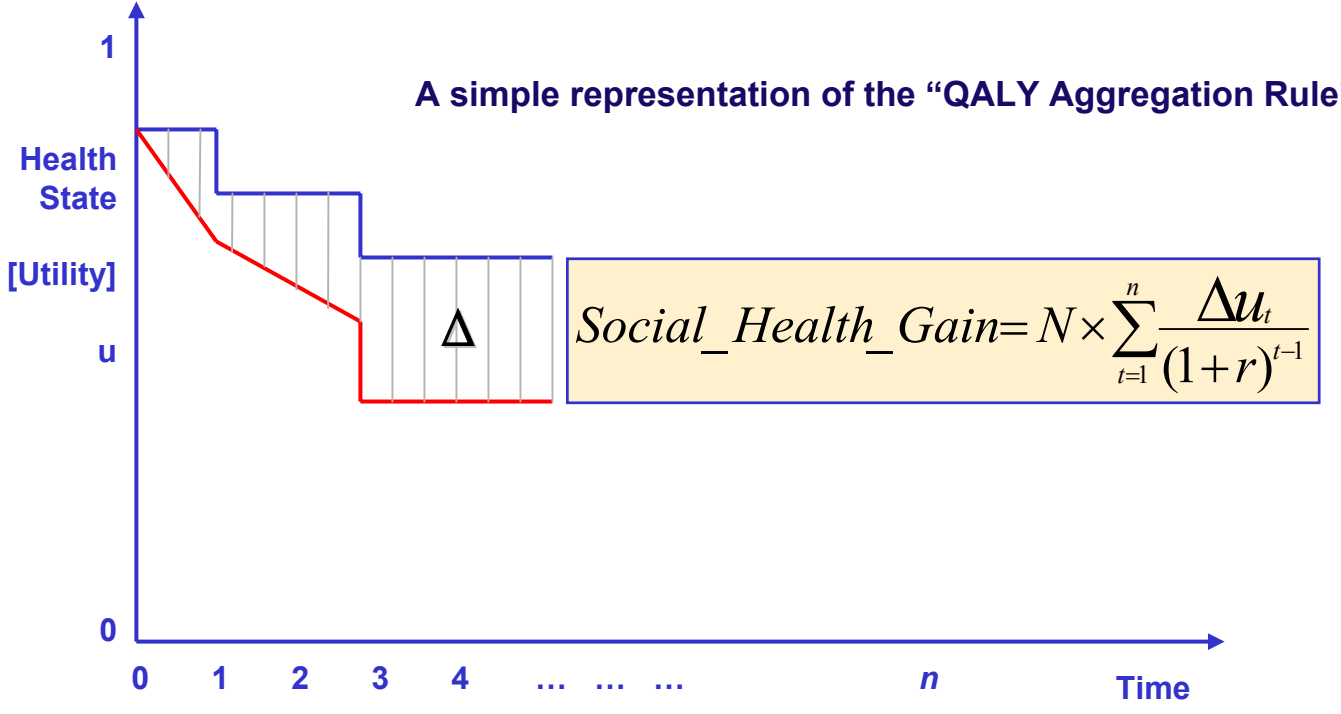
[1]**D. Feeney and G.W. Torrance (1989)**
but there are reasons to suspect that the utility of health states may be influenced by wealth – cf. C. Donaldson et al. (2002)

"The underlying **premise** of CEA in health problems is that for any given level of resources available, **society** (or the decision-making jurisdiction involved) **wishes to maximize the total aggregate health benefit** conferred."[3]

[3]**M.C. Weinstein and W.B. Stason (1977)**

Extrawelfarism

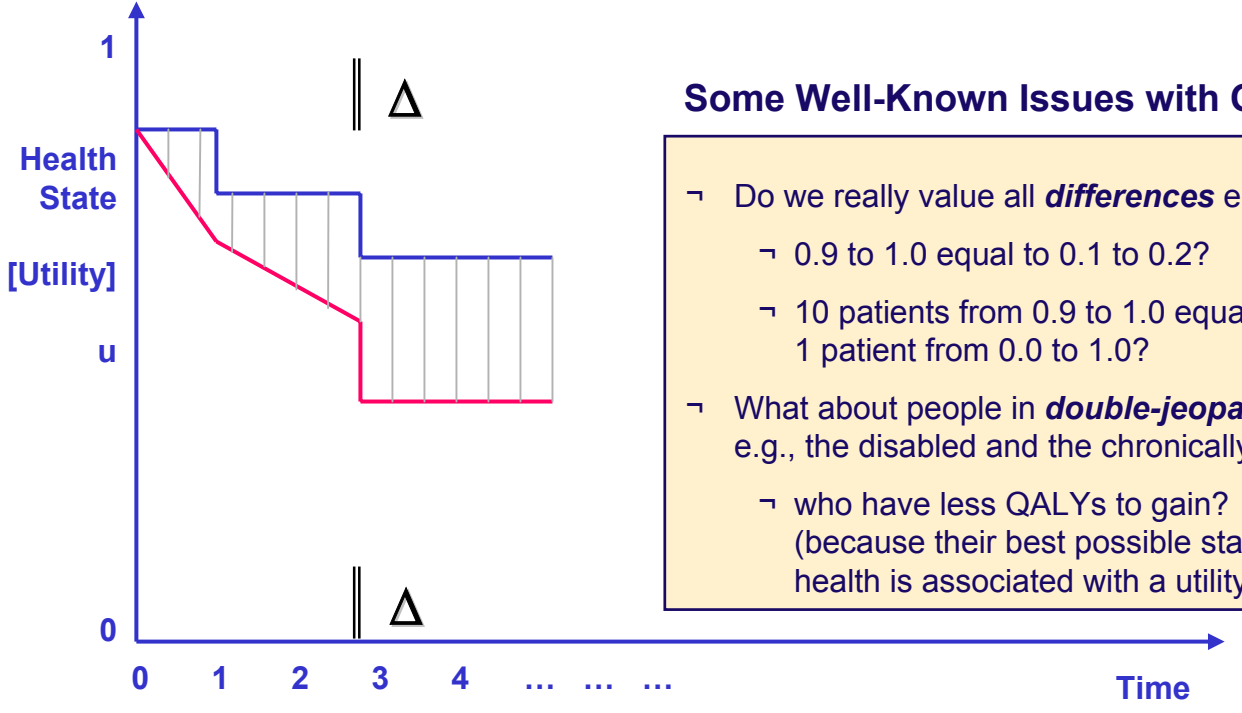
QALY Maximization



A simple representation of the "QALY Aggregation Rule"

$$Social\_Health\_Gain = N \times \sum_{t=1}^{n} \frac{\Delta u_t}{(1+r)^{t-1}}$$

Aggregation of Quality-Adjusted Life Years (QALYs)



Some Well-Known Issues with QALYs

- ¬ Do we really value all ***differences*** equally?
  - ¬ 0.9 to 1.0 equal to 0.1 to 0.2?
  - ¬ 10 patients from 0.9 to 1.0 equal to 1 patient from 0.0 to 1.0?
- ¬ What about people in ***double-jeopardy***, e.g., the disabled and the chronically ill,
  - ¬ who have less QALYs to gain? (because their best possible state of health is associated with a utility u<1)

The QALY aggregation rule is "descriptively flawed".[1]

[1]cf. P. Dolan et al. (2005), M. Schlander (2005)

Social WTP: Valuation of Quality-Adjusted Life Years (QALYs)

Does Context Matter?

- **Empirical evidence** supports a role of the following[1]:
  - **Severity** of initial health state
    - Level of impairment in addition to improvement (difference)?
  - **Rule of rescue**
    - Identifiable individuals (but is being "visible" morally relevant?)
  - **Potential** for health improvement
    - e.g., the permanently disabled and chronically ill? (who have less QALYs to gain)
  - Patients with **high-cost illnesses**

[1]cf. recent reviews by P. Dolan et al. (2005), J. Richardson and J. McKie (2005), M. Schlander (2005); further considerations include (but are not limited to) age, responsibility for dependants, and number of patients or program size.

Extrawelfarism

Guidance based on the EQ-5D

- **Some problems with walking and with usual activities, no other problems** (EQ-5D state 21211)
  - Utility gain from prevention (1 – 0.810 =) 0.190
- **Fatal heart attack**
  - Utility gain from prevention (1 – 0 =) 1.000
- **Issue**

  Is preventing fifty cases of "some problems with walking and with usual activities, no other problems" **as valuable as** preventing ten cases of fatal heart attack?

Extrawelfarism: QALY League Tables

Ranking of Interventions by Cost per QALY…

Intervention	Estimated ICER
¬ **Sildenafil** for erectile dysfunction	¬ < ~ 3,600 £ / QALY[1]
¬ **Methylphenidate** for ADHD in children	¬ < ~ 7,000 £ / QALY[2]
¬ **Riluzole** for motor neuron disease	¬ ~ 38,500 £ / QALY[3] (34,000–43,500 £/QALY[3])
¬ **Beta interferon** for multiple sclerosis	¬ ~ 120,000 £ / QALY[4] (69,000–580,000 £/QALY[4])
¬ **Laronidase** for mucopolysaccharidosis 1	¬ > 330,000 £ / QALY[4]

[1]E.A: Stolk et al. (2000); [2]S. King et al. (2004); [3]G. Ginsberg and S. Lowe (2002), NICE (2001), A. Stewart et al.(2000); [4]NICE (2006)

"QALY League Tables" Revisited

Deconstructing Counterintuitive Cost-per-QALY Rankings

- **(In)Famous example from the Oregon Health Plan (OHP):**
  - Capping a tooth for **150** (not one!) patients was ranked higher than an appendectomy for **one** person.
  - But did this ranking reflect our "powerful proclivity to rescue endangered life"?[1]
- **Some issues not adequately addressed by CEA/CUA:**
  - What priority should be given to the worst off? (those with the most serious and/or immediate conditions)
  - When should small benefits to a large number of persons outweigh large benefits to a small number of persons?
  - How can the conflict between fair individual *chances* and best aggregated outcomes be resolved?[2]

[1]cf. D.M. Eddy (1991) and D.C. Hadorn (1991); [2]For a more complete account of these and related ethical issues, cf. D. Brock (2004, 2006).

A proposed test of "reflective equilibrium"[1]

The Person Trade-Off (PTO) Method

- **Direct assessment of social preferences:**
  - Respondents indicate the number of people in one health state they would need to able to treat (with a specified outcome)

    to make them indifferent to
  - treating a given number of people in another health state (again with a specified outcome)[1]
- **Deconstructing the Person Trade-Off:**
  - Severity of the pre-intervention health state ("level")
  - Severity of the post-intervention health state ("level")
  - Health gain as a result of intervening ("difference")
  - Number of persons treated ("dimension")

[1] cf. E. Nord (1993); E. Nord et al. (1994)

Reliance on QALYs
as a "universal and comprehensive" measure of (health-related) benefits?

Societal WTP as an Alternative Metric?

- **Hypothetical Acute Pain Relief Scenario**[1]
  - Assume a surgical intervention for a small group of patients (say, n=1,000 cases per year) results in postoperative pain associated with a health state "worse than dead" (with a utility of -0.2), lasting for one day.
  - Assume further a new postoperative pain treatment results in pain relief leading to a health state with a utility of 0.8 at a total incremental cost of £ 250.
  - This treatment is associated with an ICER (**cost per QALY gained**) of £ 250 / {[(0.8 – (-0.2)] x (1/365)} = **£ 91,250**.
  - Given the size of the program, the **budgetary impact** (from the perspective of the health care scheme) is **£ 250,000 p.a.**
- **Would we be willing to pay for this intervention?**

[1]Note that this scenario may be less hypothetical than it might seem at first glance!
cf. M. Stadler, M. Schlander, M. Braeckman et al. (2004)

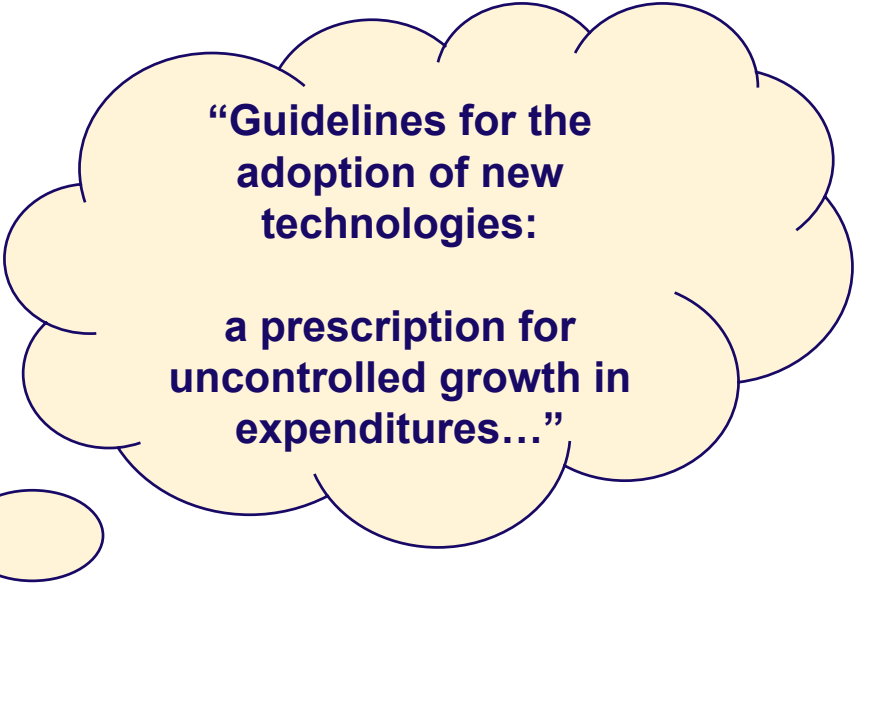
Reliance on QALYs
as a "universal and comprehensive" measure of (health-related) benefits?

Some Concerns concerning QALY Aggregation

- **An Empirically Flawed Decision Rule**
  - The Consistency Argument – A Thinly Disguised Normative Claim
- **Severity of Condition**
  - Capacity to Benefit Empirically of Secondary Importance Only, Compared to Level of Impairment (!?)
  - Priority for Life Saving Interventions and Rule of Rescue
- **The Value of Duration (of Life / of Benefit)**
  - Constant Proportional Trade-Off?
- **Mapping of Individual Utility and Societal Value?**
  - Cost-per-QALY League Tables?
  - From Sildenafil … to Orphan Treatments
  - Small Benefits for Many Outweighing Important Benefits for Few
- **ICER Benchmarks and Opportunity Cost?**

Economic evaluation of new medical technologies

An Early Warning

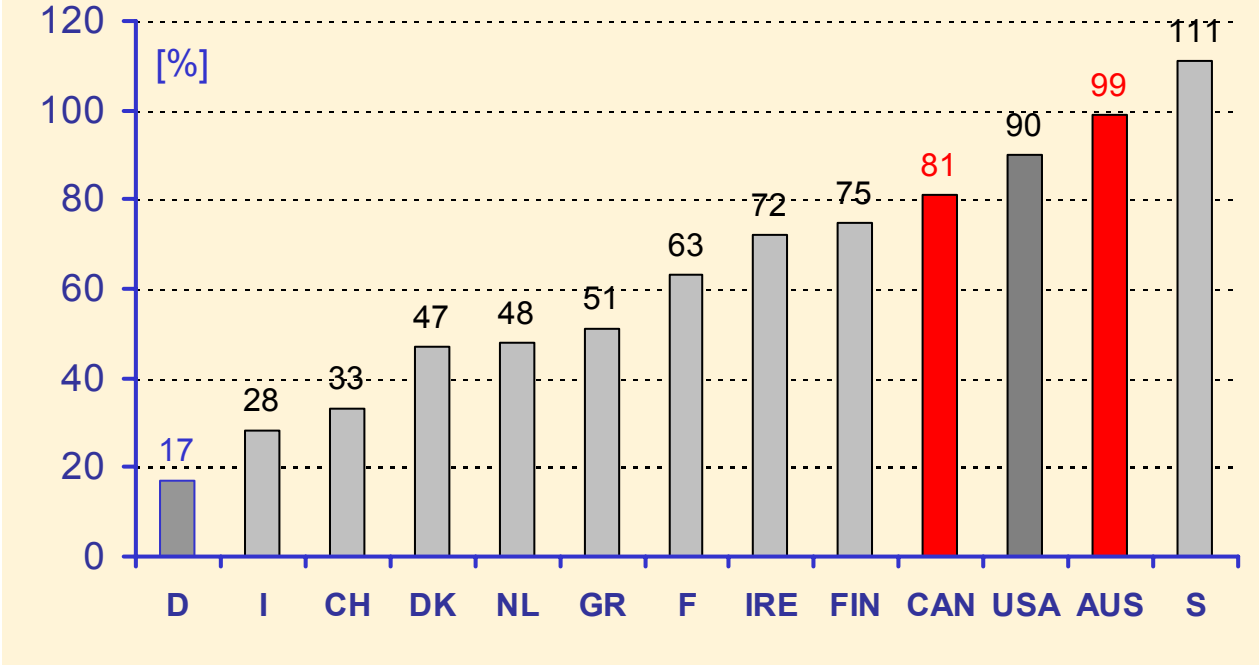


Amiram Gafni and Stephen Birch (1993)

Pharmaceutical price regulation: impact on pharmaceutical spending dynamics

Total Pharmaceutical Spending

(real per-capita growth 1990-2001)[1]



[1]Source: OECD Health Data 2003; Australia and Switzerland: 1990-2000; Germany: 1992-2001; Schlander (2004)

Real-Life Usefulness of Standard Economic Evaluation in Health Care

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

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)

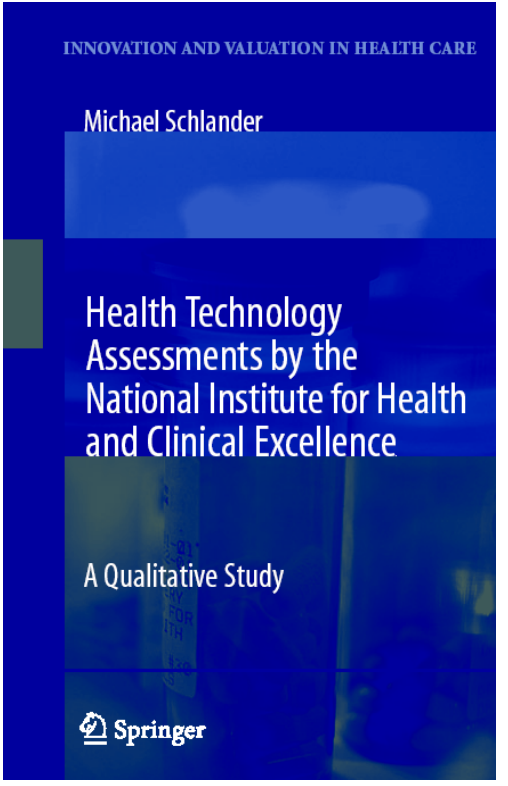
[1]cf. Griebsch et al. (2005);
[2]King et al. (2004, 2006); NICE (2006); Schlander (2007)

[3]Stant et al. (2003, 2007)

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 …**

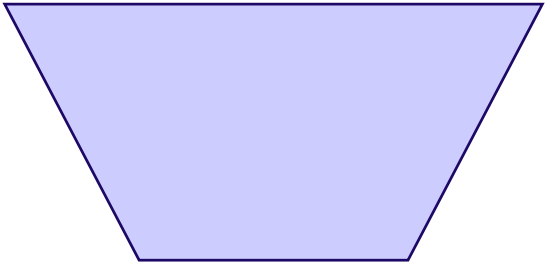


Over-restrictive use of evidence due to over-reliance on QALYs as a "universal and comprehensive" measure of effectiveness?

NICE Technology Appraisal No. 98 (ADHD)

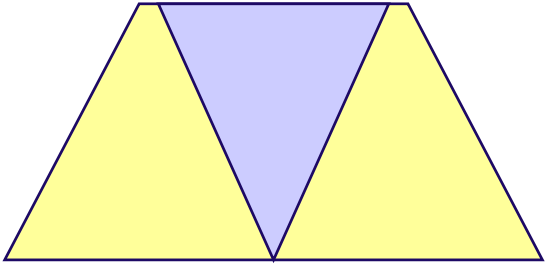
Shrinkage of Evidence Base[1]

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)?

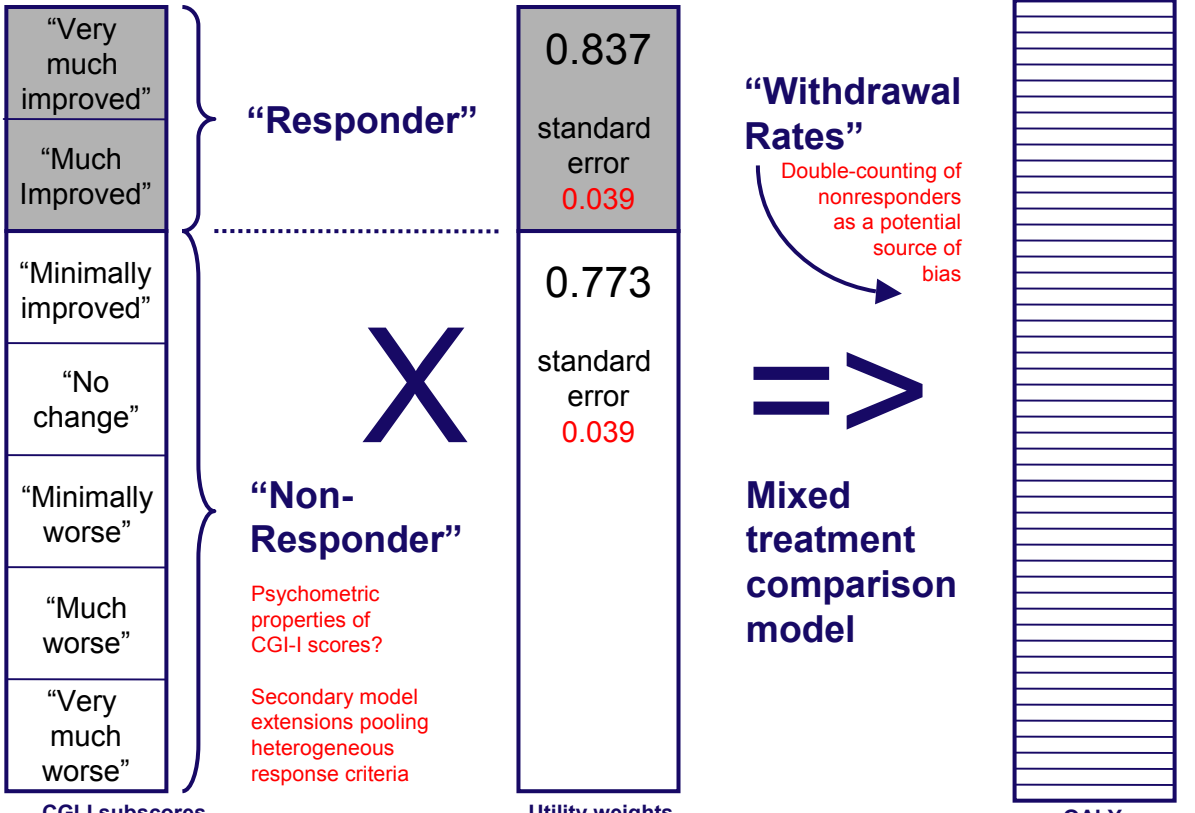
05 (+1) RCTs
(3-8w duration)

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

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

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

36+1 Treatment sequences
(12 months)



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

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

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)

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)

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

THANK YOU FOR YOUR ATTENTION!

Contact

www.innoval-hc.com
www.michaelschlander.com

michael.schlander@innoval-hc.com
ms@michaelschlander.com

INNOVAL HC
Institute for Innovation & Valuation in Health Care