GALEN INSTITUTE
Washington, DC
October 30, 2008

Comparative Effectiveness

A Perspective from Europe

Michael Schlander

www.innoval-hc.com
www.michaelschlander.com
ms@michaelschlander.com

INNOVAL HC
Institute for Innovation & Valuation in Health Care

Quid agis, medice?

"What are you doing, Doctor?"

GALEN INSTITUTE
Washington, DC
October 30, 2008

"Comparative Effectiveness"

or

Looking for "Value for Money"

Health Technology Assessments (HTAs)
including economic evaluation

quidquid agis,
prudenter agas,
et semper recipe finem

"Whatever you do,
do it thoughtfully,
and consider the end (outcome)"

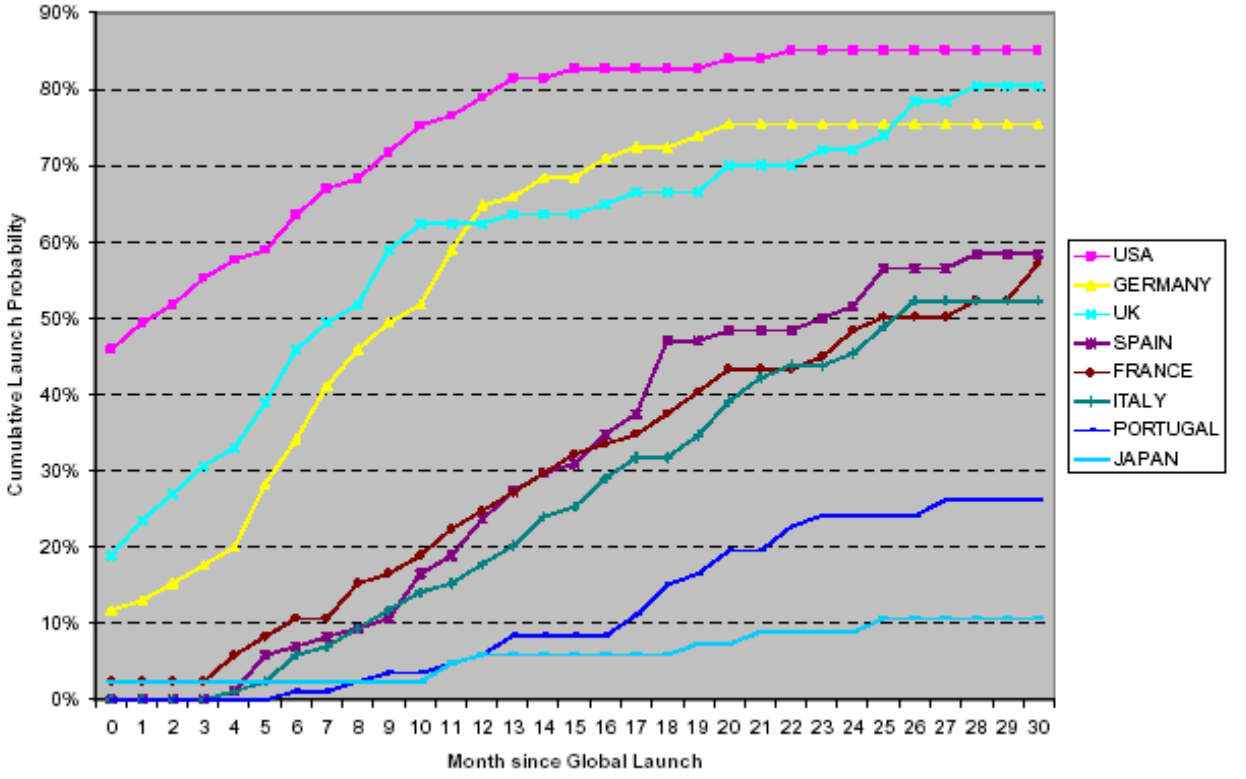
INNOVAL HC
Institute for Innovation & Valuation in Health Care

MARKET REGULATION

- Health Technology Assessments (HTAs)
- Economic Evaluation

Regulating reimbursement and market access[1]

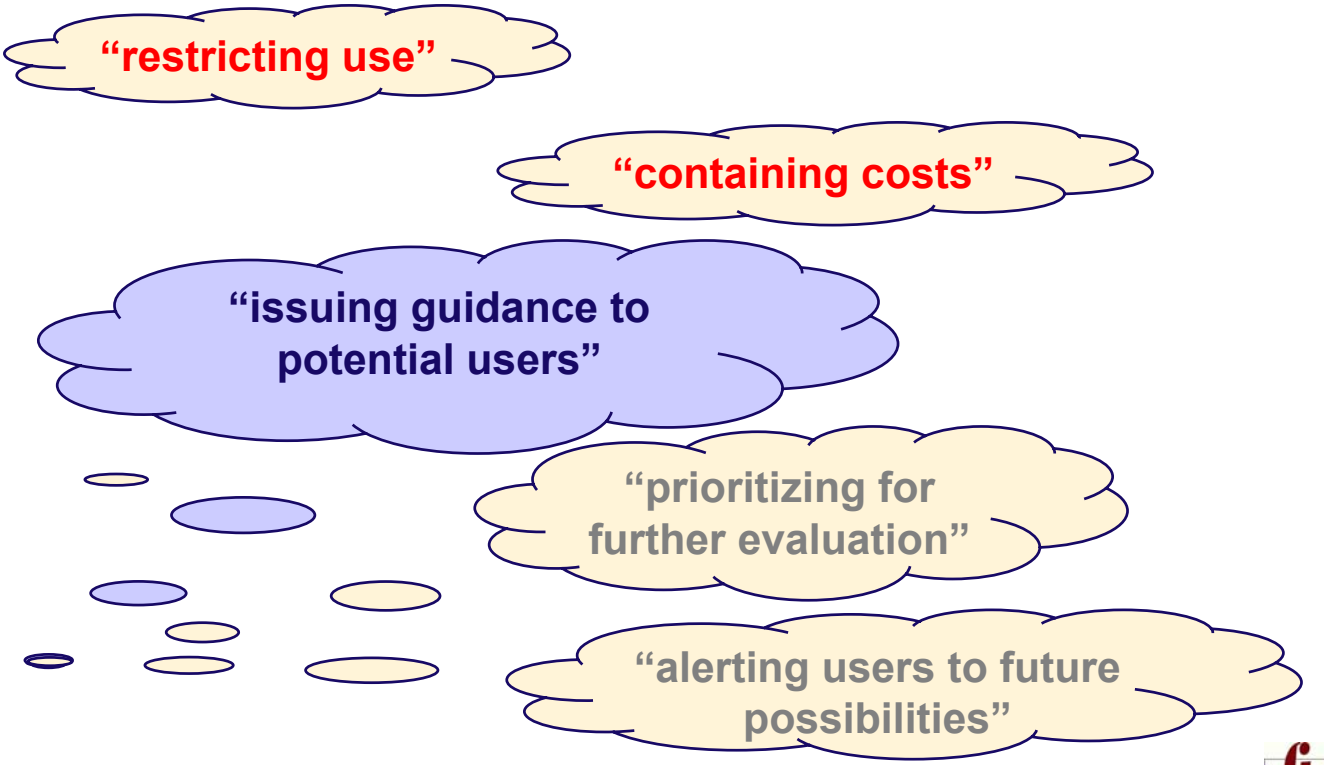
New Product Availability



[1]P. Danzon et al. (2006); OFT (2007)

A broad range of expectations (and fears) …

What are Technology Assessments for?



Economic Evaluation and Health Technology Assessment (HTA)

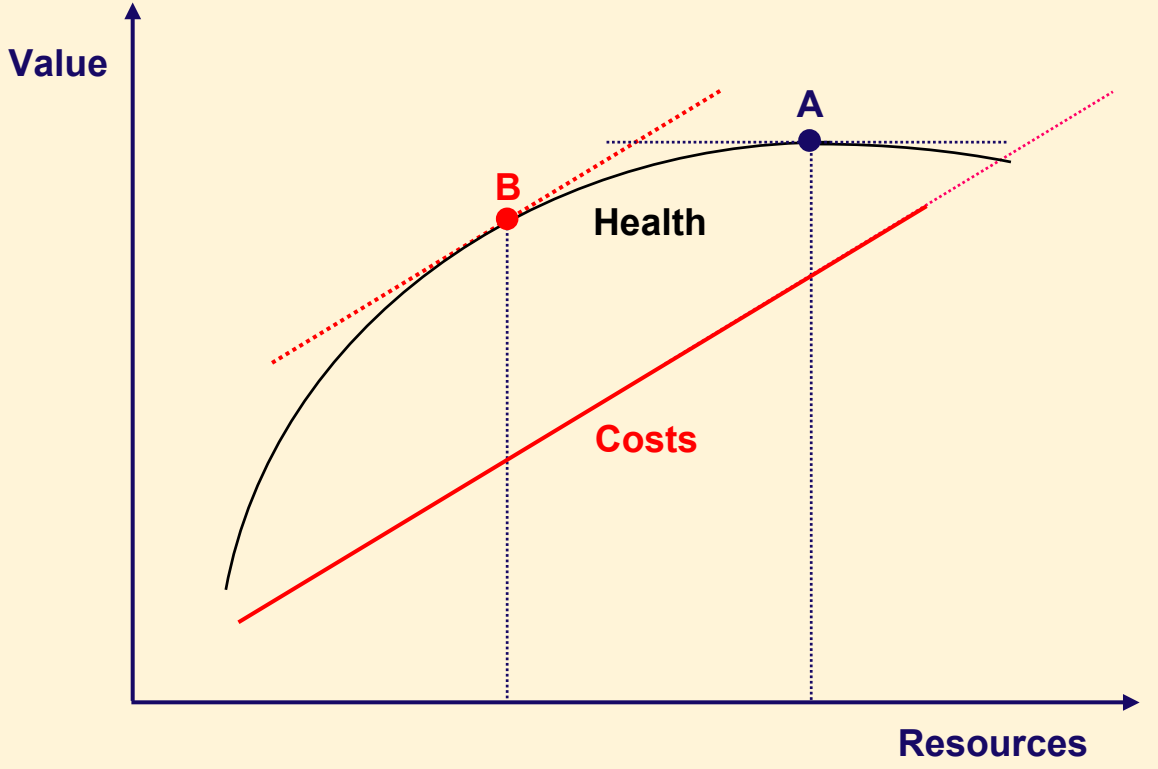
Key Questions Addressed

1. **Safety**
   - ¬ **Does it harm?** (controlled conditions)
2. **Efficacy**
   - ¬ ***Can* it work?**[1] (controlled conditions)
3. **Effectiveness**
   - ¬ ***Does* it work and is it safe?**[1] (normal practice)
4. **Efficiency**
   - ¬ **Do its benefits outweigh its costs?** (*often*: "Is it cost-effective"?)

[1]cf. D. Schwartz and J. Lellouch (1967); [2]EBM: "evidence-based medicine"

Background

Evidence Based Medicine (A) & *Economic* Evaluation[1] (B)



[1]cf. Victor R. Fuchs: "Health Care and the United States Economic System", *The Milbank Memorial Fund Quarterly*, April 1972: 211-237.

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**

Foundations:
Economic efficiency

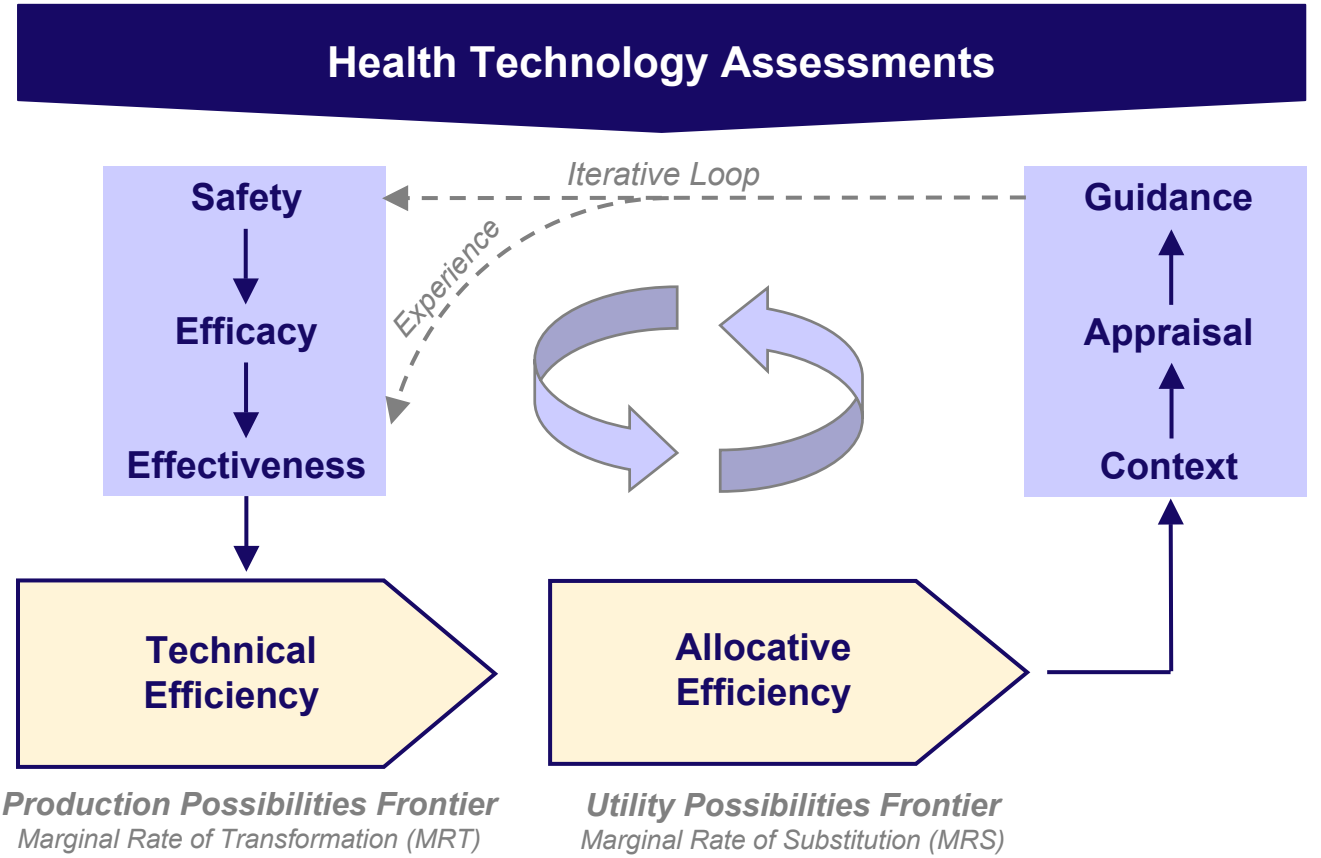
Technical Efficiency

- **Ability to produce the maximum possible output from a given set of inputs**
- **Does not routinely imply choosing between different patient (group)s**
  – *hence individual persons*

Allocative Efficiency

- **Choosing the most cost-effective set of programs for the given level of expenditure**
  (i.e.,optimal choice of input proportions, given their respective prices)
- **Does imply allocating resources across different patient (group)s**
  – *hence individual persons*

Economic Evaluation as an Integral Part of Health Technology Assessments



Principle Common to Evidence-Based Medicine and Economic Evaluation:

Using Best Currently Available Evidence

Economic Evaluation as an Integral Part of Health Technology Assessments

Health Technology Assessments

Clinically Relevant Metric

Clinical? QALYs?

Technical Efficiency

Production Possibilities Frontier

Marginal Rate of Transformation (MRT)

Comprehensive and Universal Metric

WTP? QALYs?

Allocative Efficiency

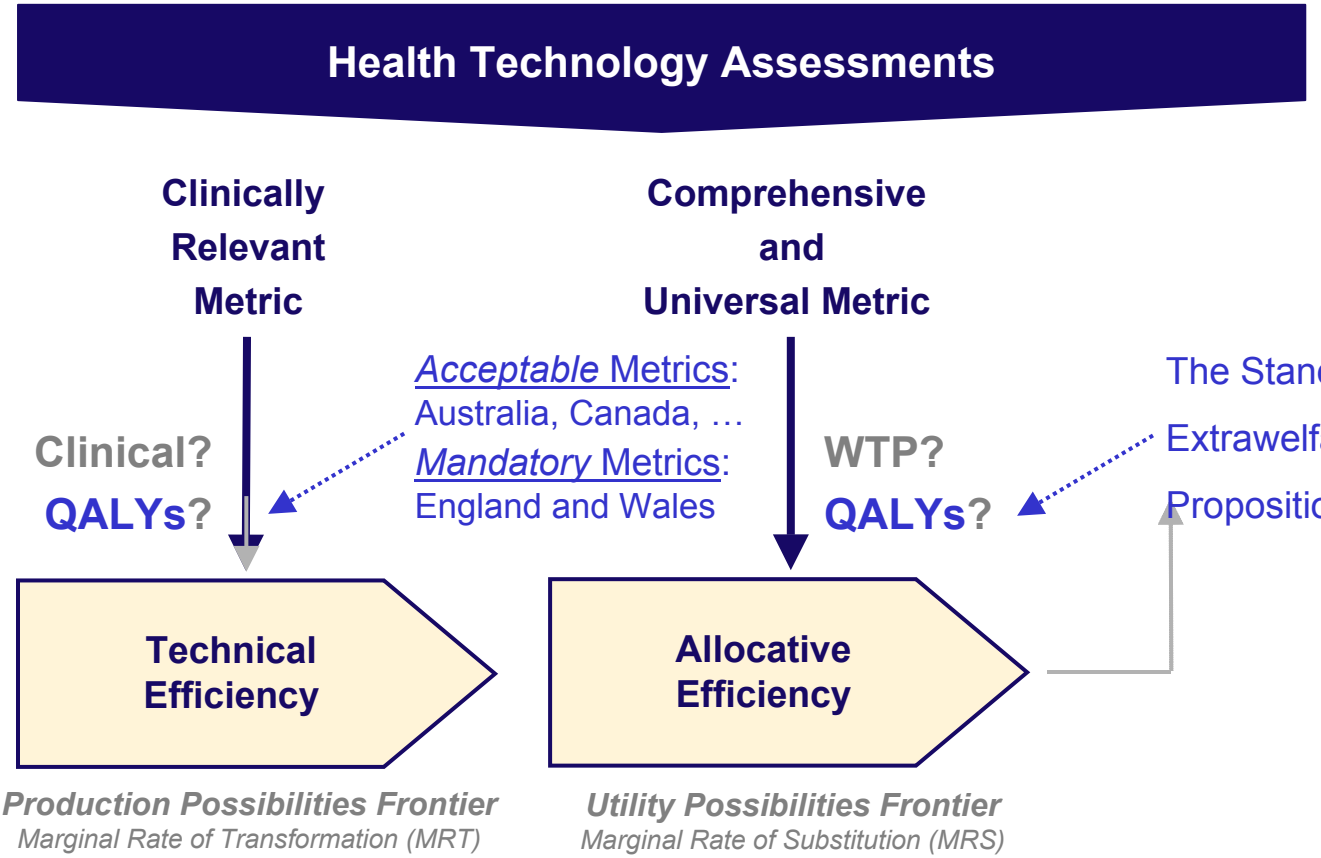
Utility Possibilities Frontier

Marginal Rate of Substitution (MRS)

Principle Common to Evidence-Based Medicine and Economic Evaluation:

Using Best Currently Available Evidence

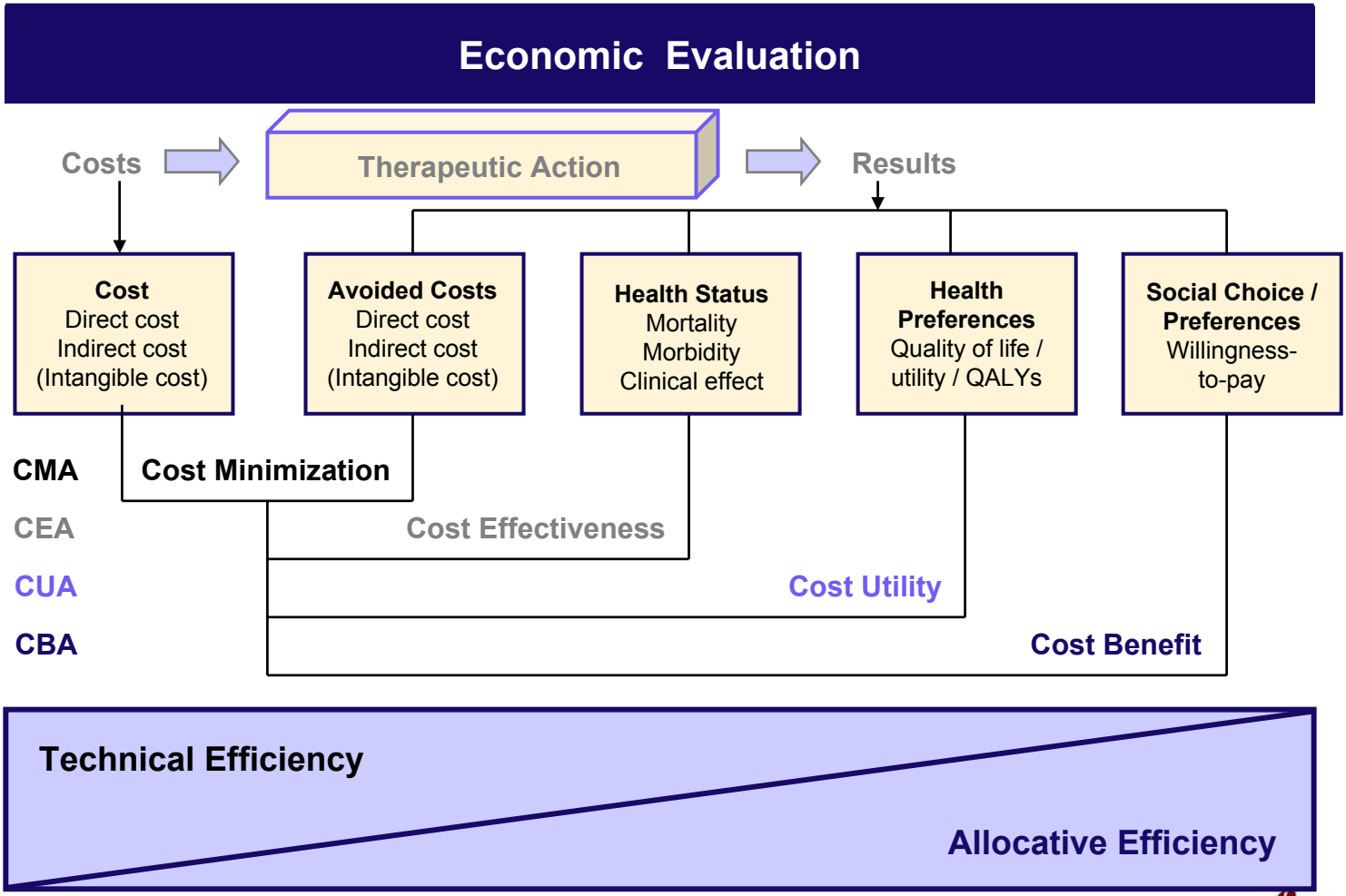
Economic Evaluation as an Integral Part of Health Technology Assessments



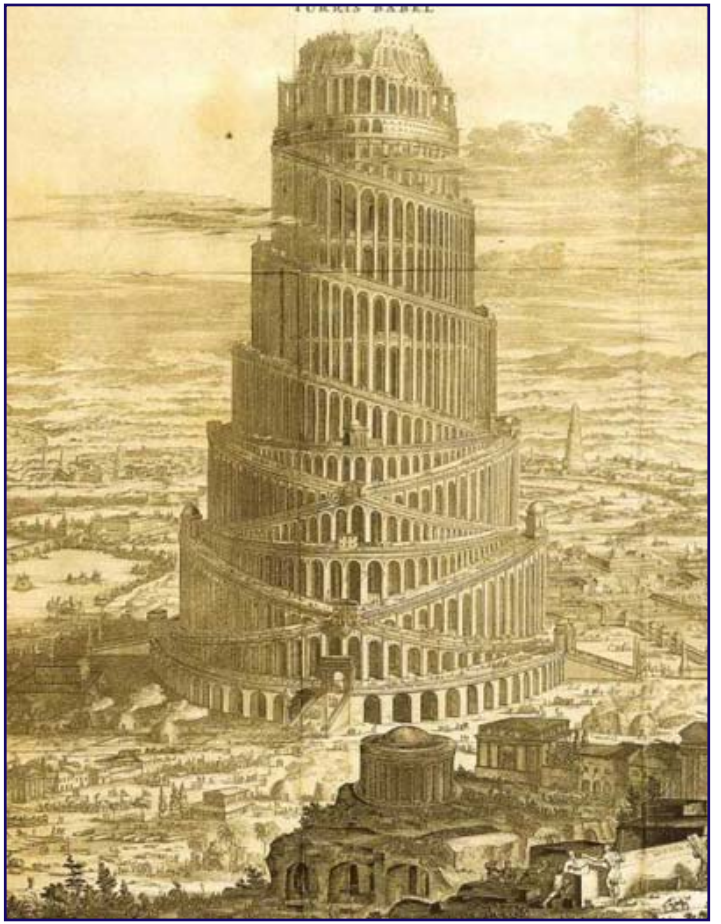
Principle Common to Evidence-Based Medicine and Economic Evaluation:

Using Best Currently Available Evidence

COMPARATIVE ECONOMIC EVALUATION



A Canadian Policy Analysis[1]



A Tower of Babel …

- Referral to many different and often incommensurate things…
- **A key paradox**:

  The discourse about values is both very important and very ambiguous…
- Stakeholders may be tempted to react to this problem with either

  reductionism
  (focusing on one particular definition of values to the neglect of other relevant types)

  or

  nihilism…
  (either rejecting all values analyses as equally unreliable, or accepting all as equally credible)

Illustration by Athanasius Kircher

[1]M. Giacomini et al. (2004)

Economic Welfare Theory and "Utility"

Valuation

"Political economy has to take as the ***measure of utility*** of an object the maximum sacrifice which each consumer would be willing to make in order to acquire the object

…

the only real utility is that which people are *willing to pay* for."[1]

- **Contemporary Textbooks of Microeconomics:**
  - "The **value** [of a product] to a given consumer is defined as the maximum amount that the consumer would be **willing to pay** for that [product]."[2]

[1] **Jules Dupuit (1844)**

[2] Steven E. Landsburg: *Price Theory and Applications*, 5th ed., Mason, OH: South-Western 2002, p. 238.

In particular, two assumptions of economic welfare theory have attracted criticism from a group of health economists (often referred to as "extrawelfarists")

An Extra-Welfarist Critique[5]

1. **"The monetary measurement [of benefits in cost-benefit analysis] inherently favors the wealthy over the poor."**[1]
   - "Extra-welfarists *and many decision-makers in the real world of health care* are willing to accept an approach that considers outcomes equitably (as CEA using QALYs does), rather than accept an approach in which choices are heavily influenced by ability to pay."[2]
2. **"Extra-welfarists identify 'health' as the principle output of health services."**[3]
   - Then, in effect (*at least in theory*[4]), health is treated as an independent argument in the welfare function. Now, health can no more be substituted by income or consumption.

[1]M.R. Gold et al. (1996), p.26; [2]M.C. Weinstein and W. Manning (1997), p. 127; [3]A.J. Culyer (1989), p. 51; [4]C. Donaldson et al. (2002);
[5]Thomas Rice (1998, 2002) has provided a systematic critique of welfare theory as a foundation of health economics.

A formal treatment

Evaluation Types (1)[1]

CBA

$$B_1 > C_1$$

B, benefit
C, (opportunity) cost

$$P_1 \bullet E_1 > C_1$$

P, price (valuation) of effect
E, effect

$$\frac{P_1 \bullet E_1}{C_1} > 1$$

$$\frac{P_1 \bullet E_1}{C_1} > \frac{P_2 \bullet E_2}{C_2}$$

Alternative formulation,
introducing a **budget constraint**
which limits how much costs can be expended.

CEA

$$\frac{E_1}{C_1} > \frac{E_2}{C_2}$$

Eliminating the pricing of effects, thus **introducing the requirement of $P_1 = P_2$** (which is considered valid in a CEA since one is comparing a common effect E with the two interventions[1]).

Thus, *formally* CEA can be regarded as a special type of CBA under restrictive assumptions: 1. a single effect must be the outcome of interest, and 2. this effect must be exactly the same for both interventions.

[1]From R.J. Brent (2003); note that this formal treatment is greatly simplifying the differences between CBA, CEA, and CUA.

A formal treatment

Evaluation Types (2)[1]

CEA

$$\frac{E_1}{C_1} > \frac{E_2}{C_2}$$

$$\frac{C_1}{E_1} < \frac{C_2}{E_2}$$

Formally CEA can be regarded as a special type of CBA under restrictive assumptions: 1. a single effect must be the outcome of interest, and 2. this effect must be exactly the same for both interventions.

CUA

$$\frac{C_1}{QALY_1} < \frac{C_2}{QALY_2}$$

If we want to compare entirely different effects (as with headache pain relief and the precision of a diagnostic test), and if we do not want to use prices explicitly, then all effects need to be converted into a common unit. This is usually the QALY.

This (CUA) is a restricted version of CEA (and thus of CBA), adding **E = QALY** for each intervention, in addition to $\mathbf{P_1 = P_2 = P}$, with P now relating to the **price of a QALY**.

CMA

$$C_1 < C_2$$

In cost-minimization analysis (CMA), consequences play no part in the evaluation as they are assumed to be identical: $E_1 = E_2$.

Note: Unless consequences are identical across interventions, a CMA would not constitute a valid evaluation of these interventions.

[1]From R.J. Brent (2003); note that this formal treatment is greatly simplifying the differences between CBA, CEA, and CUA.

Foundations:
Two prevailing philosophies[1]

Welfare Economics

- **Seeking (potential) Pareto improvements**
- **Focused on efficient allocation of scarce resources**[2]
  - Cost-benefit analysis incorporating the efficiency rationale behind markets
  - Social objective assumed to be to maximize (aggregate) consumer satisfaction ("utility")
  - Grounded in economic welfare theory
  - Strength of preferences expressed by Willingness to Pay (WTP)[2]

Decision Support

- **Decision analysis as a tool to support social objectives**
- **In practice,** [usually] **focused on** [aggregated] **health maximization**
  - Can, in principle, accommodate a variety of objectives and perspectives
  - Background in operations research
  - Striving to adopt the perspective of a 'rational' decision-maker
  - Distributive concerns representing a research frontier, not actual practice

[1]cf. R.F. Sugden, A. Williams: *The Principles of Practical Cost-Benefit Analysis*. Oxford University Press (1978); cf. also G. Torrance (2006)
[2]Note that, at least in principle, CBA can accommodate the impact of prior distribution (wealth, income; "ability to pay")

ENGLAND AND WALES: NICE

- Reliance on QALYs
- Robustness

An old German saying ...

"Wer am Wege baut, hat viele Meister"[1]

"A house built by the wayside is either too high or too low."

[1] **Martin Luther (1530)**

"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*?"

"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."

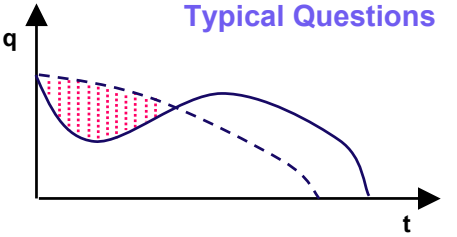
Quality-Adjusted Life Years (QALYs) as a measure of (health-related) outcomes[1]

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

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

QALYs as a utility measure of health-related consequences

Need for a cost-effectiveness benchmark

The Cost-Effectiveness Decision Rule

$$ICER = \frac{\Delta C}{\Delta E} \stackrel{!}{=} \frac{\Delta C}{\Delta QALY} < \lambda$$

Economic evaluation of new medical technologies:

"The Silence of the Lambda"

An Early Warning

"Guidelines for the adoption of new technologies:

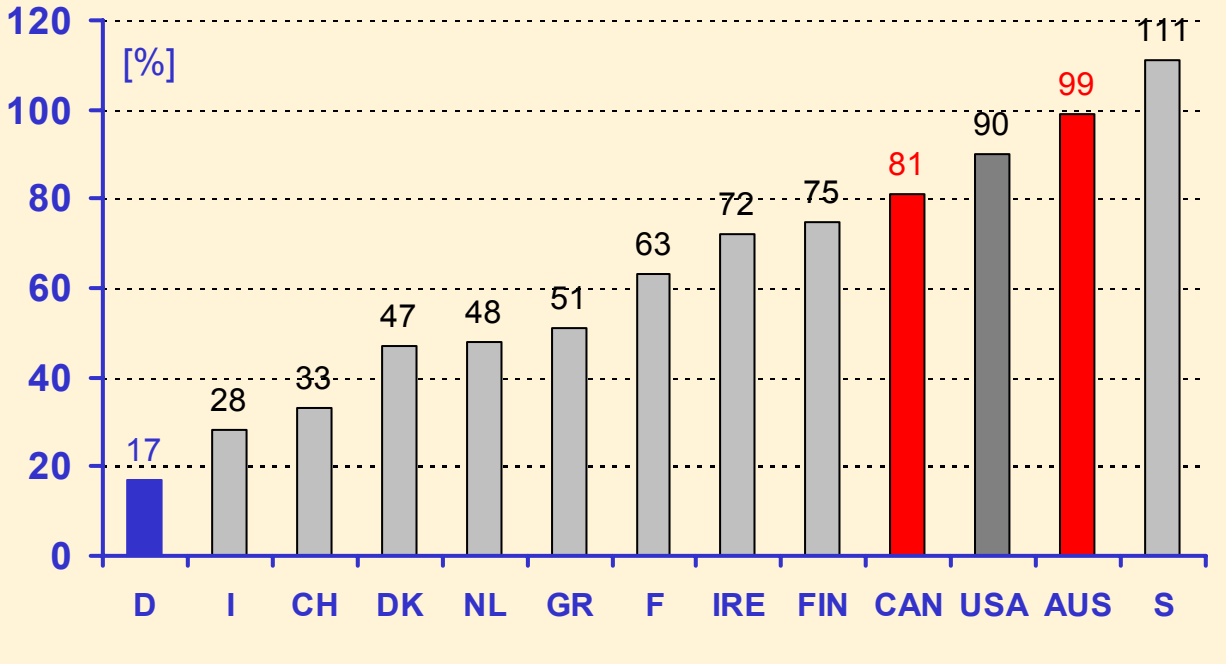
a prescription for uncontrolled growth in expenditures…"

Amiram Gafni and Stephen Birch (1993)

Australia and Canada introduced cost-effectiveness analysis in 1992 and 1994

Total Pharmaceutical Spending

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



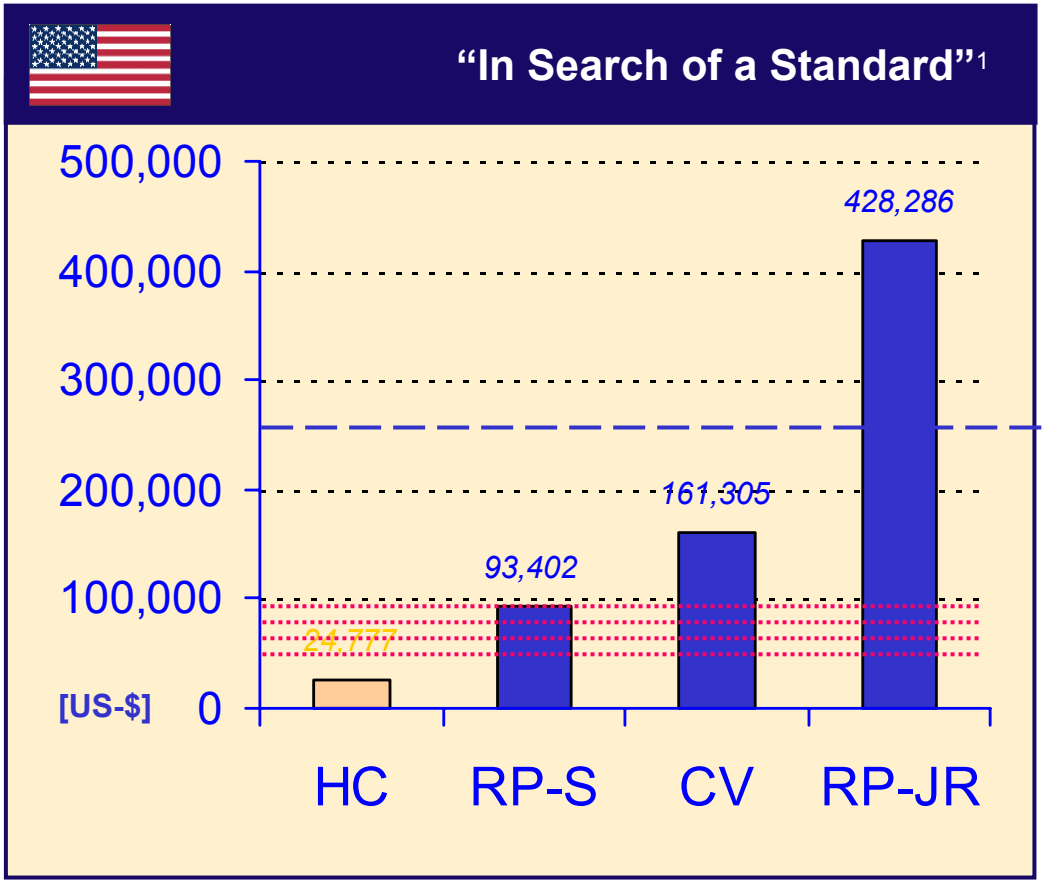
[1]Source: OECD Health Data 2003; Australia and Switzerland: 1990-2000; Germany: 1992-2001; from Schlander (2004)

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)

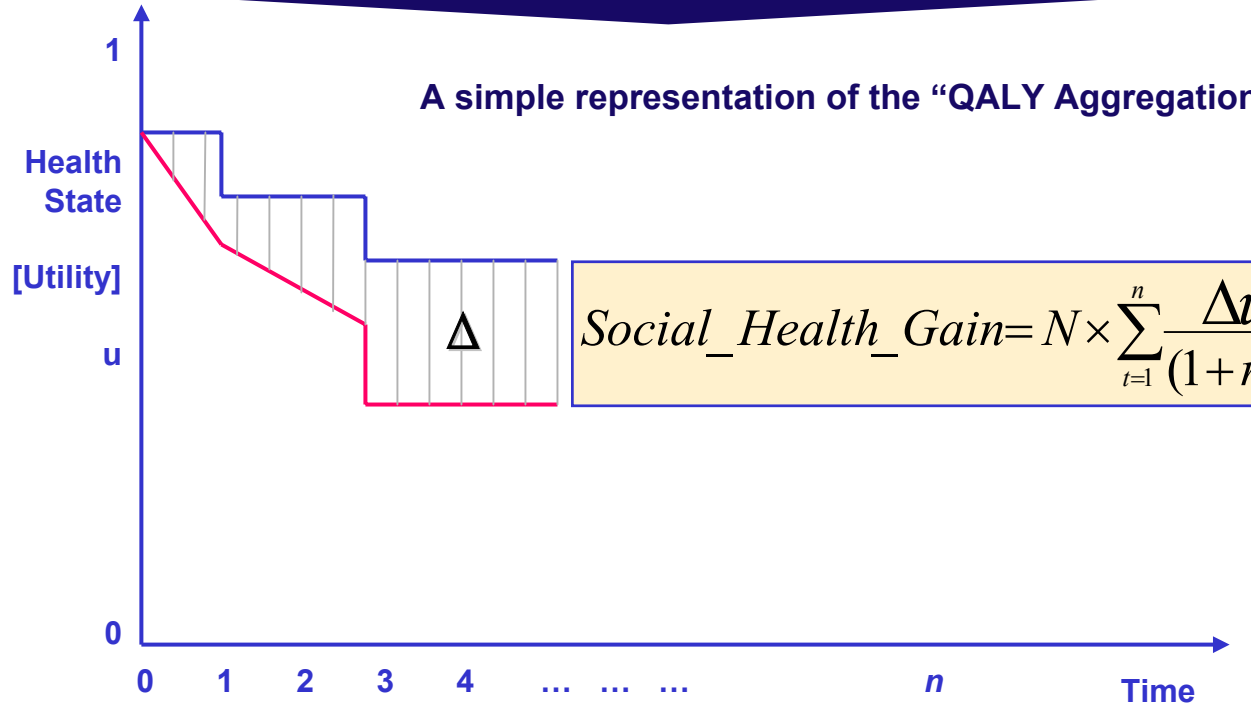
"Gaining a QALY may be worth more than analysts generally assume."[1] (?)



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

The concept of a cost per QALY "threshold" rests on the linear QALY aggregation assumption

The Conventional Unit of Health Outcomes: QALYs



A simple representation of the "QALY Aggregation Rule"

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

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]

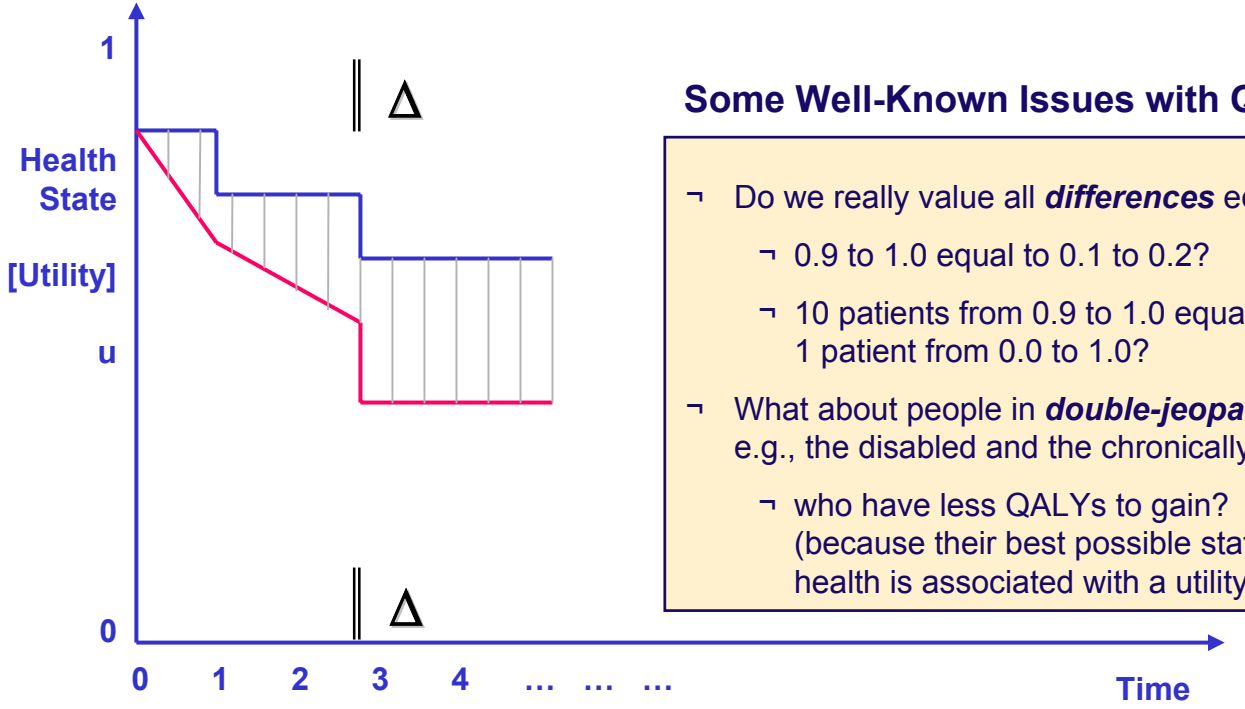
[2]**Anthony J. Culyer (1997)**

[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)**

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 five cases of "some problems with walking and with usual activities, no other problems" **as valuable as** preventing one case of fatal heart attack?

"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).

"QALY League Tables" Revisited

Ranking of Interventions by Cost per QALY ICERs

Interventions:	ICERs:
¬ **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	¬ **> 330,000 £ / QALY**[5]

[1]E.A: Stolk et al. (2000); [2]S. King et al. (2004); [3]G. Ginsberg and S. Lowe (2002), NICE (2001), [4]A. Stewart et al.(2000); [5]NICE (2006)

"QALY League Tables" Revisited

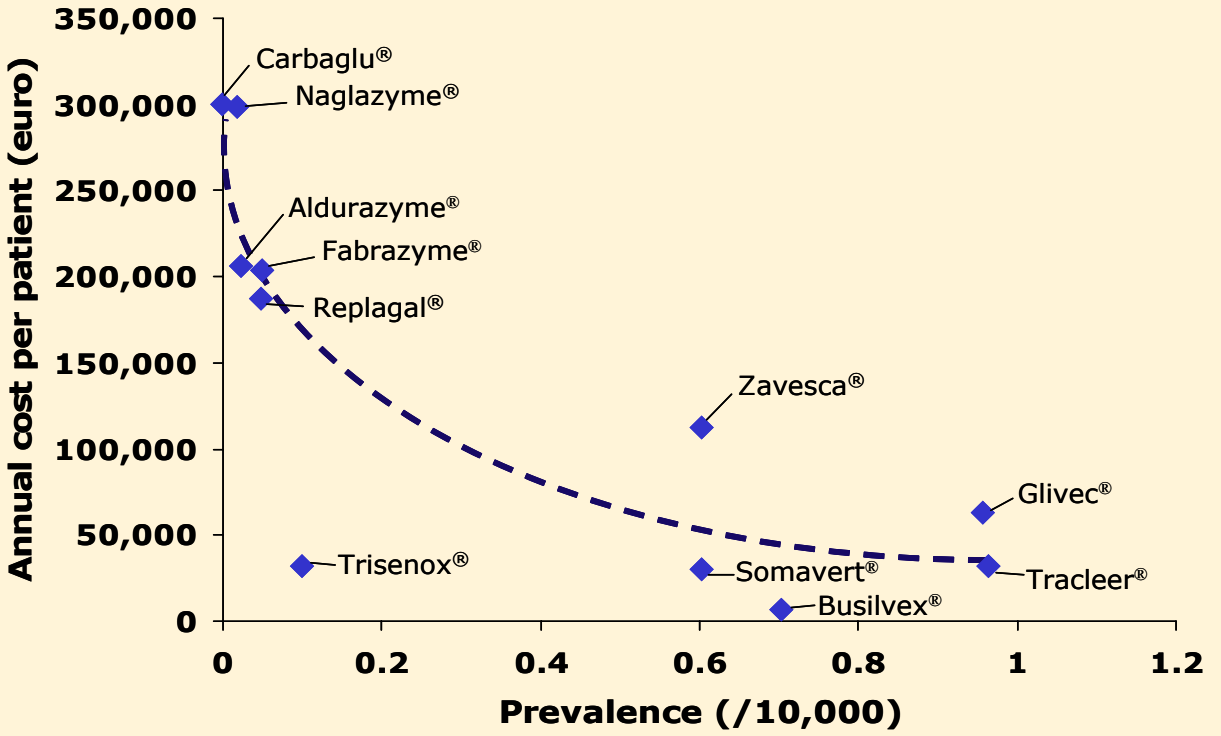
A Greater Role for Budgetary Impact Analysis?

Some ICERs for "Orphan" Treatments

Condition	Prevalence	Product	ICER ("preliminary estimated £ per QALY")
M. Gaucher (Type I and III)	270	Imiglucerase (Ceredase^{R})	391,200
MPS Type 1	130	Laronidase Aldurazyme^{R})	334,900
M. Fabry	200	Agalsidase beta Fabrazyme^{R})	203,000
Hemophilia B	350	Nonacog alpha (BeneFIX^{R})	172,500
M. Gaucher (Type I)	270	Miglustat	116,800

A Greater Role for Budgetary Impact Analysis?

"Orphan" Treatments: No Distinct Subcategory



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)

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")

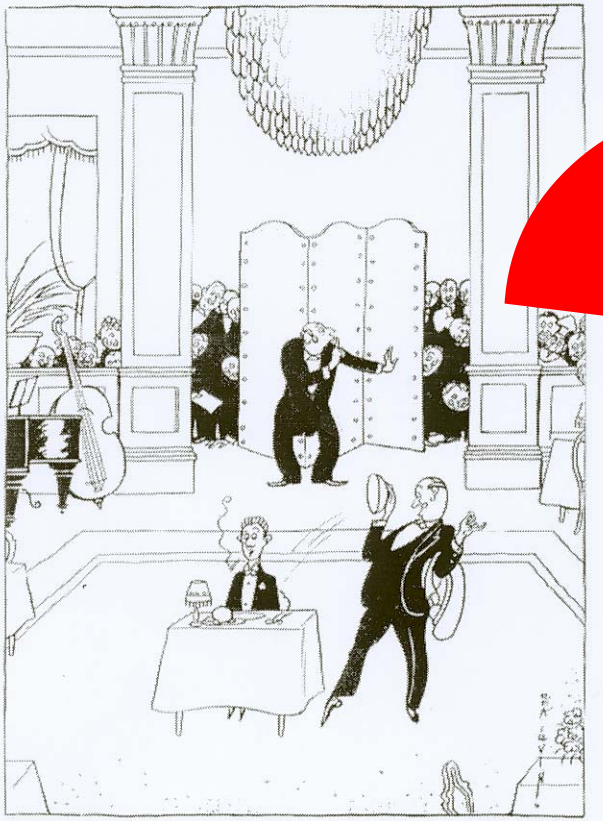
[1]cf. E. Nord (1993); E. Nord et al. (1994)

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

Some Concerns concerning QALY Maximization

- **An Empirically Flawed Decision Rule !**
  - The Consistency Argument – A Thinly Disguised Normative Claim
- **Severity of Condition**
  - Capacity to Benefit of Secondary Importance!
  - 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, Program Size, and Opportunity Cost?**

"What More Could Anyone Ask For?"



THE QALY SURPRISE

THE NEW YORKER 1925

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**

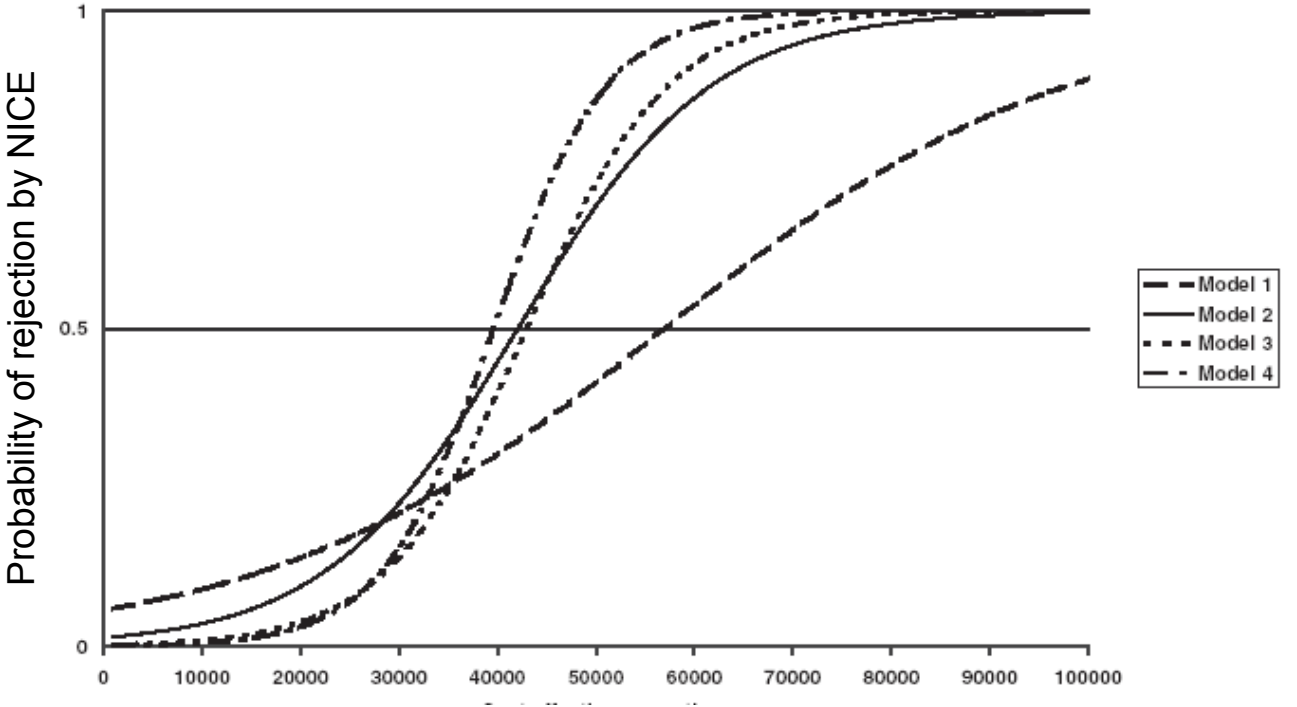
[1]non-exhaustive lists; [2]cf. G.W: Torrance (1976) [3]and in-between; [4]conflicting empirical data

ENGLAND AND WALES: NICE

- Reliance on QALYs
- Robustness

A NICE example of cost-effectiveness benchmarks in practice

'Probabilistic' NICE Cost-Effectiveness 'Benchmarks'[1]



Incremental Cost-Effectiveness Ratio (ICER: cost per QALY gained)

[1]N. Devlin and D. Parkin (2004)

UK Cancer Experts Deplore NICE Decision on Kidney Cancer Drugs

Medscape Medical News

4 new cancer drugs should not be used in the treatment of advanced and/or metastatic renal cell cancer. This draft recommendation, issued on August 7, is now open for consultation; a further review is planned for September 10.

The 4 products involved are bevacizumab (*Avastin*, Roche/Genentech), sorafenib (*Nexavar*, Bayer), sunitinib (*Sutent*, Pfizer), and temsirolimus (*Torisel*, Wyeth). **Although the drugs have been shown to extend patients' lives by some months, NICE ruled that they were not cost effective and hence should not be available on the National Health Service (NHS).**

…

"**It just can't be that everyone else around the world is wrong about access to innovative cancer care and the NHS right in rationing it so severely**," they comment. The signatories include some of the most prominent cancer specialists in the United Kingdom, and the group of 26 is headed by Karol Sikora, MBBCh, PhD, medical director of CancerPartnersUK, professor of cancer medicine at Hammersmith Hospital, in London, and former chief of the World Health Organization Cancer Programme.

UK government backs NICE

Pharmaceutical BUSINESS Review

The UK government's response to a parliamentary committee's report on NICE, the healthcare technology assessment agency for England and Wales, was lukewarm and it **refused to modify the NICE's role or its operating procedures concerning healthcare**, reported PJB news.

The House of Commons health committee report, which was delivered in January, 2008 found […] irregularities in the NICE's guidance concerning the national health service.

But **the government commended NICE's role in promoting cost-effective health care** and dismissed several of the committee's recommendations, as operational matters for NICE itself.

A High Profile

"What Could Be Nicer Than NICE?"[1]

- **Pearson and Rawlins (2005)[2]:**

  "*The conditions seem ripe for a NICE in the United States …*"

- **Smith (2004)[3]:**

  "*The triumph of NICE*":

  "NICE is conquering the world … and may prove to be one of Britain's greatest cultural exports along with Shakespeare, Newtonian physics, The Beatles, Harry Potter, and the Teletubbies …"

- **WHO (2003)[4]:**

  "Published technology appraisals are already being used as international benchmarks …"

[1]A. Williams (2004) "Learning from Europe?"

NICE

NHS National Institute for Health and Clinical Excellence

- Three (distinct) "**Centres of Excellence**":

- **Centre for Public Health Excellence**
  - Public health guidance
    on the promotion of good health and the prevention of ill health

- **Centre for Health Technology Evaluation**
  - Technology appraisals (recommendations on the use of new and existing medicines and treatments within the NHS)
  - Interventional procedure guidance (evaluates the safety and efficacy of such procedures where they are used for diagnosis or treatment)

- **Centre for Clinical Practice**
  - Clinical guidelines
    (recommendations, based on the best available evidence, on the appropriate treatment and care of people with specific diseases and conditions)

[1]http://www.nice.org.uk; last accessed September 13, 2006

NICE Technology Appraisals

Aim of Technology Appraisals

"NICE is asked to look at particular drugs and devices when the availability of the drug or device *varies* across the country.

This may be because of different local prescribing or funding policies, or because there is confusion or uncertainty over its value.

Our advice ends the uncertainty and helps to standardise access to healthcare across the country."[1]

[1]http://www.nice.org.uk/page.aspx?o=202425; last accessed September 13, 2006

NICE Technology Appraisal Process

- **Three (to four) phases**
  - **Scoping**
  - **Assessment**
  - **Appraisal**
  - **Appeal** (if lodged by one or more consultees)
- **Frequently acclaimed features**
  - NICE objective of appraising the evidence in a way that is **"objective, unbiased, and methodologically sound"**[1]
  - An appraisal process that can be described as being **"inclusive, consultative, transparent"**[1]

[1]C. Longson, ISPOR Annual Meeting, Arlington, VA, May 20, 2001

NICE Technology Appraisal Process

1. Scoping

- ¬ DoH develops *remit*; NICE develops draft *scope*
- ¬ Ministers select topics suitable for referral

- ¬ Consultation on draft remit and draft scope with consultees, commentators, & Assessment Group
- ¬ Scoping workshop and invitation by NICE to stakeholders to discuss the appraisal scope

- ¬ Final remit produced by DoH and WAG; final scope produced by NICE
- ¬ Ministers make *final decision on referral*
- ¬ NICE issues *final remit and scope*

[1]NICE (2004)

NICE Technology Appraisal Process

2. Assessment

- Assessment Group (AG) formally commissioned to prepare Assessment Report (AR) based on its *assessment protocol*
- Submissions by manufacturers and sponsors
- Preparation of *Assessment Report* (AR) ("reference case" and template defined by NICE, content and quality responsibility of its authors)
- AR sent to consultees and commentators, with confidential information removed
- *Economic model* considered confidential (intellectual property of assessment groups)

[1]NICE (2004) "Learning from Europe?"

NICE Technology Appraisal Process

3. Appraisal

- Appraisal Committee
  (AC, a standing advisory committee of NICE)
  considers *Evaluation Report*
  (including AR) and comments from consultees on AR
  (including the AG's response to comments, if any)
- AC prepares *Appraisal Consultation Document* (**ACD**);
  following instructions by the AC, a NICE project team
  drafts the ACD
- ACD distributed to consultees and commentators
- AC reviews comments on ACD and prepares
  Final Appraisal Determination (**FAD**) document

[1]NICE (2004)

NICE Technology Appraisal Process

4. Appeal

(optional)

- FAD distributed and published as ***NICE Guidance*** unless one or more *consultees* lodge an *appeal* within 15 working days from receipt of the FAD
- Appeal is permissible on the following grounds:
  - NICE has failed to act *fairly* and in accordance with its published *procedures*,
  - the FAD is *perverse* in the light of the evidence submitted, with "perverse" meaning that the FAD is "*obviously and unarguably wrong*, in defiance of logic, or so absurd that no reasonable Appraisal Committee could have reached such conclusions", or
  - NICE has *exceeded its powers*.
- *New evidence* or simply *disagreement* with a FAD will not be accepted in this last stage of the appraisal process.
- Nor is it possible to reopen arguments and issues on which a determination by NICE has been reached.

[1]NICE (2004)

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)

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)

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]

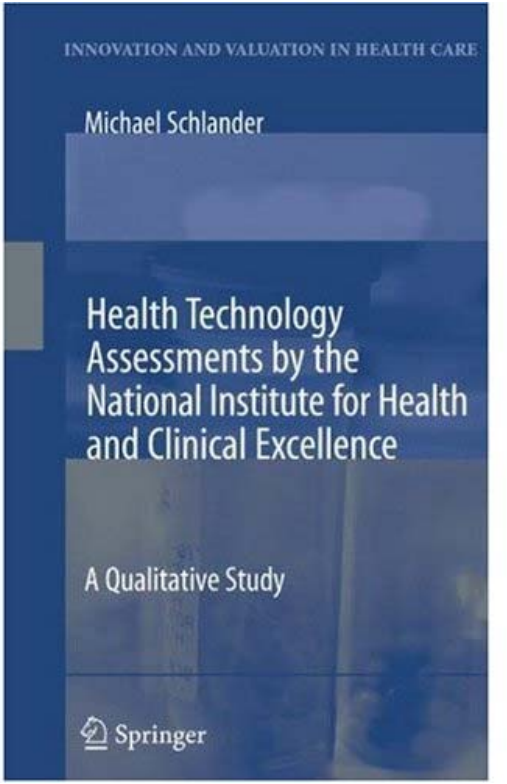
- **Dennis Stant et al. (Groningen, NL):**
- Data of a previously conducted economic evaluation assessing the cost-effectiveness of the HIT intervention in patients with schizophrenia were used to compare
  - analyses based on the primary health outcome (PANSS);
  - results based on various other health 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)

Using Best Currently Available Evidence

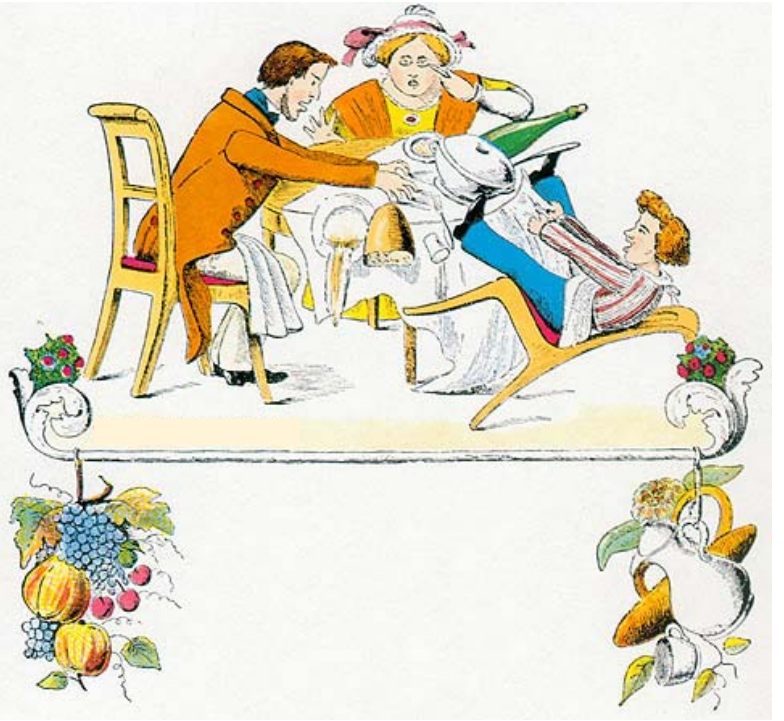
NICE Technology Appraisal No. 98 (ADHD)

- 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 …**



The Disorder

Attention-Deficit/Hyperactivity Disorder (ADHD)



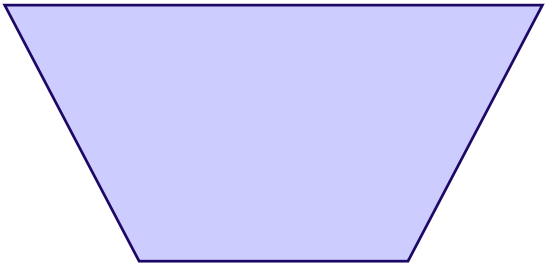
Source: www.sagen.at

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]

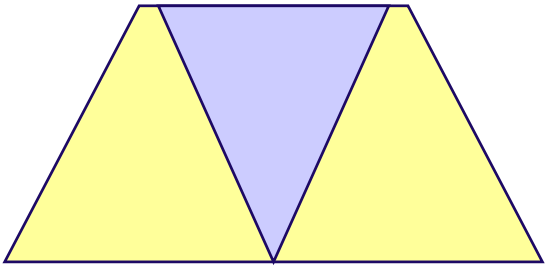
2,908 RCTs — **Literature search**



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)?

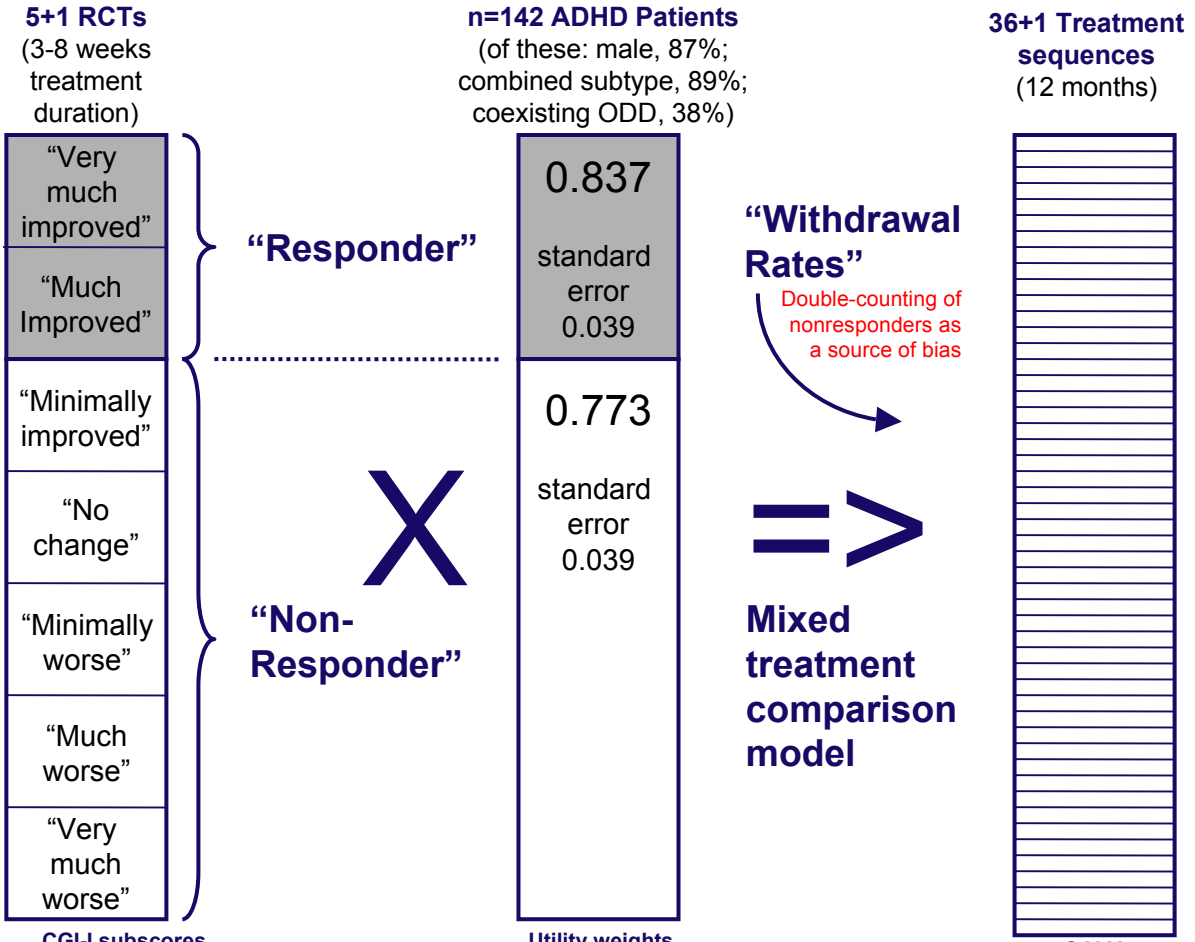
Database analyses (retrospective)?

Economic models (cost-effectiveness analyses)?

05 (+1) RCTs

(3-8w duration)

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



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)

NICE 2004: Main Conclusions of Assessment

- "**Drug therapy seems to be superior to no drug therapy.**
- **No significant differences** between the various drugs in terms of **efficacy** or side effects were found – mainly due to lack of evidence.
- The additional benefits from **behavioral therapy** (in combination with drug therapy) are uncertain"[1].
- "Given the lack of evidence for any differences in **effectiveness** between the drugs, the [economic] model tends to be driven by drug cost, which differ considerably"[1].
- "For a decision taken now, with current available data, **the results of the economic model clearly identify an optimal treatment strategy**"[2] and "this analysis showed that a [...] strategy of 1st line dexamphetamine, followed by 2nd line methylphenidate immediate-release for treatment failures, followed by 3rd line atomoxetine for repeat treatment failures was optimal."

[1]Assessment report, p. 20; King et al., 2004; [2]AR, p.261

NICE 2006: Appraisal Summary

- Where drug treatment is considered appropriate, methylphenidate, atomoxetine, and dexamphetamine are recommended within their licensed indications.
- There are no significant differences between individual drugs in terms of efficacy or side effects
  – *a conclusion derived as a consequence of paucity of evidence used for assessment:*
- Given the limited data used to inform response and withdrawal rates, it is not possible to distinguish between the different strategies on the grounds of cost-effectiveness.
- If there is a choice of more than one appropriate drug, the product with the lowest cost should be prescribed.

NICE ADHD Technology Assessment 2006 – A Critique (1)

- **Narrow scope**
  - Excluding psychosocial interventions (and long-term sequelae)
  - Role of diagnostic criteria and coexisting conditions not addressed (though included in scope)
- **Data selection for assessment**
  - Idiosyncratic interpretation and/or violation of search criteria
  - Reliance on CGI-I subscores for primary economic analysis, economic model departing from clinical effectiveness review.
  - Reliance on short-term data (3-8 weeks in primary model) to extrapolate long-term outcomes (one year; extensions up to 12 years)
- **Efficacy versus effectiveness distinction**
  - Compliance differences effectively "assumed away" for modeling, with potential implications for *all* medications with improved administration schedules.
  - "Real-world evidence", however, is suggestive of a substantial impact of noncompliance and nonpersistence on treatment effectiveness, notably in ADHD

A Critique (2)

- **Data synthesis across studies and endpoints**
  - Remaining evidence was insufficient to assess relative value of treatment options
  - Synthesis of response rates derived from heterogeneous endpoints (CGI-I / CGI-S vs. narrow-band symptom scales; definitions of response and subscales used)
  - Synthesis of data from heterogeneous studies (including, but not limited to, pooling data from pragmatic "real-world" studies and from double-blind RCTs)
- **Economic model**
  - Not transparent, at times enigmatic description (inclusion of studies, data extracted from studies [e.g., MTA], implausible QALY estimates)
  - Interpreting symptom scales explicitly as "quality of life instruments"
  - Extended time horizon of 12 years without considering long-term sequelae (confounded by technical anomalies, e.g., discount rates applied)
- **Appraisal**
  - The Appraisal Consultation Document noted the ADHD core signs of inattention, hyperactivity, and impulsiveness, the difference between ICD-10 and DSM-IV definitions, and the potential influence of comorbidity on therapeutic outcomes in ADHD, although the Assessment Report had not adequately addressed those[1]

[1]Of note, the appraisal process resulted in a correction of the "clear conclusion" of the assessment report

Separation of Clinical and Economic Evaluation

Differences in scope

Selection of clinical studies
(interpretation of 3-weeks-duration criterion; absence of consideration of carry-over effects in crossover trials)

Dissociation between effectiveness review and cost-effectiveness evaluation of technology assessment,
the latter not using findings of the systematic review (use of hyperactivity scores versus CGI-I subscale scores)

Disorder-specific outcome measures excluded from economic evaluation, contributing to the neglect of available **clinical long-term evidence**
(including absence of literature review on clinical measurement instruments and on long-term outcomes)

Broad use of secondary endpoints of clinical studies
as an input for probabilistic cost-effectiveness evaluations
(intended to capture stochastic uncertainty)

Distinction between efficacy and effectiveness
(including absence of compliance literature review)

Reasoning that utility values obtained from patients "may be more relevant to this review", raising doubt whether the clinical problem was fully understood by analysts

High Level of Standardization

Exclusive focus on cost-utility analyses

- At the expense of cost-effectiveness evaluations
- Reliance on utility estimates of limited validity
- For calculation of quality-adjusted life years (QALYs),
  linking utility estimates based on complex health state descriptions
  with response estimates based on clinical global impressions subscales
- Inability to identify differences between treatments

Highly restrictive use of clinical evidence for economic evaluation

- Clinical long-term studies
- Commonly used effectiveness measures
- Mathematical precision of quantitative meta-analysis not in tune with imprecision of dichotomized input data (mostly CGI-based “response rates”, or data pooled from heterogeneous sources) from small-scale short-term clinical studies
- Need to use data from clinical studies
  that had been excluded from effectiveness review for quality concerns

Need for (or Absence of Effective) Quality Assurance

Deviation of assessment from NICE reference case

- Discount rates used for long-term economic model
- Discussion of appropriate sources of utility estimates

Issues related to technical quality of assessment

- Multiple deviations from specified search criteria (relevant randomized clinical studies; relevant health economic evaluations; interpretation of 3-weeks cut-off criterion; inclusion of studies previously rejected for quality concerns)
- Pooling of heterogeneous studies for quantitative synthesis (e.g., efficacy vs. effectiveness; clinical effectiveness measures, treatment intensity, concomitant psychosocial treatment, etc.)
- Not controlling for potential confounding effects (e.g., effectiveness measures used and treatment strategies)
- Mismatch between clinical global impressions (and other response criteria used) and health state descriptions used for utility estimates

Has NICE Got It Right *Consistently*?

- **Apparently, the answer is "No (t yet)".**
  - **Lack of Robustness**: The current NICE approach to health technology appraisals, although often considered exemplary from an international perspective, may become overstretched by complex clinical problems.
- **Suggested underlying reasons include:**
  - Insufficient integration of clinical and economic evaluation.
  - High level of standardization, contributing to a relatively rigid application of the cost-utility (cost-per-QALY) concept, at the expense of alternative methods of health economic evaluation.
  - Provisions for (or lack of) sufficient quality assurance for technology assessments.
  - Some process-related issues (primarily related to the relevance condition of A4R and the use of "QALY egalitarianism" as fundamental equity position, contributing to NICE's strong focus on QALYs).

Does this observation invalidate the approach taken by NICE?

- **Again, the suggested answer is "No"** ("Not completely").
  - One qualitative case study (n=1) does not allow inferences about N>100 technology appraisals by NICE.
  - Of note, the NICE appraisal **process** enabled correction of some of the observed limitations of the technology assessment.
  - Nevertheless, qualitative research exploring specific issues *in-depth* may create **hypotheses** that deserve further research.
  - There are indeed some indications that certain problems observed in the present case might **not** have been **a singular occurrence**.
  - Given the (intended) **impact** of NICE guidance, the limitations associated with the assessment of ADHD treatment strategies are considered serious enough to warrant further inquiry.

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 seems little reason for analysts to be self-congratulatory because 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 is overstated.[3]

[1]Williams (2004); [2]Dolan et al. (2005); cf. Schlander (2005)

[3]relating to both technical and allocative efficiency

An influential proposal for health care priority setting by Norman Daniels and James Sabin, officially adopted by NICE[1]…

Focus on "Due Process": Accountability for Reasonableness[2]

- **Publicity**
  - Decisions and their underlying **rationales** must be **publicly accessible**.
- **Relevance**
  - These rationales must rest on **evidence, reasons, and principles** that plan managers, clinicians, patients, and consumers agree are pertinent to deciding how to meet diverse needs under resource restraints.
- **Revisability and appeals**
  - A mechanism must allow challenges to limit-setting decisions, help resolve those challenges, and allow revisions in light of further evidence and arguments.
- **Enforcement**
  - A voluntary or public regulatory process must ensure that decision makers fulfill the first three conditions.

[1]M. Rawlins and A. Dillon (2005);
[2]N. Daniels and J.Sabin (1997, 1998ff.)

Accountability for Reasonableness (A4R)[1]

A4R Condition	Key Features	Key Limitations
Publicity	**Overall process** well-defined structure; detailed timelines, key documents continuously published; predictable opportunities for stakeholders to provide input	Selection of topics for appraisal (sometimes)
	Assessment Phase Assessment Protocol and Assessment Report published	"Commercial-in-co… withheld. Economic model … ("intellectual
	Appraisal Phase Appraisal Committee meeting agendas published meeting minutes published ACD, FAD published	Uninforma… Committe… Criteria beyon… neither codifi…
	Appeal Phase (optional) Appeal Panel holding public hearings; detailed meeting minutes	
Relevance	**Fairness Condition** High level of procedural fairness within NICE framework; NICE seeking input from Citizens Council on social value judgments	No codified criteria for fairness; "efficiency-first" approach
	Integration of clinical and economic perspectives	Poor alignment of scopes (for technology appraisals and clinical guideline development). Sometimes (?) poor integration of both perspectives
Revisions and appeal	**NICE definition of "appeal"** differs from that of A4R; appeals may be lodged by consultees only	Conditions for appeal more restrictive than A4R recommendations; this appears unlikely to be fully compensated for by opportunities for stakeholder participation
Enforcement	**Consistency of technical quality of assessment reports**	Absence of effective quality assurance system for technology assessments
	Implementation	Mixed record of guidance implementation in the NHS

"NICE's use of cost-effectiveness as an exemplar of a deliberative process …"

"The use of cost-effectiveness by NICE: No(t yet an) exemplar of a deliberative process …"

A.J. Culyer (2006)
Health Economics, Policy and Law 1: 299-318

M. Schlander (2008)
Journal of Medical Ethics 34 (7): 534-539

[1]M. Schlander (2007, 2008)

An International Perspective

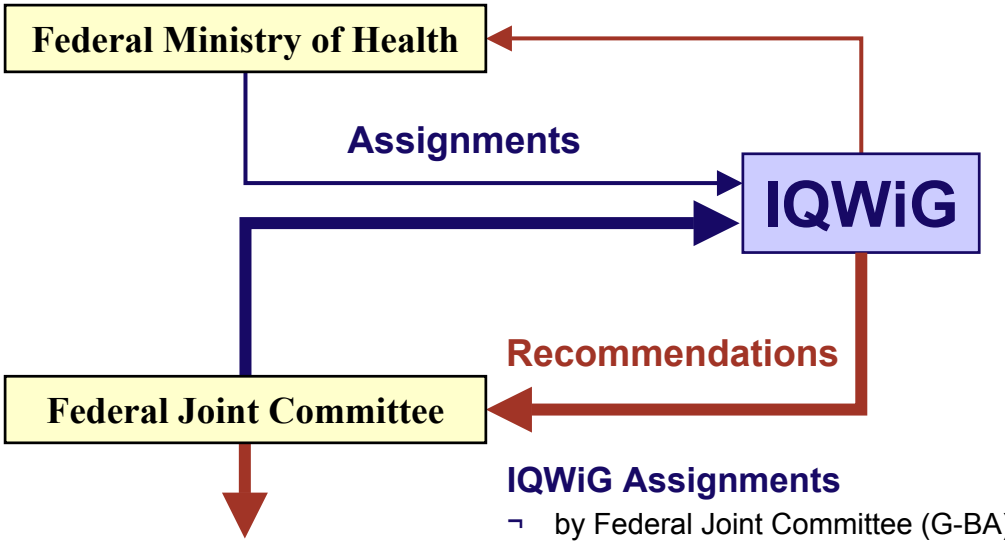
- **International policy makers, looking at NICE as a potential role model, might wish to consider...**
  - **... Objectives** of health care provision (and collective financing); which criteria are appropriate to determine "allocative efficiency" in line with social values?
  - **... Institutional context** of the respective health care system.
  - **... Reliance on QALYs** as an appropriate outcome measure?
  - **... Which technology appraisal processes**? (With few exceptions, it is suggested here that NICE might indeed serve as a role model in that respect.)
  - **... Timing** of technology appraisals?
  - **... Multidisciplinary assessment teams**?
  - **... Quality assurance** of technology assessments?
  - **... Implementation issues**...

GERMANY: IQWiG

Definitely no ultimate solution, but perhaps a feasible beginning?

IQWiG in Germany (2004/2008)

Assignment of Tasks



IQWiG Assignments

- by Federal Joint Committee (G-BA)[1]
- by Federal Ministry for Health (BMG)
- IQWiG may be independently active in pertinent topics related to medical care ("*Generalauftrag*")

[1]Application through the Federal Joint Committee possible for (a) patient organizations, (b) organizations of the German health care self-administration system; not possible for (a) companies, (b) private persons, and (c) interest or lobby groups

Definitely no ultimate solution, but perhaps a feasible beginning?

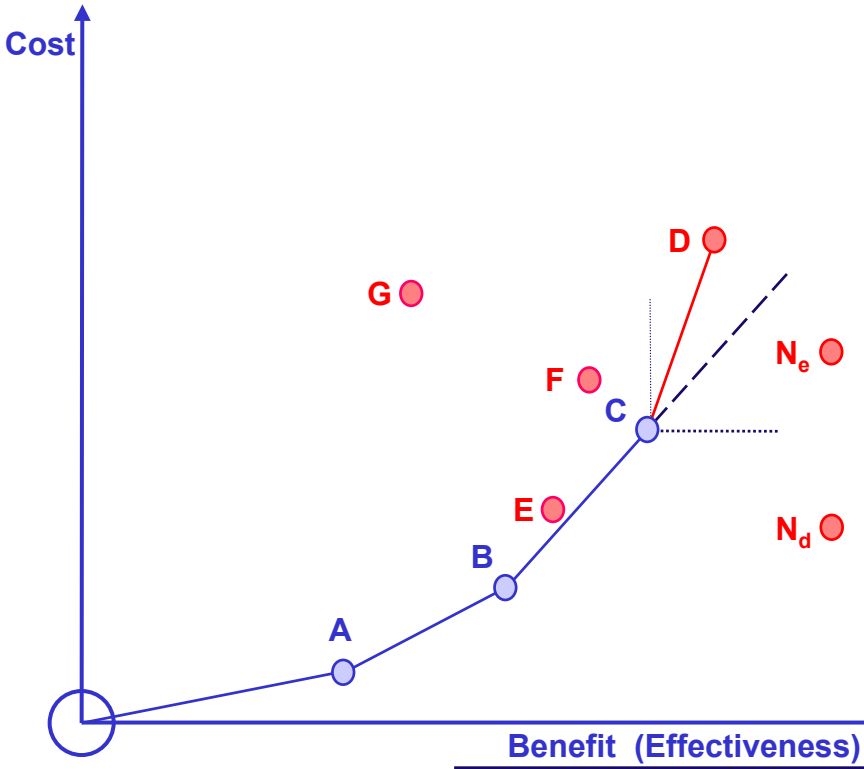
IQWiG 2008: Draft Methods for Economic Evaluation

Consultation Documents V 1.0 and V 1.1 (Jan. 24, 2008 / Oct. 09, 2008)

- **'Cost-Benefit Analysis'** [in line with '*international standards*'?]
  - providing that previous "Benefit Analysis" was positive
- **Cost-Consequence Analysis (?)**
  - No reliance on cost-utility analysis using QALYs
- **'Efficiency Frontier' Concept**
  - Focus on 'technical efficiency' – applying the textbook decision rule of CEA
- **Budgetary Impact Analysis**
  - Notion of 'affordability'
- **Approach Designed to Increase Transparency**
  - But legal mandate to support maximum reimbursement prices (!?!)
- **Feasibility Testing Ongoing**

Definitely no ultimate solution, but perhaps a feasible beginning?

Technical Efficiency as a Pragmatic Alternative?



Efficiency Frontier Approach

- **Are There Alternative Treatments** for the Condition in Question?
- Which **Alternatives** Have Been Reimbursed in the Past?
- **Dominance** of New Treatment "N_d"? => Reimbursement
- **Extended Dominance** of New Treatment "N_e"? => Reimbursement
- **Issue**: Were the Decisions Made in the Past Justified?