Cost-effectiveness studies from observational data

John Brazier

Professor of Health Economics

Health Economics and Decision Science, ScHARR, University of Sheffield, UK

Prepared for the Educational Programme on HRQL and PROs in clinical trials and observational studies, Collage voor Zorgverzekeringen, Eekholt 4, 1112 KH Diemen, 1 December 2006.

Acknowledgements: I would like to thanks Jon Nicholl for proving material for the presentation and to colleagues for their helpful comments and suggestions

The University Of Sheffield.

ScHARR
School of Health and Related Research

Overview of presentation

- Examine the role of observational data in cost effectiveness models
- Review the concerns with using observational evidence to estimate effectiveness and methods for minimising these risks
- Examine other uses of observational data – natural history, costs and utilities

Cost effectiveness

- Cost-effectiveness

  Assessed in terms of the incremental (or extra) cost per quality adjusted life years (QALYs) of the new intervention compared to the existing management
- Cost effectiveness is increasingly assessed using economic models

What is a cost effectiveness model?

- Uses a structured pathway of care for patients with a condition (e.g. decision tree or markov model)
- Assesses the impact of an intervention on the pathway of care (i.e. typically based on a meta analysis of evidence of effectiveness)
- Translates the effectiveness into costs and QALYs

What does a cost effectiveness model need?

- Data on the natural history of the condition from a representative population of the relevant health system (i.e. disease progression)
- Evidence on effectiveness (e.g. relative hazard rates)
- Resources used by patients in each disease states and their costs
- Quality of life and its associated health state utility value

Sources of data for models

- RCTs
- Observational studies
- Technical elicitation

Cost effectiveness often use all three sources of data, but:

- When and how should OS be used instead of RCT evidence
- When and how should OS be used as a complement to RCT

Why use randomised clinical trials?

- RCTs are the best source of unbiased efficacy data
- Concurrent data collection improves internal validity
- Evaluation at an early stage of the life cycle of technologies

What are the limitations of clinical trials?

Often trials do not provide the data needed to address the decision because:

- Choice of comparison therapy
- Length of follow-up
- Surrogate outcomes (e.g. disease progression and not HRQL)
- Atypical care – leading to protocol driven costs and outcomes
- Atypical patient populations
- Limited generalisability
- Inadequate sample size

Observational data provide another important source of data – perhaps the only source of evidence

What are observational studies?

- Studies in which allocation is not random are sometimes collectively termed Observational studies.
- Observational studies may or may not be ‘experiments’.
- In non-experimental OS the researcher observes and records what happens.
- In experimental OS the researcher can control aspects of the design such as allocation, timing, follow-up, outcome measures.

How can observational data be used?

Effectiveness:

- To estimate the ‘effectiveness’ of an intervention
- To estimate longer term outcomes
- To estimate rare outcomes

Other applications

- To provide the natural history of a condition
- To generate resource use, cost and utility data

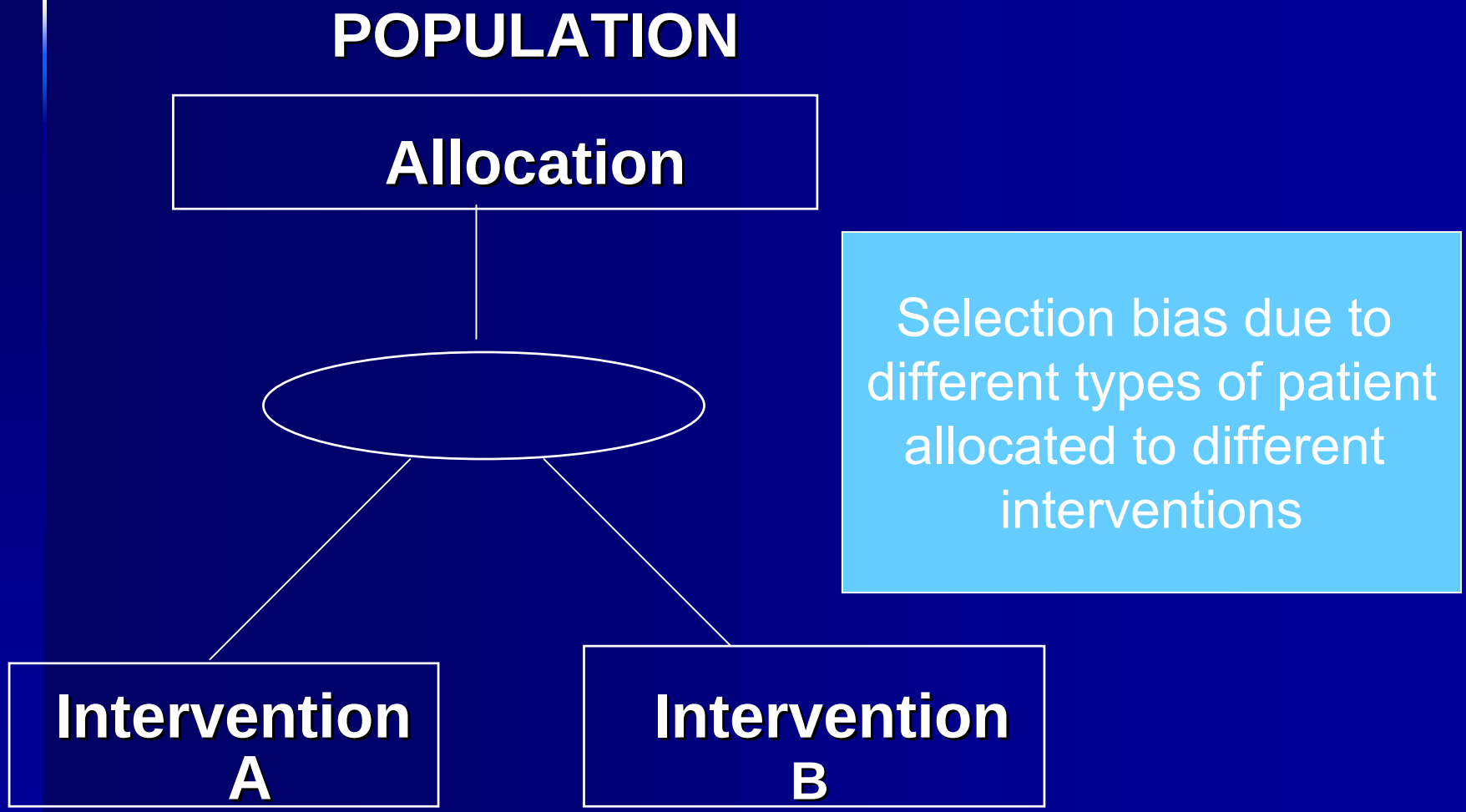
This presentation begins with the first of these.

Individual vs. population studies

Individual studies - patients from a population being selected for one therapy or another (e.g. fluids or no fluids) because of risk, characteristics, etc.

Population studies - in which two unselected populations which may have different characteristics are compared.

Selection bias in individual allocation studies



Selection bias due to different types of patient allocated to different interventions

Selection by risk: GP Maternity Units vs. Consultant Units (Clarke, BMJ; 306: 825-)

Perinatal mortality/1000 by place of delivery:	GP	Consultant
unadjusted	3.3	9.4 – 12.6
adjusted for case-mix	3.8	10.1 – 12.4

GP Maternity Units vs. Consultant Units (Clarke, BMJ; 306: 825-)

Perinatal mortality/1000 by place of delivery:	GP	Consultant
unadjusted	3.3	9.4 – 12.6
adjusted for case-mix	3.8	10.1 – 12.4
Perinatal mortality/1000 by place of booking	8.8	9.3 – 11.7

Implications

- Selection by risk of expected outcome (indication) and then comparing outcomes is fatal.
- Case-mix adjustment cannot control for selection by subtle or unmeasured indications.

Alternative methods of individual selection

By characteristic

- Such as being picked up by helicopter or ground ambulance
- Can try case mix adjustment, but problems may remain….

By circumstance (quasi randomisation)

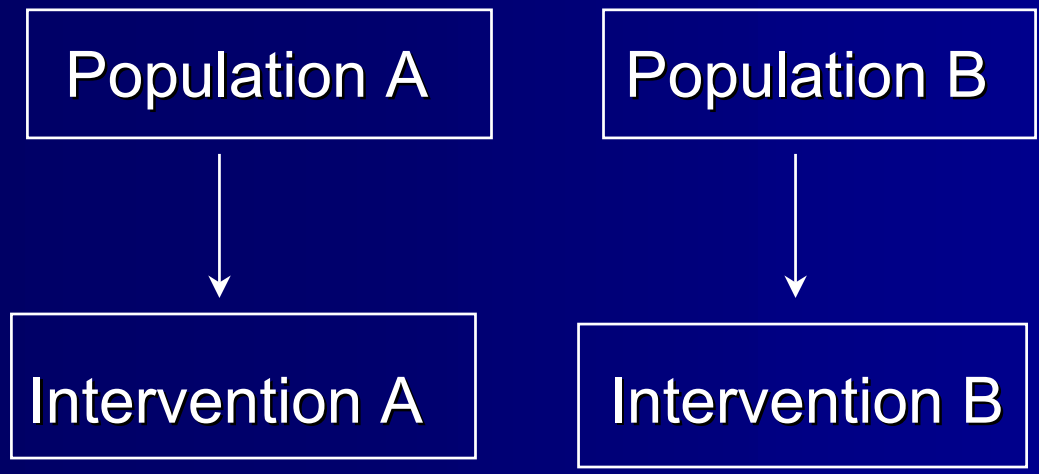
- Type of paramedic (ALS) vs. EMT (BLS) for trauma at the scene.
- Selection by (nearest) available ambulance.
- Case mix adjustment may remove bias

Effectiveness: Bias in observational studies

- All OS are more or less subject to ‘selection bias’ – due to the potential allocation of different types of patient (etc) to different interventions.
- Non-experimental OS are likely to have additional biases (in ascertainment, follow-up, outcome measurement, etc).
- Experimental OS may have these biases due to poor design – as may RCTs.

Good design in OS must address all the quality issues in RCTs + selection bias.

2. Bias in population allocation studies



No ‘selection’. Bias due to case-mix differences of populations

Randomly chosen samples from two identical populations receiving different interventions = Random allocation to different interventions from a single population

Population allocation: cohort studies

- Population OS are very useful for evaluating services
- But what is being evaluated when two populations are compared are ALL the service differences between the populations

Accounting for Population 'Case-mix' Differences

By analysis - Don't introduce selection. Compare whole populations whichever service they have – The **Intended Service** (IS) approach º ITT, and adjust for case-mix

By Design - Control for population case-mix by comparing populations before the new service was introduced

- Controlled before and after studies (i.e. populations not receiving interventions are followed to adjust for 'drift')

Propensity scores in before and after studies – By J Nicholl

- Propensity scores indicate the likelehood of receiving the new treatment (after it has been introduced)
- Apply score to before data to identify those who would have received intervention (i.e. the counter factual)
- Control provided by those identified not to receive intervention before with those who did not after to estimate ‘drift’
- estimate treatment effect as the impact on the regression coefficients over the two periods

ALL THESE METHODS NEED GOOD REGISTERS!

Example of OS: Registers

To be useful at all registers must

1. Include individual data on consecutive cases (no selection).
2. Have full standardised follow-up.
3. Accurate recording.
4. Include all known characteristics affecting outcome

Routine fall short of this. What about registers?

Quality of registers?

- Haug et al (Eur. J. Vascular and Endovascular Surgery, 2005; 571-578) compared AAA repair in a Norwegian national vascular surgery registry vs. a national administrative registry.
- Vascular registry contained
  - 84% of cases
  - 72% of hospital deaths

"Cannot exclude the possibility that some departments are less willing to report unfavourable results"

Difficulties with registries

- Ethics and governance
- Data collection – may not collect the right data
- Data quality – often high levels of missing data
- Data analysis and reporting

Registries never work without clinical ownership

Can’t work without proper resourcing

Other uses of observational data: generalising from trials

- Apply estimates of relative impact from RCTs to natural history data or local data on types of patients, outcome and use of resources
- Helps to generalise from RCT

Wailoo et al (2006): developed a CE model for US medicare patients with RA for 4 biologic drugs

- The probability of a patient achieving ACR20 or ACR50 was estimated from a meta analysis of trials
- Impact on disability of different levels of response in short and longer term estimated from US National Databank for Rheumatic Disease adjusting for patient characteristics

Other uses of observational data: generalising from trials

Robinson et al. (2005): developed a CE model of alternative strategies for the initial medical management of non-ST elevation acute coronary syndrome

- Relative risk reductions of GPAs in short term estimated from trials applied to UK data on baseline probabilities of death, non-fatal MI , revascularisation and major bleeding
- Long term costs and QALYs estimated with data from a Heart Attack Register

Other uses of observational data continued

- Confirm outcome beyond trial follow-up period
- Estimate rare events – outcomes or resource use

For these two applications there are the same design and analysis issues

HRQL and Utilities data

- HRQL measures often not used in clinical trials or the measured used are not suitable for generating utilities
- Observational studies provide an opportunity to fill this gap

Example: Age related macular degeneration

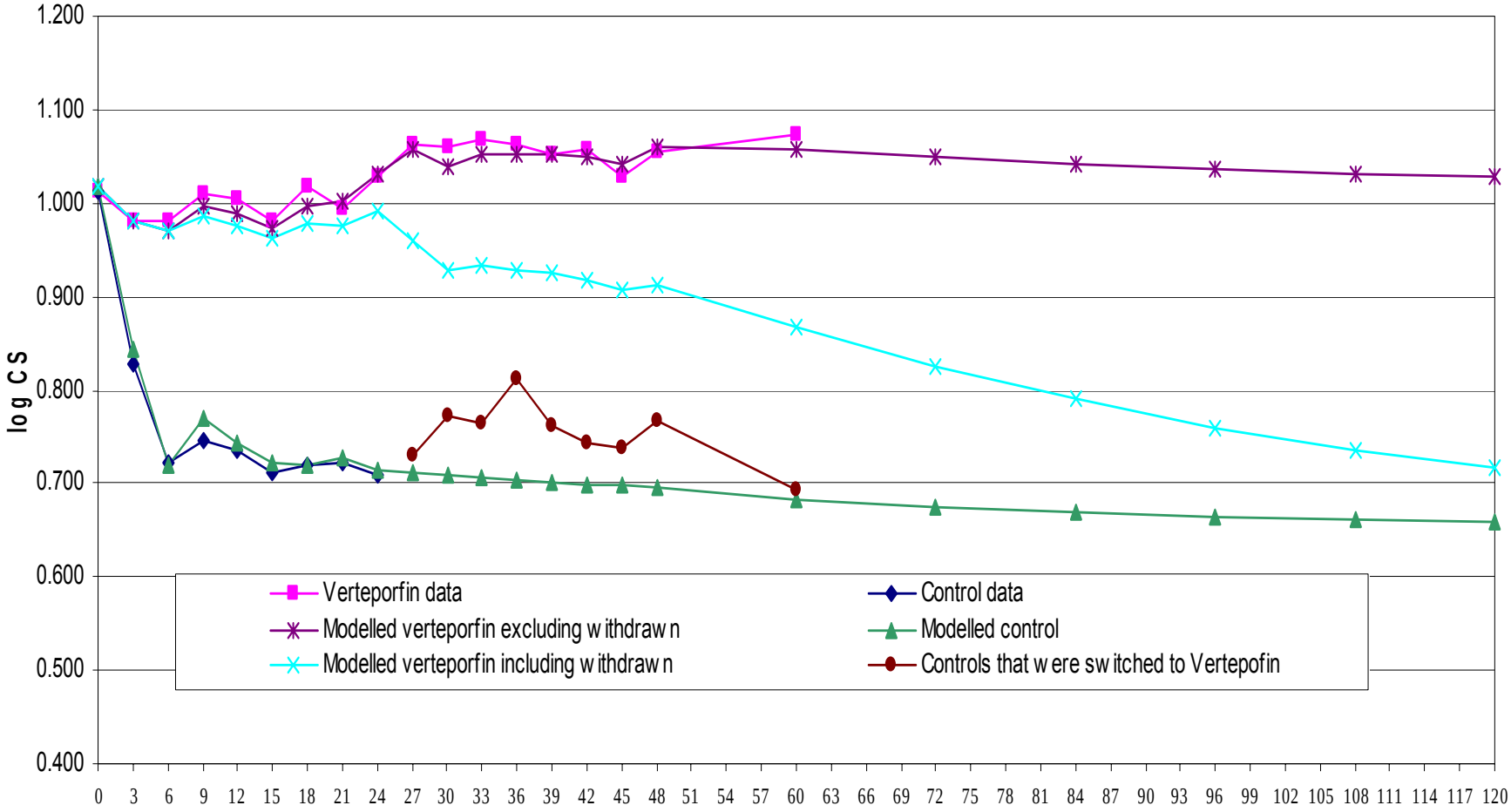
Cost effectiveness in AMD

Case study: Cost-effectiveness of Verteporfin PDT in patients with predominantly classic neovascularisation

Requires the following information:

- Clinical effectiveness of intervention measured by contrast sensitivity (and visual acuity) in a trial
- Costs associated with treatment and visual impairment (in particular legal blindness)
- Quality of life of patients in different levels of visual impairment

Clinical outcome



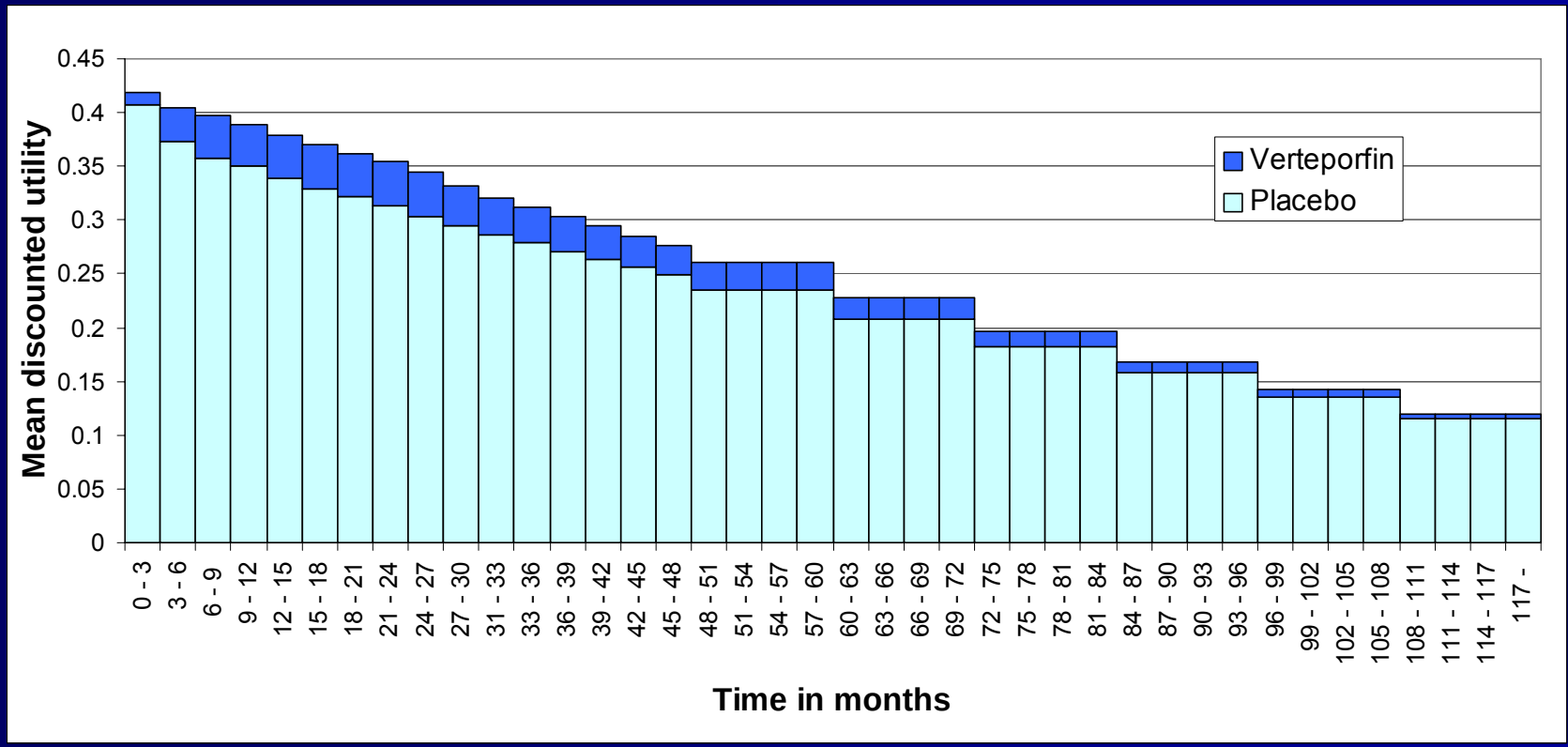
How can we convert this effectiveness data into costs and QALYs?

Mean (SD) scores of QoL measures by CS

Contrast sensitivity (binocular, log units)	N	TTO	HUI3	SF-6D	EQ-5D
<0.30	67	0.58 (0.32)	**0.25 (0.25)**	0.65 (0.11)	0.70 (0.20)
0.30 thru 0.90	67	0.56 (0.32)	**0.30 (0.26)**	0.64 (0.14)	0.70 (0.24)
0.91 thru 1.3	48	0.70 (0.28)	**0.42 (0.24)**	0.68 (0.14)	0.78 (0.16)
>1.30	26	0.83 (0.25)	**0.53 (0.31)**	0.73 (0.16)	0.70 (0.28)
eta^2		0.09*#	0.14*#	0.05*#	0.03

* *p<0.05 between groups*
\# *p<0.05 linear trend*

Results: Putting together clinical trial data with utilities



Utilities: mapping

Trials often use HRQL measures that do not generate the utility values required for QALYs

- One solution would be to map from the HRQL measure (e.g. EORTC-QL30) on to a preference-based measure (such as EQ-5D)
- Uses an observational data set containing the HRQL and the preference-based measures
- Regression techniques used to model relationship between the preference-based single index and:

- Total score or
- Dimension scores or
- Item responses or
- Item levels – allows for a categorical scale

Can consider range of transformations and interaction terms

Utilities: mapping continued

Advantages:

- It can be quick and in some circumstances it may be adequate (and seems to be accepted by some agencies, such as NICE)

Disadvantages:

- Requires substantial overlap between measures
- Assumes all important domains are covered by generic measure
- Second best to either: 1) using a generic preference-based measure or 2) developing a preference-based version of the HRQL

Utilities: which measure?

The choice of measure depends on the answer to the following questions:

- How to describe health?

Generic (such as EQ-5D or HUI3) vs. Condition specific measures

- How to value health state?

Three cardinal methods have been used to generate values: visual analogue scale (VAS), standard gamble (SG), time trade-off (TTO) or one of the ordinal methods

- Whose values?

General vs. patient population

Some agencies advocate a ‘reference case’

Overview

- Quality of design is as important as in RCTs.
- Case-mix adjustment is necessary but problematic.
- Individual allocation studies with selection
  - by risk = X
  - by characteristics = ?
  - by circumstances = √
- Population OS studies compare all differences between populations, so case-mix adjustment should be used.
- Controlled before and after population allocation studies are best.

Overview continued

- OS provide useful data to generalise form trials – such as natural history or local rates for key events in model
- Longer term and rare outcomes
- Utilities and costs – through mapping onto health states in model or HRQL measures

Observational studies should not replace RCTs where these are possible