Statistical methods for cost-effectiveness analysis: a personal history

Andrew Briggs
Professor of Health Economics

The last 25 years have seen a large increase in the contribution that health economic analysis has made in national and international decisions about health care provision. Andy Briggs has been working at the interface between medical statistics and health economics throughout this period. In this talk he gives a personal history of that journey with an emphasis on how statistical thinking has improved the methods of health economic evaluation over that period. Looking to the future, there remains much potential for statistical methods to continue to improve the way in which we evaluate the cost-effectiveness of health care interventions and to improve health care decision making as a result.
Statistical methods for cost-effectiveness analysis: a personal history

- Representing uncertainty in cost-effectiveness analysis
- Clinical trials versus decision models: a false dichotomy?
- Statistical decision theory
- Survival analysis
- Comparative effectiveness and the rise of Network Meta Analysis

Representing uncertainty in CEA
CIs on the CE plane: the confidence box

R<sub>U</sub>=£34,341/LY
R=£15,000/LY
R<sub>L</sub>=£6,552/LY

CIs on the CE plane: the confidence ellipse (zero cov)

R<sub>U</sub>=£32,542/LY
R<sub>L</sub>=£6,914/LY
CIs on the CE plane:
the confidence ellipse (corr=+0.6)

CIs on the CE plane:
the confidence ellipse (corr=-0.6)
Two worlds: the real world and the bootstrap world

Real World

unknown probability distribution

observed random sample

$x = (x_1, x_2, ..., x_n)$

S($x$) statistic of interest

Bootstrap World

empirical distribution

bootstrap sample

$x^* = (x^*_1, x^*_2, ..., x^*_n)$

S($x^*$) statistic of interest

Bootstrapping the ICER

Four stage process:

1. Bootstrap $n_c$ cost/effect pairs from the control group: calculate means
2. Bootstrap $n_T$ cost/effect pairs from the treatment group: calculate means
3. Calculate the bootstrapped ICER from these bootstrapped means
4. Repeat many times to create the bootstrap estimate of the ICER sampling distribution
CIs on the CE plane: the bootstrap interval

\[
\begin{align*}
\Delta C - R\Delta E &\sim \text{ Normally} \\
normalsize{\frac{\Delta C - R\Delta E}{\sqrt{\text{var}(\Delta C) + R^2 \text{var}(\Delta E) - 2R \text{cov}(\Delta C, \Delta E)}}} &\sim \mathcal{N}(0,1) \\
R^2 \left[ \Delta E^2 - z_{a/2}^2 \text{var}(\Delta E) \right] \\
-2R \left[ \Delta E \cdot \Delta C - z_{a/2}^2 \text{cov}(\Delta E, \Delta C) \right] \\
+ \left[ \Delta C^2 - z_{a/2}^2 \text{var}(\Delta C) \right] \\
= 0
\end{align*}
\]

Standard quadratic equation in \( R \)

\[
\begin{align*}
\left[ \Delta E \cdot \Delta C - z_{a/2}^2 \text{cov}(\Delta E, \Delta C) \right] \\
\Delta E^2 - z_{a/2}^2 \text{var}(\Delta E) \\
\pm \sqrt{\left[ \Delta E \cdot \Delta C - z_{a/2}^2 \text{cov}(\Delta E, \Delta C) \right]^2 - \left[ \Delta E^2 - z_{a/2}^2 \text{var}(\Delta E) \right] \left[ \Delta C^2 - z_{a/2}^2 \text{var}(\Delta C) \right]} \\
\Delta E^2 - z_{a/2}^2 \text{var}(\Delta E)
\end{align*}
\]

A parametric approach: Fieller’s theorem
Non-significant differences?

The estimated distribution of the ICER
Problems with negative with negative ratios

\[ R_c = \£15,000/LY \]

Uncertainty on the CE plane: using the decision rule

\[ R_c = \£0/LY \]
\[ R_c = \£5,000/LY \]
\[ R_c = \£15,000/LY \]
\[ R_c = \£30,000/LY \]
\[ R_c = \£50,000/LY \]
\[ R_c = \£100,000/LY \]
Cost-effectiveness acceptability curves

The 50% point corresponds to the point estimate of cost-effectiveness (£15,000/LY)

The curve cuts the vertical axis at the P-value (one-sided) for the cost difference (P=0.07)

The curve is tending to 1 minus the P-value (one-sided) for the effect difference (P=0.05)

80% interval exists (£2,000/LY, £72,000/LY)

95% interval is undefined
Trials versus models: false dichotomy?

Whither ‘trial-based’ analyses?

- Failure to compare all relevant options
- Truncated time horizon
- Lack of relevance to the decision context
- Failure to incorporate all evidence
- Inadequate quantification of uncertainty

Source: Sculpher et al, 2006, Health Economics
Requirements of economic evaluation for decision making

- Clear statement and measurement of the objective function
- Consistent perspective
- Appropriate specification of the decision problem
- Appropriate time horizon
- All relevant evidence
- Relevant to the decision context
- Appropriate characterisation of uncertainty

Source: Sculpher et al, 2006, Health Economics

An iterative approach to economic appraisal

Source: Sculpher et al, 2006, Health Economics
UKPDS Outcomes Model v1

- Series of linked risk equations from UKPDS study
- Capable of predicting (quality adjusted) life expectancy and lifetime cost

DOI 10.1007/s00125-004-1527-z

A model to estimate the lifetime health outcomes of patients with Type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model (UKPDS no. 68)

P. M. Clarke1, 2, 3 • A. M. Gray1 • A. Briggs1 • A. J. Farmer2 • P. Fenn1 • R. J. Stevens4 • D. R. Matthews5
I. M. Stratton1, 2 • R. R. Holman1 • on behalf of the UK Prospective Diabetes Study (UKPDS) Group

Diabetes related mortality

EVENT FATALITY (odd ratio)

DIABETES MORTALITY (by subsequent years)

First year of event

Diabetes related mortality
Statistical decision theory
Expected Net Benefit of Sampling (ENBS)

- Is it cost-effective to run a new study? If so, to choose the best design (e.g. sample size per arm).
- Difference between monetarised gain from collecting further data from chosen study design, and the cost of that study
- ENBS (n) = population EVSI (n) – cost (n)
As a result of randomizing some patients to treatments that are suboptimal—particularly to placebo [15]. In fact, these calculations can be used to determine exactly which treatments should be included in the trial, on the basis of existing evidence. EVI calculations may show that randomizing to a placebo arm is not justified by the additional information gained. Figure 4 illustrates a network of evidence on bipolar disorder treatments, involving five active treatments and placebo. Using a simple CEA model, similar to the model used in NICE Technology Assessment 66 [19], we have experimented on various forms of EVSI calculation. Figure 5 shows a plot of the EVSI against the monetarized value of a QALY (willingness-to-pay), when 1500 patients are allocated in equal numbers to each arm. Each line represents a trial that includes different treatment comparisons. The results show that an infinitely sized trial including all seven treatments provides scarcely any more information than a trial of 1500 patients involving the four best treatments: olanzapine, haloperidol, lithium and valproate semisodium. Note that in this context, best means highest in expected net benefit. Placebo does not appear in this list as it is relatively ineffective, quetiapine because it is relatively costly. This approach could be used to determine whether there is value in funding multi-arm trials of biologic therapies, and if so which arms to include.

Some readers may be interested in seeing the formulae for the different forms of EVI calculation. These can be found in the appendix (available as supplementary data at Rheumatology Online), where they are set out in a heavily annotated form.

CEA models and VOI calculations

Clearly, VOI calculations begin with a CEA model. It needs to be a CEA model of the specific type—becoming increasingly used—that fully expresses parameter uncertainty. Models that do not require individual patient simulation are favoured, because the high computational demands of EVPPI and EVSI calculations are then increased exponentially by a third level of simulation, as well as a level of optimization to identify the optimal sample size. While it is true that methods exist that mitigate this problem, they require special skills and software. Although solutions to computing problems can usually be found, it needs to be emphasized that VOI calculations, like many other techniques, require careful interpretation and sensitivity analysis [12]. It may require a large number of calculations looking at different scenarios, and possible alternative portfolios of research before a trial design can be confidently generated.

The prime requirement, however, is a CEA model that represents a consensus view of the natural history of the disease and the evidence. Clinicians commonly express serious doubts about the assumptions that are made just to achieve a CEA, and are often dismayed that, after...
Parametric survival - distribution selection

- Latimer (2011) recommends selecting the most appropriate parametric model based on both within-trial fit, and external and clinical validity

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Within trial period</strong></td>
<td></td>
</tr>
<tr>
<td>AIC &amp; BIC statistics</td>
<td>Assess the relative fit of parametric models whilst accounting for the number of parameters</td>
</tr>
<tr>
<td>Cox-Snell residuals</td>
<td>Assess how closely a parametric function follows the Kaplan-Meier function</td>
</tr>
<tr>
<td>Cumulative hazard plot</td>
<td>Assess the behavior of the hazard function over time and the plausibility of the proportional hazards assumption</td>
</tr>
<tr>
<td>Log-cumulative hazard plot</td>
<td>Assess the behavior of the hazard function over time and the plausibility of the proportional hazards assumption</td>
</tr>
<tr>
<td>Quantile-quantile (Q-Q) plot</td>
<td>Assess how closely an accelerated-failure time treatment effect model fits the data</td>
</tr>
<tr>
<td>Visual inspection</td>
<td>Assess how closely a parametric function follows the Kaplan-Meier function and the clinical plausibility of the prediction in relation to other endpoints</td>
</tr>
<tr>
<td><strong>Extrapolation period</strong></td>
<td></td>
</tr>
<tr>
<td>Monthly event probabilities</td>
<td>Compare event probabilities based on each parametric function and external longer term observational data</td>
</tr>
<tr>
<td>Visual inspection</td>
<td>Assess how closely the tail of a parametric function fitted to the active treatment arm(s) concur with external longer term observational Kaplan-Meier data</td>
</tr>
</tbody>
</table>
Cure modelling

- For some outcomes, a proportion of patients may never experience the event.
  - e.g. patients who are cured of their disease may never experience disease recurrence
- Identified graphically by a plateau in the Kaplan-Meier where the hazard equates to the general population hazard
- Standard parametric models may be incapable of fitting to these hazard functions
- May also be an interest in estimating the proportion of patient who are cured, referred to as the “cure fraction”

\[ S(t) = S^*(t)[\pi + (1 - \pi)S(t_u)] \]

- \( t \) is time
- \( S^*(t) \) is the survival for the general population
- \( \pi \) is the cure fraction
- \( S(t_u) \) is the survivor function for uncured patients
- Need to choose a parametric form for \( S(t_u) \)
  - Weibull, Lognormal, Generalised gamma
- Available using the `strsmix` command in Stata and other packages (SAS, R)

Lambert (2007)
Parametric cure models – application

Standard model

Cure model

Treatment switching

- Patients switch from control to treatment arm following primary endpoint
  - Common issue in oncology trials

- As a subset of patients in the control arm receive the benefit of treatment, overall survival times for patients who switched treatment are overestimated.
  - Failure to adjust for treatment crossover can result in an underestimated relative treatment effect.
  - Adjustment for treatment crossover aims to reduce bias in relative treatment effect estimates resulting from such treatment switching.
  - Methods that preserve randomisation are needed to provide unbiased treatment effect estimates in the presence of treatment switching.
Methods for adjusting for treatment crossover

• NICE guidelines provide a range of approaches to consider when faced with treatment switching, including:
  • Rank Preserving Structural Failure Time Models (RPSFTM)
  • Iterative parameter estimation (IPE) algorithm
  • Inverse Probability of Censoring Weights (IPCW) method

Comparative effectiveness
Network Meta-Analysis

Closed loops in network: combination of direct and indirect evidence

Network meta-analysis

(>2 studies in network)
Comparative effectiveness without a network?

Comparing existing but unconnected data:

- Naïve comparison
- Match adjusted indirect comparisons (MAIC)
- Simulated Trial Comparison (STC)

Single arm studies:

- Historical controls
- Synthetic controls
- Self control
Statistical methods for cost-effectiveness analysis: looking to the future

• Real world data, big data and personalised medicine
  • Application of ML and AI
  • Use of innovative methods for causal inference (MR / other IV approaches)
  • SEM (modelling components of CEA?)

• Refinement of existing methods
  • Methods for statistical estimation of counterfactuals
  • Application of existing VOI methods (under-used)
  • Novel survival analysis approaches (multi-state survival / competing risks)
  • Uncertainty in the face of multiple methods challenges (bootstrapping)