CMA or CEA for non-inferiority trials and those observing no significant difference in efficacy?

Helen Dakin and Sarah Wordsworth
Health Economics Research Centre, Department of Public Health, University of Oxford, Old Road Campus, Headington, Oxford, UK

Abstract

Aims: To explore the advantages and disadvantages of conducting full economic evaluations of non-inferiority trials and those observing no significant difference and to assess whether there are any situations where it is appropriate to conduct cost-minimisation analysis (CMA) rather than cost-effectiveness analysis (CEA) or when it may be inefficient to collect economic data.

Methods: Following the Briggs and O'Brien 'death of CMA' paper, we re-examined the rationale for CMA, focusing on non-inferiority trials and those finding no significant difference in efficacy. We present examples of simulated and trial data to illustrate the advantages and disadvantages of CMA and quantify its inherent biases.

Results: Besides arguments concerning irrelevance of inference and the possibility of observing a significant positive net benefit despite non-significant differences in costs and effects, we show that conducting CMA will bias measures of uncertainty (e.g. value of information or probability that treatment is cost-effective). CMA may lead to either overestimation or underestimation of uncertainty: even for non-inferiority trials. Although bias will be negligible for non-inferiority studies comparing treatments that differ enormously in cost, it is generally necessary to collect and analyse data on both costs and efficacy (including utilities) to assess this bias.

Conclusions: Our analyses show that the remit of CMA is even narrower than previously thought. Full economic evaluations evaluating the joint distribution of costs and benefits are almost always required to avoid biased estimation of uncertainty and allow data on costs, utilities, cost-effectiveness and value of information to inform future research and healthcare decision-making.

1. Introduction

1.1. Background

Increasing numbers of economic evaluations are being performed alongside trials, some of which are explicitly designed and powered to test hypotheses about non-inferiority or equivalence.[a] Although Donaldson et al[1] have suggested that it is not possible to pre-specify the type of economic evaluation in advance, many health economists in the process of applying for research funding or putting together a statistical analysis plan nonetheless specify the type of evaluation to be performed. For trial-based economic evaluations, in

[a] For brevity, studies designed to test hypotheses of non-inferiority and those designed to assess equivalence are referred to hereafter as "non-inferiority studies".

which the health care interventions being assessed are expected to differ in effects, cost-effectiveness analysis (CEA)[b] is clearly preferable to cost-minimisation analysis (CMA). However, for non-inferiority trials, the decision of which form of economic evaluation is appropriate remains debatable for many health economists. Although cost-minimisation analysis (CMA) was historically regarded as a full form of economic evaluation,[2] many health economists now consider CMA inappropriate for studies not designed as non-inferiority or equivalence trials, following the publication of the Briggs and O'Brien 'Death of CMA' paper.[3] However, some health economists and journal editors nonetheless question the rationale for conducting full economic analyses (CEA) of trials designed to detect differences between treatments but which observe no significant difference in efficacy.

This paper therefore aims to re-examine the advantages and disadvantages of conducting a full economic evaluation of non-inferiority trials and those observing no significant difference in efficacy and to assess whether there are any situations in which CMA is still appropriate. The paper is organised as follows. First we give some background to the CMA vs CEA debate, covering the advantages and disadvantages of both techniques. We then describe a literature search of published CMAs and CEAs on NHS EED, followed by an empirical investigation of the potential biases within CMA and CEA. The results of the literature search and empirical work are next reported, followed by discussion and conclusions. Finally, we pose questions for the HESG audience.

1.2. Previous research for and against CMA

Historically, it was recommended that CMA should be used for economic evaluations of trials finding no statistically significant difference in effectiveness.[2 4] Separate and sequential hypothesis tests would be conducted on costs and effects to determine whether incremental cost-effectiveness should be estimated. CMA has the obvious advantage of simplicity and ease of analysis. The results of CMAs are (arguably) also easier for clinicians to interpret since costs and effects can be presented separately. However, this simplicity is also a major disadvantage of CMA, and in an era where uncertainty may be a major consideration in interpreting results,[5 6] 'simple' CMA may not be as informative as a CEA.

[b] Within this paper, we use the term "cost-effectiveness analysis" in a more general (or US) sense to refer to both cost-utility analyses (CUA) and cost-effectiveness analyses in which cost-effectiveness ratios are calculated.

The 'blue book' by Drummond *et al* illustrates the changing guidance on the appropriate role for CMA.[2 4 7] In their first edition (1987), CMA was regarded as a full form of economic evaluation to be used when competing interventions have 'identical' consequences or 'equivalent effectiveness'.[4] In 2001, Briggs & O'Brien published their 'Death of CMA' paper, which focused on how analysts determine whether interventions have the 'same outcome' under uncertainty.[3] They argued that it is inappropriate to perform separate and sequential hypothesis tests on costs and effects to determine whether incremental cost-effectiveness should be estimated. Instead, the analysis should focus on the joint density of cost and effect differences to quantify uncertainty surrounding the incremental cost-effectiveness ratio (ICER) and present these data on cost-effectiveness acceptability curves (CEACs). Using two clinical examples (one on implantable cardioverter defibrillators for patients at risk of ventricular arrhythmia and the other on two forms of heparin [one hospital based and the other administered at home] for deep vein thrombosis [DVT]), they showed that CMA is very rarely an appropriate method of analysis when sampled data on costs and effects are available and concluded that it is only appropriate when a randomised controlled trial (RCT) has been designed to test the explicit hypothesis of equivalence between two therapies. The authors went on to say that 'the more comprehensively one defines outcome or effectiveness, the less likely that equivalence between treatments will be established'.[3] Following that paper, Johnston et al highlighted the importance of conducting CEA on final (rather than intermediate) endpoints regardless of statistical significance, using an empirical example in which the intermediate endpoint was statistically significant but the final endpoint (life years) was not.[8] By the third and current edition of Drummond in 2005, CMA was no longer regarded as a full form of economic evaluation and was considered inappropriate for most situations,[7] with the 2001 Briggs and O'Brien paper[3] being cited within the rationale against use of CMA. CMA has therefore gone from being considered a full form of economic evaluation, to now being seen as an incomplete form of analysis to be used only in very limited situations.

The concept of 'irrelevance of inference'[9] provides additional arguments against CMA. Claxton argued that hypothesis testing is essentially arbitrary and irrelevant to decision-making; instead, he argued that the treatment with the highest expected net benefits should be used regardless of the uncertainty around this decision or whether the difference in efficacy (or net benefits) reaches conventional levels of statistical significance; instead of influencing decisions about which treatment we should use now, the uncertainty around the point

estimate should only guide decisions on what future research is required.[9] This argument further undermines the idea of using sequential hypothesis testing to decide on the appropriate form of analysis and suggests that the decisions should be based on the observed differences in effectiveness and cost regardless of whether they reach statistical significance or whether the trial is powered to show equivalence or non-inferiority. Furthermore, the role of uncertainty in decisions about further research raises the question of whether measures of uncertainty or value of information (VoI) are biased by use of CMA.

1.3. Aims

The aim of this paper is to assess whether the use of CMA has changed since the 'death of CMA' paper[3] was published and to explore additional arguments for and against conducting full economic evaluations of non-inferiority trials and those observing no significant difference in efficacy. In particular, we aim to assess whether CMA produces biased estimates of uncertainty measures and whether there are any situations in which CMA is appropriate.

2. Methods

2.1. Literature search on prevalence of CMA

In order to assess the impact that the Drummond[7] and Briggs & O'Brien publications[3] have had on the use of CMA in recent years, we examined how the proportion of economic evaluations that are CMAs has changed over the last 10 years. We investigated this by searching the NHS HEED database (including DARE, NHS EED and HTA) from 1998 to 2008, using the separate search terms 'cost-effectiveness analysis', 'cost-utility analysis' and 'cost-minimisation analysis'.

2.2. Empirical study

2.2.1. Premise

In order to assess the impact of conducting CMA instead of CEA within non-inferiority trials and those observing no significant difference, we compared the point estimates and uncertainty estimates that would be generated by CMA and CEA of simulated data (Section 2.2.2) and three clinical trials (Section 2.2.3, Table 1).

Table 1: Summary of the datasets used in the study

No.	Data source	Concept illustrated
1	3 simulated datasets	Identification of the situations when there may be a significant difference in net benefits despite no significant differences in costs or effects
2	SUIT: RCT conducted by the authors comparing 3 forms of infertility care (n=580)	Illustrates irrelevance of inference and potential for differing conclusions between CMA and CEA
3	GNOME[10]: Placebo-controlled RCT conducted by the authors (n=217)	Illustrates a case where irrelevance of inference arguments may not hold
4	3 simulated datasets	Highlight situations where bias in uncertainty estimates from CMA is particularly large
	3 simulated datasets	Highlight situations where bias in uncertainty estimates from CMA is minimal
5	Briggs & O'Brien DVT example[11] [12]: Published RCT comparing two forms of heparin (n=300)	Real-life example of one situation (simulation 4d in Table 3), where bias within CMA is minimal and CMA may be appropriate

Although any external evidence should generally be combined with evidence from the trial in question (where it exists), we assume within each of the examples presented here that the trial in question represents the only available evidence on incremental costs/efficacy (as is the case for SUIT and GNOME). Allowing for any external evidence would have increased analytical complexity, but would not have affected the conclusions regarding bias.

The general premise behind CMA is the assumption that the difference in effectiveness is equal to zero with no uncertainty. Subsequently, within CMA, the incremental net benefit (INB) from giving the new treatment rather than its comparator equals zero minus the difference in cost, regardless of ceiling ratio (R_c).

Within CMA, the probability that INB is positive (i.e. that treatment is cost-effective) is therefore equal to the probability that treatment is less costly than its comparator. Although CMA implicitly assumes that incremental efficacy is known to be exactly zero (making collection of additional efficacy data unnecessary), collection of additional cost data (either through trials or other methods) may nonetheless be efficient; subsequently the VoI for CMA can be calculated in the same way as for CEA.

2.2.2. Generation and analysis of simulated data

Simulated data were generated using Microsoft Excel 2003 to compare the results of CMA and CEA in different situations. Each hypothetical example was assigned values for: the incremental cost (Δ_C) and incremental QALYs (Δ_E) for treatment versus its comparator; the standard errors of their means (SEM, $\sqrt{\text{var}(..)}$); and the correlation coefficient (ρ) for the

relationship between incremental costs and incremental QALYs. The parameters for each example were chosen to illustrate particular points and to explore how the results of CMA and CEA differ depending on the direction and magnitude of the point estimate, the uncertainty around costs and the correlation between costs and effects.

These values were used to generate the simulated data and calculate mean INB and its SEM using Equations 1-3

$$INB = R_c \quad \Delta_E - \Delta_C \tag{1}$$

$$\mathrm{cov}(\Delta_E, \Delta_C) = \rho \cdot \sqrt{\mathrm{var}(\Delta_E) \cdot \mathrm{var}(\Delta_C)} \tag{2}$$

$$SEM(INB) = \sqrt{\mathrm{var}(INB)} = \sqrt{R_c^2 \cdot \mathrm{var}(\Delta_E) + \mathrm{var}(\Delta_C) - 2 \cdot R_c \cdot \mathrm{cov}(\Delta_E, \Delta_C)} \tag{3}$$

To generate scattergraphs on the cost-effectiveness plane, 1,000 correlated pairs of incremental costs and benefits comparable to bootstrap replicates were simulated in Microsoft Excel 2003 using the NORMINV function and the CORAND array function within the "Simtools" plug-in developed by the Decision Analysis Society.[13]

Since the simulated pairs of incremental costs and benefits were assumed to be normally distributed, the probability of treatment being cost-effective at the ceiling ratio R_c ($Pr(INB>0 \mid R_c)$) and the probability of treatment being cost-saving ($Pr(\Delta_C<0)$) were calculated analytically based on a cumulative normal distribution (Φ)

$$\Pr(\mathrm{INB} > 0 \mid \mathrm{R_c}) = \Phi\left(\mathrm{INB}_{\mathrm{R_c}} / \sqrt{\mathrm{var}(INB_{R_c})}\right) \tag{4}$$

$$\Pr(\Delta_C < 0) = \Phi\left(\Delta_C / \sqrt{\mathrm{var}(\Delta_C)}\right) \tag{5}^{c}$$

These values were calculated for a range of ceiling ratios to generate CEACs. The error probability was used to quantify the risk associated with decisions in a way that can be compared across examples, regardless of whether the new treatment is superior or inferior to its comparator. The error probability was defined as the probability that the treatment with the highest net benefits (or lower costs) would in fact generate lower net benefits (or higher costs) than its comparator. When INB was negative (or $\Delta_C>0$), the error probability equaled the probability that treatment was cost-effective (or cost-saving). However, when INB was

[c] Within Microsoft Excel, Equations (4) and (5) were formulated as
=1-NORMDIST(0,INB,SQRT(varINB),TRUE) (4)
=NORMDIST(0,dC,SQRT(vardC),TRUE) (5)

positive (or $\Delta_C<0$), the error probability equaled one minus the probability of treatment being cost-effective (or cost-saving).

The expected value of perfect information (EVPI)[9 14] was also calculated to explore how the value of future research differs between CMA and CEA and assess whether use of CMA could cause decisions on further research to be based on biased information. EVPI per patient was calculated analytically based on methods described by Willan and Pinto[14]:

$$EVPI = \left(\text{var}(INB)/2\pi\right)^{1/2} \cdot \exp\left(-INB^2/2\,\text{var}(INB)\right) - |INB|\left(\Phi(-|INB|/\sqrt{\text{var}(INB)}\right) \quad (6),$$ [d]

where INB for CMA equalled $-\Delta_C$. For all simulations, population EVPI was based on 50,000 patients benefiting from treatment each year for the next 10 years; values in future years were discounted at 3.5% per annum.

2.2.3. Case studies

The Scottish Unexplained Infertility Trial (SUIT)[15] was a three-arm, parallel group, pragmatic RCT comparing two treatments for women with unexplained infertility: intrauterine insemination (IUI, which could enhance pregnancy rates by helping to overcome the cervical barrier) and oral clomifene citrate (a drug which is believed to correct subtle ovulatory dysfunction). These treatments were compared against expectant management, which is effectively a 'do nothing' arm. The trial included 580 couples attending fertility clinics across five hospitals in Scotland. Inclusion criteria included infertility for over two years, confirmed ovulation, patent fallopian tubes, and motile sperm. A health service perspective was used and the main outcome measure was live births. Cost-effectiveness was assessed using CEA, although a CMA was initially planned.

GNOME comprised a double-blind, placebo-controlled RCT evaluating topical intranasal corticosteroids with placebo in the treatment of 217 children with otitis media with effusion (OME, or "glue ear") in primary care.[10] Cost-effectiveness was assessed using both cost-utility analysis (CUA) and a CEA that calculated the cost per additional child cured, although the current analysis focuses on the results of CUA.

[d] Within Microsoft Excel, EVPI was calculated using the equation =((varINB/(2*PI()))^0.5)*EXP(-(INB^2)/(2*varINB))-ABS(INB)*NORMDIST(-ABS(INB)/SQRT(varINB),0,1,TRUE)

For both case studies, non-parametric bootstrapping was used to generate uncertainty estimates since total and incremental costs and effects were skewed and differed significantly from normality. For SUIT, 1,000 bootstrap replicates were drawn from the trial data; SEMs around incremental costs, incremental effects and INB were based on the standard deviations (SD) across bootstrap replicates. However, within GNOME (as described previously[10]), missing data were imputed using multiple imputation[16] and bootstrapping was conducted separately on the resulting five imputed datasets, with SEMs being calculated using the Rubin equation.[16]

In both studies, CEACs were based on the proportion of all bootstrap replicates with negative Δ_C or positive net benefits at different ceiling ratios. Similarly, the EVPI/patient[9] was estimated numerically by subtracting the total net benefit for the option we would choose based on current information from the maximum net benefit we would obtain with perfect information (the average of the maximum net benefit for each bootstrap replicate). Within SUIT, population EVPI assumed that 273,401 couples in the UK could benefit from treatment each year, while population EVPI estimates for GNOME assumed that 16,068 children could benefit each year; within both studies, future values were discounted at 3.5% per annum and a 10-year time horizon was used for EVPI.

In addition to the two case studies with patient-level data, the DVT example cited by Briggs and O'Brien[3] was re-analysed based on the mean incremental costs and benefits reported in the paper and their standard errors; since the correlation between costs and effects was not reported, this was assumed to be zero. Within EVPI calculations, 61,000 people in England (0.1%[17]) were assumed to require DVT treatment each year.

3. Results

3.1. Literature search

Our search of the NHS EED database showed that CMAs are still being performed since the publication of the Death of CMA paper[3] and 2005 Drummond book,[7] although their numbers appear to be dropping slightly compared to CEA, especially since 2005 (Table 2). However, the observed reduction is not as great as we would expect given the strong arguments against CMA that are made in these widely used publications.

Table 2: Number of CMAs and other economic evaluations in NHS EED by year

Year	All economic evaluations	CEA		CUA		CMA	
	No. studies	No. studies	%	No. studies	%	No. studies	%
1998	1,756	645	37	89	5	42	2.4
1999	1,804	571	32	81	4	65	3.6
2000	1,875	650	35	99	5	61	3.3
2001	2,095	647	31	94	4	51	2.4
2002	2,318	654	28	129	6	59	2.5
2003	2,079	601	29	131	6	38	1.8
2004	2,008	641	32	175	9	58	2.9
2005	2,271	730	32	215	9	61	2.7
2006	2,417	643	27	237	10	29	1.2
2007	2,190	592	27	257	12	21	1.0
2008	1,715	492	29	213	12	16	0.9
Total	**22,528**	**6,866**	**30**	**1720**	**8**	**501**	**2.2**

3.2. Empirical study

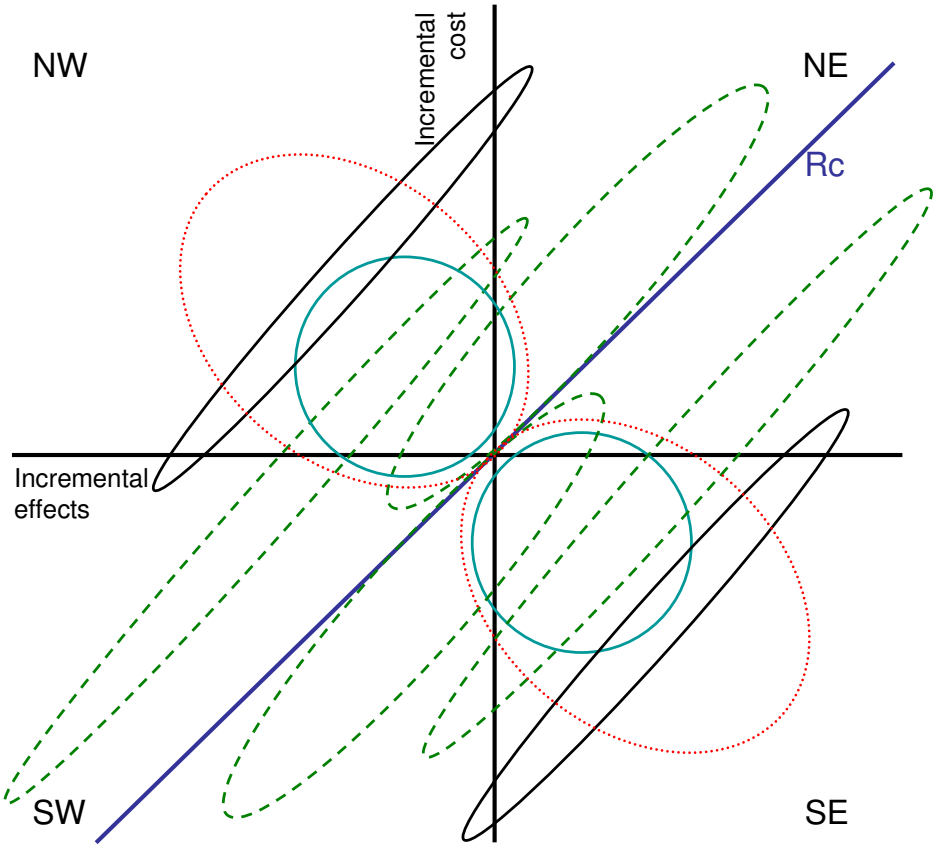
3.2.1. Significantly positive INB despite no significant difference in costs or effects

As described previously,[18] one argument for conducting CEA on studies finding no significant difference in efficacy is that it is possible for net benefits to differ significantly between treatments despite no significant differences in costs or effects. We therefore used simulated data to identify the situations when this can occur.

This analysis showed that results of this type are most likely to occur when the point estimate lies within the northwest (NW, dominated) or SE (dominant) quadrants of the cost-effectiveness plane (Figure 1). When the point estimate lies in the NW or SE quadrants, significant differences in net benefit can occur despite non-significant differences in costs or effects regardless of whether the correlation between incremental costs and effects is positive (black elipse), negative (red elipse with dotted lines) or if there is no correlation (blue circle). If we considered inference to be *relevant* to decision-making and adopted treatments only if they showed significantly higher net benefits than their comparator, CEA evidence showing the INB to be significantly greater than zero, would lead us to adopt a new treatment that dominated its comparator; by contrast, CMA evidence showing non-significantly lower costs would be considered insufficient evidence to adopt the new treatment. However, if we considered inference to be irrelevant, both CMA and CEA would be considered sufficient to adopt the new treatment in these cases (since the expected INB and expected net savings are both positive), although CMA and CEA may produce different estimates of the value of collecting further information (Table 3, example 1a).

By contrast, when the point estimate lies in the NE or SW quadrants, significant differences in net benefits without significant costs or benefits can only occur when costs and effects have a strong, positive correlation and large variance (green elipses with dashed lines). In these cases, the conclusions of CEA and CMA may differ even if we consider inference to be irrelevant to decision-making. For example, if treatment costs an additional £300/patient (SEM: £600) and gains one (SEM: 0.515) additional QALY (Table 3, example 1b), CUA will find treatment to be cost-effective with 98% certainty at a £20,000/QALY ceiling ratio, whereas CMA will find treatment to be less costly with 31% certainty. However, situations such as this are likely to be rare in practice.

Figure 1: Hypothetical cases where net benefits differ significantly despite no significant difference in costs or benefits. Each ellipse represents a 95% confidence ellipse.



3.2.2. SUIT: Irrelevance of inference

SUIT found IUI to be significantly more costly than expectant management ($p<0.001$) and resulted in six additional live births per 100 women treated (bootstrap p-value: 0.09; Figure 2). Clomifene citrate was strongly dominated by both IUI and expectant management, being more costly and less effective. For brevity, only the results for the comparison between IUI and expectant management are shown here.

Table 3: Summary of the case studies and quantification of the bias associated with CMA.

Case study	Δ cost/pt (SEM)	Δ effect/pt (SEM)†	Covariance (correlation coefficient)	INB (SEM)‡	Error probability‡		Population EVPI‡ (millions)		Optimal treatment		Direction/ magnitude of bias
					CUA	CMA	CUA	CMA	CUA	CMA	
No significant difference in costs or effects, but significant difference in net benefits											
1a: NW quadrant, +ve correlation§	£500 (£290)	-0.5 (0.27)	31.32 (0.4)	-£10,500 (£5,291)*	2.36%	4.23%	£20.14	£2.15	C	C	VoI lower, Pr(error) higher for CMA
1b: NE quadrant, +ve correlation, +ve INB§	£300 (£600)	1 (0.515)	185.4 (0.6)	£19,700 (£9,952)*	2.39%	30.85%	£38.40	£51.08	T	C	VoI and Pr(error) higher for CMA
1c: NE quadrant, +ve correlation, -ve INB§	£20,000 (£12,000)	0.003 (0.515)	3,708 (0.6)	-£19,940 (£10,088)*	2.40%	4.78%	£39.24	£102.40	C	C	VoI and Pr(error) higher for CMA
Case studies											
2: SUIT (IUI v. expectant management)	£319 (£19)*	0.06 (0.04) live births	-0.18 (-0.24)	£239 (£427)	28%	0%	£184.46	£0.36	T	C	VoI and Pr(error) negligible for CMA, but high for CEA
3: GNOME (CUA)[10]	£11 (£107)	-0.017 (0.024)	-0.163 (-0.063)	-£344 (£509)	24.19%	46.25%	£24.24	£13.64	C	C	VoI lower, Pr(error) higher for CMA
Additional simulations											
4a: +ve ΔE, sig +ve ΔC§	£500 (£100)*	1 (0.8)	56 (0.7)	£19,500 (£15,930)	11.05%	0.00%	£366.02	£0.00	T	C	VoI and Pr(error) lower for CMA
4b: small +ve ΔE, ΔC≈0§	£0.01 (£300)	0.2 (0.11)	9.9 (0.3)	£4,000 (£2,129)	3.02%	50.00%	£10.71	£51.51	T	C	VoI and Pr(error) higher for CMA
4c: large +ve ΔE, ΔC≈0§	£0.01 (£300)	3 (1.65)	148.5 (0.3)	£60,000 (£32,911)	3.41%	50.00%	£190.75	£51.51	T	C	VoI lower and Pr(error) higher for CMA
4d: Very large ΔC and highly significant, NE quadrant§	£4,500 (£450)*	0.06 (0.03)	0 (0)	-£3,300 (£750)*	0.00%	0.00%	£0.00	£0.00	C	C	Bias negligible
4e: Very large and highly significant ΔC, SE quadrant§	-£6,250 (£50)*	0.25 (0.1385)	0 (0)	£11,250 (£2,750)*	0.00%	0.00%	£0.01	£0.00	T	T	Bias negligible
4f: ΔC and ΔE both exactly equal to 0§	£0 (£100)	0 (0.25)	0 (0)	£0 (£5,001)	50.00%	50.00%	£859.66	£17.17	Equivalent		No bias in Pr(error). VoI underestimated
Published clinical trials											
5: DVT trial from Briggs & O'Brien[3]	$3045 ($520)*	0.0057 (0.029) thromboembolic events	Not reported (assumed to be 0)	-$2,988 ($598)*	0.00%	0.00%	$0.00	$0.00	C	C	Bias negligible

* $p<0.05$

† Effects are shown in QALYs unless otherwise stated.

‡ For all CUA, net benefits, EVPI and the error probability were based on a ceiling ratio of £20,000/QALY gained. For SUIT, a ceiling ratio of £10,000/live birth was used,[19] while a ceiling ratio of CAN$10,000/thromboembolic event was used for the DVT example reported by Briggs and O'Brien.[3]

§ Simulated data.

Although the probability of being cost-effective never exceeds 87% (below conventional levels of significance), our point estimate of £5,604/live birth is substantially below the value placed on this outcome in stated preference studies (£10,000-£1 million[19 20]). Based on statistical inference, we could infer that IUI is significantly more costly than expectant management, but that there is insufficient evidence to consider it cost-effective. However, if we consider inference to be irrelevant, we would conclude based on CEA that IUIs are extremely cost-effective compared with expectant management, whereas CMA would give the opposite conclusion (suggesting that IUIs should not be used due to their higher costs). Furthermore, CMA and CEA produce very different estimates of the value of collecting further evidence, with population EVPI ranging from £0.4 million for CMA to £184 million for CEA at a £10,000/birth ceiling ratio.

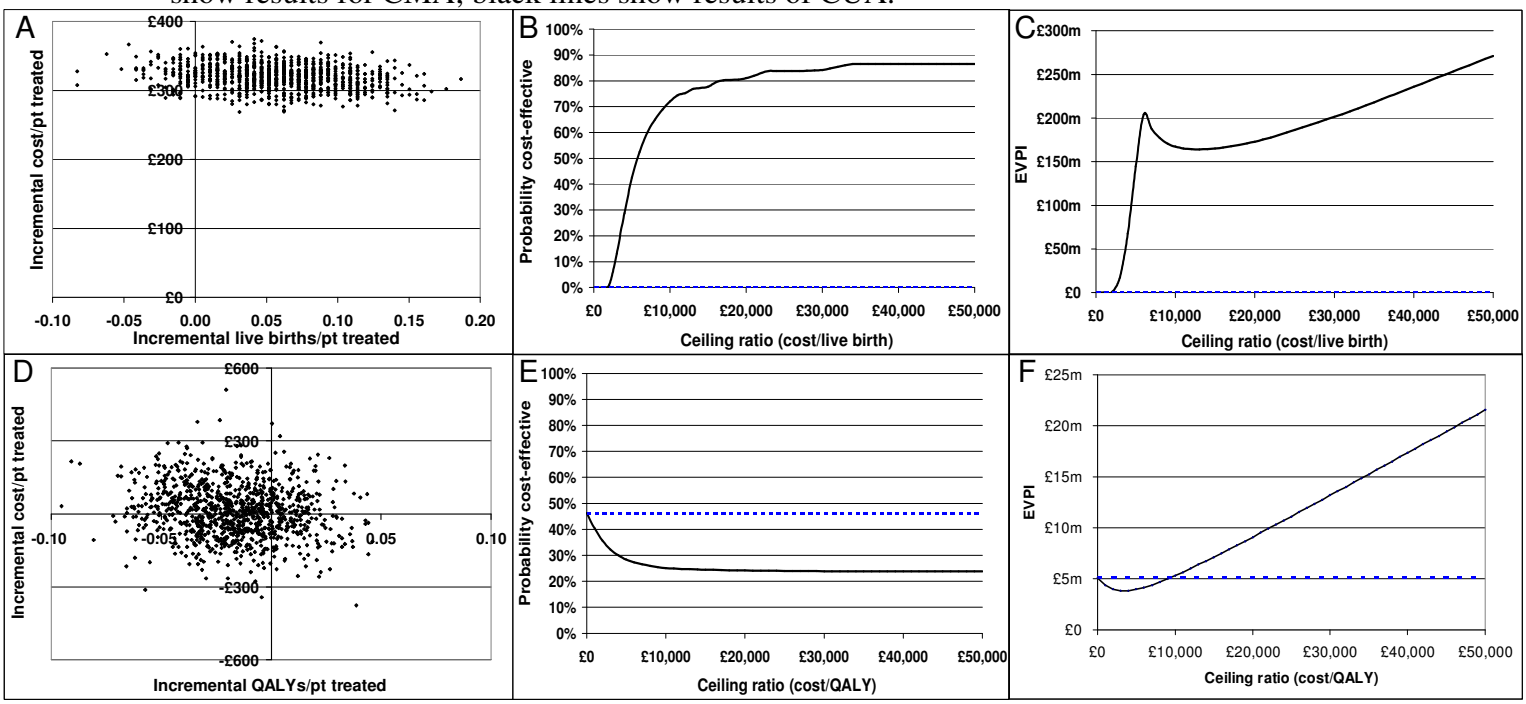
3.2.3. GNOME: A situation where inference may not be irrelevant

The GNOME trial observed no significant differences between steroids and placebo for costs, QALYs, utilities, side-effects, the primary clinical endpoint (cure at one month) or any secondary outcome measure.[10] CUA showed steroids to be dominated by placebo, generating non-significantly fewer QALYs and non-significantly higher costs (Figure 2). Although patients receiving steroids accrued slightly higher costs in all sensitivity analyses, incremental effectiveness varied substantially depending on the endpoint used. In particular, the base case CEA found the proportion of children cured at either one or three months to be non-significantly higher for steroids and found steroids to cost £347 per additional child cured.

On an 'irrelevance of inference' basis, the conclusion from the GNOME CUA (that no treatment dominates steroids) matched those of CMA, although CMA would nonetheless have overestimated the error probability and underestimated VoI at a £20,000/QALY threshold (Table 3). However, the GNOME CEA highlights a situation where inference may not be entirely irrelevant. Although the CEA results suggest that steroids are cost-effective at a £1,000/cure threshold, it may not be appropriate to license a drug or recommend it as a cost-effective treatment option on the basis of this evidence, since we can be (at best) only 65% certain that treatment increases the chance of cure *compared with placebo*, there is a huge degree of uncertainty around incremental costs and benefits and the results are highly sensitive to changes in outcome measure or assumptions. However, the costs associated with licensing and/or adopting a new treatment and (if necessary) reversing the decision based on new evidence may be explicitly captured within the economic evaluation or VoI

calculations,[21 22] enabling decisions about adoption and further evidence to be based on the most realistic estimates of expected costs and benefits, without relying on hypothesis testing.

Figure 2: Results of SUIT and GNOME studies. Results of SUIT are shown in panels A-C: (A) Cost-effectiveness plane, (B) CEAC, (C) EVPI. Results of GNOME CUA and CMA are shown in panels D-F: (D) Cost-effectiveness plane, (E) CEAC, (F) EVPI. Blue dashed lines show results for CMA; black lines show results of CUA.



3.2.4. Bias within uncertainty estimates

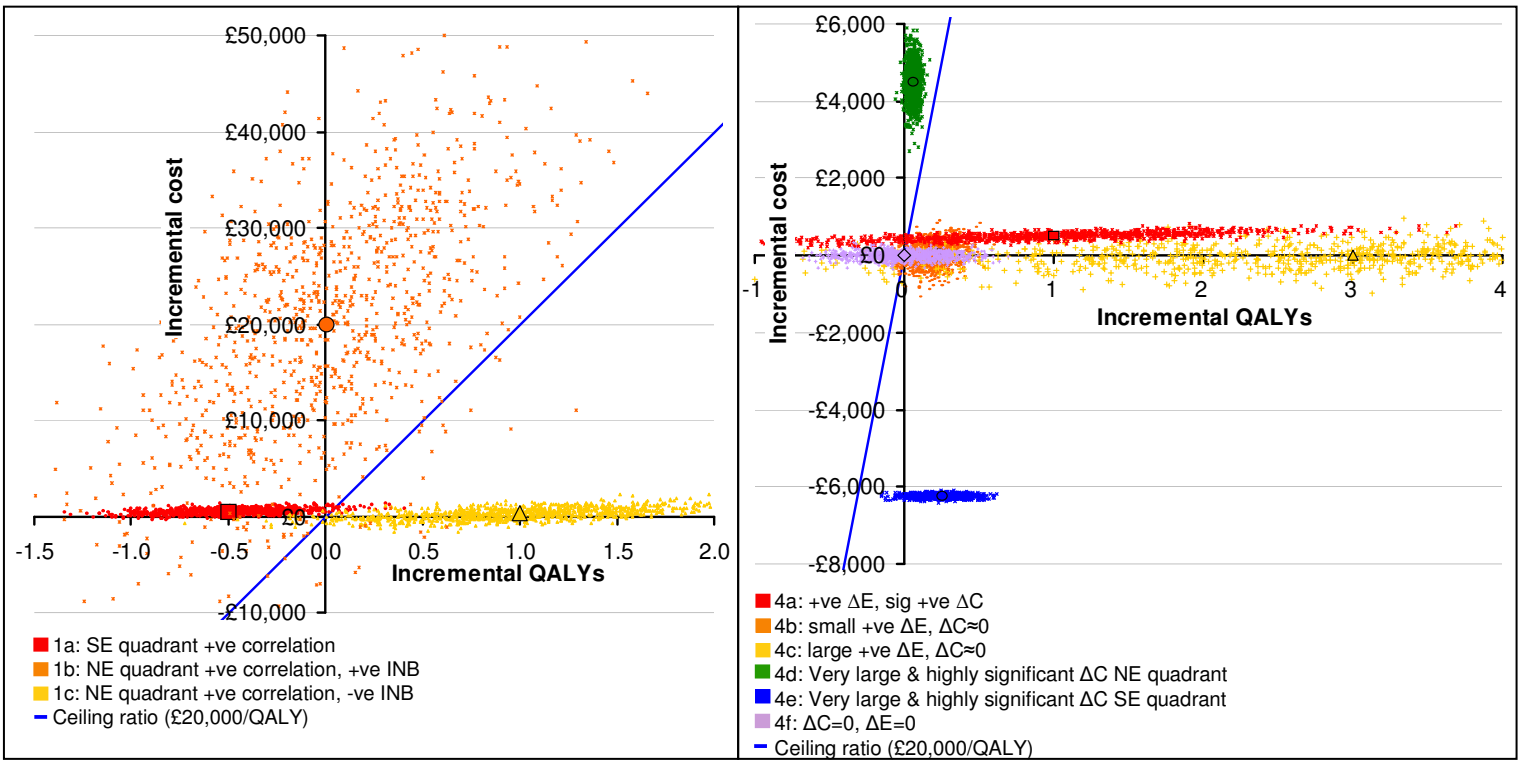
The examples described in Sections 3.2.1-3.2.3 demonstrate that in addition to producing different conclusions in several cases, CMA produces estimates of the error probability and EVPI that may be substantially higher or lower than those of CEA (Table 3). Since CMA comprises a simplification of CEA that makes the implicit assumption that incremental effects equal zero with no uncertainty, the uncertainty estimates generated using CEA must represent the "correct" values. Subsequently, the estimates of error probability and EVPI for CMA represent biased estimates of uncertainty.

A number of other simulated examples were explored to identify the situations that maximise or minimise the bias within CMA (Table 3, Figure 3). CMA and CEA were found to produce different conclusions regarding the optimal treatment when the results lie in one of two areas of the cost-effectiveness plane: (a) when the point estimate lies in the NE quadrant and INB is positive; and (b) when the point estimate lies in the SW quadrant and INB is negative.

However, CMA and CEA may produce substantially different estimates of the error probability and EVPI regardless of quadrant.

Furthermore, CMA may overestimate the error probability while underestimating the EVPI (or vice versa) as these two measures assess different aspects of uncertainty. This can be seen in the differing results of Examples 4b and 4c, where a 15-fold increase in ΔE and its SEM increases the EVPI for CUA by 18-fold but has minimal effect on the error probability. Whereas the error probability shows the risk of making the wrong decision (e.g. the chance that the true INB is negative or the chance that the true ΔC is positive), EVPI shows the expected net benefits lost through making the wrong decision. Mathematically, the error probability varies with the ratio of SEM/mean, whereas the EVPI is affected by both the ratio of SEM/mean and the absolute magnitude of INB or incremental costs.

Figure 3: Cost-effectiveness planes showing results for the simulated case studies shown in Table 3



Since perfect information is unattainable, EVPI reflects only the theoretical maximum that we should be willing to pay for information and actual decisions on future research should be based on the expected value of *sample* information (EVSI). We therefore also estimated the EVSI for each example using the analytical methods developed by Willan and Eckermann,[14] [23-25] in order to confirm the findings from EVPI analyses. This demonstrated that CMA also

produces biased estimates of EVSI whenever EVPI estimates are biased and that EVSI followed the same trends as EVPI, with CMA overestimating the EVSI and the optimal number of patients within clinical trials whenever EVPI is overestimated (and vice versa). The results of these analyses and further details on the methods used are available from the authors on request.

Although uncertainty estimates from CMA will always be biased, the bias was found to be minimal in cases where the error probability and VoI both approach zero. For studies with no significant difference in effects, this can occur when the difference in cost is so large compared with its SEM and the likely difference in efficacy that there is negligible chance of the more costly treatment being cost-effective. This result may arise in any of the four quadrants, even if the difference in efficacy is not statistically significant (Table 3, Examples 4d-e). Since CMA produces biased estimates of uncertainty in all other situations, cases such as this are likely to comprise the only situations when it is may be valid to use CMA, although even in these cases, CMA may not be appropriate if the trial is not powered to exclude a clinically significant difference in efficacy (i.e. if was not designed to test equivalence or non-inferiority) or if there is no clear reason to expect the treatments to have equivalent efficacy.

CMA was found to provide an unbiased estimate of the error probability in cases where incremental costs and benefits equal exactly zero and are symmetrically distributed (Table 1, Example 4f). However, CEA remains the most appropriate form of analysis in these situations, as CMA may nonetheless produce highly biased estimates of EVPI.

Although CMA may be appropriate in situations like Example 4d, in practice, we cannot normally pre-specify the form of analysis since the magnitude of costs and benefits is generally unknown until trial results are collected. However, we may be able to predict results like those of Example 4d for some non-inferiority trials in which clinical equivalence is expected and in which there are substantial differences in treatment costs that are likely to overwhelm any differences in efficacy or other costs. One example of this is the trial of anticoagulants for DVT that was cited by Briggs and O'Brien,[3] which we found to have $0.00 population EVPI and a 0.00% error probability for both CMA and CUA at a ceiling ratio of $10,000/thromboembolic event avoided (Table 3, Example 5b). However, even in this situation, CMA would produce substantial biases at very high ceiling ratios.

For non-inferiority studies where the likely difference in costs is anticipated to drive conclusions regardless of efficacy results, it may be appropriate to pre-specify a non-inferiority margin for QALYs (similar to that used in power calculations on clinical endpoints), whereby CMA will be conducted unless the more costly treatment is found to generate more than X QALYs more than the less costly treatment. The optimal value for X in these situations is debatable. It may be based on the minimally important difference for the outcome measure used, or the minimum difference in efficacy that would need to be observed to bias uncertainty estimates appreciably. Alternatively, for CUA, X could be based on the smallest difference in utility that could generally be measured for any given participant in the time trade-off methods used within the MVH survey (three months out of 10 years, or 0.025 QALYs per patient-year[26]).

4. Discussion

As the debate concerning the choice of economic evaluation in clinical trials continues, our results show that CMA is still used and that choosing CMA over CEA may lead to either overestimation or underestimation of uncertainty: even for non-inferiority trials. Although bias will be negligible for non-inferiority studies comparing treatments that differ enormously in cost, it is generally necessary to collect and analyse data on both costs and efficacy (including utilities) to assess this bias.

Whereas previous research suggested that CMA may be appropriate for all non-inferiority studies, our analyses show that CMA is only appropriate for a subset of such studies – namely those where the difference in costs is sufficiently large that no plausible difference in efficacy could change the conclusions or uncertainty estimates. In all other situations, it is appropriate to conduct CEA in order to produce valid conclusions about cost-effectiveness and unbiased estimates of the uncertainty around decisions. Although it may be easier and/or less costly to conduct CMA, once costs and benefits have been collected in order to assess differences in cost and efficacy, conducting CEA will require little additional work but will avoid bias and produce additional outputs (e.g. utilities or VoI) that may be useful for future research.

4.1. Limitations

Within this study, we have explored the degree of uncertainty within a small number of real and hypothetical examples chosen either by convenience (access to patient-level data) or to illustrate situations where the bias within CMA is particularly large or particularly small. The level of bias observed within these examples may therefore not be typical of economic evaluations in general and some of the examples cited (most notably examples 1b-c) may be unlikely to occur in practice. However, we believe that the general conclusion that CMA may (and usually will) produce biased estimates of uncertainty is likely to apply to economic evaluations in general.

For all of the examples described here, uncertainty estimates are based on individual trials. While this may be appropriate for SUIT and for GNOME, where no previous trials have compared the same molecule/intervention in comparable populations, in general it is necessary to take account of all available evidence in decisions about which treatment to adopt and whether further evidence is required. Although taking account of all available evidence may reduce uncertainty and the likelihood that there is no significant difference in efficacy, this will not prevent CMA from producing biased uncertainty estimates in cases where the difference in efficacy remains non-significant after taking account of additional evidence.

4.2. Why publish economic evaluations of trials finding no significant difference?

Despite the arguments for conducting CEA on non-inferiority studies and those finding no significant difference in effects, editors and reviewers sometimes question the value of publishing economic evaluations of such studies. We believe that it is essential for economic evaluations of both "positive" and "negative" trials to be published to reduce publication bias[8] within systematic reviews of economic evaluations (such as those conducted for NICE). Furthermore, publication enables data on costs, utilities and clinical outcomes from "negative" trials to be used by other researchers within decision-analytical models, healthcare budgeting or international comparisons; since the cost of conducting the trial has already been incurred, it is appropriate to make full use of the information collected. Additionally, VoI calculations may inform future research priorities: if the value of additional information is found to be low, publication of this fact will prevent future research time/funds being wasted on answering the same question, while publishing the results of a trial that shows the VoI to be high will provide evidence to support the rationale for new trials.

4.3. Is it always efficient to collect economic evidence?

Funding bodies such as the HTA now require economic evaluations in every trial that they conduct. However, if it were ***known*** *a priori* that the difference in costs was substantial and that no plausible difference in efficacy could make the more costly treatment cost-effective, there may be a case for relying on approximate costs rather than conducting a comprehensive analysis to estimate costs and effects accurately. An obvious question there is whether it is ever possible to be certain in advance what the difference in costs and effects will be. Even if drug costs are known and are predicted to account for the majority of costs, actual trial results may be very different as drug costs may be offset by differences in resource use resulting from differences in efficacy, safety or disease management. Delaying collection of cost data until the trial results are known is likely to increase data collection costs and reduce the accuracy of cost estimates, as well as preventing prospective collection of stochastic cost data for the trial population. By contrast, prospectively adding an economic evaluation to a trial that is already planned is relatively cheap and provides a large amount of data that can be used in future economic evaluations and for budgetary purposes even if the trial results show equivalence.

4.4. Conclusions

Our paper has revisited the debate concerning the use of CMA in economic evaluation, specifically in the context of non-inferiority trials and those observing no significant difference in efficacy. Conducting CMA rather than CEA introduces biases in uncertainty estimates that can be either positive or negative. These biases provide an argument for conducting full economic evaluations of non-inferiority trials as well as those powered only to detect differences between treatments. However, such biases will be negligible when the difference in cost is so large that no plausible difference in efficacy could change the conclusions. In non-inferiority studies of this type only, there may be a case for the continued use of CMA.

4.5. Discussion points for HESG audience

Is the battle against CMA yet won? Have HESG members encountered reviewers or editors who feel that there is no point in publishing trials with no significant difference in efficacy?

Are there any other situations when uncertainty estimates from CMA include minimal bias?

Is inference always irrelevant?

Is it appropriate to extrapolate the results of trials finding no significant differences in the within-trial analysis?

Is CMA appropriate for model-based economic evaluations?

Is VoI analysis appropriate or necessary following CMA or even for CEA?

Acknowledgements

We would like to thank Prof James Raftery for his input on the abstract and useful discussions on the ideas raised in the paper and Dr Helen Campbell, Dr Stavros Petrou and Professor Alastair Gray for comments on this paper. We thank James Buchanan for his analysis of the SUIT data and Giselle Abangma and Stavros Petrou for their role in GNOME. We would also like to thank Dr Andy Willan for developing the spreadsheet used in supplementary analyses of EVSI, Ian Williamson and the GNOME trial group for allowing us to present GNOME data here and Professor Bhattacharya and the SUIT trial group for allowing us to present SUIT data. SW is funded by an NIHR fellowship award and HD is funded by several HTA studies. Any views expressed in this paper are those of the authors alone and not any funding body.

References

1. Donaldson C, Hundley V, McIntosh E. Using economics alongside clinical trials: why we cannot choose the evaluation technique in advance. *Health Econ* 1996;5(3):267-9.
2. Drummond MF, O'Brien B, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes*. 2nd ed: Oxford University Press, 1997.
3. Briggs AH, O'Brien BJ. The death of cost-minimization analysis? *Health Econ* 2001;10(2):179-84.
4. Drummond MF, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes*. 1st ed: Oxford University Press, 1987.
5. National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisal, June 2008.
6. Devlin N, Parkin D. Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. *Health Econ* 2004;13(5):437-52.
7. Drummond MF, Sculpher MJ, Torrance GW, O'Brien B, Stoddart GL. *Methods for the economic evaluation of health care programmes*. 3rd ed. Oxford: Oxford University Press, 2005.
8. Johnston K, Gray A, Moher M, Yudkin P, Wright L, Mant D. Reporting the cost-effectiveness of interventions with nonsignificant effect differences: example from a trial of secondary prevention of coronary heart disease. *Int J Technol Assess Health Care* 2003;19(3):476-89.
9. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *J Health Econ* 1999;18(3):341-64.

10. Williamson I, Benge S, Barton S, Petrou S, Letley L, Fasey N, et al. A double-blind randomised placebo controlled trial of topical intra-nasal corticosteroids in 4-11 year old children with persistent bilateral otitis media with effusion in primary care. *Health Technology Assessment* 2009;13(37).
11. Briggs AH, Gray AM. Handling uncertainty in economic evaluations of healthcare interventions. *Bmj* 1999;319(7210):635-8.
12. O'Brien B, Levine M, Willan A, Goeree R, Haley S, Blackhouse G, et al. Economic evaluation of outpatient treatment with low-molecular-weight heparin for proximal vein thrombosis. *Arch Intern Med* 1999;159(19):2298-304.
13. SIMTOOLS add-in for Microsoft Excel (Available from: http://www.kellogg.northwestern.edu/faculty/myerson/ftp/addins.htm) [program], 2001.
14. Willan AR, Pinto EM. The value of information and optimal clinical trial design. *Stat Med* 2005;24(12):1791-806.
15. Bhattacharya S, Harrild K, Mollison J, Wordsworth S, Tay C, Harrold A, et al. Clomifene citrate or unstimulated intrauterine insemination compared with expectant management for unexplained infertility: pragmatic randomised controlled trial. *BMJ* 2008;337:a716.
16. Briggs A, Clark T, Wolstenholme J, Clarke P. Missing... presumed at random: cost-analysis of incomplete data. *Health Econ* 2003;12(5):377-92.
17. O'Shaughnessy D, Miles J, Wimperis J. UK patients with deep-vein thrombosis can be safely treated as out-patients. *QJM* 2000;93(10):663-7.
18. University of Oxford Health Economics Research Centre. Advanced methods of cost-effectiveness analysis - University of Oxford Short Course in Health Economics. 1999-2009.
19. Granberg M, Wikland M, Nilsson L, Hamberger L. Couples' willingness to pay for IVF/ET. *Acta Obstet Gynecol Scand* 1995;74(3):199-202.
20. Garceau L, Henderson J, Davis LJ, Petrou S, Henderson LR, McVeigh E, et al. Economic implications of assisted reproductive techniques: a systematic review. *Hum Reprod* 2002;17(12):3090-109.
21. Eckermann S, Willan AR. Expected value of information and decision making in HTA. *Health Econ* 2007;16(2):195-209.
22. Eckermann S, Willan AR. The option value of delay in health technology assessment. *Med Decis Making* 2008;28(3):300-5.
23. Eckermann S, Willan AR. Time and expected value of sample information wait for no patient. *Value Health* 2008;11(3):522-6.
24. Willan AR. Clinical decision making and the expected value of information. *Clin Trials* 2007;4(3):279-85.
25. Willan A. Tutorial on ASSENT3: Spreadsheet tutorial. In: Eckermann S, Willan A, editors. *Health economics from theory to practice: Informing related decisions of reimbursement, research and regulation. A three-day workshop 25-27 August 2009*, 2009.
26. Williams A. The measurement and valuation of health: A chronicle. *The University of York Discussion paper 136* 1995.