Cost-minimisation analysis versus cost-effectiveness analysis, revisited

Running header:
CEA or CMA?

Authors:
Helen Dakin and Sarah Wordsworth

Authors' affiliation and institute where work was carried out:
Health Economics Research Centre, University of Oxford, UK

Corresponding author:
Helen Dakin
Health Economics Research Centre
Department of Public Health
Old Road Campus
Headington
Oxford
OX3 7LF
Email: helen.dakin@dphpc.ox.ac.uk
Tel: 01865 289422
Fax: 01865 289271

Funding:
Funding for the GNOME trial was provided by the UK National Institute for Health Research (NIHR) Health Technology Assessment (HTA) Programme and the SUIT trial was funded by Chief Scientist Office, Scotland. SW is funded by an NIHR fellowship award and HD is funded by several HTA studies.

Ethics:
SUIT received ethics approval from the multicentre research ethics committee, Scotland (Bhattacharya *et al*. 2008) and GNOME received ethics approval from the Metropolitan Multi-centre Research Ethics Committee and 60 local research ethics committees and research governance approval from 76 primary care trusts (Williamson *et al*. 2009b).

Conflict of interest:
None

This manuscript describes original unpublished work and has not been submitted for publication elsewhere, although an earlier draft of this paper was presented at the Health Economists' Study Group meeting in January 2010 at London School of Economics and Political Sciences.

Word count: 4,608

Cost-minimisation analysis versus cost-effectiveness analysis, revisited

Running header:
CEA or CMA?

Ethics:
SUIT received ethics approval from the multicentre research ethics committee, Scotland (Bhattacharya *et al*. 2008) and GNOME received ethics approval from the Metropolitan Multi-centre Research Ethics Committee and 60 local research ethics committees and research governance approval from 76 primary care trusts (Williamson *et al*. 2009b).

Word count: 4,608
Number of figures: 1
Number of tables: 3

ABSTRACT

Aims: We aim to establish whether it is ever appropriate to conduct cost-minimisation analysis (CMA) rather than cost-effectiveness analysis (CEA).

Methods: We perform a literature review to examine how use of CMA has changed since Briggs & O'Brien announced its death in 2001. Examples of simulated and trial data are presented: firstly to illustrate the advantages and disadvantages of CMA in the context of non-inferiority trials and those finding no significant difference in efficacy and secondly to assess whether CMA gives biased results.

Results: We show that CMA is still used and will bias measures of uncertainty, causing over- or under-estimation of the value of information and the probability that treatment is cost-effective. 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 costs and efficacy (including utilities) to assess this bias.

Conclusions: CEA (including evaluation of the joint distribution of costs and benefits) is almost always required to avoid biased estimation of uncertainty. The remit of CMA in trial-based economic evaluation is therefore even narrower than previously thought, suggesting that CMA is not only dead, but should also be buried.

Key words: Economic evaluation; equivalence trial; expected value of perfect information; randomised controlled trial; cost-utility analysis.

1. INTRODUCTION

Historically, cost-minimisation analysis (CMA) was recommended for economic evaluations of trials finding no statistically significant difference in effectiveness (Drummond *et al.*, 1987, 1997) due to its simplicity and ease of analysis and interpretation. Separate and sequential hypothesis tests would be conducted on costs and effects to determine whether incremental cost-effectiveness should be estimated or whether CMA was appropriate.

In 2001, however, Briggs & O'Brien declared the 'death of CMA' and argued that this sequential analysis approach is inappropriate and that CMA is rarely valid when sampled data on costs and effects are available. They argued that researchers should instead conduct cost-effectiveness analysis (CEA)[a] to estimate the joint density of cost and effect differences and present uncertainty about cost-effectiveness on cost-effectiveness acceptability curves, regardless of whether statistically significant differences in effectiveness were observed. Subsequently, the latest (2005) edition of the Drummond *et al.* textbook no longer considers CMA as a form of full economic evaluation and regards it as inappropriate in most situations.

Nonetheless, Briggs & O'Brien (2001) argued that CMA was appropriate for randomised trials that were designed to test the explicit hypothesis of equivalence or non-inferiority[b] between two therapies. Similarly, Drummond *et al.*, (2005) commented that CMA may be justifiable for comparisons between drugs in the same pharmacological class where previous research or clinical consensus indicate equal effectiveness.

However, several additional considerations argue against use of CMA even in these situations. Firstly, the magnitude and statistical significance of effect differences often differ between endpoints (Johnston *et al.* 2003). In particular, quality-adjusted life-years (QALYs) may capture wider benefits, such as the quality of life

[a] Within this paper, we use the term "cost-effectiveness analysis" in a more general (or US) sense to refer to both cost-utility analyses and cost-effectiveness analyses measuring health outcomes in natural units.

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

improvements from home versus hospital therapy, that may differ between treatments even if equivalence is demonstrated for primary clinical endpoints (Briggs and O'Brien, 2001).

Secondly, the concept of 'irrelevance of inference' (Claxton, 1999) provides additional arguments against CMA. Claxton argued that hypothesis testing is essentially arbitrary and irrelevant to decision-making. He proposed that the treatment with the highest expected net benefit should be adopted regardless of the uncertainty around the decision or whether differences in efficacy (or net benefit) reach conventional levels of statistical significance.[c] Rather than influencing decisions about which treatment we should use now, uncertainty should primarily guide decisions on what future research is required. This argument further undermines the idea of using sequential hypothesis testing to determine the appropriate form of analysis and suggests that decisions should be based on expected net benefit regardless of whether differences in effectiveness or net benefit 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 the form of analysis affects the value of information (VoI) or the probability that treatment is cost-effective.

This paper aims to finally settle the debate concerning whether CMA still has a role for any trial-based economic evaluations. We assess how use of CMA has changed since the above guidelines were published and 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 assess whether the form of economic evaluation affects estimates of decision uncertainty.

[c] Although inference is argued to be irrelevant, the cost of adopting new treatments or reversing decisions is nonetheless important and may mean that decisions on adoption and further research must be made jointly (Eckermann and Willan, 2007). However, wider consideration of these issues is outside the scope of this paper.

2. METHODS

2.1. Literature search on prevalence of CMA

We examined how the proportion of economic evaluations that comprise CMAs has changed over the last 10 years to explore trends that may have been influenced by the Briggs and O'Brien (2001) and Drummond *et al.* (2005) publications. The Centre for Reviews and Dissemination (CRD) database (including Database of Abstracts of Reviews of Effects (DARE), NHS Health Economic Evaluation Database [NHS HEED] and Health Technology Assessment Database) was searched on 4th October 2010. Databases were searched from 1999 to 2009, using the separate search terms: 'cost-effectiveness analysis' (which includes cost-consequence analysis); 'cost-utility analysis' (CUA); 'cost-benefit analysis'; and 'cost-minimisation analysis'. We included all studies that CRD considered to be full economic evaluations and categorised studies by type of evaluation based on CRD classifications. Full text of all CMAs published in 2008 was reviewed to assess how many were based on clinical trials or (specifically) non-inferiority trials.

2.2. Empirical study

2.2.1. Premise

To assess the impact of conducting CMA instead of CEA within non-inferiority trials and those observing no significant difference, we compared the conclusions and uncertainty estimates that would be generated by CMA and CEA within two clinical trials and ten sets of simulated trial data (Table 1).

Within CEA, we would typically conclude that a treatment represents good value for money if the value of the health gains (Δ_E) associated with giving treatment rather than the comparator outweighs the added cost (Δ_C): i.e. if the incremental net benefit (INB) (Stinnett and Mullahy, 1998) is positive at the ceiling ratio (R_c) that represents society's willingness (or ability) to pay for one unit of health outcome.

$$INB = R_c \cdot \Delta_E - \Delta_C \quad (1)$$

The assumption underpinning CMA is that Δ_E equals zero with no uncertainty. Within CMA, INB therefore equals $-\Delta_C$ regardless of ceiling ratio and the

probability that INB is positive (i.e. that treatment is cost-effective) is 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 study designs) may nonetheless be efficient if the value of this information exceeds the cost of the research.

Any external evidence should generally be combined with evidence from the trial in question (where it exists). However, to simplify the analysis and reporting of results, we assume that for each example, the trial under consideration represents the only available evidence on incremental costs and effects.

2.2. Case studies

The Scottish Unexplained Infertility Trial (SUIT) (Bhattacharya *et al.*, 2008, Wordsworth *et al.*, 2011) was a three-arm, parallel group, pragmatic randomised trial comparing three strategies for managing unexplained infertility: intrauterine insemination (IUI), oral clomifene citrate and expectant management ('do nothing'). The trial included 580 couples attending fertility clinics across five hospitals in Scotland. CEA was used to calculate the cost per additional live birth. The study found IUI to dominate clomifene citrate and cost £5,604 per live birth versus expectant management.

GNOME comprised a randomised, double-blind, placebo-controlled trial evaluating topical intranasal corticosteroids with placebo in the treatment of 217 children with otitis media with effusion ("glue ear") in primary care (Petrou *et al.*, 2010, Williamson *et al.*, 2009a, 2009b). Cost-effectiveness was assessed using CUA and a CEA that calculated the cost per additional child cured. We focus here on the results of CUA, which found steroids to be non-significantly more costly and less effective than placebo.

For both case studies, uncertainty was analysed using non-parametric bootstrapping. For SUIT, 1,000 bootstrap replicates were drawn from the trial data; standard errors (SEs) for incremental costs, incremental effects and INB were based on the standard deviations across bootstrap replicates. However, within GNOME (as described by

Williamson *et al.*, 2009a), missing data were imputed using multiple imputation and bootstrapping was conducted separately on the resulting five imputed datasets, with SEs calculated using Rubin's rule (Briggs *et al.*, 2003). In both studies, cost-effectiveness acceptability curves were based on the proportion of all bootstrap replicates with negative Δ_C or positive INB at different ceiling ratios.

The expected value of perfect information (EVPI) (Claxton, 1999; Willan and Pinto, 2005) was also calculated to explore how the value of future research differs between CMA and CEA. EVPI shows the value of eliminating all uncertainty around the decision by collecting perfect information. Since perfect information cannot be obtained in practice, EVPI reflects the theoretical maximum that we should consider spending on research into the decision problem. For both trials, EVPI/patient 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). To calculate population EVPI, EVPI/patient was multiplied by the number of patients likely to benefit each year and summed over the period for which the information will be used (assumed to be 10 years). Future values were discounted at 3.5% per annum. For SUIT, we assumed 16,950 couples in the UK could benefit from treatment each year (NCC-WHC, 2004), while for GNOME we assumed 16,068 children could benefit each year (Petrou *et al.*, 2010).

2.2.3. 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. The first three examples comprised cases where net benefits differ significantly between treatments despite no significant differences in costs or effects. Seven further examples were used to illustrate how the conclusions and uncertainty estimates differ between CMA and CEA.

Each hypothetical example was assigned values for the incremental cost (Δ_C) and incremental QALYs (Δ_E) of treatment versus comparator, their SEs ($\sqrt{\text{var}(\Delta_C)}, \sqrt{\text{var}(\Delta_E)}$) and the correlation coefficient (ρ) for the relationship between

Δ_C and Δ_E. Parameters for each example were chosen 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 effect and to illustrate specific points. However, the examples are not intended to comprise an exhaustive demonstration of the effect of each parameter. These values were used to calculate mean INB and its SE using Equations 1-2.

$$SEM(INB_{R_c}) = \sqrt{\text{var}(INB_{R_c})} = \sqrt{R_c^2 \cdot \text{var}(\Delta_E) + \text{var}(\Delta_C) - 2 \cdot R_c \cdot \rho \cdot \sqrt{\text{var}(\Delta_E) \cdot \text{var}(\Delta_C)}} \quad (2)$$

Since the simulated pairs of incremental costs and benefits were normally distributed, the probability $\Pr(\text{INB} > 0 | \text{R}_\text{c})$ of treatment being cost-effective at ceiling ratio R_c and the probability of treatment being cost-saving ($\Pr(\Delta_\text{C} < 0)$) were calculated analytically using Microsoft Excel 2003 based on a cumulative normal distribution.

$$\Pr(\text{INB} > 0 | \text{R}_\text{c}) = \Phi\left(\text{INB}_{\text{R}_\text{c}} / \sqrt{\text{var}(INB_{R_c})}\right) \quad (3)$$

$$\Pr(\Delta_\text{C} < 0) = \Phi\left(\Delta_\text{C} / \sqrt{\text{var}(\Delta_C)}\right) \quad (4)$$

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 lowest costs) would in fact generate lower net benefits (or higher costs) than its comparator. The error probability equalled the probability that treatment was cost-effective (or cost-saving) when INB was negative (or $\Delta_C > 0$), and equaled one minus the probability of treatment being cost-effective (or cost-saving) when INB was positive (or $\Delta_C < 0$).

Willan and Pinto (2005)'s analytical methods were adapted to allow for treatments with negative INB and used to calculate EVPI per patient,

$$EVPI = \left(\text{var}(INB) / 2\pi\right)^{1/2} \cdot \exp\left(-INB^2 / 2\,\text{var}(INB)\right) - |INB|\left(\Phi(-|INB| / \sqrt{\text{var}(INB)}\right), \quad (5)$$

where INB for CMA equalled $-\Delta_C$. Population EVPI was based on 50,000 patients benefiting from treatment each year for all simulated datasets to facilitate comparisons across examples; a 10-year time horizon was taken and values in future years discounted at 3.5% per annum.

3. RESULTS

3.1. Literature search

Our search of the CRD database showed that CMAs are still being performed, although the proportion of economic evaluations comprising CMA has fallen from 8% in 1999, to 1% in 2009 (Table 2). The number of CMAs appears to have fallen most rapidly after the Drummond textbook was revised in 2005, although the proportion decreased steadily after the Briggs and O'Brien paper was published in 2001. By contrast, the number of CEAs and CUAs increased by around 5% per year over the same time period and the number of cost-benefit analyses remained stable. Of the 17 CMAs published in 2008, five (29%) were trials and the rest were models, costing studies or studies based on review or database information. None of the trials mentioned whether they were powered for non-inferiority or equivalence. The vast majority of publications were in peer-reviewed journals, rather than assessment reports and the journals that continue to publish CMA papers tended to be clinical or health policy journals rather than specialist health economic journals. CRD includes evaluations conducted for some but not all health technology assessment organisations, including (for example) National Institute for Health and Clinical Excellence (NICE) appraisals but not those for the Scottish Medicines Consortium; as result, searches may underestimate the prevalence of CMA by missing many manufacturer evaluations and lower quality studies.

3.2. Empirical study

3.2.1. SUIT

SUIT found that IUI was significantly more costly than expectant management ($p<0.001$) but resulted in six additional live births per 100 women treated (bootstrap p-value: 0.09; Figure 1, Table 3) (Wordsworth *et al.*, 2011). Since clomifene citrate was strongly dominated by both IUI and expectant management, only the results for IUI versus expectant management are shown.

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 (Granberg *et al.*, 1995, Garceau *et al.*, 2002)). CEA therefore suggests that IUI is cost-effective compared

with expectant management if we consider inference to be irrelevant to resource allocation decisions, whereas CMA would give the opposite conclusion (suggesting that expectant management is optimal as it is less costly). Furthermore, CMA suggests that the value of further evidence on incremental costs is negligible, whereas the population EVPI from CEA equalled £11.4 million for CEA at a ceiling ratio of £10,000/birth.

3.2.2. GNOME

The GNOME trial observed no significant differences between steroids and placebo for costs, QALYs, utilities, side-effects or the primary clinical endpoint (cure at one month; Petrou *et al.*, 2010, Williamson *et al.*, 2009a). CUA showed steroids to be dominated by placebo, generating non-significantly fewer QALYs and non-significantly higher costs (Table 3). Although patients receiving steroids accrued slightly higher costs in all sensitivity analyses, incremental effectiveness varied substantially depending on the endpoint used. In particular, the proportion of children cured at one or three months was non-significantly higher for steroids and CEA found steroids to cost £347 per additional child cured.

On an 'irrelevance of inference' basis, both CUA and CMA demonstrate that no treatment dominates steroids, although CMA would nonetheless have overestimated the error probability and underestimated VoI (Table 3). However, the GNOME CEA highlights a situation where it may be appropriate to take account of uncertainty as well as expected net benefits. Although CEA suggests that steroids are cost-effective at a £1,000/cure threshold, it may not be appropriate to adopt a drug based on this evidence, since (1) we can be (at best) only 65% certain (Figure 1) that treatment increases the chance of cure *compared with placebo*, (2) there is a huge degree of uncertainty around incremental costs and benefits and (3) the results are highly sensitive to changes in outcome measure or assumptions. In this situation, an "only in research" recommendation may be appropriate if the expected value of sample information (EVSI) exceeds the opportunity cost of further research, taking account of the costs associated with licensing and adopting a new treatment and, potentially, reversing the decision based on new evidence (Eckermann and Willan, 2007, 2008a, Towse and Garrison, 2010, Briggs *et al.*, 2010). This highlights the need for reliable,

unbiased estimates of VoI to enable efficient allocation of healthcare and research funds.

3.2.3. Bias within uncertainty estimates

For both SUIT and GNOME, the error probability and EVPI estimates from CMA were very different from those of CEA or CUA (Table 3) and CMA produced different conclusions from the SUIT and GNOME CEAs (but not the GNOME CUA). Since CMA comprises a simplification of CEA that makes the strong assumption that incremental effects equal zero with no uncertainty, the uncertainty estimates generated using CEA must represent the "correct" values since they make no such assumption. Subsequently, the estimates of error probability and EVPI from CMA represent biased estimates of uncertainty.

Ten simulated datasets were generated to explore what determines the degree and direction of bias within error probabilities and EVPI and when conclusions will differ between CMA and CEA (Table 3).

If inference is considered irrelevant, CMA and CEA will produce different conclusions regarding the optimal treatment when the results lie in one of two areas of the cost-effectiveness plane: when the point estimate lies in the north-east quadrant and INB is positive (Examples 1, 3b and 3d-f, Table 3); and when the point estimate lies in the south-west quadrant and INB is negative (Example 3g).

However, CMA and CEA may produce substantially different estimates of the error probability and EVPI regardless of quadrant: particularly when we observe significant differences in net benefit despite non-significant differences in costs or effects (Examples 3a-c) and when treatment is significantly more costly but has positive INB (Example 3d).

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 3e and 3f, where a 15-fold increase in Δ_E and its SE increases the EVPI for CUA by 18-fold but has minimal

effect on the error probability. Furthermore, although CMA produces unbiased estimates of the error probability when incremental costs and benefits are exactly equal to zero and are symmetrically distributed (Example 3j), CMA may nonetheless bias EVPI estimates in these situations. Whereas the error probability shows the risk of making the wrong decision (i.e. the chance that the true INB is negative or Δ_C is positive), EVPI allows for both the probability and the opportunity cost of making the wrong decision. Mathematically, the error probability varies with the ratio of SE/mean, whereas the EVPI is affected by both the ratio of SE/mean and the absolute magnitude of INB or incremental costs.

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 analytical methods (Willan and Pinto, 2005, Eckermann and Willan, 2008b, Willan, 2007, 2009). This confirmed the findings from EVPI, demonstrating that CMA always biases EVSI in the same direction as EVPI (details available on request).

Although uncertainty estimates from CMA will always be biased, we found the bias to be minimal when the error probability and VoI both approach zero. For studies with no significant difference in effectiveness, this can occur when the difference in cost is so large compared with its SE and the likely difference in effectiveness that there is negligible chance of the more costly treatment being cost-effective. This result may arise in any of the four quadrants (Examples 3h-i).

Since CMA produces biased estimates of uncertainty in all other situations, cases where incremental costs drive the conclusions are likely to comprise the only situations when it may be valid to use CMA. However, even in these cases, CMA may not be appropriate if the trial is not powered to exclude a clinically (or economically) 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.

4. DISCUSSION

As the debate concerning the appropriate form of economic evaluation for 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.

Whereas CMA was previously considered appropriate for all studies demonstrating non-inferiority or equivalence (Briggs and O'Brien, 2001), our analyses show that CMA is only appropriate for a subset of such studies, where the difference in costs is sufficiently large that no plausible difference in efficacy could change the conclusions or uncertainty estimates. As noted previously (Glick *et al.*, 2007), separate consideration of costs and effects may also be sufficient where one intervention is significantly more effective and significantly less costly; bias within uncertainty estimates from CMA will also be negligible in these cases since error probabilities and VoI will approach zero. In all other situations, CEA is necessary to inform decisions about current resource allocation and future research.

However, conducting CEA is also more costly than CMA, requiring data on costs and health outcomes to be collected, prepared and analysed. It may therefore be appropriate to consider whether the reduced risk of bias is worth the additional research cost of conducting CEA. In practice, however, trial-based economic evaluations are conducted alongside trials primarily conducted to assess clinical outcomes. Although it may be cost-effective to omit utility measures from studies where we expect CUA to be poor use of resources, such studies will still collect clinical data alongside costs and (once collected) the additional research time required to analyse the results using CEA not CMA is minimal.

4.1. Pre-specification of the form of analysis or non-inferiority margins

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 (Donaldson *et al.*, 1996). However, minimal bias within CMA may be predicted in advance for some non-inferiority trials in which there are substantial differences in treatment costs that are likely to overwhelm any plausible differences in efficacy or other costs.

For such studies, it might be appropriate to pre-specify a non-inferiority margin or minimally important difference for QALYs, whereby CMA will be conducted unless the more costly treatment generates >X more QALYs than its comparator. The optimal value for X in these situations is debatable. It may be based on the smallest difference that patients consider beneficial (Norman *et al.*, 2003) or the smallest difference measurable on the utility instrument used or estimated by mapping the non-inferiority margin for the primary endpoint onto QALYs. However, since saving one patient's life (which is of indisputable value to that patient) could give the same QALY difference as "minimal" quality of life improvements for all patients, the concept of a minimally important difference for QALYs may be questionable.

Furthermore, CMA could bias uncertainty estimates even if incremental QALYs were smaller than a minimally important difference defined using the methods described above. If it is essential to pre-specify the form of analysis or non-inferiority margin within a statistical analysis plan, it may instead be more appropriate to conduct CMA only if the difference in QALYs is sufficiently small in comparison with expected costs that bias within uncertainty measures is acceptably small. Such a limit would not comprise a non-inferiority margin in the usual sense and would not be based on statistical inference but on whether the form of analysis will change our decisions about adoption of healthcare technologies or future research. By contrast, the non-inferiority margins for clinical endpoints (Span *et al.*, 2006) or costs (Bosmans *et al.*, 2008) that have been suggested previously are unlikely to minimise bias within uncertainty estimates.

4.2. Wider issues

Many funding bodies 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 based on prospective trial data. An obvious question there is whether it is ever possible to be certain in advance what the difference in costs and effects will be or

whether it is more efficient to collect prospective data on costs and utilities than wait until results are known.

Despite the arguments for conducting CEA, editors and reviewers sometimes question the value of publishing full economic evaluations of studies finding no significant difference in efficacy. We believe that it is essential for economic evaluations of both "positive" and "negative" trials to be published to reduce publication bias (Johnston *et al.*, 2003) within systematic reviews of economic evaluations. 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, international comparisons or VoI calculations used to inform future research priorities. Since the cost of conducting the trial has already been incurred, it is appropriate to make full use of the information collected.

Although CMA is very rarely appropriate for trial-based economic evaluation, it may have a larger role in model-based economic evaluation: particularly when decisions must be made quickly with limited resources. Guidelines recommend that models take account of all factors that could affect the conclusions (including efficacy), even if they do not reach statistical significance (Philips *et al.*, 2004), and CMA will produce biased estimates of VoI and the error probability in model-based economic evaluations as well as trials. However, the incremental cost of conducting CEA rather than CMA will be substantially higher for model-based analyses than for trials in which costs and utilities have already been collected. In particular, building a complex model to accurately estimate costs and effects is unlikely to be cost-effective if the incremental cost is so large that no plausible difference in effectiveness could change the conclusions. However, even when it is impractical to use CEA in the base case analysis due to insufficient data, time or resources, it would be appropriate to conduct CEA in sensitivity analyses (e.g. simple "back of an envelope" calculations) to assess whether the conclusions are sensitive to differences in incremental effects.

4.3. Limitations

This study compared CMA and CEA results for 12 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. 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.

The examples described are based on individual trials and do not take account of external evidence. While this may be appropriate for SUIT and for GNOME, where no previous trials have compared the same intervention in comparable populations, it is generally appropriate to take account of all available evidence in decisions about which treatment to adopt and whether further evidence is required (NICE, 2008, Drummond *et al.*, 2008, Sculpher *et al.*, 2006). Although taking account of supplementary 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.

4.4. Conclusions

We have revisited the debate concerning use of CMA for non-inferiority trials and those observing no significant difference in efficacy and have demonstrated that conducting CMA rather than CEA introduces bias into uncertainty estimates. Biased VoI estimates could lead to inefficient decisions concerning future research or whether to recommend treatment only in research or under a risk-sharing scheme (Towse and Garrison, 2010). Due to this bias, full economic evaluations are necessary for non-inferiority trials as well as those powered only to detect differences between treatments unless the difference in cost is so large that no plausible difference in efficacy could change the conclusions. In non-inferiority trials of this type only, there may be a case for the continued use of CMA within trial-based economic evaluation.

5. ACKNOWLEDGEMENTS

We would like to thank members of the Health Economists' Study Group for their comments and discussion on the ideas raised in the paper and Helen Campbell, Stavros Petrou and Alastair Gray for comments on this paper. We thank the SUIT trial

group (particularly James Buchanan and Siladitya Bhattacharya) and the GNOME trial group (particularly Giselle Abangma, Stavros Petrou and Ian Williamson) for their work on SUIT and GNOME. SUIT was funded by Chief Scientist Office and GNOME was funded by the UK National Institute for Health Research (NIHR) Health Technology Assessment (HTA) Programme. SW is funded by an NIHR fellowship award and HD is funded by several HTA studies. The views expressed in this paper are those of the authors alone and not any funding body.

REFERENCES

Bhattacharya, S, Harrild, K, Mollison, J, Wordsworth, S, Tay, C, Harrold, A, McQueen, D, Lyall, H, Johnston, L, Burrage, J, Grossett, S, Walton, H, Lynch, J, Johnstone, A, Kini, S, Raja, A & Templeton, A. 2008. Clomifene citrate or unstimulated intrauterine insemination compared with expectant management for unexplained infertility: pragmatic randomised controlled trial. *BMJ,* **337**: a716.

Bosmans, JE, de Bruijne, MC, van Hout, HP, Hermens, ML, Ader, HJ & van Tulder, MW. 2008. Practical guidelines for economic evaluations alongside equivalence trials. *Value Health,* **11**: 251-8.

Briggs, A. 2000. Economic evaluation and clinical trials: size matters. *Bmj,* **321**: 1362-3.

Briggs, A, Clark, T, Wolstenholme, J & Clarke, P. 2003. Missing... presumed at random: cost-analysis of incomplete data. *Health Econ,* **12**: 377-92.

Briggs, A, Ritchie, K, Fenwick, E, Chalkidou, K & Littlejohns, P. 2010. Access with evidence development in the UK: past experience, current initiatives and future potential. *Pharmacoeconomics,* **28**: 163-70.

Briggs, AH & O'Brien, BJ. 2001. The death of cost-minimization analysis? *Health Econ,* **10**: 179-84.

Claxton, K. 1999. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *J Health Econ,* **18**: 341-64.

Donaldson, C, Hundley, V & McIntosh, E. 1996. Using economics alongside clinical trials: why we cannot choose the evaluation technique in advance. *Health Econ,* **5**: 267-9.

Drummond, MF, O'Brien, B, Stoddart, GL & Torrance, GW. 1997 *Methods for the economic evaluation of health care programmes,* Oxford University Press:

Drummond, MF, Schwartz, JS, Jonsson, B, Luce, BR, Neumann, PJ, Siebert, U & Sullivan, SD. 2008. Key principles for the improved conduct of health technology assessments for resource allocation decisions. *Int J Technol Assess Health Care,* **24**: 244-58; discussion 362-8.

Drummond, MF, Sculpher, MJ, Torrance, GW, O'Brien, B & Stoddart, GL. 2005 *Methods for the economic evaluation of health care programmes,* Oxford University Press: Oxford.

Drummond, MF, Stoddart, GL & Torrance, GW. 1987 *Methods for the economic evaluation of health care programmes,* Oxford University Press:

Eckermann, S & Willan, AR. 2007. Expected value of information and decision making in HTA. *Health Econ,* **16**: 195-209.

Eckermann, S & Willan, AR. 2008a. The option value of delay in health technology assessment. *Med Decis Making,* **28**: 300-5.

Eckermann, S & Willan, AR. 2008b. Time and expected value of sample information wait for no patient. *Value Health,* **11**: 522-6.

Garceau, L, Henderson, J, Davis, LJ, Petrou, S, Henderson, LR, McVeigh, E, Barlow, DH & Davidson, LL. 2002. Economic implications of assisted reproductive techniques: a systematic review. *Hum Reprod,* **17**: 3090-109.

Glick, HA, Doshi, JA, Sonnad, SS & Polsky, D. 2007 *Economic evaluation in clinical trials,* Oxford University Press: Oxford.

Granberg, M, Wikland, M, Nilsson, L & Hamberger, L. 1995. Couples' willingness to pay for IVF/ET. *Acta Obstet Gynecol Scand,* **74**: 199-202.

Johnston, K, Gray, A, Moher, M, Yudkin, P, Wright, L & Mant, D. 2003. 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,* **19**: 476-89.

NCC-WCH (National Collaborating Centre for Women's and Children's Health). February 2004 Fertility: assessment and treatment for people with fertility problems. http://www.nice.org.uk/nicemedia/live/10936/29269/29269.pdf [Accessed: 17th May 2011].

NICE (National Institute for Health and Clinical Excellence). 2008 Social value judgements: Principles for the development of NICE guidance. Second

Edition. http://www.nice.org.uk/media/C18/30/SVJ2PUBLICATION2008.pdf [Accessed: 17th November 2010].

NICE (National Institute for Health and Clinical Excellence). June 2008 Guide to the methods of technology appraisal. http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf [Accessed: 13th July 2010].

Norman, GR, Sloan, JA & Wyrwich, KW. 2003. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care,* **41**: 582-92.

Petrou, S, Dakin, H, Abangma, G, Benge, S & Williamson, I. 2010. Cost-utility analysis of topical intranasal steroids for otitis media with effusion based on evidence from the GNOME trial. *Value Health,* **13**: 543-51.

Philips, Z, Ginnelly, L, Sculpher, M, Claxton, K, Golder, S, Riemsma, R, Woolacott, N & Glanville, J. 2004. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technology Assessment,* **8 No. 36**.

Sculpher, MJ, Claxton, K, Drummond, M & McCabe, C. 2006. Whither trial-based economic evaluation for health care decision making? *Health Econ,* **15**: 677-87.

Span, MM, TenVergert, EM, van der Hilst, CS & Stolk, RP. 2006. Noninferiority testing in cost-minimization studies: Practical issues concerning power analysis. *Int J Technol Assess Health Care,* **22**: 261-6.

Stinnett, AA & Mullahy, J. 1998. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Med Decis Making,* **18**: S68-80.

Towse, A & Garrison, LP, Jr. 2010. Can't get no satisfaction? Will pay for performance help?: toward an economic framework for understanding performance-based risk-sharing agreements for innovative medical products. *Pharmacoeconomics,* **28**: 93-102.

Willan, A. 2009. Tutorial on ASSENT3: Spreadsheet tutorial. IN Eckermann, S & Willan, A (Eds.) *Health economics from theory to practice: Informing related decisions of reimbursement, research and regulation. A three-day workshop 25-27 August 2009.*

Willan, AR. 2007. Clinical decision making and the expected value of information. *Clin Trials,* **4**: 279-85.

Willan, AR & Pinto, EM. 2005. The value of information and optimal clinical trial design. *Stat Med,* **24**: 1791-806.

Williamson, I, Benge, S, Barton, S, Petrou, S, Letley, L, Fasey, N, Abangma, G, Dakin, H & Little, P. 2009a. 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,* **13**: 1-144.

Williamson, I, Benge, S, Barton, S, Petrou, S, Letley, L, Fasey, N, Haggard, M & Little, P. 2009b. Topical intranasal corticosteroids in 4-11 year old children with persistent bilateral otitis media with effusion in primary care: double blind randomised placebo controlled trial. *BMJ,* **340**: b4984.

Wordsworth, S, Buchanan, J, Mollison, J, Harrild, K, Robertson, L, Tay, C, Harrold, A, McQueen, D, Lyall, H, Johnston, L, Burrage, J, Grossett, S, Walton, H, Lynch, J, Johnstone, A, Kini, S, Raja, A, Templeton, A & Bhattacharya, S. 2011. Clomifene citrate and intrauterine insemination as first-line treatments for unexplained infertility: are they cost-effective? *Human Reproduction,* **26**: 369-75.

Table 1: Summary of the datasets used in the study

Example	Data source	Concept illustrated
1	SUIT: RCT conducted by the authors comparing 3 forms of infertility care (n=580) (Wordsworth *et al.*, 2011, Bhattacharya *et al.*, 2008)	Difference in effectiveness approaches statistical significance. Illustrates potential for differing conclusions between CMA and CEA
2	GNOME (Petrou *et al.*, 2010, Williamson *et al.*, 2009a): Placebo-controlled RCT on intra-nasal steroids for otitis media conducted by the authors (n=217)	Differences in costs and effectiveness are negligible and non-significant.
3	10 simulated datasets	Highlight situations where differences in uncertainty estimates between CMA and CEA are particularly large or small or when there may be a significant difference in net benefits despite no significant differences in costs or effects

CEA, cost-effectiveness analysis; CMA, cost-minimisation analysis; RCT, randomised controlled trial.

Table 2: Number of economic evaluations in the Centre for Reviews and Dissemination (CRD) database by year

Year	Number (%) of studies on CRD					Relevant publications
	Full economic evaluations	**Cost-minimisation analyses**	**Cost-effectiveness or cost-consequence analyses**	**Cost-utility analyses**	**Cost-benefit analyses**	
1999	758 (100%)	64 (8.4%)	571 (75.3%)	81 (10.7%)	42 (5.5%)	(Claxton, 1999)
2000	850 (100%)	60 (7.1%)	650 (76.5%)	99 (11.6%)	41 (4.8%)	-
2001	831 (100%)	51 (6.1%)	648 (78%)	94 (11.3%)	38 (4.6%)	(Briggs and O'Brien, 2001)
2002	892 (100%)	55 (6.2%)	654 (73.3%)	129 (14.5%)	54 (6.1%)	-
2003	806 (100%)	36 (4.5%)	601 (74.6%)	131 (16.3%)	38 (4.7%)	(Johnston *et al.*, 2003)
2004	914 (100%)	54 (5.9%)	643 (70.4%)	175 (19.1%)	42 (4.6%)	-
2005	1,050 (100%)	61 (5.8%)	737 (70.2%)	217 (20.7%)	35 (3.3%)	(Drummond *et al.*, 2005)
2006	948 (100%)	26 (2.7%)	647 (68.2%)	238 (25.1%)	37 (3.9%)	(Span *et al.*, 2006)
2007	949 (100%)	19 (2%)	617 (65%)	268 (28.2%)	45 (4.7%)	(Glick *et al.*, 2007)
2008	1,007 (100%)	17 (1.7%)	664 (65.9%)	297 (29.5%)	29 (2.9%)	(Bosmans *et al.*, 2008)
2009*	548 (100%)	3 (0.5%)	363 (66.2%)	162 (29.6%)	20 (3.6%)	-
Total	**9,553 (100%)**	**446 (4.7%)**	**6,795 (71.1%)**	**1,891 (19.8%)**	**421 (4.4%)**	**-**

* The number of economic evaluations is lower in 2009 due to the time lag between publication and inclusion of abstracts in CRD.

Table 3: Summary of the case studies and simulated datasets

Dataset	Δ cost/pt (SE)	Δ effect/pt (SE)†	Covariance (correlation coefficient)	INB (SE)‡	Error probability‡		Population EVPI‡ (millions)		Optimal treatment		Direction/ magnitude of bias
					CUA	CMA	CUA	CMA	CUA	CMA	
Case studies											
1: SUIT (IUI v. expectant management) (Wordsworth *et al.*, 2011)	£319 (£18)*	0.06 (0.04) live births	-0.18 (-0.24)	£239 (£427)	28%	0%	£11.44	£0.00	T	C	VoI and Pr(error) negligible for CMA, but high for CEA
2: GNOME (CUA) (Petrou *et al.*, 2010)	£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
Simulated data: No significant difference in costs or effects, but significant difference in net benefits											
3a: North-west quadrant, -ve correlation	£2.00 (£1.10)	-2 (1.1)	-0.36 (-0.3)	-£4.00 (£1.77)	1.21%	3.45%	£0.01	£0.01	C	C	Pr(error) higher for CMA, VoI negligible
3b: North-east 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
3c: North-east 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
Simulated data: large bias within error probabilities and value of information											
3d: North-east quadrant, large +ve Δ_E, sig +ve Δ_C	£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
3e: North-east quadrant, small +ve Δ_E, $\Delta_C \approx 0$	£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
3f: North-east quadrant, large +ve Δ_E, $\Delta_C \approx 0$	£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
3g: South-west quadrant,	-£0.01	-3 (1.65)	-148.5 (-0.3)	-£60,000	3.49%	50.00%	£457	£120	C	T	VoI lower and Pr(error)

Dataset	Δ cost/pt (SE)	Δ effect/pt (SE)†	Covariance (correlation coefficient)	INB (SE)‡	Error probability‡		Population EVPI‡ (millions)		Optimal treatment		Direction/ magnitude of bias
					CUA	CMA	CUA	CMA	CUA	CMA	
large –ve Δ_E, $\Delta_C \approx 0$	(£300)			(£33091)							higher for CMA
Simulated data: minimal bias within error probabilities and/or value of information											
3h: North-east quadrant, very large Δ_C and highly significant	£4,500 (£450)*	0.06 (0.03)	0 (0)	-£3,300 (£750)*	0.00%	0.00%	£0.00	£0.00	C	C	Bias negligible
3i: South-east quadrant, very large and highly significant Δ_C,	-£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
3j: Δ_C and Δ_E 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

Abbreviations: Δ_C, incremental cost; Δ_E, incremental effectiveness; CMA, cost-minimisation analysis; CEA, cost-effectiveness analysis; CUA, cost-utility analysis; EVPI, expected value of perfect information; INB, incremental net benefit; IUI, intrauterine insemination; Pr(error), error probability (probability that the treatment with the highest net benefits is not cost-effective [or not cost-saving]); SE, standard error.

* $p<0.05$

† Effects are shown in QALYs unless otherwise stated.

‡ For all CUA, INB, EVPI and Pr(error) were based on a ceiling ratio of £20,000/QALY gained (NICE, 2008). For SUIT, a ceiling ratio of £10,000/live birth was used (Granberg *et al.*, 1995).

Figure 1: Results of SUIT and GNOME studies. Results of SUIT are shown in panels A-C: (A) Cost-effectiveness plane, (B) cost-effectiveness acceptability curve, (C) expected value of perfect information (EVPI). Results of GNOME cost-utility analysis (CUA) and cost-minimisation analysis (CMA) are shown in panels D-F: (D) Cost-effectiveness plane, (E) cost-effectiveness acceptability curve, (F) EVPI. In all figures, dashed lines show results for CMA; solid lines show results of CUA or cost-effectiveness analysis.