Article:
Stevenson, M., Tappenden, P. and Squires, H. (2014) Methods for handling uncertainty within pharmaceutical funding decisions. International Journal of Systems Science, 45 (1). pp. 60-68. ISSN 0020-7721

https://doi.org/10.1080/00207721.2012.723056

Current methods for handling uncertainty within pharmaceutical funding decisions: A position paper

Matt Stevenson

School of Health and Related Research,

The University of Sheffield: m.d.stevenson@shef.ac.uk

Paul Tappenden

School of Health and Related Research,

The University of Sheffield: p.tappenden@shef.ac.uk

Hazel Squires

School of Health and Related Research,

The University of Sheffield: h.squires@shef.ac.uk

Abstract

This paper provides a position statement regarding decision making under uncertainty within the economic evaluation of pharmaceuticals, with a particular focus upon the National Institute for Health and Clinical Excellence context within England and Wales. This area is of importance as funding agencies have a finite budget from which to purchase a selection of competing health care interventions. The objective function generally used is that of maximising societal health with an explicit acknowledgement that there will be opportunity costs associated with purchasing a particular intervention. Three elements of decision uncertainty are discussed within a pharmaceutical funding perspective: methodological uncertainty; parameter uncertainty and structural uncertainty, alongside a discussion around value of additional information. Methods development has focussed primarily on handling parameter uncertainty and a clear reference case has been developed for consistency across evaluations, however, uncertainties still remain. Less attention has been given to methods for handling structural uncertainty; the lack of adequate methods to explicitly incorporate this aspect of model development may result in the true uncertainty surrounding health care investment decisions being underestimated. Research in this area is ongoing.

Key Words: Uncertain data, robust procedure, health economics, value of information, mathematical modelling.

Background

This manuscript was prepared for discussion at the 25th Mini-EURO Conference on Uncertainty and Robustness in Planning and Decision Making which had the aim of integrating methodologies used by different research communities. Accordingly, we have set out the methodologies

used within decision making in the evaluation of health technologies, (a relatively new discipline), mainly in terms of the National Institute for Health and Clinical Excellence (NICE) context within England and Wales, in order that researchers in other fields understand how health economists deal with uncertainty. This consists of: an introduction to health economic evaluation; NICE technology appraisals; a taxonomy of uncertainty used in health technology assessments; representing uncertainty around the objective function; alternative decision making criteria , and value of additional information.

Introduction to health economic evaluation

The finite budget for health care in conjunction with the number of potential treatments results in a position where not all treatments can be funded. Consequently there is a need to quantify the opportunity costs in terms of the health outcomes lost by curtailing existing services in order to fund the new intervention. Attempts to maximise societal health from a limited budget has led to funding bodies such as, being established to provide guidance in England and Wales on whether pharmaceuticals should be recommended for routine use on the NHS. At the present time, such recommendations carry a mandate for funding by the health service. The Pharmaceutical Benefits Advisory Committee, the Canadian Agency for Drugs and Technology in Health and the Scottish Medical Consortium, hold similar positions in Australia, Canada and Scotland respectively. In all jurisdictions the decision will be made in the presence of uncertainty.

Due to the collective experience of the authors as a NICE appraisal committee member, and as authors of reports for consideration by NICE representing both academia and pharmaceutical perspectives, NICE is the key reference point. Additionally NICE represents an exemplar in that its decision making approach is largely focussed on the formal consideration of opportunity cost under conditions of uncertainty. Some of the key concepts will apply to other funding agencies.

NICE technology appraisals

NICE makes recommendations to the NHS within England and Wales on selected new and existing medications, treatments and procedures. It is an independent body which aims to produce guidance for health professionals using the best available evidence, with a focus on transparency regarding the decision making process. The decision for or against the use of a given health technology is made by an independent appraisal committee, appointed for a 3-year term consisting of members drawn from the NHS, patient and carer organisations, academia and pharmaceutical and medical devices industries and which is free from political influence. There are two types of technology appraisal: the single technology appraisal (STA) and the multiple technology appraisal (MTA). An STA is specifically designed for the appraisal of a single health technology, with a single indication, whereby most of the relevant evidence lies with one manufacturer; in this instance the manufacturer submits a report and model which is critiqued by an independent academic group. In an MTA, where multiple interventions are being appraised, the academic group constructs a mathematical model; the methods and results of which are discussed in comparison with those produced by the manufacturers. The STA process was developed to produce more timely guidance, but guidance may be delayed if additional analyses are required following the committee meeting to resolve concerns regarding the robustness of the submitted evidence. The preliminary guidance is subject to public consultation which is considered before final guidance is developed.

The appraisals undertaken by NICE are typically posed in the form *"should we buy programme A or* programme B for the treatment of disease *C in population D?"* These are substance focussed motivations (Morgan and Henrion 1992) in that there is a specifically formulated policy question that will be directly considered. The evaluation criterion is the incremental cost per Quality Adjusted Life Year (QALY) gained, which is a common metric intended to provide a level playing field such that decisions can be taken consistently across treatments and disease areas.

The QALY combines increased life expectancy and improvements in health status by assigning to each period of time a utility ranging from 0 to 1, corresponding to the health-related quality of life during that

period. A utility score of 1 corresponds to "perfect health", whilst a utility of 0 corresponds to a health state judged to be equivalent to death (Weinstein and Statson, 1977). Utility scores lower than zero represent states of health that are judged to be worse than death. Health utilities may be elicited using a range of preference-based methods which, in general, seek to trade-off an individual's preferences for one state of health over another under conditions of uncertainty. The QALY approach thus adjusts survival according to the individual's level of health-related quality of life. A person who survives 10 years at a utility of 0.8 will gain eight QALYs. The benefits of a treatment that increases survival at a utility of 0.8 (from 10 to 20 years) or improves the quality of the 10 years (from 0.8 to 0.9) can be valued in terms of the QALY gain (i.e. gains of eight and one, respectively).

The incremental cost per QALY ratio is compared with defined thresholds in the decision making process. Guidance from NICE has suggested that treatments with an incremental cost per QALY of below £20,000 per QALY are likely to receive funding, whilst those above £30,000 per QALY are unlikely to receive funding, (NICE, 2008) although a higher threshold may be used for a subset of interventions which meet 'end of life' criteria. (NICE, 2009). The central underlying assumption is that the successive implementation of this threshold approach will move towards a health maximising solution.

A taxonomy of sources of uncertainty within health technology assessments.

Economic modelling involves projecting forward to estimate profiles of expected costs and health outcomes for each intervention under consideration over a time horizon which ideally captures all differences in costs and health outcomes between the intervention(s) and the comparator(s). Profiles of cost and health outcomes (survival, quality of life or both) are then compared incrementally and synthesised to produce an expected incremental cost per QALY ratio. Health economic models often draw together evidence concerning the natural history of a disease, epidemiology, treatment effectiveness, adverse events, resource use and costs. Uncertainty surrounding the use of such models, and more broadly around the decision problem to be addressed by the model, is pervasive; its importance is manifested in the possibility that the adoption decision may be incorrect. An incorrect decision would lead to inefficient allocation of resources and hence sub-optimal health gains for society.

This paper outlines current approaches to handling three different but overlapping elements of decision uncertainty:

1. Methodological uncertainty – uncertainty surrounding the methods used to underpin the evaluation e.g. specification of the objective function, discount rates, the time horizon and the measurement and valuation of costs and health outcomes
2. Parameter uncertainty – uncertainty surrounding the true mean values of model parameters
3. Structural uncertainty – uncertainty surrounding the conceptual and mathematical representation of a decision problem within a model.

This taxonomy has emerged as a common language within the field of health economics. For completeness, an alternative interdisciplinary taxonomy of the sources of uncertainty is discussed within a health economic context.

1) Methodological uncertainty

Within health technology evaluations for NICE methodology uncertainty is handled using a Reference Case (NICE, 2008) The Reference Case stipulates the metric of the cost per QALY and also prescribes a common discount rate applied to handle preferences in the timing of accrued costs and outcomes, relevant comparators for the model, a consistent focus on costs and outcomes incurred by the NHS and Personal Social Services and specific approaches to measuring and valuing the health outcomes and costs associated with health interventions. In addition, the Reference Case requires that uncertainty around

model parameters should be handled within a multivariate probabilistic sensitivity analysis (PSA). It is acknowledged that there will be uncertainty in the values and approaches chosen by NICE for the Reference Case, but that this is outside of the control of the modeller within this context and will not be discussed further here.

2) Parameter uncertainty

The appropriate handling of parametric uncertainty has been the focus of considerable methodological development in health economics for over 20 years. Uncertainty exists in key parameters that are incorporated into mathematical models such as the efficacy of treatment, the natural history of patients who do not receive the intervention, relationships between intermediate and final outcomes and the level of health-related quality of life associated with being in a particular health state, for example having sustained a hip fracture. Even where well conducted randomised controlled trials have been undertaken, uncertainty will still be present due to finite trial sizes and insufficient follow up to consider all impacts upon future costs and health outcomes.

PSA uses Monte Carlo methods to simultaneously sample from all uncertain distributions associated with each parameter in order to produce distributions of incremental costs and health outcomes. This uncertainty can then be interpreted in terms of the probability that the given technology is cost-effective over a range of acceptable cost-effectiveness thresholds. Conducting PSA is fundamentally important because if the model is non-linear then the results calculated from the PSA may be markedly different to those calculated using the mean value (as recommended by NICE) estimate from each distribution. (Claxton et al., 2005) Typically crude Monte Carlo techniques have been employed to perform PSA, although alternative approaches using meta-modelling (Stevenson et al., 2004) and through calculating an efficient ratio of the number of individuals sampled to the number of PSA parameters (O'Hagan et al., 2007) have been proposed.

Early health economic models assumed independence when sampling from the distributions for each parameter, however the use of the Cholesky decomposition, which explicitly takes into account correlation between variables and the dirichelet distribution which appropriately samples the proportions of patients that fall into mutually exclusive and exhaustive groups are now commonly used. Further details on these distributions are provided in Briggs et al., 2006.

Three developments in addressing parameter uncertainty that are particularly pertinent within the discipline are explicitly discussed. These are mixed treatment comparisons or Network Meta-Analyses, fitting statistical curves to censored survival data and methods for dealing with patient cross-over within a trial.

Network meta-analysis is a generalisation of standard pair-wise meta-analysis that derives estimates of treatment effects from a synthesis of direct and indirect evidence (Caldwell et al., 2005), (Lu and Ades, 2006) which ensures that the output from a combination of different evidence sources are consistent. This contrasts with direct head-to-head evidence (for which the NICE Reference Case has a strong preference), which can (in an extreme) produce discrepancies where intervention A is assumed to be more efficacious than intervention B, intervention B is assumed to be more efficacious than intervention C yet intervention C is assumed more efficacious than intervention. Furthermore it has been shown that assuming direct evidence produces more accurate estimates of relative efficacy than the estimates estimated through a network meta-analysis where the two are discordant is not always correct. Madan et al., (In Press) provide an example where the initials estimate of efficacy obtained from a network meta-analysis was discordant with the direct evidence, but following subsequent trials (where the estimates from both methods were now in agreement) it was shown that initially the estimate from the network meta-analysis was more accurate.

It is typical that a clinical trial is terminated before all patients have had an event. In this instance a substantial proportion of the observations are censored and a number of statistical distributions could be

used to provide a plausible fit to the data. The importance of this was shown in a NICE appraisal of sorafenib for hepatocellular cancer. Both the lognormal and weibull distributions produced a reasonable fit to the observed data and had similar Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values; however, differences in the tails of the distributions produced markedly different cost-effectiveness ratios (£64,754 and considerably higher (commercial-in-confidence) respectively)

Within a clinical trial, it is possible that an intervention is shown to be substantially more efficacious than a comparator (possibly placebo) arm before the trial has completed. In this circumstance, patients within the comparator arm are often crossed-over and receive the intervention. Whilst ethically just, this creates problems in performing intention to treat analyses (where patients are classified solely on their initial treatment allocation) due to confounding, as the comparator would very likely do better because of the cross-over than would have been the case had patients continued with their initial treatment, which can have dramatic influences on the resulting cost-effectiveness ratio. Simple correction methods such as exclusion or censoring of observations are likely to lead to severe selection bias and more complex methods such as rank preserving structural failure time (RPSFT) models (Robins and Tsiatis, 1991) and inverse probability of censoring weights (IPCW) (Robins and Finkelstein, 2000) have been used in submissions to NICE. The RPSFT uses the randomisation assumption to estimate treatment effect such that counterfactual survival (a function of observed survival time, observed treatment and the treatment effect) would have been equal in randomised groups had no experimental treatment been given to any patients. The IPCW approach treats crossover patients as informatively censored, and applies time-varying weights to uncensored survival probabilities based on the probability of patients crossing over given their covariate history. The IPCW approach is reliant on the assumption of no unmeasured confounders, and upon the existence of a reasonable number of uncensored observations. Research is currently ongoing to determine which method is most applicable in health technology evaluations.(reference?)

3) Structural uncertainty

Structural uncertainty relates to whether all relevant processes are represented in the model, that is, what is included and excluded and how the relationships between inputs and outputs are captured. Recent empirical work has highlighted marked heterogeneity in the development of health economic models (Tappenden et al., 2009). Methods for handling structural uncertainty are relatively underdeveloped in the field of health economics, yet structural uncertainties can often have a greater impact upon the model results than parameter uncertainties. This means that the estimated uncertainty within the model outputs resulting from PSA is likely to be underestimated, as such estimates are reliant on the assumption that the model structure is optimal. The NICE Methods Guide (NICE, 2008) states that 'structural assumptions should be clearly documented and the evidence and rationale to support them provided. The impact of structural uncertainty on estimates of cost effectiveness should be explored by separate analyses of a representative range of plausible scenarios'. These methods can provide an indication around the impact of uncertainties regarding key structural assumptions upon the model results. However, the overall model uncertainty cannot be captured using these methods.

A 'model averaging' method has been proposed (Bojke et al., 2009) for handling the propagation of structural uncertainty through the model. This involves parameterising all identified structural uncertainties within the model and weighting each according to their likelihood, often estimated using elicitation of expert judgement. These 'structural uncertainty' parameters would be varied within the PSA in the same way as other model parameters. The model results could therefore account for both parameter and structural uncertainties. However, the feasibility of this method for complex problems with many structural uncertainties has not been demonstrated. Moreover, structural uncertainties not identified within the model construction process, which may markedly change the results, are ignored. A 'discrepancy' approach has been advocated (Kennedy and O'Hagan, 2001) which involves adjusting the model outputs using an error term to account for the difference between the model outputs and reality, however, it is unlikely that the required data will be available to the modeller. Research is currently ongoing (Strong et al., 2009) to evaluate the feasibility of this method. Both the model averaging approach and the discrepancy approach require further research before they can be used routinely in health care decision

models. In parallel, it is essential that methodologies for developing the initial model structure are appropriate and transparent. However, current approaches for structural development are not well defined, and as a result this process is inconsistent between modellers (Chilcott et al., 2009). Further research is therefore also required in this area.

Comparison of our taxonomy of uncertainty with that of Morgan and Henrion

Compared with previous taxonomies of uncertainty the one we have defined as relevant for health technology assessments is small [PT – NOT SURE "SMALL" IS APPROPRATE. WOULD BE HAPPIER SAYING THAT WHILST THE TAXONOMIES DIFFER IN THEIR CATEGORISATIONS, THEY ACTUALLY CONSIDER VERY SIMILAR THINGS], primarily as only the key areas related to health technology assessment have been defined. For completeness we have compared our taxonomy with that of Morgan and Henrion (Morgan and Henrion 1990) which has a much wider breadth and which details seven sources of uncertainty: Statistical variation; subjective judgement; linguistic imprecision; variability; inherent randomness; disagreement; and approximation.

Statistical variation and approximation are included within our taxonomy as 'parameter uncertainty' and 'structural uncertainty' respectively. The need for subjective judgement is reduced as the trials required for pharmacotherapy licensing are conducted in the appropriate population. Where the results need to be translated to a different population (for age, ethnicity or disease severity reasons) the cost per QALY is typically considered by an appraisal committee with the aid of sensitivity analyses. Linguistic imprecision is generally not an issue due to the formation of a tightly defined scope, which explicitly lists the population, intervention, comparators and outcomes to be considered [PT – BUT THE SCOPE DOESN'T DICTATE WHAT THE MODEL SHOULD LOOK LIKE. DOESN'T THIS ALSO APPLY WITH RESPECT TO MODEL DEVELOPMENT DECISIONS?. Variability and inherent randomness are not considered in health technology assessments [PT – STRONGLY DISAGREE - VARIABILITY CERTAINLY IS], whilst disagreement is typically not relevant as whilst there may be inherent uncertainty in the biological action of an intervention, the key outcome measures (such as death or number of fractures) will not be affected and are reported in the clinical trials [PT – AGREE WITH HAZEL – DISAGREEMENT IS BASICALLY STRUCTURAL UNCERTAINTY].

Representing uncertainty around the objective function

Cost effectiveness acceptability curves (CEACs) (Fenwick et al., 2001) explore the likelihood of optimality for each intervention and are used to display the uncertainty in the cost effectiveness results. CEACs plot the proportion of Monte Carlo simulations in which an intervention is estimated to be most cost-effective using different cost per QALY thresholds. An illustrative CEAC is shown in Figure 1, which highlights that for a specified population, as the threshold increases the proportions of simulations where silver-donating antimicrobial bandages were estimated to be more cost-effective than standard dressings increased. (Michaels et al., 2009)

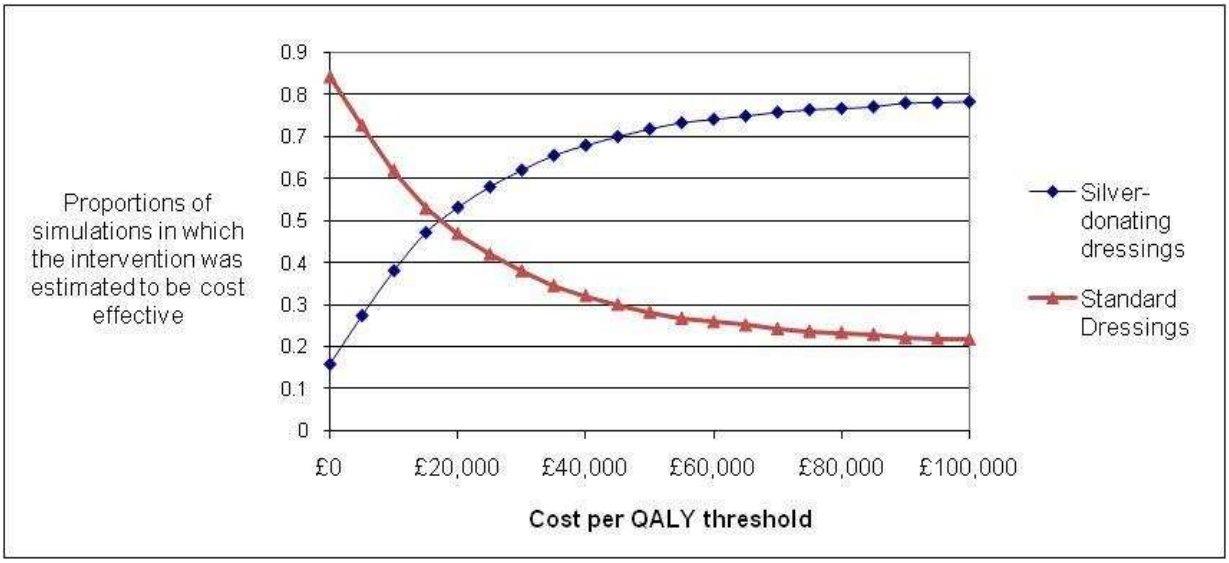

Figure 1: An illustrative CEAC

Importantly, the CEAC does not indicate which treatment is most cost-effective; this must be determined by calculating expected incremental cost per QALY values and comparing with an appropriate threshold. The proportion of times that the intervention that is estimated to be most cost-effective at a given threshold was actually most cost-effective in the PSA is displayed on a cost effectiveness acceptability frontier (CEAF). (Fenwick et al., 2001) Within CEAFs only the most cost-effective treatment at each threshold is plotted against the proportion of parameter configurations in which that intervention was estimated to be most cost-effective. In the example provided in Figure 1, the CEAF would follow the 'standard dressings' curve until the expected incremental cost per QALY value associated with silver donating dressings was reached, at which point the CEAF would use the 'silver donating dressings' curve.

Uncertainty surrounding the cost-effectiveness ratio is typically reported as the 95% range of the results produced by the PSA, with a percentile methodology used selecting the 2.5% and 97.5% values from the ranked cost per QALY values [PT – STRONGLY DISAGREE. CONFIDENCE INTERVALS ARE NOT OFTEN REPORTED DUE TO PROBLEMS IN PRODUCING THESE FOR A RATIO]. It is not uncommon when there is large uncertainty in the relative efficacy of a treatment compared with a comparator for this 95% confidence interval to range from dominated (a higher cost and less QALYs than the comparator) to dominating (a lower cost and higher QALYs than the comparator). Far less often reported, but arguably of more value, is the uncertainty in the mean cost per QALY ratio, which could be used to determine if sufficient PSA iterations have been performed. One approach used to estimate this is jackknifing (Inglehardt, 1975), which is summarised in Law (Law, 2007), which provides an upper and lower estimate of the mean cost per QALY, and also reduces the statistical bias that comes from classical estimates of a non-linear function such as the cost per QALY ratio.

Alternative decision making criteria

The NICE Reference case suggests a number of scenarios in which the appraisal committee has discretion to use an alternative cost per QALY threshold than £20,000. These include: the degree of certainty around the cost per QALY, with the committee being more cautious about recommending a technology where there is large uncertainty; when there are strong reasons to believe that the change in utility associated with a treatment has been inadequately captured; and if the technology has an innovative nature, which has not been adequately captured in the QALY measure. Formal scoring and ranking methods such as multi-creiteria decision analyses are not used within health economics.

In addition, NICE guidelines suggest that a higher cost per QALY threshold may be applied for some interventions aimed at end of life care. To meet the 'End of Life' criteria the following must be met: the treatment is indicated for patients with a short life expectancy, normally less than 24 months, there is sufficient evidence that the treatment provides a life extension of at least three months compared with current NHS treatment and that the treatment is licensed or otherwise indicated for small patient populations. (NICE 2009)

Value of additional information

Value of additional information techniques are becoming more widespread in health economic evaluation. These methods aim to quantify the value of undertaking further research in order to reduce, or even eliminate parameter uncertainty. As such, these methods can be used as a means of prioritising and planning future research to ensure that research funding is targeted at those areas in which reducing current uncertainty is expected to be of most value. The simplest form is that of Expected Value of Perfect Information (EVPI) (Claxton and Posnett, 1996), which is defined as the maximum investment a decision-maker would be willing to pay to eliminate all parameter uncertainty from the decision problem. EVPI is initially calculated in terms of a defined unit (typically per patient) and then multiplied by the number of people expected to benefit from eliminating all parameter uncertainty to form an estimate of total EVPI. EVPI per person is relatively high where there is large uncertainty in the adoption decision; conversely where there is only a small probability of error and the impact of an incorrect decision is small the EVPI per person will be relatively low. EVPI can be directly calculated from the PSA. Global EVPI is calculated as the average of the individual EVPI values.

An adaptation of EVPI evaluates the maximum value of removing all uncertainty in one, or a subset of parameters and has been defined as the expected value of partial perfect information (EVPPI); (Felli and Hazen, 1998). However, this is computationally more expensive to perform as it requires two nested Monte Carlo sampling levels (Brennan and Kharroubi, 2007). In the absence of methods to reduce such computational expense, such as meta-modelling or efficient sampling processes, the EVPPI process can be prohibitive (Tappenden et al, 2004). A more advanced method for determining the value of added information, the expected value of sample information (EVSI) has been proposed (Ades et al., 2004) which explicitly takes into account that some uncertainty will remain even with large sample sizes. The EVSI methodology simulates the results from the proposed research and synthesises the simulated data with prior knowledge to form a posterior distribution: the larger the trial size the more the posterior distribution resembles the simulated data. The posterior distribution is then used in PSA. The optimal trial size from the options evaluated can then be estimated based on the costs of conducting the trial and the expected net benefit of the sampled information. The application of EVSI is becoming more common. (Stevenson et al., 2009, Stevenson et al., 2011) [PT – THERE IS A REVIEW OF THE USE OF VOI – THINK IT WAS 2005 IN EITHER VIH OR PHARMACOECONOMICS]. The validity of the results estimated from these techniques is dependent upon the appropriate characterisation of uncertainty.

Conclusion

The above discussion has provided a synopsis of current approaches to handling uncertainty in health economic models. The emphasis of methods development has been placed on parameter uncertainty and a clear reference case has been developed, although uncertainty still exists. Less attention has been paid to methods for formally incorporating structural uncertainty. Until adequate methods are developed to explicitly include structural uncertainty, true decision uncertainty may be underestimated, which may have implications for decision making both in terms of whether an intervention should be publically funded and whether additional research should be undertaken.

Acknowledgment

The authors would like to acknowledge the input of Nicholas Latimer with regard to the section on cross-over methods.

References

Ades AE, Lu G, Claxton K. Expected values of sample information calculation in medical decision making. Med Decis Making 2004; 24:207-227.

Bojke L, Claxton K, Sculpher M, Palmer S. Characterizing structural uncertainty in decision analytic models. A review and application of methods. Value in Health 2009; 12: 739 – 749.

Brennan A, Kharroubi SA. Efficient computation of partial expected value of sample information using Bayesian approximation. J Health Econ 2007; 26:122-148.

Briggs A, Claxton K**,** Sculpher M. Decision Modelling for Health Economic Evaluation. Oxford: Oxford University Press 2006

Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ 2005 Oct 15;331(7521):897-900

Claxton K, Posnett J. An economic approach to clinical trial design and research priority-setting. Health Economics 1996; 5:513–24.

Claxton K, Sculpher M, McCabe C, Briggs A, Akehurst R, Buxton M et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. Health Economics 2005; 14:339-47.

Chilcott J, Tappenden P, Rawdin A, Johnson M, Kaltenthaler E, Papaioannou D, Shippam A. Avoiding and identifying errors in health technology assessment models. Health Technol Assess. 2010; 14 (25) 1-79.

Felli JC,Hazen GB. Sensitivity analysis and the expected value of perfect information. Medical Decision Making. 1998; 18:95–109.

Fenwick E, Claxton K and Sculpher M. Representing uncertainty: The role of cost-effectiveness acceptability curves. Health Economics 2001; 10: 779–787.

Inglehart DL. Simulating stable stochastic systems,V: Comparison of ratio estimators. Naval Res Logist Quart 1975; 22:553–565.

Kennedy M and O'Hagan A. Bayesian calibration of computer models (with discussion). Journal of the Royal Statistical Society, Series B, 2001; 63: 425 – 464.

Law A. Simulation modelling & analysis. Fourth edition. 2007 McGraw-Hill (Singapore)

Lu G, Ades A. Assessing evidence consistency in mixed treatment comparisons. Journal Of The American Statistical Association 2006;101:447-59.

Madan J, Stevenson MD, Ades T, Cooper KL, Whyte S, Akehurst R. Consistency between direct trial evidence and Bayesian Mixed Treatment Comparison: Is head-to-head evidence always more reliable? Value In Health (In Press)

Michaels JA, Campbell B, King BM, Palfreyman SJ, Shackley P, Stevenson MD. A prospective randomised controlled trial and economic modelling of anti-microbial silver dressing bandages versus non-adherent control dressings for venous leg ulcers: The Vulcan trial. Health Technol Assess 2009; 13(56) 1 - 138

Morgan MG and Henrion M. Uncertainty. 1990. Cambridge University Press

National Institute for Health and Clinical Excellence. Appraising life-extending, end of life treatments. 2009 http://www.nice.org.uk/media/E4A/79/SupplementaryAdviceTACEoL.pdf

National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisals. 2008 http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf

O'Hagan A, Stevenson M, Madan J. Monte Carlo probabilistic sensitivity analysis for patient level simulation models : Efficient estimation of mean and variance using ANOVA. Health Economics 2007;16(10):1009-23

Robins JM and Finkelstein DM. Correcting for non-compliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. Biometrics 2000;56:779-788

Robins JM and Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. Communications in Statistics – Theory and Methods 1991;20;8:2609-2631

Stevenson MD, Oakley J, Chilcott JB. Gaussian process modelling in conjunction with individual patient simulation modelling. A case study describing the calculation of cost-effectiveness ratios for the treatment of osteoporosis. Med Decis Making 24 (2004) 89-100

Stevenson MD, Oakley JE, Lloyd Jones M, Brennan A, Compston JE, McCloskey EV, Selby PL. The cost-effectiveness of an RCT to establish whether 5 or 10 years of bisphosphonate treatment is the better duration for women with a prior fracture. Med Decis Making 2009; 29(6): 678-689

Stevenson MD Lloyd Jones M. M**.** The cost effectiveness of an RCT comparing alendronate with Vitamin K_1. Med Decis Making 2011; (31) 43-52

Strong M, Pilgrim H, Oakley J, Chilcott J. Structural uncertainty in health economic decision models. CHEBS discussion paper. 2009

Tappenden P, Chilcott JB, Eggington S, Oakley J, McCabe C. Methods for expected value of information analysis in complex health economic models: developments on the health economics of interferon-β and glatiramer acetate for multiple sclerosis. Health Technol Assess 2004; **8**(27) 1-92

Tappenden P, Chilcott JB, Brennan A, Pilgrim H. Systematic review of economic evidence for the detection, diagnosis, treatment and follow-up of colorectal cancer in the UK. International Journal of Technology Assessment in Health Care 2009; 25(4):470-478

Weinstein MC and Statson WB. Foundations of cost effectiveness analysis for health and medical practices. N Eng J Med 1977; 296: 716.