Article:
Kaltenthaler, Eva, Tappenden, Paul and Paisley, Suzy (2012) Reviewing the evidence to inform the population of cost-effectiveness models within health technology assessments. HEDS Discussion Paper, 12/07. (Unpublished)

HEDS Discussion Paper

No. 12.07

Reviewing the evidence to inform the population of cost-effectiveness models within health technology assessments.

Dr Eva Kaltenthaler, BSc, MSc, PhD[1] ,Dr Paul Tappenden, BA, MSc, PhD[1] ,Ms Suzy Paisley, BA, MA, PhD[1]

1. Health Economics and Decision Science, School of Health and Related Research, University of Sheffield
2. Department of Economics, University of Sheffield

Title Reviewing the evidence to inform the population of cost-effectiveness models within health technology assessments.

Dr Eva Kaltenthaler, BSc, MSc, PhD[1]

Dr Paul Tappenden, BA, MSc, PhD[1]

Ms Suzy Paisley, BA, MA, PhD[1]

[1]School of Health and Related Research (ScHARR), University of Sheffield, Regent Court, 30 Regent Street, Sheffield S1 4DA, UK

Corresponding author:

Dr Eva Kaltenthaler, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield S1 4DA, UK; Telephone: 44 114 2220810; E-mail: e.kaltenthaler@sheffield.ac.uk; Fax: 44 114 2724095

Abstract

Background

Health technology assessments (HTA) typically require the development of a cost-effectiveness model which necessitates the collection of information, in addition to clinical effectiveness evidence, to populate the model parameters. The reviewing activity associated with model development should be transparent and reproducible but can result in a tension between being both timely and systematic. Little procedural guidance exists in this area. The purpose of this paper is to provide guidance, informed by focus groups on what might constitute a systematic and transparent approach to reviewing information to populate model parameters where little procedural guidance exists.

Methods

A series of focus groups were held with HTA experts in the UK including systematic reviewers, information specialists and health economic modellers in order to explore these issues. Framework analysis was used to analyse the qualitative data elicited during the focus groups.

Results

Suggestions included the use of rapid reviewing methods and the need to consider the trade-off between relevance and quality. The need for transparency in the reporting of review methods was emphasised. It was suggested that additional attention should be given to the reporting of parameters deemed to be more important to the model or where the decision regarding the choice of evidence was not clear cut.

Discussion

These recommendations form part of a Technical Support Document produced for the NICE Decision Support Unit in the UK. It is intended that these recommendations will help to ensure a more systematic, transparent and reproducible process for the review of model parameters within HTA.

Introduction

Health technology assessment (HTA) reports used to inform evidence based decisions concerning the use of healthcare interventions typically involve the development of a systematic review of clinical effectiveness and the development of a cost-effectiveness model. By its very nature, the development of the health economic model requires information beyond clinical efficacy such as health utilities, resource use and costs amongst others. In addition, the model structure may require evidence to inform judgments concerning the plausibility of relationships between intermediate and final endpoints. The way in which this is done can have a fundamental impact on the results of the model and ultimately the decision outcome (Coyle and Lee, 2002). Some of these information needs are illustrated in Figure 1 below.

Figure 1: Types of evidence used to inform models

There are a number of issues that need to be considered when reviewing evidence to inform the specification and population of cost-effectiveness models. These include the timelines for HTA, which may be restrictive, and the need for methods which are systematic, transparent and reproducible in order to minimse the risk of bias and therefore produce more

robust results. If model results are to be considered credible, then researchers need to be transparent about how that model came about and why certain inputs should be considered reliable. Sources of evidence may include: randomised controlled trials (RCTs), observational evidence and other clinical studies, registry databases, elicitation of expert clinical judgement, existing cost-effectiveness models, routine data sources and health valuation studies. Whilst reviewing processes are often used to identify data for economic models, it is less usual for model reports to describe and justify how they have identified and synthesised the evidence beyond the efficacy data or for reports to set out criteria against which the relevance and quality of the evidence are assessed. (Cooper et al, 2007). Although some of the issues surrounding reviewing evidence for models have been discussed previously (Cooper et al, 2007; Cooper et al, 2005; Philips et al, 2004; Coyle & Lee, 2002; Shemilt, et al, 2008; Marsh 2010; Paisley 2010) there remains very little formal guidance with respect to best practice in this area. Briggs et al (2012) in their ISPOR-SMDM Modelling Good Research Practice report, recommend that analysts should conform to the broad principles of evidence-based medicine and avoid "cherry picking" the best single source of evidence. Coyle and Lee (2002) demonstrated that using different sources of data can have a large impact on the results and highlighted that there is lack of agreement as to what constitutes good evidence for specific data inputs in economic models. It has further been argued by Chilcott et al (2010) that one potential source of errors in health technology assessment models is the separation of the information gathering, reviewing and modelling functions.

Methodological guidance regarding the reviewing of evidence to inform model parameters, apart from clinical effectiveness from the National Institute for Health and Clinical Excellence (NICE) states: "For all parameters (including effectiveness, valuation of HRQL and costs) *a* systematic consideration of possible data sources is required" (NICE 2008). This absence of clarity presents a considerable challenge to organisations submitting evidence to NICE as a full systematic review is not required for each parameter yet it is not clear what a "systematic

consideration" is. A recent Technical Support Document from the NICE Decision Support Unit (Kaltenthaler et al, 2011) considers the requirements and provides methodological guidance for identifying and reviewing evidence to inform models of cost-effectiveness, in particular model parameter estimates, in the NICE Technology Appraisal Process. Part of the TSD provides guidance on methods for reviewing model parameter data in a systematic fashion. It draws distinctions between systematic reviews and reviewing in the context of informing model parameters and demonstrates how the key components of systematic review methods can be used to systematise and make explicit the choices involved in selecting evidence to inform models. The purpose of this paper is to summarise the guidance, informed by the focus groups provided in the TSD regarding the reviewing of evidence to inform model parameter estimates with suggestions as to what might constitute a systematic and transparent approach where there is not a requirement to use conventional systematic review methods but where little procedural guidance exists.

Methods

A series of focus groups were used to gather information on the issues around reviewing for model parameters and provide the basis of the recommendations covered in the TSD. An initial focus group was held with 17 researchers who had extensive experience in HTA including five systematic reviewers, two information specialists and nine health economic modellers in January 2010. The researchers were all from the School of Health and Related Research (ScHARR) at the University of Sheffield. A range of different people with different areas and levels of expertise were purposively invited to attend the focus group in order to reflect the breadth of input into the model development process. A topic guide was developed to structure the discussion within the focus group and was informed through discussion with experts in the field of HTA. The topic guide included questions covering:

- current practice
- adequate information
- timing at which reviewing activity takes place
- ideal practice
- areas for further research

A subsequent seminar was held at ScHARR in June 2010 and used as a member checking device. All of the 17 researchers, among others were invited to the seminar where further discussions were held on each of the key themes identified in the focus group.

The key issues identified through the initial focus group and seminar were presented at a workshop held on 7 February 2011 which included 12 participants from UK universities considered to be experts in the field of HTA including seven modellers, one health economist, one statistician, two information specialists and two reviewers. The workshop consisted of three focus group sessions. The topic guide for the workshop focus group covered the following topics:

- model development

- time constraints
- sufficient evidence
- communication and team work
- problem structuring
- identification of evidence
- reviewing methods
- recommendations for reporting

Ethics approval for all focus groups was obtained from the University of Sheffield. The focus groups were facilitator-led (EK) and were all recorded using digital media with the recordings transcribed verbatim. Qualitative Framework analysis (Ritchie and Spencer, 1994) was used to draw out the key themes and subthemes from the transcribed data. Coding was checked by a second researcher.

The information gained through the information gathering activities described above informed the development of guidance on the reviewing of evidence to populate model parameters. A report of the draft recommendations were shared with all workshop participants for comment.

Results

The key themes related to the reviewing of model parameters identified from the focus groups were: (1) selection and prioritisation of data to inform parameter estimates; (2) reviewing methods; (3) minimising bias, (4) hierarchies of evidence; (5) study selection; (6) assessment of evidence and (7) evidence synthesis and analysis. These themes informed the statements presented below. More detail of the focus group findings are available from Kaltenthaler et al (2012a) and Kaltenthaler et al (2012b), while more detail on the resulting recommendations can be found in the TSD (Kaltenthaler et al, 2011).

Theme 1: Selection and prioritisation of data to inform parameter estimates

Every model parameter will need to be estimated, therefore the choices made regarding the values selected need to be explained and justified. The choice of estimate will often be made according to some trade-off or weighing up of the available options, rather than according to rigid, pre-defined criteria. This may be because an estimate is required and there will usually at best be a range of options, all of which may fall short of what would be considered ideal to differing degrees. The nature of the trade-off between selecting alternative parameter values will often include elements relating to quality versus relevance for each option. Procedures associated with undertaking systematic reviews can be used to make the process of choosing evidence systematic and transparent. However, given the differences between models and systematic reviews, the purposes for which these procedures are undertaken and the sequence in which they are undertaken may differ. In addition, time and resource constraints will also impact on how they are undertaken. These processes need to be transparent, justifiable and replicable. It is important to prioritise parameters and focus reviewing resources on those most likely to impact on model outputs, bearing in mind that the importance of parameters is subject to change during the course of the modelling process. Although some parameters will be identified as important to the model early on in the process, the importance of some other parameters will only be identified later in the process.

Theme 2: Reviewing methods

Due to time and resource constraints it may be necessary to use rapid review methods to identify and select evidence to inform certain model parameters. Rapid review methods are not ideal as there is the potential for missing relevant information. It is therefore essential that methods are reported in a transparent manner and that the limitations and potential biases are addressed. Some rapid methods used for reviewing clinical effectiveness evidence may be applicable for the reviewing of model parameters, including the use of restricted review questions. Other potentially relevant rapid review methods in this context include reduced formal quality assessment, data extraction of key outcomes only and reduced levels of synthesis. Transparent reporting of methods is essential.

Theme 3: Minimising bias

A variety of potential biases may be introduced through the process of reviewing evidence to inform model parameters values. This may include biases introduced through the use of less thorough searching and reviewing methods as well as biases through the purposive selection of evidence to create more or less favourable. One option to reduce such bias is to ensure and to demonstrate that more than one member of the team is involved with making decisions where choices about values need to be made. This is partly because there may be more than one plausible option, and a joint decision may provide a more robust and systematic approach to considering the advantages and disadvantages of each. Those involved in this decision making process may include clinical advisors, information specialists, systematic reviewers and modellers on the team.

Theme 4: Hierarchy of evidence

Types of evidence used to populate models will vary considerably. Hierarchies of evidence sources may be of use as a means of judging the quality of individual parameter estimates and aid the study selection process. However, whilst hierarchies of evidence may be useful,

there are other issues to consider, for example the quality of the individual studies. The Grading of Recommendations Assessment, Development and Evaluation (GRADE) system offers some promise in this regard and provides a framework for rating the quality of evidence collected (or derived) from all potential sources of all data components that may be used to populate model parameters, including research based and non-research based sources.

Theme 5: Study selection

The definition of what is required may be based on an initial understanding of what will constitute "relevant" evidence. The objective is to identify a set of possible options from which choices will be made. One option is that initially strict selection criteria may be applied. If no relevant studies are identified, the selection criteria may be broadened. It is important to explain the process used and why it was chosen in order to justify the choices and maintain transparency. For many parameters there may be very few sources and potential studies to use or alternatively many good quality studies to choose from. If several potentially relevant studies are identified, slightly stricter selection criteria can be applied. For large numbers of sources, study selection using standard systematic review processes of screening for titles, abstracts and full texts can be made. It is important to be as transparent as possible about the judgements being made when selecting studies for example, stating which studies were deemed to be most relevant to UK clinical practice.

Evidence for model parameters will need to be assessed on the basis of relevance to the context of the decision problem, as well as quality. By assessing relevance first, a large number of studies may be eliminated. Criteria for relevance are ideally established *a priori*. However, it is important to recognise that it is not possible to have pre-specified criteria for every parameter as information needs will change and information that was not expected may be identified iteratively. Relevance criteria may therefore change throughout the project, hence flexibility is essential. What remains important is that the criteria or factors

which inform the choice of evidence remain transparent. Anticipated evidence requirements, as perceived during the earlier stages of model development, may be identified by adopting an explicit stage of conceptual model development before embarking on the mathematical model. When the final model is developed it is important to be clear how this deviated from the initial plan and why.

Theme 6: Assessment of evidence

After appraising studies for relevance, they can then be assessed for quality, preferably using standardised quality assessment tools. In this context, quality assessment may be difficult due to the absence of standardised methods for all types of information used to populate the model. Also, some studies may be poorly reported. It may be possible to establish quality assessment criteria *a priori.* An example of this is data collected for utility studies which may include study recruitment procedures, inclusion and exclusion criteria, description of the background characteristics of the sample population from whom values are obtained, response rates and follow-up data (Brazier et al, 2010). Other issues to consider may include the type of reporting (self or proxy), follow-up rates, number of patients, location, methods of elicitation among other issues. Establishing very broad *a priori* criteria may be necessary initially making quality assessment closely lined with selection of evidence. Criteria may change according to the availability and relevance of existing evidence. For example, "there were five options and we chose one because of the reasons a, b and c." This also captures the necessary trade-off between relevance and quality. It is important to be clear about the factors or criteria that drive the choice and to examine the implications of that choice. This level of transparency will allow judgements to be made as to whether or not a reasonable choice has been made. As it can be very time consuming to judge the quality of all potentially relevant studies, adjusting them according to relevance and rigour may not be practical. Some types of data are of potentially very poor quality and it can be very difficult to identify appropriate sources of information, for example for cost

data. These are not limitations of the cost effectiveness model but rather of the evidence base and as such these evidence gaps should be exposed and reported.

Data to be extracted from studies may include study date, information on disease area and patients (age, sex, co-morbidities), study methods, outcomes and other important descriptive details. This can be set out *a priori* and presented in a way to make it easy for the reader to compare and contrast the characteristics of the available studies from which a selection has been made, for example using tables and/or graphs. This level of detail is not appropriate for all parameter values but should be reserved for those decisions whereby none of the available studies are clearly superior or whereby evidence available to inform a particular parameter or set of parameters is notably weak. When extracting data from studies it is important to provide information for all of the potentially relevant studies. By providing a summary of all potentially relevant studies, the reader is able to assess the study differences and heterogeneity more accurately and to examine the spread of evidence. Information from the studies which are not selected may be used to inform the sensitivity analysis. Inconsistencies between different estimates should be represented. Although the results presented may be wide when using the available studies, it is important to show how the range of values between disease stages or different baseline event rates for example, are driving the model results. It is recognised that these suggestions may be quite time consuming and there may be time and reporting constraints within a technology assessment report. However, the overriding objective should be to present the information as transparently as possible.

Theme 7: Evidence synthesis and analysis

For many types of model parameters, the issue of synthesis may not be considered relevant due to study heterogeneity. Often only one or two values are appropriate for use in populating a model parameter. The issue of synthesis obviously becomes important when

there are more than one or two potentially relevant studies. A decision needs to be made as to whether complex synthesis methods may provide a meaningful value for a parameter. In some instances however, it may be simpler and more defensible to select the value from the most appropriate and relevant study as opposed to using a weighting system for pooling estimates. However, in instances whereby a quantitative synthesis is not undertaken, this should be justified explicitly. The choice of available evidence should be made clear and the implication of choosing one source from a number of available options explored through sensitivity analyses.

Discussion

This paper presents seven key themes exploring issues around the reviewing of evidence to populate model parameters informed by a series of focus groups with experts in health technology assessment in the UK. Selection and prioritisation of data to inform parameter estimates was considered to be important by the participants. There was agreement that reviewing effort should be prioritised around the important model parameters and reviewing methods chosen commensurate with the parameter's importance. Caution was advised however as the importance of certain model parameters may change as other parts of the model are developed and refined. Also mentioned was the applicability of rapid reviewing methods. As suggested by Watt et al (2008) and Ganann et al (2010), rather than developing a formalised methodology to conduct rapid reviews, which may be inappropriate and oversimplified, emphasis should be placed on the transparent reporting of methods. Having more than one team member involved in choosing appropriate parameter estimates was considered an option in order to minimise bias.

Hierarchies of evidence as suggested by Coyle et al (2010) were considered to be a potentially useful tool for guiding the choice of evidence to inform parameter values although concerns were raised as hierarchies only consider the quality of the evidence type. GRADE may be a useful tool in that it allows flexibility in the quality assessment process to include additional considerations alongside internal validity, including (crucially for most data components used to populate model parameters) applicability to the specific decision problem at hand, which is part of the 'indirectness' criterion in GRADE (Shemilt et al, 2010). Sources of data that could be incorporated include national disease registers, claims, prescriptions or hospital activity databases, or standard reference sources such as drug formularies or collected volumes of unit costs (Guyatt et al, 2011, Brunetti et al, 2012 Shemilt et al, 2012).

Both quality and relevance were considered to be important when undertaking study selection. Quality assessment of the evidence was thought to present some challenges due to the absence of quality assessment tools for many types of evidence used to populate cost effectiveness models. With regard to choices made related to evidence synthesis and analysis, these need to be made clear and explicitly justified. Study selection processes should be clearly reported and there should be transparency around what judgements have been made regarding study selection.

The issues raised in this research are important in the field of health technology assessment. There is potential for these issues to have a big impact on the development of the cost effectiveness model and thus on the decision making process. There is a need for agreed standardised practice in this area while still maintaining flexibility and adaptability to suit the needs of individual health technology assessments. As Cooper et al (2007) states "it is imperative that evidence for all model parameters is identified systematically, quality assessed and where applicable pooled using explicit criteria and reproducible methods." The findings from this research support this statement. A systematic, transparent and reproducible process is essential for the development of cost effectiveness models to support health technology assessments. Further research is needed in this area and includes the need for the development of appropriate rapid reviewing methods; quality assessment tools for non-standard sources of evidence; investigation into the use of hierarchies of evidence and GRADE; development of the methods used for the selection of evidence and development of reporting standards.

References

Brazier, J., Papaioannou, D., Cantrell, A., Paisley, S., Herrmann, K. Identifying and reviewing health state utility values for populating decision models. In: Shemilt I. Mugford M., Vale L., Marsh K., Donaldson C., eds. *Evidence-based decisions and economics: health care, social welfare, education and criminal justice.* Wiley-Blackwell; Oxford: 2010.

Briggs A, Fenwick E, Karnon J, Paltiel AD, Schulpher, M, Weinstein M.2012 Model Parameter Estimation and Uncertainty Analysis: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6 (DRAFT) http://www.ispor.org/workpaper/modeling_methods/model-parameter-estimation-uncertainty.asp.

Brunetti, M., Shemilt, I., Pregno, S., Vale, L., Oxman, A., Lord, J. et al. GRADE guidelines: 10. Special challenges - quality of evidence for resource use. *Journal of Clinical Epidemiology*. 2012; (in press)

Chilcott JB, Tappenden P, Rawdin A, Johnson M, Kaltenthaler E, Paisley S, Papaioannou D, Shippam A. Avoiding and identifying errors in health technology assessment models: qualitative study and methodological review. Health Technology Assessment 2010; 14 (25).

Cooper N, Sutton AJ, Ades AE , Paisley S, Jones DR et all. Use of evidence in economic decision models: practical issues and methodological challenges. Health Economics 2007; 16: 1277-1286.

Cooper, N., Coyle, D., Abrams, K., Mugford, M., Sutton, A. Use of evidence in decision models: An appraisal of health technology assessments in the UK to date. *Journal of Health Services Research and Policy* 2005; 10(4):245-250.

Coyle D & Lee K Evidence-based economic evaluation: how the use of different data sources can impact results. In Donaldson C, Mugford M and Vale L, eds. In Evidence-based Health Economics. London: BMJ Books, 2002.

Coyle, D., Lee, K., Cooper, N. Use of evidence in decision models. In: Shemilt I.,Mugford M., Vale L., Marsh K., Donaldson C., eds. *Evidence-based decisions and economics: health care, social welfare, education and criminal justice.* Wiley-Blackwell; Oxford: 2010.

Gannon, R., Ciliska, D., Thomas, H. Expediting systematic reviews: methods and implications of rapid reviews. *Implementation Science* 2010; 5(56).

Guyatt, G., Oxman, A., Schunemann, H., Tugwell, P., Knottnerus, A. GRADE guidelines: A new series of articles in the Journal of Clinical Epidemiology. *Journal of Clinical Epidemiology*. 2011; 64 (4): 380-382.

Kaltenthaler E, Tappenden P, Paisley S.2012a Reviewing the evidence used in cost-effectiveness models in health technology assessment: a qualitative investigation of current concerns and future research priorities. HEDS Discussion Paper number 12/01, 2012; http://www.shef.ac.uk/polopoly_fs/1.165496!/file/12.01.pdf

Kaltenthaler E, Essat M, Tappenden P, Paisley S. 2012b A qualitative study to explore the identification and reviewing of evidence used in cost-effectiveness models in health technology assessments HEDS Discussion Paper 2012 http://www.shef.ac.uk/scharr/sections/heds/dps-2012.

Kaltenthaler E, Tappenden P, Paisley S, Squires H. NICE DSU Technical Support Document 13: Identifying and reviewing evidence to inform the conceptualisation and population of

cost-effectiveness models. Report by the NICE Decision Support Unit, May 2011. http://www.nicedsu.org.uk/TSD%2013%20model%20parameters.pdf

Marsh K Chapter 2: The role of review and synthesis methods in decision models. In: Shemilt I, Mugford M, Vale L, Marsh K, Donaldson C, eds., In Evidence-based decisions and economics: health care, social welfare, education and criminal justice. Oxford: Wiley-Blackwell, 2010.

NICE Guide to the methods of technology appraisal 2008 http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf

Paisley S. Classification of evidence used in decision-analytic models of cost effectiveness: a content analysis of published reports. *IJTAHC*. 2010; 26(4):458-462.

Philips Z , Ginnelly L, Schulpher M, Claxton K, Golder S, Riemsma R, Woolacott N, Glanville J. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. Health Technology Assessment 2004; 8(36).

Ritchie J, Spencer L. Qualitative data analysis for applied policy research. Bryman A, Burgess RG, eds. , In Analysing qualitative data. London: Routledge, 1994.

Shemilt, I., Vale, L., Ruiz, F., Drummond, M., Brunetti, M., Lord, J. GRADE guidelines: 17. Going from evidence to recommendations - evidence profiles and decision models for economic evaluation. *Journal of Clinical Epidemiology*. 2012; (in press).

Shemilt, I., Mugford, M., Vale, L., Marsh, K., Donaldson, C., Drummond, M. Evidence synthesis, economics and public policy. *Research Synthesis Methods* 2010; 1(2):126-135.

Shemilt I, Mugford M, Byford S, Drummond M, Eisenstein E, Knapp M, Mallender J, McDaid D, Vale L, Walker D. Chapter 15. Incorporating economics evidence, In: Higgins JPT, Green S (editors) Cochrane Handbook for systematic reviews of interventions. Chichester (UK): John . Wiley and Sons, 2008:

Watt A , Cameron A, Sturm L, Lathlean T, Babidge W, Blamey S, Facey K, Hailey D, Norderhaug I, Maddern G. Rapid reviews versus full systematic reviews: an inventory of current methods and practice in health technology assessment. International Journal of Technology Assessment in Health care 2008; 24(2): 133-139.