Avoiding and identifying errors in HTA models: qualitative study and methodological review

Avoiding and identifying errors in health technology assessment models: qualitative study and methodological review

J Chilcott,* P Tappenden, A Rawdin, M Johnson, E Kaltenthaler, S Paisley, D Papaioannou and A Shippam

School of Health and Related Research (ScHARR), Regent Court, Sheffield, UK

*Corresponding author

Executive summary

Health Technology Assessment 2010; Vol. 14: No. 25
DOI: 10.3310/hta14250

Health Technology Assessment
NIHR HTA programme
www.hta.ac.uk

Executive summary

Background

The National Institute for Health and Clinical Excellence (NICE) in England and Wales and similar structures elsewhere are required to make health policy decisions that are relevant, evidence-based and transparent. Decision-analytic modelling is well placed to support this process. The key role that models play is, however, reliant on their credibility. Credibility in models depends on a range of factors including the coherence of the model with the beliefs and attitudes of the decision-makers, the decision-making framework within which the model is used, the validity of the model in being an adequate representation of the problem in hand and the quality of the model. A recent study investigating the quality of models used to support national policy-making in Australia reported that 203 of 247 models reviewed were considered by the investigators to be flawed in some respect.

Errors in mathematical decision models or simulation exercises are a natural and unavoidable part of the software development process. However, little attention has been paid to the processes involved in model development. Good practice guidance either acknowledges the absence of methodological and procedural guidance on model development and testing or makes no reference to the issue. Numerous error avoidance/identification strategies could be adopted, potentially impacting upon the whole range of disciplines involved in the decision support process, including information specialists, health economists, statisticians, systematic reviewers and operational research modellers. However, it is difficult to evaluate the merits of strategies for improving the credibility of models without first developing an understanding of error types and causes. This study seeks to understand the nature of errors within HTA models, to describe current processes for minimising the occurrence of such errors and to develop a first classification of errors to aid discussion of potential strategies for avoiding and identifying errors.

Aim and objectives

The study aims to describe the current comprehension of errors in the HTA modelling community and to generate a taxonomy of model errors to facilitate discussion and research within the HTA modelling community, on strategies for reducing errors and improving the robustness of modelling for HTA decision support.

The study has four primary objectives:

1. to describe the current understanding of errors in HTA modelling, focusing specifically on:
   i. types of errors
   ii. how errors are made
2. to understand current processes applied by the technology assessment community for avoiding errors in the development, debugging models and critically appraising models for errors
3. to use HTA modellers' perceptions of model errors together with the wider non-HTA literature to develop a taxonomy of model errors
4. to explore potential methods and procedures to reduce the occurrence of errors in models.

In addition, the study describes the model development process as perceived by practitioners working within the HTA community; this emerged as an intermediate objective for considering the occurrence of errors in models.

Methods

The study involved two parallel methodological strands. The first strand involved a methodological review of literature discussing model errors. The second strand comprised in-depth qualitative interviews with 12 HTA modellers, including representatives from academic and commercial modelling sectors. All qualitative data were analysed using the Framework approach. Descriptive and explanatory accounts were used to interrogate the data within and across themes and subthemes.

The themes identified within the analysis are:

- organisation, roles and communication
- the model development process
- definition of error
- types of model error
- strategies for avoiding errors
- strategies for identifying errors
- barriers and facilitators.

Results

Current understanding of modelling errors in the HTA community

There is a general consensus that an important part of the definition of what constitutes an error is its impact on decision-making. This indicates that a pragmatic approach is generally taken across the HTA modelling community. Despite this implied common outlook, there was no common language used in the discussion of modelling errors and inconsistency in the perceived boundaries of what constitutes an error. When asked explicitly about the definition of model error, there was a tendency for interviewees to exclude matters of judgement from being errors and focus on 'slips' and 'lapses'. However, discussion of slips and lapses comprised less than 20% of the discussion on types of errors. When considering how individual elements of the modelling process might contribute to flaws in decision-making, interviewees devoted 70% of the discussion to softer elements of the process of defining the decision question and conceptual modelling, mostly the realms of judgement, skills, experience and training.

Although the original focus of the study concerned model errors, when considering methods of improving modelling practice it may be more useful to refer to modelling risks rather than the more black and white term modelling errors. Several interviewees discussed concepts of validation and verification, with notable consistency in interpretation. Verification was taken to mean the process of ensuring that the computer model correctly implemented the intended model, whereas validation meant the process of ensuring that a model is fit for purpose (hence subsuming verification). The qualitative analysis highlights considerable variation in modelling practice across the HTA modelling community, particularly in terms of the demonstration of explicit conceptual modelling before implementation. Methodological literature suggests that overall validity comprises conceptual model validity, verification of the computer model, and operational validity of the use of the model in addressing the real-world problem. In the absence of explicit conceptual modelling, the concept of overall model validity breaks down.

The methodological literature on verification and validation of models makes reference to the Hermeneutic philosophical position that recognises that objectivity is unachievable and suggests that meaning is created through intersubjective communication. The literature proposes that it is the interaction between the modeller and client in developing mutual understanding of a model that establishes a model's significance and its warranty. This position highlights that model credibility is the central concern to decision-makers in using models as an input to the decision-making process. This highlights the point that the concept of model validation should not be externalised from the decision-makers and the decision-making process.

A taxonomy of HTA modelling errors

Interviewees collectively demonstrated examples of all major error types identified in the literature on errors in end-user developer spreadsheet systems. Broad error domains include: (1) errors in the description of the decision problem; (2) errors in model structure; (3) errors in the use of evidence; (4) errors in the implementation of the model; (5) errors in the operation of the model; and (6) errors in the presentation and understanding of results. Each error domain contains a breakdown of error types and their potential root causes. The HTA errors classifications were compared against existing classifications of model errors identified within the literature.

Current strategies for avoiding errors

The qualitative analysis suggests that a range of techniques and procedures are currently used to avoid errors in HTA models. Importantly, there is some overlap between methods used to identify errors and methods used to avoid errors in models. Strategies for error avoidance are loosely defined as either processes or techniques; the former relate to issues in the model development process, whereas the latter relate to techniques of implementation. Generally, the 'techniques' are explicit and can be interpreted as relating to *how* something should be done, for example, implementing a specific

model layout. Conversely, the 'processes' recognise an unfulfilled requirement and acknowledge that something should be done as part of the model development process, yet in many cases this is not accompanied by a clear strategy for achieving the required goal. Current methods for avoiding errors include: engaging with clinical experts, clients and decision-makers to ensure mutual understanding, producing written documentation of the proposed model, explicit conceptual modelling, e.g. using diagrams and sketches, stepping through skeleton models with experts, ensuring transparency in reporting, adopting standard housekeeping techniques, and ensuring that those parties involved in the model development process have sufficient and relevant training. Clarity and mutual understanding were identified as key issues. Current strategies supporting these aspects of model development are expressed as process requirements, for example, establishment of long-term clinical input and iterative negotiation with the decision-maker or client may be used to avoid errors in the description of the decision problem. Although a number of techniques were suggested by the interviewees, for instance, sketching out clinical pathways, their use appears to be partial, and the extent of their use appears to vary considerably between individual modellers. Despite an acknowledgement of the importance of these methods, their current implementation is not framed within an overall strategy for structuring complex problems.

Current strategies for identifying errors in HTA models

Methods for identifying errors in HTA models include checking face validity, assessing whether model results appear reasonable, black-box testing strategies, testing internal consistency and predictive validity, checking model input values, double-programming and peer review. These strategies largely relate to specific techniques (rather than processes) that may be applied by third-party scrutiny. However, the specific target of the techniques, i.e. the types of error that the technique is intended to identify, is not always clear. The majority of methods may be used to identify *symptoms* of errors; however, the root cause may be entirely unclear. This represents a considerable challenge in the peer review of models. The same may be true of certain black-box validation techniques; only the tests of the underlying logic of the model guarantee the presence of an error.

Those error identification methods which do map directly to specific aspects of the taxonomy are diagnostic in nature; mismatches in model results and expectations are indicative of the presence of model error. Those aspects which map to any or all points in the taxonomy are effectively non-specific model screening methods; the presence of differences between models and prior expectations are not necessarily the result of a model error.

Recommendations

- Published definitions of overall model validity comprising conceptual model validation, verification of the computer model, and operational validity of the use of the model in addressing the real-world problem are consistent with the views expressed by the HTA community and are therefore recommended as the basis for further discussions of model credibility.
- Discussions of model credibility should focus on risks, including errors of implementation, errors in matters of judgement and violations – violations being defined as puffery, fraud or breakdowns in operational procedures.
- Discussions of modelling risks should reflect the potentially complex network of cognitive breakdowns that lead to errors in models and subsequent failures in decision support. Existing research concerning the cognitive basis of human error should be brought into the examination of modelling errors.
- There is a need to develop a better understanding of the skills requirements for the development, operation and use of HTA models.
- The qualitative interviews highlighted a number of barriers to model checking. However, it was indicated that increasing time and resources would not necessarily improve model checking activities without a matched increase in their prioritisation.
- The authors take the view, as supported within the methods literature, that it is the interaction between the modeller and client in developing mutual understanding of a model that establishes a model's significance and its warranty. This position highlights that model credibility is the central concern to decision-makers in using models. It is crucial then that the concept of model validation should not be externalised from the decision-makers and the decision-making process.

Research recommendations

Verification and validation

Most modellers instinctively take a pragmatic approach to developing model credibility. Further research on the theory of model verification and validation is required to provide a solid foundation for (1) the model development process and (2) processes for making evidence-based policy and guidance.

Model development process

Further research is required in the model development process. Two specific areas were identified:

- techniques and processes for structuring complex HTA models, developing mutual understanding and identifying conflicting perceptions between stakeholders in the decision problem
- development of the model design process and mechanisms for reporting and specifying models.

Errors research

There is little evidence to suggest that models are improving in reliability. Further research is required to define, implement and evaluate modifications to the modelling process with the aim of preventing the occurrence of errors and improving the identification of errors in models. Mechanisms for using National Institute for Health Research-funded model developments to facilitate this research could be pursued, for example, by providing research funding for the specification and evaluation of enhanced modelling methods within National Institute for Health Research-funded studies.

Publication

Chilcott JB, Tappenden P, Rawdin A, Johnson M, Kaltenthaler E, Paisley S, *et al.* Avoiding and identifying errors in health technology assessment models: qualitative study and methodological review. *Health Technol Assess* 2010;**14**(25).

NIHR Health Technology Assessment programme

The Health Technology Assessment (HTA) programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The research findings from the HTA programme directly influence decision-making bodies such as the National Institute for Health and Clinical Excellence (NICE) and the National Screening Committee (NSC). HTA findings also help to improve the quality of clinical practice in the NHS indirectly in that they form a key component of the 'National Knowledge Service'.

The HTA programme is needs led in that it fills gaps in the evidence needed by the NHS. There are three routes to the start of projects.

First is the commissioned route. Suggestions for research are actively sought from people working in the NHS, from the public and consumer groups and from professional bodies such as royal colleges and NHS trusts. These suggestions are carefully prioritised by panels of independent experts (including NHS service users). The HTA programme then commissions the research by competitive tender.

Second, the HTA programme provides grants for clinical trials for researchers who identify research questions. These are assessed for importance to patients and the NHS, and scientific rigour.

Third, through its Technology Assessment Report (TAR) call-off contract, the HTA programme commissions bespoke reports, principally for NICE, but also for other policy-makers. TARs bring together evidence on the value of specific technologies.

Some HTA research projects, including TARs, may take only months, others need several years. They can cost from as little as £40,000 to over £1 million, and may involve synthesising existing evidence, undertaking a trial, or other research collecting new data to answer a research problem.

The final reports from HTA projects are peer reviewed by a number of independent expert referees before publication in the widely read journal series *Health Technology Assessment*.

Criteria for inclusion in the HTA journal series

Reports are published in the HTA journal series if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the referees and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search, appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

The research reported in this issue of the journal was commissioned by the HTA programme as project number 07/54/01. The contractual start date was in May 2007. The draft report began editorial review in March 2009 and was accepted for publication in November 2009. As the funder, by devising a commissioning brief, the HTA programme specified the research question and study design. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the referees for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

ISSN 1366-5278

Published by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk), on behalf of NETSCC, HTA.
Printed on acid-free paper in the UK by Henry Ling Ltd, The Dorset Press, Dorchester.