

## **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



May 2010  
10.3310/hta14250

**Health Technology Assessment**  
**NIHR HTA programme**  
[www.hta.ac.uk](http://www.hta.ac.uk)





### **How to obtain copies of this and other HTA programme reports**

An electronic version of this title, in Adobe Acrobat format, is available for downloading free of charge for personal use from the HTA website ([www.hta.ac.uk](http://www.hta.ac.uk)). A fully searchable DVD is also available (see below).

Printed copies of HTA journal series issues cost £20 each (post and packing free in the UK) to both public **and** private sector purchasers from our despatch agents.

Non-UK purchasers will have to pay a small fee for post and packing. For European countries the cost is £2 per issue and for the rest of the world £3 per issue.

How to order:

- fax (with **credit card details**)
- post (with **credit card details** or **cheque**)
- phone during office hours (**credit card** only).

Additionally the HTA website allows you to either print out your order or download a blank order form.

### **Contact details are as follows:**

Synergie UK (HTA Department)  
Digital House, The Loddon Centre  
Wade Road  
Basingstoke  
Hants RG24 8QW

Email: [orders@hta.ac.uk](mailto:orders@hta.ac.uk)

Tel: 0845 812 4000 – ask for ‘HTA Payment Services’  
(out-of-hours answer-phone service)

Fax: 0845 812 4001 – put ‘HTA Order’ on the fax header

### **Payment methods**

#### *Paying by cheque*

If you pay by cheque, the cheque must be in **pounds sterling**, made payable to *University of Southampton* and drawn on a bank with a UK address.

#### *Paying by credit card*

You can order using your credit card by phone, fax or post.

### **Subscriptions**

NHS libraries can subscribe free of charge. Public libraries can subscribe at a reduced cost of £100 for each volume (normally comprising 40–50 titles). The commercial subscription rate is £400 per volume (addresses within the UK) and £600 per volume (addresses outside the UK). Please see our website for details. Subscriptions can be purchased only for the current or forthcoming volume.

### **How do I get a copy of HTA on DVD?**

Please use the form on the HTA website ([www.hta.ac.uk/htacd/index.shtml](http://www.hta.ac.uk/htacd/index.shtml)). *HTA on DVD* is currently free of charge worldwide.

---

The website also provides information about the HTA programme and lists the membership of the various committees.

# **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

**Declared competing interests of authors:** none

Published May 2010

DOI: 10.3310/hta14250

---

This report should be referenced as follows:

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).

*Health Technology Assessment* is indexed and abstracted in *Index Medicus/MEDLINE*, *Excerpta Medica/EMBASE*, *Science Citation Index Expanded (SciSearch®)* and *Current Contents®/Clinical Medicine*.

# 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.

The views expressed in this publication are those of the authors and not necessarily those of the HTA programme or the Department of Health.

Editor-in-Chief: Professor Tom Walley CBE  
Series Editors: Dr Martin Ashton-Key, Dr Aileen Clarke, Professor Chris Hyde,  
Dr Tom Marshall, Dr John Powell, Dr Rob Riemsma and Professor Ken Stein  
Editorial Contact: edit@southampton.ac.uk

ISSN 1366-5278

## © 2010 Queen's Printer and Controller of HMSO

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (<http://www.publicationethics.org/>). This journal may be freely reproduced for the purposes of private research and study and may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising.

Applications for commercial reproduction should be addressed to: NETSCC, Health Technology Assessment, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Published by Prepress Projects Ltd, Perth, Scotland ([www.prepress-projects.co.uk](http://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.



## Abstract

### **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

**Background:** Health policy decisions must be relevant, evidence-based and transparent. Decision-analytic modelling supports this process but its role is reliant on its credibility. Errors in mathematical decision models or simulation exercises are unavoidable but little attention has been paid to processes in model development. Numerous error avoidance/identification strategies could be adopted but 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.

**Objectives:** The study aims to describe the current comprehension of errors in the HTA modelling community and generate a taxonomy of model errors. Four primary objectives are to: (1) describe the current understanding of errors in HTA modelling; (2) understand current processes applied by the technology assessment community for avoiding errors in development, debugging and critically appraising models for errors; (3) use HTA modellers' perceptions of model errors with the wider non-HTA literature to develop a taxonomy of model errors; and (4) explore potential methods and procedures to reduce the occurrence of errors in models. It also describes the model development process as perceived by practitioners working within the HTA community.

**Data sources:** A methodological review was undertaken using an iterative search methodology. Exploratory searches informed the scope of interviews; later searches focused on issues arising from the interviews. Searches were undertaken in February 2008 and January 2009. In-depth qualitative interviews were performed with 12 HTA modellers from academic and commercial modelling sectors.

**Review methods:** 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: organisation, roles and communication; the model development process; definition of error; types of model error; strategies for avoiding errors; strategies for identifying errors; and barriers and facilitators.

**Results:** There was no common language in the discussion of modelling errors and there was inconsistency in the perceived boundaries of what constitutes an error. Asked 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', but discussion of slips and lapses comprised less than 20% of the discussion on types of errors. 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. The original focus concerned model errors, but it may be more useful to refer to modelling risks. Several interviewees discussed concepts of validation and verification, with notable consistency in interpretation: verification meaning the process of ensuring that the computer model correctly implemented the intended model, whereas validation means the process of ensuring that a model is fit for purpose. Methodological literature on verification and validation of models makes reference to the Hermeneutic philosophical position, highlighting that the concept of model validation should not be externalised from the decision-makers and the decision-making process. Interviewees demonstrated examples of all major error types identified in the literature: errors in the description of the decision problem, in model structure, in use of evidence, in implementation of the model, in operation of the model, and in presentation and understanding of results. The HTA error classifications were compared

against existing classifications of model errors in the literature. A range of techniques and processes are currently used to avoid errors in HTA models: engaging with clinical experts, clients and decision-makers to ensure mutual understanding, producing written documentation of the proposed model, explicit conceptual modelling, 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. However, their current implementation is not framed within an overall strategy for structuring complex problems.

**Limitations:** Some of the questioning may have biased interviewees responses but as all interviewees were represented in the analysis no rebalancing of the report was deemed necessary. A potential weakness of the literature review was its focus on spreadsheet and program development rather than specifically on model development. It should also be noted that the identified literature concerning programming errors was very narrow despite broad searches being undertaken.

**Conclusions:** 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. Such discussions should focus on risks, including errors of implementation, errors in matters of judgement and violations. Discussions of modelling risks should reflect the potentially complex network of cognitive breakdowns that lead to errors in models and existing research on the cognitive basis of human error should be included in an 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. Interaction between modeller and client in developing mutual understanding of a model establishes that model's significance and its warranty. This highlights that model credibility is the central concern of decision-makers using models so it is crucial that the concept of model validation should not be externalised from the decision-makers and the decision-making process. Recommendations for future research would be studies of verification and validation; the model development process; and identification of modifications to the modelling process with the aim of preventing the occurrence of errors and improving the identification of errors in models.



# Contents

<b>List of abbreviations</b> .....	vii	Errors in operation of the model and model reporting .....	38
<b>Executive summary</b> .....	ix	Review of taxonomy/classification literature .....	39
<b>1 Introduction</b> .....	1	Placing the interview taxonomy of errors in the context of literature on error classifications .....	44
Background .....	1	Conclusion .....	47
Aim and objectives .....	2	<b>6 Strategies for avoiding errors</b> .....	49
<b>2 Methods</b> .....	3	Overview .....	49
Methods overview .....	3	Mutual understanding .....	49
In-depth interviews with HTA modellers ..	3	Model complexity .....	53
Literature review search methods .....	5	Housekeeping .....	55
Validity checking .....	6	Guidelines .....	57
The organisation of the interview sample .....	6	Skills and training .....	58
<b>3 The model development process</b> .....	7	Explanatory analysis concerning methods for preventing errors .....	60
Overview .....	7	Placing the findings of the interviews in the context of the literature on avoiding errors .....	62
Understanding the decision problem .....	7	<b>7 Strategies for identifying errors</b> .....	63
Conceptual modelling .....	9	Overview .....	63
Use of information in model development .....	11	Check face validity with experts .....	63
Implementation model .....	13	Do the results appear reasonable? .....	63
Model checking .....	14	Testing model behaviour .....	64
Reporting .....	15	Is the model able to reproduce its inputs? ..	66
Synthesised model development process .....	15	Can the model replicate other data not used in its construction? .....	66
Discussion of model development process .....	17	Compare answers with the answers generated by alternative models .....	67
<b>4 Definition of an error</b> .....	19	Peer review of models .....	68
Overview .....	19	Double checking of input values .....	69
Intention, deliberation and planning .....	19	Double-programming .....	69
Extent to which the model is fit for purpose .....	19	Explanatory analysis concerning methods for preventing errors .....	71
Simplification versus error .....	21	Placing the findings of the interviews in the context of the literature on identifying model error .....	71
Impact of error on the model results and subsequent decision-making .....	21	<b>8 Barriers and facilitators</b> .....	73
Placing the interview findings in the context of literature on model verification and validation .....	22	Barriers .....	73
<b>5 A taxonomy of errors</b> .....	25	Facilitators .....	74
Introduction .....	25	<b>9 Discussion and conclusions</b> .....	75
Error in understanding the problem .....	26	Strengths and weaknesses of the study ....	75
Error in the structure and methodology used .....	30	Current understanding of modelling errors in the HTA community .....	76
Error in the use of evidence .....	33		
Error in implementation of the model .....	36		

A taxonomy of HTA modelling errors .....	77	<b>Appendix 1</b> Avoiding and identifying errors in health technology assessment models .....	87
Current strategies for avoiding errors .....	77	<b>Appendix 2</b> Search strategies .....	91
Current strategies for identifying errors in HTA models .....	78	<b>Appendix 3</b> Charts .....	95
Recommendations .....	78	<b>Health Technology Assessment reports published to date</b> .....	109
Research recommendations .....	79	<b>Health Technology Assessment programme</b> .....	131
<b>Acknowledgements</b> .....	81		
<b>References</b> .....	83		





## List of abbreviations

EuSPRIG	European Spreadsheet Risks Interest Group	NICE	National Institute for Health and Clinical Excellence
HAQ	Health Assessment Questionnaire	PBAC	Australia's Pharmaceutical Benefit Advisory Committee
HTA	Health Technology Assessment	PSA	probabilistic sensitivity analysis
ICER	incremental cost-effectiveness ratio	QALY	quality-adjusted life-year
NCCHTA	National Coordinating Centre for Health Technology Assessment	RFP	Request For Proposal
		SCS	Society for Modelling and Simulation International

All abbreviations that have been used in this report are listed here unless the abbreviation is well-known (e.g. NHS), or it has been used only once, or it is a non-standard abbreviation used only in figures/tables/appendices, in which case the abbreviation is defined in the figure legend or in the notes at the end of the table.





# 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.

# Chapter I

## Introduction

### 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 recognised as being well placed to support this process.<sup>1</sup> The key role that models play is, however, reliant on their credibility. Credibility in decision 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 as an adequate representation of the problem in hand and the quality of the model.

The predominant software platform for Health Technology Assessment (HTA) models is the spreadsheet. In studies of generic spreadsheet system development, there is abundant evidence of the ubiquity of errors and some evidence of their potential to impact critically on decision-making. The European Spreadsheet Risks Interest Group (EuSPRIG) maintains a website recording research on risks in spreadsheet development and on public reports of errors ([www.eusprig.org/](http://www.eusprig.org/) accessed 20 February 2009). In 1993, Cragg and King<sup>2</sup> undertook an investigation of spreadsheet development practices in 10 organisations and found errors in 25% of the spreadsheets considered. Since 1998, Panko<sup>3</sup> has maintained a review of spreadsheet errors and this suggests that spreadsheet error rates are consistent with error rates in other programming platforms, and when last updated in 2008 error rates had not shown any marked improvement.

To date, there has been no formal study of the occurrence of errors in models within the HTA domain. However, high-impact errors have been recorded, perhaps the earliest being the case of the sixth stool guaiac test, where Neuhauser and Lewicki<sup>4</sup> estimated an incremental cost of \$47 million per case of colorectal cancer detected. Brown and Burrows<sup>5</sup> subsequently identified an error in the model and generated a comparatively

modest corrected estimate. The grossly inflated incorrect figure has, however, subsequently been used (and is still in use) by authors including Culyer<sup>6</sup> and Drummond *et al.*<sup>7,8</sup> in their seminal texts to demonstrate the importance of incremental analysis in health economics. The failure in this case was traced back to an error in the interpretation and subsequent modelling of diagnostic test characteristics. In a study in 2000 investigating the quality of models used to support a national policy-making process, Australia's Pharmaceutical Benefit Advisory Committee (PBAC) reported that around 60% were flawed in some way and 30% demonstrated problems in the modelling.<sup>9</sup> Although it is tempting to think that things may get better over time, a recent repeat retrospective analysis reported in 2008 demonstrated that in fact the situation had appeared to deteriorate with 82% (203 out of 247) of models reviewed being considered by PBAC to be flawed in some respect.<sup>10</sup>

Errors in software implementing mathematical decision models or simulation exercises are a natural and unavoidable part of the software development process.<sup>11</sup> This is recognised outside the HTA domain and accordingly there exists research, for example in the computer sciences field and in the operational research field, in model testing and verification. In contrast, the modelling research agenda within HTA has developed primarily from a health-services research, health economics and statistics perspective and there has to date, been very little attention paid to the processes involved in model development.<sup>12</sup> The extent of this shortcoming is reflected in the fact that guidance on good practice either acknowledges the absence of methodological and procedural guidance on model development and testing processes, or, makes no reference to the issue.<sup>13-17</sup>

A wide range of strategies could be adopted with the intention of avoiding and identifying errors in models. These include improving the skills of practitioners, improving modelling and decision-making processes, improving modelling methods, improving programming practice and improving

software platforms. Within each of these categories there are many further options, so improving the skill base of practitioners might include: the development and dissemination of good practice guides, identification of key skills and redesign of training and education, sponsoring of skills workshops, and determination of minimum training/skill requirements. These interventions again might impact across the whole range of disciplines involved in the decision support process, including information specialists, health economists, statisticians, systematic reviewers and not to forget, modellers themselves. Outside the HTA domain, Panko identifies a wide range of initiatives for best practice in spreadsheet development, testing and inspection, including such things as guidelines for housekeeping structures in spreadsheets.<sup>18</sup> However, Panko also goes on to review the evidence on the effectiveness of these interventions and identifies first a lack of high-quality research on effectiveness and second that what evidence does exist is at best equivocal.<sup>18</sup> This, taken together with the evidence that models are not improving over time, indicates that caution should be exercised in recommending quick-fix, intuitively appealing solutions.

In the absence of an understanding of error types and the causes of errors, it is difficult to evaluate the relative merits of such initiatives and to develop an efficient and effective strategy for improving the credibility of models. 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 aim of this study is 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 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.



# Chapter 2

## Methods

### Methods overview

The project involved two methodological strands:

1. A methodological review of the literature discussing errors in modelling and principally focusing on the fields of modelling and computer science.
2. In-depth qualitative interviews with the HTA modelling community, including:
  - Academic Technology Assessment Groups involved in supporting the NICE Technology Appraisal Programme (hereafter referred to as Assessment Groups)
  - Outcomes Research and Consultancy Groups involved in making submissions to NICE on behalf of the health-care industry, (hereafter referred to as Outcomes Research organisations).

### In-depth interviews with HTA modellers

#### Overview of qualitative methods

Interviews with 12 mathematical modellers working within the field of HTA modelling were undertaken to obtain a description of current model development practices, to develop an understanding of models errors and strategies for their avoidance and identification. From these descriptions, issues of error identification and prevention were explored.<sup>19</sup>

#### Sampling

The interview sample was purposive, comprising one HTA modeller from each Assessment Group contracted by the National Coordinating Centre for Health Technology Assessment (NCCHTA) to support NICE's Technology Appraisal Programme, as well as HTA modellers working for UK-based Outcomes Research organisations. The intention of the sampling frame was to identify diversity across modelling units. Characteristics of the interview sample are detailed at the end of this chapter.

### Qualitative data collection

Face-to-face in-depth interviews<sup>19</sup> were used as the method of data collection; this is particularly appropriate given their focus on the individual and the need to elicit a detailed personal understanding of the views, perceptions, preferences and experiences of each interviewee. A topic guide was developed by the research team that enabled the elicitation of demographic information, as well as views and experiences of the modelling process and issues around modelling error (see Appendix 1). Interviews began with a discussion of background details and progressed to a more in-depth exploration of modellers' experience and knowledge. The topic guide was piloted internally with one of the authors (AR) to ascertain clarity of the questions; the topic guide was subsequently revised. The topic guide was designed to facilitate the flow of the interviews, although the interviews were intentionally flexible and participant-focused. During the interviews, interviewees were asked about their personal views rather than acting as representatives of their organisation.

During each interview the participant was asked to describe their view of the model development process. This description was sketched onto paper by a second researcher and subsequently reviewed with the participant who was asked to confirm the sketch or 'map' as a representative and accurate record of their view. This diagrammatic personal description of the model development process was instrumental within the interviews, representing a personalised map against which to locate model errors according to the perceptions of the interviewee, thereby guiding the content and agenda of the remainder of the interview.<sup>20</sup> 'Prompting' was used to 'map' and 'mine' interviewee responses while 'probing' questions were used to further elaborate responses and to provide richness of depth in the interviewees' responses.

All interviews were undertaken between September and October 2008 by three members of the research team (PT, JC and AR). All three

interviewers have experience in developing and using health economic models. One of the interviewers (PT) received formal training in in-depth interviewing techniques and analysis of qualitative data, this training being shared with the modelling team working with the advice of a qualitative reviewer (MJ). All but two interviews were paired, involving a lead interviewer and a second interviewer. The role of the second interviewer was to facilitate the elicitation of the model development process by developing charts and to ensure relevant issues were explored fully by the lead interviewer. The remaining two interviews were undertaken on a one-to-one basis. Each interview lasted approximately 1½ hours.

### Data analysis and synthesis

Interviews with participants were recorded and transcribed verbatim; all qualitative data were analysed using the Framework approach.<sup>21</sup> Framework is an inductive thematic framework approach particularly suited to the analysis of in-depth qualitative interview data, involving a continuous and iterative matrix-based approach to the ordering and synthesising of qualitative data. The first step in the analysis involved familiarisation with the interview data, leading to the identification of an initial set of emergent themes (e.g. definition of model error), subthemes (e.g. fitness for purpose) and concepts. Transcript data were then indexed according to these themes, both by hand and using NVivo® software (QSR International, Southport, UK). Full matrices were produced for each theme, detailing data from each interviewee across each of the subthemes. Data within each subtheme was then categorised, ensuring that the original language of the interviewee was retained, and classified to

produce a set of dimensions that was capable of both describing the range of interview data and discriminating between responses. Discussions within the team were carried out during coding and categorisation to obtain consensus before the next stage. Descriptive and explanatory accounts were used to interrogate the data within and across themes and subthemes. The key stages of analysis and synthesis are described in more detail below.

### The coding scheme

The coding scheme covers seven main themes together with a range of subthemes. This coding system was initially based on the topic guides used within the interviews. Upon further interrogation of the interview transcript data, a revised coding scheme was developed through discussion between the three interviewers. *Table 1* presents a brief description of each theme together with a brief description of its content.

The first theme (Organisation, roles and communication) includes mainly demographic information that is used to describe the sample of participants and compare data between groups. Although this is not central to the qualitative synthesis, it does provide relevant information that helps to interpret and explain variations in stated views between respondents.

The second theme (see Chapter 3) relates to data from the model development process maps and interview transcript data. A meeting was held with all the authors to analyse and synthesise evidence from the process maps. This meeting assisted the analysis by providing an overview of the data and informing decisions regarding the subsequent qualitative process. The analysis of the

**TABLE 1** Description of themes for the qualitative analysis

Theme	Description of theme	Chapter
1. Organisation, roles and communication	Background of the interviewee, roles and experience, use of specific software platforms, working arrangements with other researchers	2
2. The model development process	Key sets of activities and processes in the development of health economic models	3
3. Definition of error	The interviewee's perception of what does and what does not constitute a model error	4
4. Types of model error	Specific types of model error discussed within the interviews	5
5. Strategies for avoiding errors	Approaches to avoiding errors within models	6
6. Strategies for identifying errors	Approaches to identifying errors within models	7
7. Barriers and facilitators	Potential interventions to avoid and identify errors within models and constraints on their use	8

model development charts unearthed a potential typology of modelling processes followed by practitioners. Generalities could be made to some extent around particular steps taken during the modelling process, and so these were explicitly drawn up on a separate generic map/list during discussions between the authors. In addition, a comprehensive descriptive analysis of the perceived model development process was undertaken using the interview transcript data. Finally, a stylised model development process was developed; this model attempts to both capture and explain key stages in the model development process as well as important iterations between stages, based on the perceived processes of individual interviewees. This was used to explain some of the nuances in the data and variations in processes between respondents.

The third theme (see Chapter 4) was analysed using the Framework approach. Key dimensions of what is perceived to be and what is not perceived to be an error were drawn out from the interview transcript material. Literature concerning the verification and validation of models from outside the HTA field was used to facilitate the interpretation of qualitative evidence on the characteristics of model errors.

The fourth theme (see Chapter 5) involved a descriptive analysis of interview data according to the initial framework analysis coding scheme: error in understanding the problem; error in the structure and methodology used; error in the use of evidence; error in implementation of the model; error in operation of the model; and error in presentation and understanding of results. As with other emergent themes within the interview data, the descriptive analysis was developed through a process of categorisation and classification of comments that made direct or indirect reference to error types.<sup>21</sup> The structure of the taxonomy emerged from the descriptive analysis of interview data. The taxonomy identifies three levels:

- the error domain, that is the part of the model development process in which the error occurs
- the error description, illustrating the error in relation to the domain
- the root cause error, which attempts to identify root causes of errors and to draw out common types of error across the modelling process.

Non-HTA literature concerning modelling error classifications and taxonomies was identified from searches and used to facilitate the interpretation of the qualitative data.

The fifth and sixth themes (see Chapters 6 and 7) present an analysis of techniques and processes for avoiding and identifying errors in HTA models, as discussed by interviewees. Within both chapters, a descriptive analysis is presented together with an explanatory analysis, which relates current error identification and avoidance methods to the classification of model errors. The seventh theme (see Chapter 8) briefly presents a descriptive analysis of interviewees' views on barriers and facilitators to error checking activities.

## Literature review search methods

At the outset the scope and fields of literature relevant to the study were not known. The purpose of the searches was twofold; first exploratory searches in advance of the interviews were undertaken to inform the development of the scope and content of the in-depth interviews, second the searches after the interviews were undertaken to expand the discussion on key issues raised by the interview analysis and to place the interview data in the context of the broader literature.

An iterative approach was taken to searching the literature. This is recognised as being the most appropriate approach to searching, in the case of reviews of methodological topics, where the scope of relevant evidence is not known in advance.<sup>22</sup> Based on information-seeking behaviour models<sup>23</sup> the exploratory search strategy used techniques including author searching, citation searching, reference checking and browsing.

The scoping search was undertaken to establish the volume and type of literature relating to modelling errors within the computer science and water resources fields both identified as potentially rich fields. Highly specific and focused searches were carried out in *Computer and Information Systems Abstracts* and *Water Resources Abstracts*.

The interviews and scoping search of the literature identified the need for further reviews including: first, a topic review of error avoidance and identification; second, a review of model verification and validation; and third a review of error classifications and taxonomies. The searches were undertaken in February 2008 and January 2009. The search strategies and sources consulted are detailed in Appendix 2.

## Validity checking

The report was subjected to internal peer review by a qualitative researcher, mixed methodologist, modellers not involved in the interviews and a systematic reviewer. The final draft report was distributed to all interviewees to obtain their feedback on whether the findings are representative of their views, hence providing respondent validation.<sup>24</sup>

## The organisation of the interview sample

This section briefly details the background, expertise, experience and working arrangements of the 12 individuals included in the interview sample.

### Organisation

Of the 12 respondents, four work primarily for industry and eight are employed to produce reports for NICE and are university-based. Of those interviewees working within Assessment Groups, two also discussed work undertaken for commercial clients. The respondents reported a variety of work types including economic evaluations, as well as primary and secondary research. Commercial work included health outcomes and pricing and reimbursement work as well as cost-effectiveness modelling. The size of the organisations varied with a range of 12 to 180 staff employed. Three interviewees discussed working internationally.

### Composition of the research teams

The respondents reported a variety of working arrangements, to demonstrate the range of organisational arrangements these included:

- a unit with 30 health economists with close links to another unit with 60–70 health-service researchers
- a small team of two dedicated modellers with links to a team who undertake systematic reviews
- a unit including operational researchers, health economists, health-service researchers and literature reviewers with direct access to clinical experts
- a large international organisation covering many different disciplines
- a team of 14–15 health economists with no mention of systematic reviewers.

### Background of the interviewees

Five of the 12 respondents came from an economics or health economics background, two from a mathematical background, one of whom previously worked as an actuary. Five respondents have an operational research or modelling background. All interviewees included in the sampling frame have considerable experience developing health economic models and were purposefully selected on this basis.

### Platforms and types of models

All respondents apart from one, mentioned Microsoft EXCEL<sup>®</sup> as one of the preferred modelling software platforms, although two mentioned potential limitations. TREEAGE was frequently mentioned and SIMUL8 was mentioned by two respondents. Other software programs and applications discussed by interviewees included CRYSTAL BALL, STATA, WINBUGS, DELPHI, VISUAL PASCAL and SPSS. Ten of the twelve respondents mentioned the use of Markov models. Also reported were microsimulation/discrete event simulation, decision analysis and decision tree models.

# Chapter 3

## The model development process

### Overview

This chapter presents a descriptive analysis of individual respondents' perceptions of the model development process. The purpose of eliciting information concerning current model development processes within the interviews was twofold. First, it allows for the development of an understanding of the similarities and variations in modelling practice between modellers that might impact on the introduction of certain types of model error. Second, the elicitation of interviewee's personalised view of the process was used to form the structure of the remaining portion of the interview, with the model development charts therefore forming a map against which specific types of model error could be identified and discussed.

The descriptive analysis was undertaken according to the coding scheme, and examined six key stages of the model development process: (1) understanding the decision problem; (2) use of information; (3) implicit/explicit conceptual modelling; (4) software model implementation; (5) model checking activities; and (6) model reporting. It should be noted that these six stages were not specified a priori, but rather they emerged from the qualitative synthesis process. A seventh theme concerning iterations between stages of model development was also analysed. The descriptive synthesis of the model development process is presented together with an inferred stylised model development process describing both the range and diversity of responses.

### Understanding the decision problem

For all 12 interview participants, the first stage in the model development process involved developing an understanding of the decision problem. Broadly speaking, this phase involved a delicate balancing/negotiation process between the modeller's perceived understanding of the problem, the clinical perspective of the decision problem, as well as the clients' needs and the

decision-makers' needs. This phase of model development may draw on evidence, including published research and clinical judgement, and experience within the research team. In some instances the research question was perceived to follow directly from the received decision problem, whereas other respondents expressed a perception that the research question was open to clarification and negotiation with the client/decision-maker. In this sense the research question represents an intellectual leap from the perception of the decision problem. One respondent noted that this aspect of model development may not be a discrete step which is finalised before embarking on research:

The whole process starts at the point where we are thinking about a project but it will continue once a project has actually started, almost until the final day when the model is finalised that process will continue...we will not finalise the structure of the model and the care pathway perhaps until quite close to the end of the whole project because we will be constantly asking for advice about whether or not we have it correct or at least as good as we can get it.

[I8]

Similarly, another respondent highlighted that the decision problem may not be fixed even when the model has been implemented, indicating that modelling may have a role in exploring the decision problem space. Hence, for instance, modelling can answer questions such as 'Is technology A economically attractive compared to technology B' or 'Under what circumstances is A economically attractive compared to B'. These are fundamentally different decision problems however the decision-makers may change their focus throughout the course of an analysis and are dependent on the outcomes of the analysis.

...commonly at the analysis stage, you have the opportunity or there's the possibility of changing the question that you're asking. So, if the results of the model are in any way unexpected or in fact, if they're in any way insightful, then you might want to refine or

amend the outcome or the conclusion you're going to try and support. So if we start off with an analysis of a clinical trial population that is potentially cost-effective, in the course of the model, you find that the product is particularly appropriate for a subgroup of patients, or there are a group of patients where the cost-effectiveness is not attractive, then we'd refine the question.

[I6]

Several interviewees noted the importance of understanding the decision problem and defining the research question appropriately; in particular, one respondent indicated the fundamental importance of understanding the decision problem before embarking on the implementation of any model:

...it all stems from what is the decision problem, to start off with...no matter what you've done, if you've got the decision problem wrong...basically it doesn't matter how accurate the model is. You're addressing effectively the wrong decision.

[I7]

### **Written documentation of the decision problem and research question**

Several interviewees mentioned that explicit documentation of the perceived decision problem and the research question forms a central part of the proposal within consultancy contracts. The discussions suggested that such documentation extends beyond the typical protocols produced within the NICE Technology Appraisal Programme, including a statement of the objective, a description of the patient population and subgroups, and a summary of key data sources. Several respondents highlighted that although the decision problem may be set out in a scope document, the appropriate approach to addressing the decision problem may be more complex.

...you've read your scope you think you know what it's about. But then you start reading about it, and you think, 'Oh! This is more complicated than I thought!'

[I3]

This point was further emphasised by another respondent, who highlighted that there was a need to understand not only the decision-makers'

perceived decision problem, but also what the research question should be (so questioning the appropriateness of that perception) and understanding wider aspects of the decision problem context outside its immediate remit:

...our entire initial stage is to try and understand what the question should be and what the processes leading up to the decision point and the process that goes from that decision point.

[I8]

The issue of generating written documentation of the proposed model to foster a shared agreement of the decision problem was raised only by two interviewees, both of whom work for Outcomes Research organisations:

It's almost like a mini report really. It's summarising what we see as the key aspects of the disease, it's re-affirming the research question again – what's the model setting out to do. It's pitching the model and methodology selected and why, so if it's a Markov [model] it's kind of going on to detail what the key health states are. We try and not make it too detailed because then it just becomes almost self-defeating but you are just trying to, again, with this, we are always trying to get to a point at the end of that particular task where we have got the client signed in and as good a clinical approval I guess, as we can get.

[I10]

For several respondents, an important dimension of understanding the decision problem involved understanding the perceived needs of the decision-maker and the client, mediated by the ability of the modelling team to meet these. A common theme which emerged from the interviews with Outcomes Research modellers was an informal iterative process of clarification of the decision-maker needs; that the potential cost of not doing so is the development of a model which does not adequately address the decision problem. The analysis suggested some similarities in the role of initial documentation between Assessment Groups and Outcomes Research organisations. In particular, modellers working for the pharmaceutical industry highlighted the use of a Request For Proposal (RFP) document and the development of a proposal; these appear to broadly mirror the scope and protocol documents developed by NICE and the Assessment Groups, respectively.

There was some suggestion that RFPs were subject to clarification and negotiation whereas scope documents were not.

The first thing is to get an idea of exactly what the client wants. On a number of occasions you get a RFP through which is by no means clear what the request is for and it won't be very helpful to rush off and start to develop any kind of model on that kind of platform...you can often get a situation where you answer the question that they have asked and they then decide that was not the question they had in mind so certain processes of ensuring clarity are useful.

[I12]

### Methods for understanding the decision problem

Methods used by modellers to understand the perceived decision problem varied between respondents, including 'immersion' in the clinical epidemiology and natural history of the disease, understanding the relationship between the decision point and the broader clinical pathway, as well as understanding relevant populations, subgroups, technologies and comparators.

...you start by, by just immersing yourself in whatever you can find that gives you an understanding of...all the basics.

[I3]

...it's the process of becoming knowledgeable about what you are going to be modelling... reading the background literature, knowing what the disease process is, knowing the clinical...pathways typically that patients experience within the situation you are modelling. Going to see clinical experts to ask questions and find out more and gradually hone in on an understanding on the clinical area being studied in a way that enables you to begin to represent it systematically.

[I4]

### Conceptual modelling

Nine of the twelve interviewees either alluded to, or explicitly discussed, a set of activities involving the conceptualisation and abstraction of the decision problem within a formal framework before implementing the mathematical model in a software platform.

### Conceptual modelling methods

The extent to which conceptual modelling is demonstrated in practice appears to be highly variable between participants; variation was evident in terms of the explicitness of model conceptualisation methods and the extent to which the conceptual model was developed and shared before actual model programming. The interview data for three respondents implied that no distinct formal or informal conceptual modelling activities took place within the process, that the model was conceptualised and implemented in parallel. One of these respondents highlighted an underlying rationale for such an approach.

...I'd rather go with a wrong starting point and get told it's wrong than go with nothing.

[I2]

Explicit methods for conceptual modelling included developing written documentation of the proposed model structure, assumptions, the use of diagrams or sketches of model designs and/or clinical/disease pathways, memos, representative mock-ups to illustrate specific issues in the proposed implementation model, and written interpretations of evidence.

The extent to which the conceptual model was complete before embarking on programming the software model varied among interviewees; several respondents indicated that they would not begin implementing the software model until some tangible form of conceptual model had been agreed, or in other cases, 'signed-off' by the client or experts. Across the nine interviewees who did undertake some degree of explicit conceptual modelling, the main purposes of such activities included fostering agreement with the client and decision-maker, affirming the research question, pitching and justifying the proposed implementation model, sense-making or validity checking, as well as trying ideas, getting feedback, 'throwing things around', 'picking out ideas' and arranging them in abstraction process. Broadly speaking, such activities represent a communication tool, that is, a means of generating a shared understanding between the research team, the decision-maker and the client.

...we'll get clinical advisors in to comment on the model's structure and framework, comment on our interpretation of the evidence, and...we think we've got a shared understanding of that clinical data but are we really seeing it for what it is?

[I10]

...for us, whether we've written it or not, we've got an agreed set of health states, assumptions, and...an agreed approach to populating the parameters that lie within it...which I think is the methods and population of your decision model.

[I7]

The extent to which the conceptual model is formally documented varied between the interviewees. One respondent represented a deviant case in the sense that whereas the majority of the model development process concerns understanding the decision problem and (potentially) implicit conceptual modelling, very little time is spent on actual model implementation.

So every aspect of what you then need to, to programme in and populate it is...in people's brains to various degrees...if you get all that agreed, then the actual implementation in EXCEL should be pretty straightforward. But if that process has taken you 90% of your time, then...you build a model pretty quickly.

[I7]

For three interviewees [I5, I2 and I12] there was no formal distinction between conceptual modelling activities and software model implementation, rather they occur in parallel. Among these three interviewees, there was a degree of consistency in that all three respondents discussed the development of early implementation models (sometimes referred to by interviewees as 'quick models' or 'skeleton models') as a basis for eliciting information from clinical experts, for testing ideas of what will and will not work in the final implementation model, or as a means of generating an expectation of what the final model results are likely to be:

So there is an element of dipping your toe in the water and seeing what looks like it's going to work. It's actually being prepared to say no this is not going to work.

[I12]

I would generally get a rough feel for the answer within three or four days which is almost analogous to the pen and paper approach which is why I do that. If you know that 20,000 people have this disease which is costing  $x$  thousand per year and you know your drug which costs  $y$  pounds will prevent half of it

and you know what utilities are associated with it you can get a damn good answer.

[I2]

## Identifying key model factors

Several interviewees discussed the processes they undertake in developing the structure of the model. One interviewee described this activity as a means of:

...identifying the fundamental aspects of the decision problem and organising them into a coherent framework.

[I4]

Much of the discussion concerned identifying and specifying relationships between key outcomes, identifying key states to be differentiated, identifying transitions, capturing the importance of patient histories on subsequent prognosis, and addressing the need for more complex descriptions of health states. Several interviewees suggested that this process was at least to some degree, data-driven, for example, having access to patient-level trial data may influence the structure of a survival model and the inclusion of specific coefficients. Several interviewees discussed the inevitability of simplification; one suggested the notion of developing 'feasible models', 'best approximations' and 'best descriptions of evidence'.

...I think it's a judgement call that modellers are constantly forced to make. What level of simplification, what level of granularity is appropriate for the modelling process? I think what's very important is to continually refer back to the decision that you are hoping to support with your model. So don't try and answer questions that aren't going to be asked...

[I4]

...the aim is to produce the feasible model that best approximates what we think we're going to need...on occasions we will just have to say yes we would love to put this into a model but the trials have not recorded this...

[I5]

One respondent had a different viewpoint concerning how to identify key factors for inclusion in the mathematical model, indicating an entirely data-led approach whereby the decision to include certain elements of the decision problem in the



implementation model was determined by its influence on the model results:

...what to keep in and what to chuck out...it depends on how quickly you've got the data. If you've got all of the data then keep everything in, if you've already got it, because you can tell with an EVSI [expected value of sample information] or an EVI [expected value of information] what is important or not even if you or you fit a meta-model to it or the ones that are capped fall straight out...

[I2]

...you are only including these things because you think they will affect the ICER [incremental cost-effectiveness ratio], if they don't affect the ICER you shouldn't be including them.

[I2]

Clearly, decisions concerning what should and should not be included in the model, how such factors should be captured and related to one another, and the appropriate level of complexity and granularity are highly complex. Indeed, two respondents highlighted that this aspect of model development was hard to teach and was learned through experience.

I don't think...that I can sit here and write out how you build a model. I think it's something which comes with experience.

[I7]

### Formal a priori model specification

One respondent [I1] highlighted the absence of an important stage of the model development process common within software development projects: that of model design and analysis, whereby the proposed model is formally specified and its software implementation is planned before the model is programmed. In this sense, the design and analysis stage represents a direct link between the conceptual model and the implementation model. Such activity would usually include producing a formal model specification in advance of any programming, including details of how the model will be programmed, where parameters will be stored and linked, housekeeping issues and an a priori specification of how the model will be checked and tested.

### Use of information in model development

For many of the interviewees, the use of evidence to inform the model development process was broader than the generation of systematic reviews of clinical effectiveness. The use of evidence used to determine the model structure was also commonly discussed. Further, the use of evidence in informing the model was seldom restricted to a single stage in the development process; for several respondents, evidence was used to understand, shape and interpret all aspects of model development.

### Use of previous models

Views concerning the appropriate use of previous models to inform the model under development varied between study participants. One subtle but potentially important difference in use of previous models concerned the extent to which such models developed for a different decision problem, decision-maker and purpose, would be used to inform the structure of the current model. For some respondents, previous models were examined as a means of 'borrowing' model structures, whereas for others, the model represents a starting point for thinking about an underlying conceptual framework for the model:

...suppose the first thing we will do actually is see if there are any existing models that we can build from and if there are and we think they are any good then we have a structure there.

[I5]

...it's not a question of what people have done in the past and whether they did a good job or not, it's seeing what we can borrow from those that can be relevant.

[I12]

The interview data suggest the notion of being 'data-led' in model structuring. One may infer two risks associated with being dependent on previous model structures: one uses an existing model which is not adequately structured to address the decision problem at hand, or one misses out on developing an understanding and agreement of the decision problem through conceptual modelling (see Conceptual modelling). This inference appears to be supported by the views of a further respondent:

If you set two modelling teams the same task I think it's quite likely they would come up

with very different models, which is one of the aspects of familiarisation. Unless of course there were lots of existing models that they could draw on and fall into familiar grooves in terms of the way you model a particular process so that often happens you look at what has already been done in terms of the modelling process and that is part of the familiarisation process and you say oh yeah that makes perfect sense we will do the same again.

[I4]

I think there is a danger in that as well you know if you just slavishly adopt a previous structure to a model and everybody does the same there's no potential for better structures to develop or for mistakes to be appealed so I think there is a danger in that in some ways.

[I4]

### Use of clinical/methodological expert input

The extent to which clinical experts were involved within the model development process varied considerably between interviewees. Several interviewees indicated that they would involve clinical experts in the research team itself, from the very outset of the project, to help develop an understanding of the decision problem and to formulate the research question; this was particularly the case where in-house clinical expertise was not available.

...we try to get in clinical experts as part of the team and they will usually also be authors on the report and it tends to be very much a joint process of structuring the problem.

[I5]

At the other end of the spectrum, one respondent suggested that they would only involve clinical experts *after* having built a skeleton model.

I would build the model, just build it straight off, I mean it depends on how quick you work but if you build it quickly then you can have all that in place before you talk to clinicians or other peer reviewers...

[I2]

### Relationship between clinical data and model structure

The majority of the interviewees highlighted the existence of a complex iterative relationship

between model structuring processes and data identification and use; whereby the model structure has a considerable influence on the data requirements to populate the model, whereas the availability of evidence may in turn influence the structure of the model (whether implemented or conceptualised). In effect, the 'worldview' or *weltanschauungen*<sup>25</sup> of the modeller influences the perception of what evidence is required, and the identification and interpretation of that evidence in turn, influences and adapts that *weltanschauungen*.

I divided it between design[ing] and populat[ing]...the populating being searching for, locating the information to actually parameterise those relationships. And then going back and changing the relationships to ones that you can actually parameterise from the data that's available, and then changing the data that you look for to fit your revised view of the world.

[I6]

Two respondents indicated that previous working arrangements had led to the systematic review and modelling activities being perceived as distinct entities that did not inform one another, so hindering the iterative dynamic highlighted above. However, both of these respondents indicated a change in this working culture, suggesting a joint effort between the modeller and the other members of the research team:

I think things are...kind of trying to be changed. I'm not seeing a systematic review... as something separate from modelling; now they are working together and defining what [it] is...that we are looking for together.

[I1]

Several respondents highlighted the importance of this iterative phase of model development and its implications for the feasibility of the resulting mathematical model.

The most difficult ones aren't just the numbers; they are where there's a qualitative decision. You know, is there any survival benefit? Yes or no? If there is a survival benefit, it's a different model?...So it's structural.

[I6]

...it's no good having an agreed model structure that's impossible to populate with data so you obviously give some thought as to how you are going to make it work that's part

of the agreement process. You outline what the data requirements are likely to be and step through with the people collecting the data...

[I4]

One respondent highlighted the difficulties of this relationship, suggesting a danger for models to become data-led rather than problem-led.

We try as a matter of principle not to let data blur the constraints of models too much but in practice sometimes it has to.

[I5]

One respondent highlighted the use of initial searches and literature summaries as a means of developing an early understanding of likely evidence limitations, a process undertaken alongside clinical experts, before model development. Another respondent highlighted what he considered a key event in the model development process concerning iterations between the model structure and the evidence used to inform its parameters:

We have a key point which we call data lockdown because that's key to the whole process, if you are working to a tight timescale you need to make sure the modellers have enough time to verify their model with the data.

[I4]

## Differences between the model structure and the review data

A common emergent theme across respondents was the idea that the data produced by systematic reviews may not be sufficient or adequate for the model, meaning that either the data would need transforming into a format in which the model could 'accommodate' the evidence (for example premodel analysis such as survival modelling), or the developed model itself would need to be restructured to allow the data to be incorporated into the model.

...but it wouldn't be just a question of understanding the clinical data, it's a question of what it's going to mean for the modelling.

[I12]

A further point raised by one respondent concerned the Assessment Groups' perception of what constitutes evidence for the model; this

particular respondent indicated that his institution is likely to represent a deviant case in this respect.

I wouldn't say we had a procedure, but an underlying principle...as a department we don't distinguish between direct and indirect evidence. We just think it's all evidence.

[I7]

[We] are less beholden to statistical significance; other people are more conventional in their sort of...approaches. One...set of researchers may say that...there is no statistical significance of heterogeneity...[that] we can assume that all this evidence can be pooled together may not be considered by...another set of researchers to be an appropriate sort of thing to do...

[I7]

Interestingly, almost all discussions concerning the use of evidence in populating models focused entirely on clinical efficacy evidence; very little discussion was held concerning methods for identifying, selecting and using non-clinical evidence to inform other parameters within models, for example costs and health-related quality of life. Across the 12 interviewees, it was unclear who holds responsibility for identifying, interpreting and analysing such evidence or how (and if) it influences the structure of the model, and how such activities differ from the identification and use of clinical efficacy data.

## Implementation model

All respondents discussed a set of activities related to the actual implementation of a mathematical model on a software platform. Such activities either involve the transposition of the prespecified soft conceptual model framework into a 'hard' quantitative model framework involving various degrees of refinement and adaptation, or the parallel development of an implicit conceptual model and explicit implementation model.

### Model refinement

Several respondents discussed activities involving model refinement at the stage at which the model is implemented, although the degree to which such activities are required was variable. The interview data suggest that this aspect of model development was a key feature for the three respondents who do not separate conceptual and implementation

model processes. As noted earlier (see Conceptual modelling), although these individuals do not draw a distinction between the conceptual and the implementation, they do appear to develop 'skeleton models' to present to clinical experts, which evolve iteratively and take shape over time.

I will actually literally take them through a patient path and actually run through, do a step-by-step running of the model in front of them.

[I5]

However, one respondent highlighted the potential dangers of developing software models without having fully determined the underlying structure or having the agreement of experts:

...whether it's adding something in or taking something out there is the worry at the back of your mind that it's going to affect something else in a way that perhaps you don't observe... there is a danger if you make the decision to include or exclude definitively that you might regret it later on, it might not be a question of states it might be a question of which outcomes to model on as well...

[I12]

The interview data suggested that model refinement activities were less iterative and burdensome for those respondents who had the conceptual model agreed or 'signed-off' before implementation; this was particularly true for one respondent:

...then build an EXCEL model...it's an iterative process; you don't want to keep on rebuilding your model,...I can't imagine what I'd do in EXCEL which would make me rethink the way I structure my model...most of the big issues I see are all about the sort of thought processes behind defining that decision problem, defining the structure, defining the core set of assumptions...if we can get agreement about that, the implementation of it is really straightforward.

[I7]

At the other extreme, however, one respondent mentioned rare occasions in which the research team had developed a model, decided it was not adequate and subsequently started from scratch.

Obviously it doesn't very often happen that we have built a whole model and have had to tear

up a whole model but if necessary we would rather do that than carry on with a model that isn't what we intended.

[I5]

## Model checking

Given the extent of discussion around model checking during the interviews, a detailed analysis of current methods for checking models is presented (see Chapters 6 and 7). Model checking activities varied and appear to occur at various stages in the model development process. One interviewee suggested that model checking begins once the model has been built; however, for other respondents, the checking process happens at various distinct stages. One respondent indicated that model checking activities are undertaken throughout the entire model development process. Whereas, subject to certain nuances in the precise methods employed, much of the discussion around current model checking activities focused on testing and understanding the underlying behaviour of the model and making sure that it makes sense:

...it's about finding out what the real dynamics of the model are so what are the key parameters that are driving the model, where do we need to then emphasise our efforts in terms of making sure our data points are precise? What are the sensitivities, in which case what subgroups might we need to think about modelling? Discussing how the model is behaving in certain situations. What the likely outcome is. If the ICER is near to the critical thresholds for the decision-maker then where we need to pay attention if the ICER is way off to the side, if you've got an output that's dominated or if you have got an ICER that's half a million pounds it seems likely that unless the model is wildly wrong the decision is going to be fairly clear cut. So if you are operating at the decision-making threshold how precise and careful you have to be and what kind of refinement you need therefore to build into the model.

[I4]

Other model checking activities included examining the face validity of the model results in isolation or in comparison to other existing models, internal and clinical peer review, checking the input data used in the model and checking premodel analysis can replicate the data, and

checking the interpretation of clinical data. Many of these approaches involve a comparison of the actual model outputs against some expectation of what the results should be; the interview data were not clear how such expectations were formed.

Either the intuition or the modelling or the data is wrong and we tend to assume that it is only one of them...I think you tend to assume that once they're [clinical experts] not surprised by the thing then, that means you have got it right.

[I5]

Two respondents indicated a minimalist approach to model checking activities:

So we do just enough...just enough but not as much as you'd want to do.

[I10]

...it depends whether you're error searching for what I term significant errors or non-significant errors, I probably would stop and the non-significant errors that are probably in a larger percentage of models than people would like to know would remain.

[I2]

## Reporting

The majority of respondents referred to model reporting as the final step in the model development process. This stage typically involved writing the report, translating the methods and results of the model and other analysis; and engaging with the decision. Although little of the interview content concerned this phase of the process, two important points were raised within the interviews. First, one respondent highlighted that the reporting stage may act as a trigger for checking models; where the results are unexpected or queried by internal or external parties. The second point concerned the importance of the interpretation of the model by decision-makers and other process stakeholders:

I think the recommendation we would make is we need to pay more attention to understanding how our models are understood and how we present them, a lot of work that can be done in model presentation.

[I4]

...reportable convention standards can be very important in ensuring everyone has a clear view of what's being said. There are ways in which model outputs can be more transparently depicted and the key messages conveyed to users more clearly, often some can get lost in the text or hidden away somewhere either intentionally or unintentionally.

[I4]

## Synthesised model development process

Figure 1 shows a stylised representation of the model development process generated through a synthesis of all interview transcripts and from the charted descriptions of the model development process elicited from the interviewees. The stylised model development process is comprised of five broad bundles of activities including; understanding the decision problem, conceptual modelling, implementing the model within a software platform, checking the model, and finally engaging with the decision. It should be noted that these activities do not perfectly match the coding scheme used to undertake the Framework analysis. The general remit of these sets of activities is as follows.

**Understand the decision problem** This phase involves immersion in research evidence, definition of research question, engaging with clinicians, engaging with decision-makers, understanding what is feasible, understanding what the decision-maker wants/needs and engaging with methodologists.

**Conceptual modelling** It is a commonplace to state that all mathematical models are built upon a conceptual model. At its most banal it is impossible to implement a model without first conceiving the model. The key issue in the decision-making endeavour and reflected in the modelling process is communication: specifically the process of developing a shared understanding of the problem. Conceptual modelling is the process of sharing, testing, questioning and agreeing this formulation of the problem; concerned with defining the scope of a model and providing the inputs to the process of systems analysis and design associated with defining a solution to the problem. The scope of a model deals with defining the boundary and depth of a model, to identify the critical factors

that need to be incorporated in the model. Models are necessarily simplifications of our understanding of a problem. The conceptual model allows a description of one's understanding of the system that is broader than the description of the system captured in the model. An implemented model will therefore be a subset of a conceptual model. This hierarchy allows simplifications represented in the model to be argued and justified. This process of simplification is the process of determining relevance.

The arrow leading from 'understanding the decision problem' to 'implementation model' reflects the views of three interviewees who implied that they either do not build a conceptual model or that conceptualisation and implementation are simultaneous activities.

**Implementation model** This phase involves the implementation of the model within a software platform.

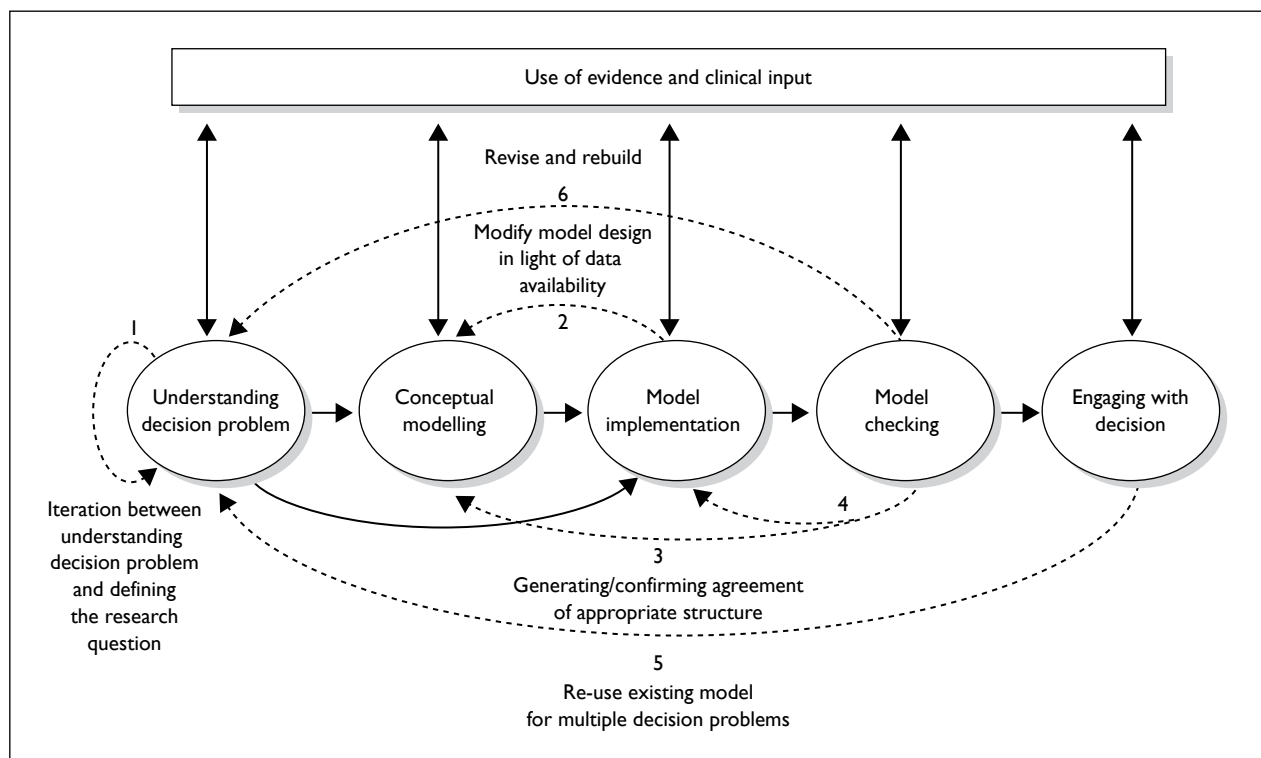
**Model checking** This stage includes all activities used to check the model. This includes engaging with clinical experts to check face validity, testing extreme values, checking logic, checking data

sources etc. A detailed description of current model checking activities is presented in Chapters 6 and 7.

**Engage with decision** This phase concerns the reporting and use of the model with and by the decision-maker.

Activities concerning the use of evidence, peer review and other clinical consultation may arise within any or all of the five model development phases.

Several points are of note in the interpretation of the diagram. First, as the diagram has been developed through a synthesis of all interview data, respondents did not adopt a uniformly standard model development process, and respondents discussed different activities to varying degrees of detail, the lines of demarcation between main sets of activities were not entirely clear. Second, the diagram is intended to be descriptive rather than prescriptive; it describes current model development processes rather than a normative judgement about how models should be developed. The dashed arrows in *Figure 1* indicate significant iterations between key sets of model development



**FIGURE 1** Stylised model development process showing key iterations between sets of activities.

activities. It should also be noted that each individual phase is likely to involve iterations within sets of activities, e.g. there may be several versions of an implementation model as it is being developed.

The iterative nature of developing mathematical models is well documented. The interview data suggest six key points of model iteration within the development process. The first substantial iteration (marked as number 1 on *Figure 1*) concerns developing an understanding of the decision problem; this may involve iterations in terms of striking a balance between what the analyst believes are the decision-makers needs, what is feasible within the project resources and negotiations therein.

The second notable iteration (marked as number 2 on *Figure 1*) involves looping between the development of an explicit conceptual model and its implementation in a software platform. Several respondents highlighted the inevitable existence of a circular loop between these two sets of activities whereby the intended conceptual model structure is redefined in light of limited evidence available to populate that structure, and whereby the evidence requirements are redefined in light of the intended model structure.

The interview data suggest two further loops relating to iterations from the model checking stage to revising or refining the conceptual and/or implementation model (marked as numbers 3 and 4 on *Figure 1*). The interview data suggest that any of the large number of checking activities currently undertaken have the capacity to result in backwards iterations to earlier steps.

The fifth key iteration of note (marked as number 5 on *Figure 1*) concerns the use of an existing model for multiple decision problems. In such instances, there is a loop between the final and first stages of the process. Two respondents highlighted examples of this iteration: one in terms of the use of 'global models' for use in different decision-making jurisdictions, and one in terms of the ongoing development of independent models across multiple appraisals. As noted earlier (see Use of information in model development), several respondents discussed the use of existing models as the basis for developing a model structure. In such circumstances, there is a danger that adopting an existing model developed by another party and applying this to a new decision problem could effectively represent a loop back to a conceptual

model without a comprehensive understanding of the decision problem.

The final iteration (marked as number 6 on *Figure 1*) loops from model checking to understanding the decision problem. The three respondents who highlighted that implementation and conceptualisation activities occur in parallel all indicated the possibility of revising and rebuilding the model from scratch. The same suggestion was not indicated by the remaining nine respondents. At the other extreme, one respondent indicated that there was no significant backwards iteration once the implementation model was under way, i.e. the conceptual model was not amended once agreed.

## Discussion of model development process

The analysis presented within this chapter highlights considerable variation in modelling practice between respondents. This may be in part explained by the apparent variations in expertise, background and experience between the respondents. The key message drawn from the qualitative analysis is that there is a complete absence of a common understanding of a model development process. Although checklists and good modelling practice have been developed, these perhaps indicate a general destination of travel without specifying how to get there. This represents an important gap and should be considered a priority for future development.

The stylised model development process presented in *Figure 1* represents a summary of all interviewees' stated approaches to modelling. It should be noted that the diagram does not entirely represent the views of any single individual, but is sufficiently generalisable to discriminate between the views of each interviewee; in this sense, any modeller should be able to describe their own individual approach to model development through reference to this diagram. As noted above, one particularly apparent distinction that emerged from the qualitative synthesis was the presence or absence of explicit conceptual modelling methods within the model development process. Although it is difficult to clearly identify a typology on these grounds from a mere 12 interviewees, the qualitative data are indicative of such a distinction. Related to this point is the issue of the a priori specification of the model design and analysis, which represents the leap from the conceptual model to the

implementation model. Some respondents did discuss the preparation of materials which go some way in detailing proposed conceptual models, platforms, data sources, model layout, checking activities and general analytical designs; however, the extent to which these activities are specified before implementation appears to be limited.

A common feature of modelling critiques that is reflected in the interviews is the statement that different teams can derive different models for the same decision problem. This arises from the largely implicit nature of existing conceptual modelling processes described by the interviews. It is suggested that focusing on the process of conceptual modelling may provide the key to addressing this critique. Furthermore, the quality of a model depends crucially on the richness of the underlying conceptual model. Critical appraisal checklists of models all currently include bland references to incorporating all important factors in a disease yet do not address how these ‘important

factors’ can be determined. It is the conceptual modelling process that is engaged with identifying and justifying what are considered important factors and what are considered minor or irrelevant factors for a decision model.

The conceptual modelling process is centrally concerned with communication and this can be supported by many forms including, conversation, meeting notes, maps (causal, cognitive, mind etc.), skeletal or pilot models. However, care needs to be exercised in using methods that focus on the problem formulation rather than formulation of the solution.<sup>26</sup>

The description of the model development processes adopted by the interviewees within this chapter is directly used to inform the development of the taxonomy of model errors (see Chapter 5) and to provide a context for analysing strategies for identifying and avoiding model errors (see Chapters 6 and 7).



# Chapter 4

## Definition of an error

### Overview

This chapter presents the description of key dimensions and characteristics of model errors and attempts to draw a boundary around the concept of 'model error', i.e. '*what is it about a model error that makes it an error?*'. The descriptive analysis highlighted both overt and subtle variations in respondents' perceptions concerning what constitutes a model error. In seeking to explain, assign meaning and draw a boundary around the concept of model error, the respondents identified a variety of factors including:

- non-deliberate and unintended actions of modellers in the design, implementation and reporting of a model
- the extent to which the model is fit for purpose
- the relationship between simplifications, model assumptions and model errors
- the impact of error on the model results and subsequent decision-making.

The above factors are discussed in this chapter, the interview findings are then placed in the context of the literature on model verification and validation.

### Key dimensions of model error

*Table 2* presents the key characteristics and dimensions of model error as suggested by the 12 interview respondents (note: this has attempted to retain the natural language of respondents but does involve a degree of paraphrasing by the authors).

*Table 2* highlights a diverse set of characteristics of model errors as discussed by interviewees. Evidently, there is no clear consensus concerning what constitutes a 'model error'. Conflicting views were particularly noticeable in factors concerning model design, for example model structures, assumptions and methodologies. The respondents' explicit or implied construct of a model error appeared to be considerably broader than 'technical errors' relating to the implementation of the model.

### Intention, deliberation and planning

The interview data strongly suggested a general consensus that model errors were unintended, unplanned or non-deliberate actions; instances whereby the modeller would have done things differently had they been aware of a certain aspect or underlying characteristic of the model; or instances whereby the model produced unexpected or inconsistent results. The unintended actions of the modeller and the non-deliberate nature of the modellers' actions when developing models appeared to be central to the interviewees' definitions of model error.

### Extent to which the model is fit for purpose

A commonly cited dimension of a model error concerned the extent to which the model could be considered fit for purpose. In particular, respondents highlighted that a model could be unfit for its intended purpose in terms of the underlying conceptual model as well as the software implementation model, so representing distinct errors relating to model design as well as errors relating to the execution of the model. Respondents implied a relationship between the concept of model error and the extent to which a model is fit for purpose, suggesting that the presence of model errors represents a threat to the model's fitness for purpose, whereas the development and use of a model that is unfit for its intended purpose is a model error in itself.

Several interviewees drew reference to concepts of verification and validity as a means through which to define the concept of model error. Where interviewees used these terms there was a consensus regarding their definition, with verification referring to the question 'is the model right?' and validation referring to 'is it the right model?'. Although there was general consensus that the term model error included issues concerning the verification of a model, there was less clarity about

TABLE 2 Dimensions/characteristics of model error

Dimensions which are perceived to constitute a model error	Dimensions which are perceived not to constitute a model error
<p>a model that is not fit for the purpose for which it was intended</p> <p>something that causes the model to produce the wrong result</p> <p>something that causes the model to lead to an incorrect interpretation of the results/answer</p> <p>an aspect of the model that is either conceptually or mathematically invalid</p> <p>failure to accord with accepted best practice conventions – not doing something that you should have done</p> <p>use of inappropriate assumptions that are arbitrary</p> <p>use of inappropriate assumptions that contradict strong evidence</p> <p>use of assumptions that the decision-maker does not own, do not feel comfortable with or cannot support</p> <p>choices made on the grounds of convenience rather than evidence</p> <p>something that causes the model to produce unexpected results</p> <p>something you would do differently if you were aware of it</p> <p>an aspect of the model that is unambiguously wrong</p> <p>an aspect of the model that does not make sense</p> <p>an aspect of the model that does not reflect the intention of the model builder for an unplanned reason</p> <p>a model that inappropriately approaches the scope or modelling technique</p> <p>an unjustified mismatch between what the process says should be happening and what actually happens</p> <p>an inappropriate relationship between inputs and outputs</p> <p>an implementation model that does not exactly replicate the conceptual model</p> <p>something that markedly changes the incremental cost-effectiveness ratio irrespective of whether it changes the conclusion</p> <p>something in the model that leads to the wrong decision</p> <p>a mistake at any point in the appraisal process – from decision problem specification to interpretation of results</p>	<p>software bugs</p> <p>unwritten assumptions that are never challenged</p> <p>choices made on the grounds of evidence</p> <p>matters of judgement</p> <p>inconsistencies with other reports</p> <p>simplifying assumptions that are explained</p> <p>use of simpler methodologies/assumptions</p> <p>aspects of the model that do not influence the results</p> <p>small technical errors with limited impact</p> <p>overoptimistic interpretation of data</p> <p>mistakes that are immediately corrected</p> <p>reporting errors</p>

the relationship between model errors and validity. For instance, problems in the conceptualisation of the decision problem and the structure of the model were generally referred to as ‘inappropriate’ rather than ‘incorrect’, yet nonetheless these were described as model errors by some respondents.

I think it’s probably easier to know that you have made an error of verification where there’s clearly a bit of wrong coding or one cell has been indexed wrongly or the wrong data has been used for example which wasn’t intended. So those are all kinds of error that are all verification type errors. Errors in the model rather than it being the wrong model which is a much bigger area the area of validation and there people can argue and it can be open to debate.

[I4]

Further, it should be noted that one interviewee indicated that errors affecting the model validity, that is errors in the description of the decision problem and conceptual model potentially dwarfed technical errors that affected its verification.

...my belief is that if you spend...the majority of your time, as I say, getting the decision problem right, making sure you understand data and how that data relates to the decision problem and decision model, that...the errors of programming itself, those ones which are much more minor.

[I7]

A repeatedly stated dimension of model error that emerged from the interviews was the failure to adhere to best practice or failure to accord to current conventions. Alternative viewpoints

included the notions of failure to do *'the best you can do'* or failure to do *'what you should do'*. It was noted that in some instances this may be the result of a lack of skills. Further complexity was apparent in terms of the selection of modelling method and its relationship to model errors:

I don't think I would class [it] as an error when someone says that...they developed this hugely complex...kind of Markov-type model, and then said they'd much rather have done it as a simulation.

[I11]

## Simplification versus error

The qualitative synthesis suggested mixed views between the concepts of model assumptions and model errors. For example, one respondent appeared to hold a strong view that all inappropriate assumptions were model errors, highlighting the importance of the perspective of the stakeholder (or model user) in distinguishing between the two concepts:

...One man's assumption is another man's error...

[I3]

Other respondents highlighted a fine line between the necessary use of assumptions, in model simplification and the introduction of errors into the model:

If you have an inaccuracy in the model that has no, or a trivial effect, then it's not obvious to me where that stops being a simplification that you've made to make the decision problem tractable, and where it becomes an error,... something that's inaccurate and we wish to avoid...that point would be where it starts to cause a risk that the model is giving a wrong answer, or is not addressing the question that you wanted it to.

[I6]

...it's very rare that the model is completely accurate in that it does exactly what it says it's meant to be doing. And it's never the case that the model is completely accurate in that it's a full and complete picture of the world. And I'm very comfortable with that, because the whole point of doing your model is that it's a simplification, and that you're trying to extract a simplified view of the world that teaches you

something new that wasn't obvious from the data that you start with.

[I6]

Further to this discussion, one respondent highlighted a distinction between assumptions and errors by examining the underlying reason for the use of the assumption:

So you have to say, 'Well, to what extent are the choices being made being made on the grounds of evidence, and how much were they on the grounds of convenience?' Now if they're made on the grounds of convenience and in addition to that you know, from my point of view they are mathematically illiterate, then that is an error, to me. They are unsustainable.

[I3]

Contrary to this viewpoint, one respondent suggested that unreasonable assumptions and the overoptimistic interpretation of evidence did *not* constitute a model error because of their deliberate intent. An alternative suggestion for distinguishing between assumptions and model errors concerned the perspective of the decision-maker; assumptions were errors if the decision-maker does not 'own' them, does not feel comfortable with them, or cannot support them upon a wider appraisal of evidence.

## Impact of error on the model results and subsequent decision-making

A number of interviewees suggested that the existence of model errors is inevitable; that it is only important to ensure that the model is free from significant errors which have an important influence on the model outputs and the resulting policy decision.

...I think it's rare to find any model without an error in it somewhere...I think we're deluding ourselves if we think...that there are no errors in our models...I think [that] even if you identify one, that doesn't mean there aren't others, and if you don't identify one that doesn't mean there are none.

[I5]

An error is something that markedly changes the ICER regardless of whether it changes the conclusion.

[I2]

The problem is...can you really cope with that error if someone's making a policy decision based on that outcome?

[I4]

One interviewee made the interesting observation that the impact of an error should be considered relative to the overall uncertainty in results:

...if it changes [the incremental cost-effectiveness ratio] from £8266 to...£8267 technically it's an error, the question is, those technical errors will be absolutely dwarfed by the uncertainty in the efficacy of the drug, so I wouldn't consider them an error.

[I2]

This begs the question how wrong does something have to be before it becomes a model error?

Consequently, for some interviewees, this led to a very broad boundary around what would be defined as an error:

...it's a mistake at any point in the...appraisal process, from the initial specification of a decision problem through to the assumptions, the parameter inputs and the analysis that goes behind that, through to the final implementation and programming of an EXCEL-based model. And even then, an error in the interpretation of the results therein.

[I7]

## Placing the interview findings in the context of literature on model verification and validation

In 1979 the Society for Modelling and Simulation International (SCS) defined the distinction between model verification and validation.<sup>27</sup>

- Model validation is defined as 'substantiation that a computerised model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model'.
- Model verification is defined as 'substantiation that a computerised model represents a conceptual model within specified limits of accuracy'.

The definitions of model validation and verification provided by the HTA modellers interviewed in this study were best captured by the description that verification refers to the question 'is the model right?' and validation refers to the question 'is it the right model?'. These definitions are consistent with the SCS definitions. Although the interviewees' descriptions might lack the apparent rigour of the SCS definitions, it should be noted that this rigour is perhaps somewhat illusory because in both cases the SCS definitions simply recast the problem as one of defining what constitutes a satisfactory range of accuracy.

In searching for rigorous approaches to validation and verification of Operational Research models many authors draw reference to epistemology and philosophies of science. In 1967 Naylor *et al.*<sup>28</sup> rehearsed the opposition between rationalism and empiricism and proposed a three-stage approach to model verification based explicitly upon merging these principles; it should be noted that at this early stage the terms verification and validation were used interchangeably. Naylor *et al.* suggest that rationalism is represented in the initial search for a set of postulates regarding the components, variables and functional relationships that constitute a model and specifically that the search for face validity of a model by its accordance to expert judgement is a rationalist perspective on validity. However, the authors consider these initial postulates only as 'tentative hypotheses' on the behaviour of the system and suggest that these postulates should be verified empirically, and that this verification constitutes an empiricist perspective on validity. The third stage of validation suggested by Naylor *et al.*, consists of testing the ability of a model to predict, retrospectively or prospectively, the behaviour of a system and represents an application of a positive economics perspective on model validity. The three stages outlined above are closely reflected in the HTA interviewees' description of the model development process outlined in Chapter 3 and indeed in the discussion of simplification versus error contained earlier in this chapter.

This early literature on model verification and validation focuses on modelling as a knowledge creation process. This perspective has led to validity being defined with reference to a model's ability to predict observable phenomena. In certain methodological work this epistemological perspective has led to the definition of statistical criteria for the goodness of fit of a model as

a measure of its validity.<sup>29,30</sup> None of the HTA modellers interviewed pursued this perspective of validity.

In 1993 the *European Journal of Operational Research*<sup>31</sup> published a special issue on validation in modelling, with the purpose of raising debate on this topic. The papers presented in this special issue expand upon the epistemological roots of model validation methodology. A key characteristic of operational research modelling is the focus, not on creating knowledge, but rather on supporting decision-makers. Moreover, supporting decision-makers working within a broad social and political context, working with a range of criteria and with particular approaches to the decision-making process. This focus on supporting decisions underlies the importance of the pragmatic, instrumentalist or utilitarian philosophies for model validation. Under these pragmatic approaches the usefulness of a model, that is its success in action, is its measure of validity. Déry *et al.*<sup>32</sup> acknowledge Raitt<sup>33</sup> and Rorty<sup>34</sup> as first explicitly bringing this utilitarian approach to the discussion of model validation. The interviewees' principle focus on fitness for purpose as opposed to statistical accuracy of models reflects strongly this instrumentalist viewpoint on validity, all of the HTA modellers interviewed were consistent in adopting this perspective.

Kleindorfer *et al.*<sup>35</sup> in 1998 reviewed the philosophical underpinnings of model validation and extended the discussion to the implications of *Hermeneutics* as propounded by Gadamer<sup>36</sup> and more recently Bernstein.<sup>37</sup> Hermeneutics recognises that objectivity is unachievable and suggests that meaning is created through intersubjective communication. It is characterised by a description of rationality that is historically situated and practical, involving choice, deliberation and judgement. Kleindorfer's reading of Hermeneutics is used to provide a basis for the primacy of interaction between the modeller and client in developing mutual understanding of a model. It is this mutual understanding that establishes a model's significance and its warranty. Kleindorfer *et al.* propose a framework for model validation based upon the analogy of a judicial system, where the model builder is free to establish the credibility of the model through any reasonable means. They argue that most model practitioners instinctively operate in this middle ground between objectivism and relativism in attempting to achieve model credibility. Whereas the HTA

modellers interviewed recognised the importance of ensuring mutual understanding in the definition of the problem and structure of the model, none of the interviewees explicitly raised the issue of model credibility in considering the definitions of model error, although this may well have been assumed. The interviews demonstrate that HTA modellers tend to act in the manner described by Kleindorfer *et al.*, using all means at their disposal to demonstrate model validity and gain credibility, without necessarily referring to any philosophical underpinning for this mode of action.

Decision analytic modelling is essentially a Bayesian undertaking, both in the theoretical underpinnings and in the central role of the Bayes updating formula in the statistical analysis of decision problems. The Bayes formula provides a statistical mechanism for updating probabilities or beliefs in the light of new observations. There is a substantial methodological and applied literature on Bayesian methods. Most applications of the Bayesian approach in the HTA domain apply the updating procedure to refining estimates of model parameters within a model in the light of new data or in facilitating probabilistic sensitivity analysis. However, the same principle can be applied in approaching the validity of a model. In the Bayesian approach, the problem of validity is framed explicitly in terms of model credibility. Hence, the credibility, or measure of belief, in a theory or in our case mathematical model is updated in the light of new evidence.<sup>38</sup>

$$P(m|d) = \frac{P(d|m)P(m)}{P(d|m)P(m) + P(d|\text{not } m)P(\text{not } m)}$$

Where m = model and d = data.

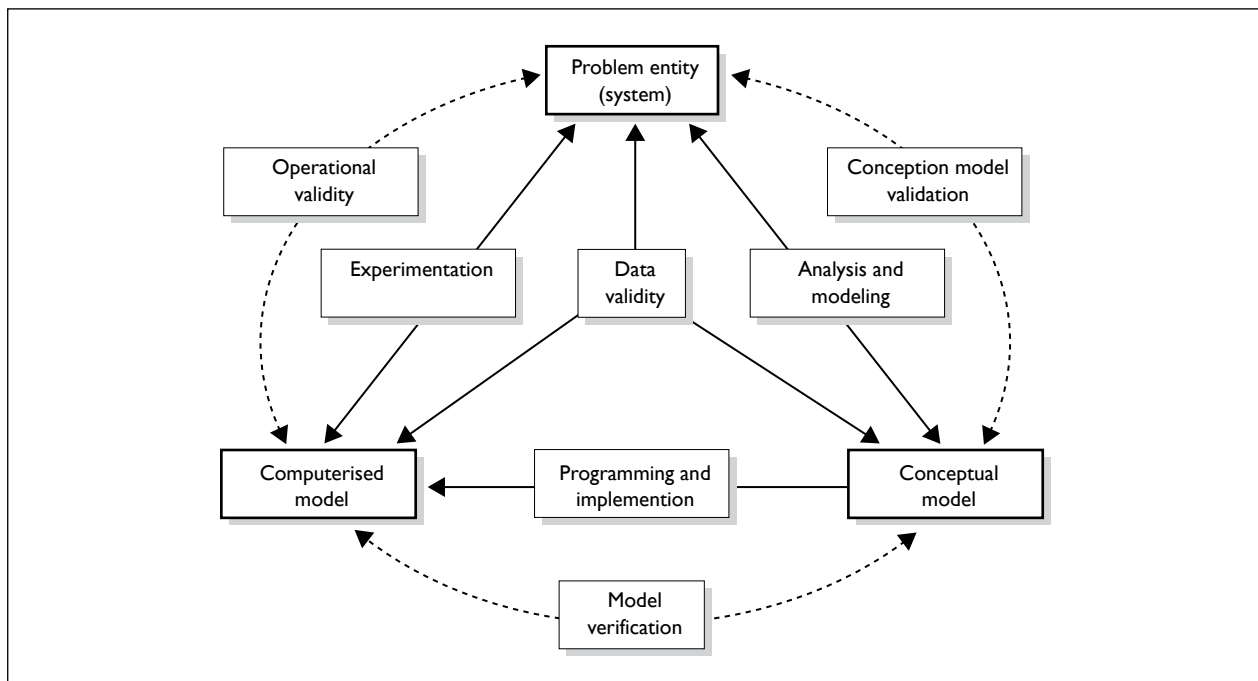
This latter approach is closely related to handling structural uncertainty in models through model averaging methods. A number of issues arise in this approach principally concerning the nature of the subjective prior probability of a theory or model, and the difficulty in accounting for the open nature of possible models, i.e. the impossibility of constructing a complete set of possible models to consider in the model averaging process or assigning an adequate assessment of the probability that none of the described models are correct. Furthermore, it should be noted that there is almost a complete absence of discussion of model verification or operational validity of the model within the decision-making process within the Bayesian literature.

Sargent<sup>39-41</sup> in a series of closely related conference proceedings describes a refreshingly practical approach to verification and validity of models without recourse to epistemology or philosophy. Sargent uses a simple model of the model development process, reproduced in *Figure 2*, to develop different aspects of model validity and verification.

In this scheme the overall validity of a modelling project comprises the validity of the conceptual model, the veracity of the implemented model to the conceptual model and the operational validity of the results and interpretation of the results in the decision-making process. It should be noted that using verification to describe the correct implementation of the conceptual model in the computerised model, has resonance with the logical empiricist philosophical approach to verification as discussed by Déry *et al.*<sup>32</sup> However, this highlights the fact that this definition of verification relies on their being an explicit and complete description of the conceptual model that acts as a specification of the computerised model. Where the description of the conceptual model is absent or incomplete, this separation between the concepts of verification and validation breaks down. As has been discussed in the previous section the absence of an adequate description of the conceptual model is a common feature of HTA modelling practice.

The methodological literature on model validation becomes sparse after the 1993 *European Journal of Operational Research* Special Issue,<sup>31</sup> somewhat ironically given its stated aims of promoting discussion. Citations of the articles contained therein (searches undertaken January 2009) consist mainly of developments of validation approaches in specific application domains, for example Oreskes *et al.*<sup>42</sup> in the earth sciences and indeed this monograph in HTA. There are few generic methodological papers<sup>41,43,44</sup> and these focus on pragmatic approaches to validation.

The HTA modellers interviewed demonstrate a pragmatic approach to model validation. Although there may certainly exist sound philosophical underpinnings for this stance, it is unclear, firstly, the extent to which this is a satisfactory position and secondly the extent to which practitioners are consciously implementing this as principled pragmatism or are implementing an unprincipled *laissez faire* approach to validation. In their seminal 1967 paper, Naylor *et al.*<sup>28</sup> described the problem of ascertaining the validity of a model as ‘perhaps the most elusive’ unresolved problem in computer simulation methods; the level of conflicting opinions and practices identified in the literature and in the interviews undertaken in this study indicates that there is still room for development in this area.



**FIGURE 2** Model of the modelling process. Reproduced with permission from the Institute of Electrical and Electronics Engineers, Sargent,<sup>41</sup> ©2004 Institute of Electrical and Electronics Engineers, Inc.

# Chapter 5

## A taxonomy of errors

### Introduction

This chapter presents a taxonomy of errors in HTA models, with the purpose of developing a common language for the development of methods and processes for identifying errors and avoiding errors in HTA models. The taxonomy provides a basis for relating current strategies for avoiding and identifying model errors to the types of errors they are intended to impact upon (see Explanatory analysis concerning methods for preventing errors in Chapters 6 and 7).

Identifying errors and avoiding errors imply two different aspects to the taxonomy. First, identifying errors implies a focus on error symptoms and error checking processes, whereas avoiding errors implies a focus on the process of model development and implementation. The taxonomy is not designed with the purpose of providing a checklist for the critical appraisal of models. As a result, where terminology such as 'error of judgement' is used, this is not meant to necessarily imply that a process is required or may exist for identifying a particular instance of judgment as being in error.

A two-stage process was used in the development of the taxonomy. First, an interview-based taxonomy was developed based upon the framework analysis of the in-depth interviews undertaken with a sample of the HTA modelling community. Before discussing individual error types, the interviewees were asked to describe their conception of

the decision support process, through model development, implementation and use. This description of each individual's perception of the model development process was then used as a vehicle for the discussion of the error types. This helped to ensure that the discussion of error types, and the balance of error types across the modelling process, is not biased towards different parts of the process by the nature of the interviews and perceptions of the interviewees. In addition, non-HTA literature regarding spreadsheet and programming error taxonomies was used as a basis for comment on the current understanding and as a means to further develop the taxonomy (see Review of taxonomy/classification literature).

*Table 3* below presents the distribution of comments from the interviews across the major themes. It is notable that over 70% of comments related to the problem definition, structuring process and the use of evidence, whereas only 17% of comments related to actual errors in the implementation of models.

This focus among HTA modellers on errors in the 'softer' side of the model development process is supported by explicit comments throughout the interviews.

...my concerns about quality of outputs and accuracy is much more around the design of the conceptual model and asking the correct questions than it is about the implementing. Which is not, that they're not both important, but that it's easier to do guidelines around

**TABLE 3** Distribution of comments across major themes

	Comments	
	n	%
Error in understanding the problem	21	9
Error in the structure and methodology used	71	31
Error in the use of evidence	70	31
Error in implementation of the model	38	17
Error in operation of the model	6	3
Error in presentation and understanding of results	20	9
Total	226	100

the technical aspects and it's something that as modellers you're more interested in, it's your distributions and your methods and your environments, and it's things that you can write down precisely and measure.

[I6]

But it's much easier, I would think, to ask the wrong question and [that] not get found, and for that not to be picked up until the end, than it is to make a major mistake in coding and for that not to be picked up.

[I6]

These concerns relate directly to the synthesis of respondent's views of what constitutes a model error (see Chapter 4), whereby inappropriate model assumptions and matters of judgement were clearly considered to represent model errors by some interviewees but not by others.

The next five sections on errors describe the interview-based taxonomy of model errors presented in *Figure 3* structured according to the six major coded themes presented in *Table 3*, noting that the discussion of 'error in operation of the model' and 'error in presentation and understanding of results' have been merged, owing to the limited number of comments in these domains and their close association. The taxonomy is structured vertically according to the major coded themes that represent different aspects of the model development process constituting error domains, these major domains are each broken down into subdomains. Types of errors described by interviewees in each domain are then discussed; these are presented in column 2 of the taxonomy. For each type of error there has been an attempt to identify potential root causes of errors, these are presented in column 3 and are grouped together again in column 4 to identify major themes in the causes of error.

## Error in understanding the problem

Well, I guess it all stems from what is the decision problem, to start off with. So...you know, no matter what you've done, if you've got the decision problem wrong...basically it doesn't matter how accurate the model is. You're addressing effectively the wrong decision.

[I7]

The key error domains in the understanding and description of the decision problem identified by the interviewees were:

- error in the definition of population and subgroups
- error in the definition of interventions and comparators
- error in the definition of outcomes.

The charted interview data were examined to identify components that make up the description of a decision problem. Interviewees explicitly mentioned the definition of comparators, interventions and populations as being important in the description of the decision problem and potentially being subject to error. Five out of seven comments mentioned the comparators and/or interventions, which were sometimes indistinguishable, and three comments identified the population as a key element. Comments included:

Their conceptualisation missed out strategies that were highly likely to be the most cost-effective.

[I2]

...we have to be very clear from the outset that the specification of a decision problem... and I guess that's got a number of different levels, which is, you know...what is the patient population that we're interested in, what are the relevant potential subgroups that lie therein.

[I7]

Errors in the definition of populations, interventions and comparators seemed to be associated with subtle aspects of their definition, for example when considering populations it is important to appreciate the differences between subgroups, or in the definition of interventions or comparators it is the treatment strategy, that is the definition of treatment sequences, or missed comparators that give rise to errors.

It is also notable that none of the interviewees mentioned the definition of the outcome as potentially problematic when discussing the description of the problem. This probably reflects an assumption that has become axiomatic among HTA modellers that decision-makers require and are interested primarily in the generic cost-effectiveness outcome of the incremental cost per



Error domain	Description of error	Root cause error	Index	Grouping of causes
<p><b>Error in description of the decision problem</b>                      Definition of populations and subgroups</p> <p>Definition of interventions and comparators</p> <p>Definition of outcomes</p>	<p>Incoherent with clinical understanding of the disease</p> <p>Inadequate description of the boundary and content of the assessment</p>	<p>Error in process – failure of communication between stakeholders</p>	10	Communication: All stakeholders
<p><b>Error in model structure and methodology used</b>                      Development of conceptual model (Epidemiology, natural history, management)</p>	<p>Error in simplifying assumption</p>	<p>Counterfactual evidence exists</p> <p>Where there is no evidence or equivocal evidence there should be transparency in:                      a) the description of any assumptions                      b) underpinning evidence or judgements                      c) the potential to impact on outcomes and decision-making</p> <p>Lack of such transparency may constitute an error in process or methods</p>	20 30	Judgement error – interpretation of evidence Communication: Model reporting
<p>Selection of modelling methods</p>	<p>Obsolete or outdated model structures</p> <p>Error in conceptual modal structure</p>	<p>Error in evidence gathering</p> <p>Misinterpretation of evidence – error in judgement, misinterpretation of statistics</p>	40 50	Methods error – evidence gathering Judgement error – interpretation of evidence
<p>Moving from conceptual model to implementation</p>	<p>Inappropriate selection of modelling method</p> <p>Error in selection of modelling software</p> <p>Premature implementation of model structure</p>	<p>Judgement error</p> <p>Modelling process error – failure to consider system analysis and design</p> <p>Judgement error under time and resource constraints</p> <p>Judgement error – modelling skills</p>	60 70 80 90	Judgement error – methods selection Methods error – system analysis and design Methods error – system analysis and design Judgement error – modelling skills

FIGURE 3 Taxonomy of errors in HTA modelling: parts I–III.

Error domain	Description of error	Root cause error	Index	Grouping of causes
<b>Error in the use of evidence</b>				
Generic evidence processes	Inadequate capturing of the evidence	Inadequate search terms Incomplete use of evidence sources Deliberate selection of evidence	100 110 120	Methods error – evidence gathering Judgement error – interpretation of evidence Direct violation
	Error in systematic review of evidence	Data extraction errors Misunderstanding through failure of communication within team	130 225	Slips Communication: Within team
Development of model structure	Errors arising from misinterpretation of data	Misunderstanding through failure of communication within team  Deliberate misinterpretation or over-optimistic interpretation	150  160	Communication: Within team  Direct violation
		Errors of judgement	170	Judgement error – interpretation of evidence
	Inadequate capturing of the evidence	Error in evidence gathering methods Inadequate methods of identifying and eliciting subjective judgement	180 190	Methods error – evidence gathering Judgement error – methods selection
Populating data inputs	Errors in statistical analysis of data	Judgement error – experience, skills and training	200	Judgement error – statistical methods
	Error in the selection of distributions for characterising uncertainty	Judgement error	210	Judgement error – modelling skills
	Use of evidence incoherent with critical appraisal – failure to adjust for study quality	Error in modelling methods	220	Judgement error – modelling skills

Error domain	Description of error	Root cause error	Index	Grouping of causes
<b>Error in implementation of the model</b>				
Coding of individual cells	Error in cell references/variable calling Error in data Error in text Error in formulae/operators	Typing errors Copying errors	230 240	Slips Slips
Coding of logic structures	Error in implementation of modelling methods	Judgement error; skills and experience	250	Judgement error – modelling skills
<b>Error in operation of the model</b>				
Setting up model runs	Error in data Error in environment variables	Typing errors Failure to update values/model switch appropriately	260 270	Slips Lapses
<b>Error in the presentation and understanding of results</b>				
Presentation of results	Errors in results Errors in tables	Transcription errors	280	Slips
Communications of results and conclusions	Error in communication of results	Misunderstanding decision-makers' perception of problem (disease, health economics, modelling, decision-making approach)	290	Communication: Model reporting
	Failure in communication through loss of model credibility	Failure to present adequate range of information on outcomes and key drivers (this can imply an error in problem description)	300	Communication: Model reporting
	Errors in drawing incorrect conclusions from results	Over interpretation of results	310	Direct violation

FIGURE 3 Taxonomy of errors in HTA modelling: parts I–III. (continued)

quality-adjusted life-year (QALY) gained. However, in discussing the presentation and understanding of results one interviewee indicated that presentation of disease-specific results was a key element in helping decision-makers to understand the nature of the model results:

So cost per QALY amalgamates a lot of things together and you might not really know what's potentially driving the results quite clearly and you might not necessarily know that there is an implication of some of the things that you have done and you are not pulling them out so you are failing to really provide all the information to a decision-maker that they should have.

[I8]

It is noteworthy that interviewees focused solely on elements of the PICO (populations, interventions, comparators, outcomes) description of the decision problem. This may indicate that interviewees considered that the PICO definition adequately captures the typical HTA decision problem.

The errors identified by the interviewees applicable across these areas were:

- definitions incoherent with the clinical understanding of the disease
- inadequate description of the boundary and content of the decision problem.

These error types are illustrated in the interview responses presented below:

...but you can see that in some of the NICE reviews...where the question was asked incorrectly or the question was framed early on in a way that was not coherent with the clinicians' understanding of the disease. And so the whole process flowed through and generated junk at the end.

[I6]

...they didn't realise I'd have to build a treatment model alongside it and then they hadn't invited the right people to come to it... [they] dropped the ball if they believed you could do a screening [model] without having a treatment model, but that wasn't our error.

[I2]

It's not an error in the model because we did what they asked us to, but they could have asked us a better question.

[I2]

The prime root cause suggested by the interviewees for errors in the description and understanding of the decision problem was a failure of communication between the stakeholders to the process. The interviewees identified three types of stakeholders:

- the decision-maker (NICE or client)
- the decision-modelling team (either an academic Assessment Group or an Outcomes Research organisation)
- clinical experts.

## Error in the structure and methodology used

Three broad domains within model structuring can be distinguished within the interviews:

- development of a conceptual model of the disease including a description of its epidemiology, natural history and management
- selection of modelling methods
- moving from the conceptual model to an implemented model.

### Development of a conceptual model of the disease including a description of the epidemiology, natural history and management

The key subthemes identified by the interviewees were:

- boundary between simplification and structural error
- obsolete or outdated model structures
- error in model structure.

A common subtheme identified by nearly half the interviewees concerned difficulties in identifying when a simplifying assumption made within the model structure constitutes an error, as discussed earlier (see Chapter 4).

Any model is going to be built on a whole set of assumptions, some of which should be written up, will be written and some of them might not be and it is all those assumptions basically are the ones which in some way may cause us an error in terms of the results we are going to get out because they are almost always a simplification in terms of how we should have been doing things. There could be very good

reasons as to why you have done it but in some senses they are going to potentially force you down to a particular conclusion when it's not necessarily the conclusion you should have got. [I8]

So if there [was] something important in terms of how long these people were going to live or what their quality of life was going to be, that would have been captured by a more detailed prognostic model, but is not being captured by the simplification, then it's an inaccuracy that you've introduced. And you try and insure yourself against that by looking at the whether there's an established literature, there's a consensus among the clinicians that response status is a good predictor for prognosis. There's not a consensus; there's a good body of published methodology that this simplification is acceptable for, for patients with similar conditions. But unless you do both models, you'll never actually know whether you've simplified something important out of it. [I6]

These two quotes demonstrate the nub of the problem in that it is difficult to determine with any confidence whether assumptions underlying the model structure might have an impact on the model results such that subsequent conclusions and recommendations are affected.

Discussions focused on the role and nature of evidence supporting such structural assumptions. There was consistent agreement among the interviewees that where available evidence contradicts an assumption then this would constitute an error. However, where evidence is absent or equivocal, interviewees were less consistent in whether it would constitute an error.

...if you've developed a structural assumption that the data would contradict, then you'd reconsider that. But if it's simply a case of absent data, then I wouldn't seek to change that, that assumption. [I9]

...if the assumption is arbitrary, or if the assumption directly contradicts strong evidence, then it is probably an error in the sense that the decision-maker would not own it. [I3]

This second quote makes the interesting observation that whether an assumption constitutes

an error depends on whether a decision-maker would have ownership of the assumption. Transparency represents a necessary condition for ownership of assumptions; as noted by the respondent [I8], not all simplifying assumptions are necessarily reported and transparent.

Where there is no evidence or equivocal evidence, the interviewees indicated that the key themes are transparency in (1) the nature of the structural assumptions; (2) underpinning evidence or judgment; and (3) the potential to impact on the model outcomes and subsequent decision-making. Although interviewees did not explicitly make the analytical leap, the implication could be that lack of such transparency may constitute an error in process.

Three interviewees [I3, I4, I9] expressed concerns about the use of obsolete and outdated model structures arising from the shifting nature of the evidence base.

Sometimes there might [be an] accumulation of minor differences, sometimes there's a paradigm change. [I3]

I think there is a danger in that as well you know if you just slavishly adopt a previous structure to a model and everybody does the same there's no potential for better structures to develop or for mistakes to be appealed so I think there is a danger in that in some ways. [I4]

The above quotes indicate that an error may lie in failing to sufficiently capture the relevant evidence base and as such constitutes an error in evidence gathering. Other errors in model structure were identified by interviewees as being caused by the misinterpretation of evidence. A cited example of such misinterpretation is in assuming that a non-statistically significant difference implies an equality between two treatment effects that can therefore be analysed using a common model structure.

Someone who treats something...not to be statistically significant who treats that as being the evidence of no difference at all seems to me to be a huge error and has consequences for the rest of the model which is a bit difficult to describe. [The] counter position of having allowed the uncertainty around that non-significant treatment effect or whatever it is

and see what the consequences are – that is an important one.

[I12]

## Selection of modelling methods

Three interviewees [I2, I3, I4] explicitly identified the selection of appropriate modelling methods as a potential area for error. Explicit examples included the selection of Markov methods in violation of Markov constraints, inappropriate use of Monte Carlo sampling methods, and the inappropriate decision concerning the use of cohort model or individual patient-level models.

You are making a decision first of all as to what modelling method to use so there is a danger there that you maybe adopt the wrong methodology. I suppose that is a potential source of validation error as you set off using Markov when the Markov constraints stop you like simulation or an individual patient sampling model or something like that so there is a potential for an error there and that may not be picked up for a long time.

[I4]

...as soon as you get interactions it becomes a lot more dangerous to work on a simple cohort model...just assuming the patients are independent.

[I2]

...even when it comes to Markov type modelling, not...Monte Carlo type things, because I find that there are very few people who routinely use these packages who actually know the finer points; who understand the potential pitfalls.

[I3]

The above comments indicate that the selection of appropriate modelling methods is a complex decision area and that the errors are associated with sometimes fine errors of judgement.

## Moving from conceptual model to implementation model

There's no such thing as the right way to model. You can choose any way you like. It's just that some are, you know, easier to use than others, more convenient. And so, you know, your conceptual model can be realised in, in an infinite number of ways.

[I3]

Nearly half the interviewees [I1, I2, I8, I9, I12] identified potential errors introduced in moving from the conceptual model to the implemented software model. The process of moving from the conceptual to the implementation of the model involves a set of activities and modelling decisions including:

- selection of modelling software platform
- development of a detailed realisation of the conceptual model, possibly including amendments, that arise in implementation of the model
- design of the implementation of modelling methods.

As discussed in Chapter 3, there is a wide range of modelling practice among the HTA community, particularly with regard to the explicit identification of conceptual modelling and model implementation as distinct activities.

Two interviewees [I2, I12] indicated that the premature selection of the modelling software and the premature implementation of the model structure (that is before the conceptual model has been fully developed) can lead to the generation of errors in models. Furthermore, these interviewees indicated that such failures in model design can lead to a vicious circle wherein errors when identified necessitate further reprogramming that is in itself highly prone to further errors.

...the addition of states halfway through the model as regularly happens...just makes your recoding of an EXCEL model incredibly, well not incredibly difficult, incredibly tedious which can mean that by being bored senseless you inadvertently introduce an error.

[I2]

The model grew from 30 megabytes to 180 in trying to fix this (that is a structural error in the model) at a late stage in an EXCEL model, if we had known this at the outset no way would we have created the model in EXCEL.

[I12]

One interviewee suggested that time and resource constraints imposed upon the modelling process and the skills of analysts were factors in decisions regarding implementation and design and therefore might be underlying causes of error.

...in terms of our skills it is subjective because arguably what you might have is a particular

modelling approach that's been adopted that is less than optimal given constraints on time and research time to what could have been produced to answer that research question. So you may think a very simple decision analytical model where more appropriately it would have been a more sophisticated model...[that] more correctly follows the care pathways or where we believe there may be changes in costs and outcomes between the interventions we are looking at.

[I8]

Sometimes exploring whether more sophisticated models would actually have made a difference and again that's...making sure you have actually got the time and there's your skill base.

[I8]

As noted in Chapter 3, one interviewee highlighted a missing step in the model development process as implemented by most HTA modellers, that is a failure to consider system analysis and design before model implementation:

...how we will be best able to implement the... program?, how will it be the structure of the programming part? And we miss all this, so we end up building a model that will solve the problem for sure, but I...I'm not sure it is the best...the optimal model.

[I1]

## Error in the use of evidence

Discussions with interviewees concerning errors in the use of evidence focussed on:

- the use of evidence in informing the development of model structure
- the role of evidence in populating data inputs to the model
- generic evidence processes.

## Use of evidence in the development of model structure

The development of model structure is previously defined as incorporating the development of the conceptual model of the disease, including the epidemiology, natural history and management of the disease, selection of modelling methods and moving from the conceptual to the implemented model. Interviewees did not necessarily distinguish

between these processes and, therefore, individual quotes concerning the use of evidence in model structuring may refer to different aspects of this process. A central theme raised by interviewees concerning errors in the development of model structure is the misinterpretation of data, for example the inappropriate generalisation of evidence from one environment to another.

If...you're modelling an intervention, and the intervention is a clinical action which is predicated on the environment in which it has to be delivered and the prevailing ethos and, and accepted norms of that clinical community, which is why, you know, an evaluation done in the States or in Brazil cannot be imported into the UK without substantial adaptation.

[I3]

The interviewees suggested that the misinterpretation of the data can be associated with simple misunderstanding of data definitions, deliberate misinterpretation or errors of judgement in interpretation.

In terms of what types of errors, well there's errors that can be made in understanding, interpreting...the data, that's quite an important area so it's errors in how the data has been used in the model.

[I4]

One root cause of misunderstanding suggested by two interviewees concerned a failure of communication within the Assessment Group; this was particularly associated with separation of the information gathering, reviewing functions and modelling functions.

Basically we have made errors before where the modellers misunderstood the data that has been given to us by the person sourcing the data so that's quite an easy and obvious error.

[I4]

It is the problem of having effectively two models, one that's doing a review of the effectiveness and one that's an economic model and they are not the same thing so they are not translating directly one to the other. One person does one thing, one person the other and they have to talk.

[I8]

Two interviewees explicitly identified deliberate misinterpretation of data as a particular form of

culpable error. One interviewee suggested that this type of error was sometimes associated with direct and possibly indirect pressure from decision-makers to produce models in the absence of direct evidence, whereas other interviewees suggested commercial modelling clients might exert indirect pressure to provide modelled estimates of outcomes.

Yes I think it is this business of being deliberate but then it tends to be over optimistic interpretation of data is really what we're criticising.

[I5]

They [NICE] made us build the model, and there was no information!

[I1]

One interviewee identified errors of judgement in making 'soft' decisions about model structure as a potential root cause of error.

...the most difficult ones aren't just the numbers; they are where there's a qualitative decision. You know, is there any survival benefit? Yes or no? If there is a survival benefit, it's a different model? You know what I mean? So it's structural.

[I3]

One interviewee identified the premature definition of model conceptualisation as a potential source of error in model structuring; this was associated with making decisions regarding the model structure that inadequately capture the evidence base, including subjective clinical judgement. The root cause of this type of error is a failure of the process for evidence gathering for use in model structuring.

...the approaches that have been derived from the basis of, of your first conceptualisation as you're setting up to wed yourself that first conceptualisation so far, and to make sure that are there changes in understanding when going from an ad hoc, non-systematic overview to something that is more systematic... That's to say one of the dangers, I guess, could be that we end up, because of the time constraints, developing an idea of the area before you go off and see your clinical experts.

[I9]

Another interviewee suggested that errors arose from the inadequate capture of subjective evidence from clinical experts, from the inappropriate

selection of experts covering the breadth of opinion in a given area, and in ensuring effective elicitation of clinical information.

Well there are obvious errors in terms of that if you don't necessarily speak to the right people. People have particular perspectives on a situation and they may tell you what they think but it may miss out important aspects and the implication of that is you have a care pathway which is biased for or against a particular intervention because you have missed out a benefit or you've missed out a problem with it.

[I8]

Statistical analysis of data, either in premodel analysis or directly within a model, was identified by four interviewees [I2, I3, I4, I7] as an area of potential errors within models. Three specific example areas were identified; ranging from complex issues of data synthesis or survival analysis to simple manipulation of data in modelling population subgroups; all of these impact upon model structuring.

...then it's actually how do you then use that data and synthesise it in a particular way? And I think there's a whole series of errors just waiting to happen at that stage in terms of assumptions, in terms of statistical sort of approaches...

[I7]

I think survival analysis is an interesting area and it's becoming a more prominent area in many of our models in the way we model survival and I think again it's not such a clear cut thing, because fitting a curve, determining the parameters for a Weibull curve in another source of potential error.

[I4]

I've seen it happen elsewhere where people have taken the average value and then for a subset multiplied it by relative risk without thinking that the average would then change.

[I2]

One of the key themes regarding survival modelling is in the selection and justification of models for extrapolating survival.

It's this classic situation of a chronic disease, and what people are doing is extrapolating straight lines. Indefinitely. OK, what have you got in the whole of the rheumatoid arthritis literature is built around, modelling for the



HAQ [Health Assessment Questionnaire] scale, which is closed at both ends. Right? So, what will happen to any trend in HAQ? Well, it won't go like a straight line. It will asymptote towards a limit. It may be the maximum of the scale or the minimum of the scale, or somewhere in between, but it will certainly, the rate of change will decrease over time, OK? What does that mean? Well, if you, if you compare a straight line extrapolated against the trend, what you, what you are doing is always overstating your benefit. Systematically. The net result is that you're virtually halving your ICER.

[I3]

...you can fit a polynomial trend line, but you don't know what happens the moment you drop off the end of the data. It could go anywhere so the argument really was to go back to causality, and go through the metabolic processes that are driving the changes, and then model those and then demonstrate effectively calibrate those against clinical evidence. And then you've got...a basis on which to project forward.

[I3]

A root cause of these errors was identified by one interviewee as errors of judgement related to the experience, skills and training of analysts undertaking this element of the modelling work.

I guess it's just the statistical techniques themselves, you know. I guess people don't understand what they're doing sometimes... it's sort of less experienced people trying to do fairly sophisticated analysis...without actually necessarily knowing...you know, exactly the statistical methods they should be using.

[I7]

...it's errors in judgement, I think, as opposed to errors in what people are doing. I just think it's...you know, whether I think you can pool... you know, different types of utility studies together is, is my sort of analytic judgement versus somebody else's. But it could quite clearly result in an error.

[I7]

### The role of evidence in populating data inputs to the model

Probabilistic sensitivity analysis (PSA) is recognised as the preferred method for estimating mean values of outcomes and describing uncertainty

in outcomes. A central requirement of PSA is the characterisation of all parameters within the model using quantified statistical distributions. However, interviewees recognised the potential for error in characterising input parameter uncertainty through the specification of adequate distributions:

I would suggest it's to do with the transparency of the inputs that are required, and the robustness of the analysis to model parameters which are not within the range that's expected.

[I6]

Two specific issues were raised by interviewees. The first concerned the use of triangular distributions to characterise parameter uncertainty.

The widespread use of, you know, so-called triangular distributions. Good grief what is a triangular distribution? You know, the fact that people who call themselves statisticians recommend using constructs like that. I don't, it just drives me mad. I mean just you know, it's illiterate, really.

[I3]

The second, noted that the characterisation of highly skewed distributions can lead to errors.

If the distributions are really skewed that's happened in the past, that's the one where I can recall where I got the answer wrong because it's got a really skewed distribution of relative risk that may have had a mean of or a median of 2.8 and an upper a log normal distribution with an upper of 31 and you don't need many [samples] within the 31 or above 20 times relative risk in order to make something cost-effective.

[I2]

Two interviewees [I6, I12] explicitly referred to problems in interpreting evidence from poorly reported or poorly executed clinical studies.

...when I go look in the literature, I have one study published in 1984 for 17 patients that doesn't even report standard deviations. So I can follow the guideline, and I can create a gamma distribution about that, with a bunch of arbitrary parameters, and that will then give me a nice neat 95% confidence interval in my output. But the quality of that information is no better, and probably worse, than the information I had to basically make up to put into the distribution.

[I6]

There is one that's occurred to me that isn't to do with model building, we are relying on what other people have done, we are dependent on their having reported things correctly and what they did in the first place having been appropriate. I guess that might be if we want to be consistent with what's being done in a clinical trial that presupposes to some degree that what was done in the clinical trial was right in the first place.

[I12]

The specific role of critical appraisal of clinical effectiveness studies is to identify the quality of execution and reporting of studies. There is an absence of recognised and accepted methods in how to incorporate such quality assessment into the quantitative interpretation of study results. One interviewee indicated that this issue is not covered adequately by existing guidelines in characterising uncertainty.

### Generic evidence processes

Interviewee comments on generic evidence issues focused on searching for evidence and systematic reviewing of evidence.

You have to be careful to look for the right sort of search terms.

[I12]

...you can look to see what they have done and see whether there is any, the same way you peer review...are there any data sources missed should they have done other stuff within analyses, sensitivities, so that all comes in, you would be stupid to blind yourself to it.

[I2]

I suppose in terms of when you're critically appraising something, has a potential effect been missed out...I mean you can look at something missing from the model that says, well, that's just because the data weren't there, people missed it, or it's something that, you know, the model's been structured deliberately in a way to produce a given outcome.

[I9]

Underlying the theme of missing data are issues of intention versus accidental error, application of adequate processes and methods, appropriate level of skills in team members and the application of judgement.

In considering generic systematic review activities, two sources of error were identified by the interviewees. First the potential for mismatch in the interpretation and definition of data between the systematic review and the decision model, and second, the potential for simple transcription errors in the data extraction activities within the systematic review.

You might have a result from the systematic review that is because of the type of data restrictions they are facing is clinically implausible so the actual estimate and level of precision they have around the estimate is just not plausible so you think what could be clinically plausible so you use that to define what are the numbers you put into the model.

[I8]

I think there are two sources of errors: one which is the sheer getting those numbers out, and making sure that no errors actually happen in terms of the transcription...and I think there's a huge source...of error at that stage...again depending on the size of your project etc., you'd hope that it would be kind of extracted by two individual people with some kind of consultation, but...I'm not sure that always happens.

[I7]

### Error in implementation of the model

Errors in implementation identified by the interviewees can be classified into the following two domains:

- coding of individual cells
- coding of logic structures within the model.

#### Coding of individual cells

The interviewees identified that cell contents can be subject to:

- error in cell referencing/variable calling
- error in values
- error in cell text
- error in formulae/operators.

The cell referencing within a spreadsheet is the primary mechanism for representing the logical structure of a decision model:

...it should map out your conceptual model it should be an identical replica to the conceptual.

[I2]

However, the most frequently discussed error in model implementation concerned incorrect cell referencing.

There's errors in the wiring of the model, references that aren't correct, coding that is wrong within the model.

[I4]

If you're in EXCEL referencing is quite often a big error, you've referenced it to the wrong cell and you don't notice [because] you reference 2000 cells in your model.

[I2]

The issue of intention was also raised as a defining characteristic of an error:

Likewise a programming error where you misreference cells, presumably you would do it differently if you did it again because you misreferenced it in the first place so that is an error.

[I12]

There was also some discussion about the relative propensity of different software platforms to logical errors through misreferencing or errors in calling variables:

...there are...the technical errors that are... related to...all these little cells in EXCEL and, er...calling a wrong variable.

[I1]

Errors in values within a model and errors in cell text were also identified as key issues.

...it wasn't the point estimate that had been entered wrong, it was the confidence interval and things like that are difficult to spot...I couldn't see what was wrong at first it was only looking at that and thinking that can't be right.

[I12]

...we've put in this process somebody separate, a health economic modeller, will get a couple of days to look into the model, check all our links, check, you know, even down to spelling in text in there. Try and break this! Give it a kicking.

[I10]

Interviewees only identified two basic root causes of errors in cell codes: simple typing errors and copying errors.

...there are so many different types I think, from, you know, basic slip-ups where, it's pretty obvious that you've made a silly mistake, a typing mistake or copying something from one cell to another wrongly, all those sort of very mechanical type problems.

[I3]

The operators and functions within a model were not explicitly identified by interviewees as being subject to implementation error. Where operators and functions were explicitly mentioned the focus concerned structural errors. However, it would be reasonable to assume that these elements of a cell code or programme code could be subject to similar root cause errors, i.e. typing error or copying error.

Well the typos is factually incorrect, the second one, the referencing is structurally incorrect but the entire thing could be built perfectly but still be wrong because your entire structure's round the bend. I could give you a model that says well for cancer the ICER is  $2 \times A + 3 \times B$  and A and B are just some numbers but mathematically that's correct, but structurally and conceptually it's a load of XXXX.

[I2]

### Coding of logic structures within the model

One interviewee gave two examples of a broader class of errors in implementation which concerned the incorrect coding of logic structures. These examples were errors of implementation of standard methods that affect the spreadsheet more broadly than simple random errors in the contents of single cells. The root cause of these types of errors was indicated to lie in the judgement, knowledge or experience of the modeller.

So the accepted convention for discounting is that you discount after the first year in whole numbers. Right? So this 12 months is not discounted, the next is discounted at 1 year and 2 years and so on. Right, well, you know, people will correctly use the correct discounting formula from point zero, so they are discounting every day from then onwards, OK? It's not technically incorrect, but it is conventionally incorrect and inappropriate if you are then going to make a comparison

with another study in the same area that's discounted in a different way. So then your answers are then not comparable.

[I3]

...when I first learned this sort of thing you always had to think very carefully about randomisation and the order in which you generate random numbers. Because you can inadvertently bias your results if you don't get it right.

[I3]

Where I've written 4.3 I actually meant 5.3 but the model's right and it's just me writing down the number wrong. That's not a modelling error, it's a reporting error.

[I2]

...because we will know what the tables are going to look like prior to the results being produced we will effectively have dummy tables, though after they will be completed they will be looked at to see if they make sense, they will be checked by the other member of the project team.

[I8]

## Errors in operation of the model and model reporting

The key domains discussed by the interviewees in this section were:

- operation of the model
- presentation of results
- communication of results and conclusions.

### Operation of the model

Two interviewees [I4, I6] identified the potential to make errors in setting up the model to generate outputs, this was particularly relevant when generating large numbers of sensitivity analyses and was primarily related to ensuring that all environmental variables, model switches and data were set to the intended values for the specific model run.

For some of the simpler models we leave the one-way sensitivity analysis to be done manually, which, of course, gives you an opportunity for errors at that stage.

[I6]

You can have errors in the way the model output is resolved and you can have errors in the way the model is used.

[I4]

Two primary causes of these errors are implied, first simple typing errors and second failure to update values of variables and model switches.

### Presentation of results

Two interviewees [I2, I8] identified the potential for making errors in the presentation of results into the text and tables of reports.

The root cause identified for the above examples concerns a simple typing error in transcribing the results from the model to the report.

### Communication of results and conclusions

One interviewee discussed the potential for errors in the communication of results to the decision-making audience.

...something that you...may not have considered actually...is error in presentation or understanding; that although the model outputs are reported correctly they may not have been reported in a way that the intended users understand properly so users may misunderstand the presented outputs and I think it's an area that is not often given enough attention.

[I4]

Hence, the communication of results is not clear-cut and the root cause of this failure lies in the modelling team's understanding of the external audience, specifically in terms of understanding the needs, requirements, experience and language of the external audience.

I think it's very easy when you are working on the inside of a project to assume that the knowledge you possess...the things you take for granted are readily accessible to the people you are communicating to and that is not always the case.

[I4]

Two interviewees discussed failures of communication through loss of model credibility with the decision-making audience. Once

credibility of the model is lost, the interviewees experienced grave difficulty in regaining it. In both instances the loss in credibility was associated with the presentation of counterintuitive results.

...we had one example of a model where we were dealing with quite a large number of technologies but the model was actually being driven by adverse events rather than by the prime reason for giving the treatment and it was quite a struggle not so much with the people within our own team but actually the clinical experts who had read the model and then had come along to NICE and it was quite a struggle explaining why we were actually getting plausible results even though we were getting widely different effectiveness results from a number of technologies that were equally effective for the primary purpose but it was because they had different profiles for adverse events and that was what was driving the model but then of course it goes to the technical leads and if there is anything that the technical leads don't understand that comes back to us so again.

[I5]

I was presenting a model and I got quite a lot of stick, 'this is meant to be a final version of a model – what is this glaring error?' And it wasn't an error in the model, it didn't arrive at something that was consistent with some external view of what it should be and it was taken to be an error in the model.

[I12]

The credibility of model results therefore depends on them either being intuitively acceptable, or for counterintuitive results to have an immediate rational explanation through presentation of an adequate range of information on outcomes and key drivers. The tailoring of the presentation of results to such circumstances of course depends on the modelling team having a perception of what the audience might be intuitively expecting. As noted earlier (see Error in understanding the problem), one interviewee explicitly mentioned the benefit of generating a larger range of disease-specific outcomes than the generic cost per QALY, explicitly to aid decision-makers in understanding the nature of the model results.

With regard to errors in drawing incorrect conclusions from results, one interviewee explicitly

referred to the potential for the overinterpretation of results in drawing conclusions.

There is interpretation isn't there and I mean there is overinterpretation of results.

[I5]

I suppose there [are] things like saying that... looking at the cost-effectiveness acceptability curve and saying well if this one is at 51% then it is the preferred option.

[I5]

## Review of taxonomy/ classification literature

### Quantity and quality of the studies identified

Searches identified eight studies<sup>45–52</sup> that presented original work on classifications or taxonomies of errors, and one critical review.<sup>53</sup> Of the eight studies all except two focused on errors in spreadsheet systems, one focused on generic programming errors<sup>48</sup> and a second focused on errors in the requirements specification process.<sup>50</sup> All of the papers were conference proceedings, with the exception of one,<sup>48</sup> which was published in a peer-reviewed journal. A brief description of the studies is presented with the spreadsheet systems papers discussed first and in chronological order. *Table 4* presents the error classifications from each of the spreadsheet studies aligned to highlight congruent themes.

The interview taxonomy is set in the context of this literature (see Placing the interview taxonomy of errors in the context of literature on error classifications).

### **Panko RH, Halverson RP. Spreadsheets on trial: a survey of research on spreadsheet risks. 1996<sup>45</sup>**

Panko and Halverson<sup>45</sup> report a review of literature including experiments in spreadsheet development, field audits of spreadsheets and observational studies of spreadsheet development. Methods for identifying the literature are not reported.

The purpose of the paper is to summarise research findings on the risks inherent in end-user spreadsheet development. The paper uses this

literature as the basis for developing a framework of spreadsheet risks and organises a discussion of the research findings around this framework.

Three dimensions of risk are identified: methodology, life cycle stage and research issues. Research issues are described as the dependent variables of the research surveyed, the principal examples given are structural concerns about spreadsheets as a development platform and errors in spreadsheets. Second, the life cycle stage is described as important because of the changing characteristics and frequencies of errors throughout the different stages of spreadsheet development and use. Five life cycle stages are identified; requirements and design, cell entry, draft stage, debugging stage and the operational stage. The methodology dimension is not discussed in detail in this report.

The Panko–Halverson 1996 classification differentiates first between qualitative errors and quantitative errors. Quantitative errors are flaws that ‘lead to incorrect bottom-line values’. Qualitative errors are described as ‘flaws that do not produce immediate quantitative errors. But they do degrade the quality of the spreadsheet model and may lead to quantitative errors’.

Panko and Halverson classify quantitative errors into mechanical, logic and omission errors. Mechanical errors refer to simple slips including typing and pointing errors. Logic errors include designing an incorrect algorithm or ‘creating the wrong formula to implement the algorithm’. Omission errors are ‘things left out of the model that should be there’.

The authors explore further possible subdivisions of logic errors. One suggested subdivision is into Eureka or Cassandra errors, the first being errors that are readily identified and proven to be errors as compared with Cassandra errors that are not identified or demonstrated to be errors. An alternative subdivision is into pure logic errors ‘resulting from a lapse in logic’ or domain logic errors resulting from a lack of domain knowledge.

### **Teo TSH, Tan M. Quantitative and qualitative errors in spreadsheet development. 1997<sup>46</sup>**

This conference proceedings paper by Teo and Tan describes a laboratory-based experiment in spreadsheet development using business school students. The literature review gives a brief outline

of spreadsheet research using a subset of the literature referenced by the earlier Panko and Halverson<sup>45</sup> paper.

The error classification cited in the paper and used in the design of the experiment is the Panko–Halverson<sup>45</sup> classification with qualitative errors further subdivided into ‘jamming errors’ and ‘duplication errors’. These additional error classes being referenced to an earlier monograph by Panko.<sup>54</sup> Jamming errors refer to the practice of entering values within a formula, and duplication errors refer to the practice of defining a variable more than once within a spreadsheet.

### **Rajalingham K, Chadwick DR, Knight B. Classification of spreadsheet errors. 2000<sup>47</sup>**

These authors present a classification based upon a ‘thorough review of literature relevant to spreadsheet development’. The methods for undertaking the review are not presented and the source materials used are not referenced.

The authors state that the purpose of the taxonomy is to facilitate analysis and comprehension of errors to devise solutions or methods of detection.

The classification, presented in *Table 4*, is defined as a binary tree. The first level divides errors into system-generated, that is, generated by bugs in the software, and user-generated errors. Panko and Halverson<sup>45</sup> is referenced in dividing user-generated errors into qualitative and quantitative errors.

Qualitative errors are divided into semantic and maintainability errors. Semantic errors are described as being related to distortion or ambiguity in the meaning of data and maintainability errors are flaws in spreadsheet design or implementation that make it hard to update or are error prone in use. It is not clear that this division of qualitative errors is exhaustive or mutually exclusive. Semantic errors are further subdivided into structural and temporal errors.

Quantitative errors are first divided into accidental errors or reasoning errors. Accidental errors are then categorised by perpetrator, of which three different types are identified, ‘developer’, ‘data inputter’ and ‘interpreter’. The constraint of using a binary classification forces the authors to group ‘data inputter’ and ‘interpreter’ together as ‘end user’. For each perpetrator a tertiary classification of errors into ‘omission’, ‘alteration’

TABLE 4 Classifications of spreadsheet errors

Reference	45	46	47	49	51	52
<b>Author</b>	Panko and Halverson	Teo and Tan	Rajalingham <i>et al.</i>	Rajalingham	Purser	Panko
<b>Date</b>	1996	1997	2000	2005	2006	2008
<b>Scope</b>	Spreadsheet errors	Spreadsheet errors	Spreadsheet errors System generated User generated	Spreadsheet errors	Spreadsheet errors	Spreadsheet errors
<b>Classification</b>	Qualitative – Jamming errors – Duplication errors  Quantitative – Mechanical errors – Logic errors – Domain errors – Pure logic errors – Eureka errors – Cassandra errors – Omission errors	Qualitative – Jamming errors – Duplication errors  Quantitative – Mechanical errors – Logic errors – Omission errors	Qualitative – Qualitative – Semantic – Structural – Temporal – Maintainability Quantitative – Accidental – Developer – Omission – Alteration – Duplication – End user – Data Inputter – Omission – Alteration – Duplication – Interpreter – Omission – Alteration – Duplication – Reasoning – Domain knowledge – Real world knowledge – Mathematical representation – Implementation – Syntax – Logic	Qualitative – Structural – Visible – Hidden – Temporal Quantitative – Accidental – Structural – Insertion – Update – Modification – Deletion – Data Input – Insertion – Update – Modification – Deletion – Reasoning – Domain knowledge – Real world knowledge – Mathematical representation – Implementation – Syntax – Logic	Qualitative – Structural – Visible – Hidden – Temporal Quantitative – Accidental – Insertion – Update – Modification – Update – Reasoning – Domain knowledge – Real world knowledge – Mathematical representation – Implementation – Syntax – Logic	Violations Blameless errors – Qualitative  Quantitative – Mistakes – Context errors: section algorithm design, requirements – Formula errors: wrong algorithm, wrong expression of algorithm – Slips and lapses – Slips: sensory-motor errors (typing, pointing) – Lapses: memory errors

and ‘duplication’ is then defined, that is leaving something out of the model, amending content of the model incorrectly or duplicating either data or structural components of a model. It is not clear where primary slips in model development, such as accidentally mistyping a formula, would fall in this categorisation.

Reasoning errors are categorised into domain knowledge errors and implementation errors. Domain knowledge errors are divided into real world knowledge errors and errors in mathematical representation. These are similar to the pure logic and domain logic errors described by Panko and Halverson.<sup>45</sup> Implementation errors arise from incomplete knowledge of the functionality of the platform being used and are divided into errors of syntax and logic.

### **Rajalingham K. A revised classification of spreadsheet errors 2005<sup>49</sup>**

In 2005 Rajalingham presented a simplified version of the earlier classification,<sup>47</sup> that incorporates two main modifications. First, the dubious classification of qualitative errors into semantic and maintainability errors is removed and the previous structural errors classification is further subdivided into visible and hidden errors. Second, the observation that developers of spreadsheets tend to be end users means that the error perpetrator classification of accidental errors is superfluous. Instead, accidental errors are classified into structural errors or data errors. Both structure and data are then subdivided into insertion and update errors and update errors are further divided into modification and deletion errors. The classification of reasoning errors is left unchanged from the earlier paper.<sup>47</sup>

### **Purser M, Chadwick D. Does an awareness of differing types of spreadsheet error aid end-users in identifying spreadsheet errors. 2006<sup>51</sup>**

Purser and Chadwick in this conference proceeding describe an investigation of the effect of experience and error type awareness on the identification of errors in spreadsheets. The investigation uses a web survey including an error identification exercise undertaken with professionals who use spreadsheets as part of their work and students from Business, Computing and Mathematics schools. As a preliminary part of the project a classification of error types is devised.

The focus on error identification leads the authors to make the observation that ‘all errors exhibit characteristics’ and further that classification of error types based upon these characteristics would provide a classification or taxonomy of particular use in error identification. However, the authors state that the characteristics of error types are ‘beyond the scope of this research’.

As an alternative, the authors review previous classifications of errors including Panko and Halverson<sup>45</sup> and generate a further revision of the classification of Rajalingham<sup>49</sup> with the objective of removing the remaining duplication within the bifurcation tree and tailoring it for the purposes of error identification.

The classification of qualitative errors and quantitative reasoning errors remain unchanged. By removing the classification of accidental errors into ‘structural’ and ‘data input’ errors the authors remove the duplication of insertion, update, modification and deletion errors.

The results of the investigation are equivocal and not all the conclusions made by the authors are entirely supported by the results of the investigation.

The authors use a quasi-experimental language to describe the investigation, proposing a series of hypotheses. However, there appears to be no serious attempt at randomisation in the study and no statistical hypothesis testing is undertaken to support the conclusions. The results suggest that awareness of the examined error types makes little difference to overall error detection rates (slight worsening); however, subgroup analysis suggests that identification of qualitative errors actually gets markedly worse and quantitative errors better. The conclusions however focus on the improvement in qualitative error together with the statement ‘and thus logically awareness of spreadsheet error types should aid the user in identifying quantitative spreadsheet errors identification’. It is possible that the central weakness of this paper arises from the authors’ failure to develop a classification of errors based upon observable error type characteristics, as they themselves identify.

### **Panko RR. Revising the Panko–Halverson taxonomy of spreadsheet risks. 2008<sup>52</sup>**

This paper, yet another conference proceeding, reiterates the three dimensions of spreadsheet risks; risk category (or error), life cycle stage and



methodology from the author's earlier work.<sup>45</sup> The paper presents the original classification<sup>45</sup> together with a further discussion of its roots in human error research.<sup>55</sup> The modified classification starts with a division of all errors into blameless errors and violations. Key examples of violations are defined as puffery, fraud and a failure to follow organisational policies regarding, for instance, spreadsheet development, testing and archiving.

Blameless errors are as before divided into qualitative and quantitative errors. Quantitative errors, again following human error research, are divided into mistakes, or errors of incorrect intention, and errors of incorrect execution.

Mistakes are divided into context errors and formula errors. Formula errors include both mistakes in the design of an algorithm and mistakes in the expression of an algorithm. Panko uses a corollary with a study examining the writing process<sup>56</sup> to demonstrate the scope of context errors. Accordingly, in writing, a hierarchy of abstraction levels is described including overall purpose, document, chapter paragraph and word. In writing each word the writer has to maintain the integrity of the entire hierarchy. Spreadsheet development has an exact corollary of abstract levels including; the decision problem requirements, spreadsheet, module, section algorithm and formula. Hence, just as in writing, in spreadsheet development the integrity of the hierarchy needs to be maintained. Panko differentiates errors in formulae and errors in context as referring to the rest of this hierarchy. Omission errors are included in these context errors.

Errors of incorrect execution are categorised into slips, that is sensory-motor errors, such as typing errors and pointing errors, and lapses, that is memory errors.

### **Ko AJ, Myers BA. A framework and methodology for studying the causes of software errors in programming systems. 2005<sup>48,57</sup>**

In this paper Ko and Myers provide a review of previous studies that have classified bugs, errors and problems in a range of programming languages. This review, together with previous studies on the cognitive difficulties in programming and general human error mechanisms, has been used to define a framework for describing the occurrence of errors in programming systems and a methodology for studying them.

The authors differentiate between failures, faults and errors. Runtime failures are defined as situations where a programme's behaviour does not comply with its design specification. A runtime fault is defined as 'a machine state that may cause a failure' and a software error is defined as 'fragment of code that may cause a fault during execution'. Following reason,<sup>58</sup> cognitive processes underpinning the occurrence of errors in software development are classified into skill breakdowns, rule breakdowns and knowledge breakdowns.

Skill breakdowns relate to inattention or overattention: inattention being associated with five types of breakdown, strong habit intrusion, interruptions, delayed action, exceptional stimulus and interleaving and overattention breakdown types being omission and repetition. To illustrate, interruptions during programming may cause attentional checks to be missed leading to an action being skipped or a goal being forgotten, unusual stimuli may be overlooked and appropriate actions not taken or attentional checks in the middle of a routine action may lead to an assumption that an action was not completed leading to repetition.

Rule breakdowns are classified into wrong rules, 'use of a rule that is successful in most contexts, but not all' and bad rules, 'use of a rule with problematic conditions or actions'. Knowledge breakdowns are categorised into bounded rationality problems; selectivity, biased reviewing and availability bias, and faulty models of the problem space, including simplified causality, illusory correlations, overconfidence and confirmation bias.

Three life cycle stages are defined, specification, implementation and runtime activities, the programming activities in each stage are considered in relation to the types of cognitive breakdowns. The authors introduce the important concept that although individual breakdowns may give rise to software errors, chains of multiple breakdowns, errors and faults may be involved in giving rise to a runtime failure.

### **Walia GS, Carver J, Philip T. Requirement error abstraction and classification: an empirical study. 2006<sup>50</sup>**

Walia, Carver and Philip present a classroom experiment investigating the occurrence of errors in the requirements specification phase of software development. As a precursor to undertaking this classroom study the authors develop a requirement

error taxonomy, outlined in *Table 5*. The use of this classification was evaluated qualitatively by the study participants and all errors identified in the study could be classified.

## Placing the interview taxonomy of errors in the context of literature on error classifications

### Purpose and design of error classifications

The error classifications reviewed have been developed with different purposes in mind. Panko and Halverson originally described their intention as providing a framework for the discussion of research on spreadsheet risks, similarly Rajalingham focuses on facilitating an analysis and comprehension of errors to devise solutions.

Strategies for reducing errors and improving the robustness of systems for decision support are wide ranging, including for instance; improving error detection, improving software systems through design of development environments and debugging tools, training and skills development of practitioners and development of modelling methods. Clearly each strategy may be associated with different characteristics of errors and therefore imply a different focus for the classification of errors within a taxonomy.

Explicitness concerning the purpose of a taxonomy is a key factor in its development and design. Hence the failure of the study by Purser and Chadwick<sup>51</sup> into error detection can be ascribed at least in part to weaknesses in their classification system. The use of a slightly modified version of the Rajalingham<sup>49</sup> system, rather than a classification based upon error characteristics or symptoms as discussed in their introduction, was inadequate for the purpose of improving error detection. Similarly, Teo and Tan<sup>46</sup> used the Panko and Halverson<sup>45</sup> classification as a basis for a laboratory study of spreadsheet development. So as to avoid bias in the experimental design the investigation is restricted to those error categories unrelated to domain knowledge. The simple classification of domain knowledge or contextual errors is insufficient to support experimental examination of these error types.

In contrast Ko and Myers<sup>48</sup> develop a classification system focused on enabling examination of root causes of errors and this classification leans heavily on research into cognitive mechanisms of human error. The Ko and Myers classification works at a far more detailed level of abstraction than is reflected in the discussion of errors in the HTA modelling interviews. Ko and Myers use the software error framework that they develop as the basis for an observational study of software programmers practice in development and debugging. The results of this study are used to inform the development of the software

**TABLE 5** Requirements errors identified by Walia et al.<sup>50</sup>

<b>People errors</b>	Communication
	Participation
	Domain knowledge
	Understanding specific application
	Process execution
	Other human cognition
<b>Process errors</b>	Inadequate method of achieving goal/objective
	Management
	Elicitation
	Analysis
	Traceability
<b>Documentation errors</b>	Organisation
	No standard usage
	Specification

development environment, including the design of debugging aids. The work of Ko and Myers<sup>57</sup> indicates the importance and relevance of research on cognitive processes underlying human errors for programming errors. The HTA modelling interviews indicate that there is scope for developing our understanding of human error processes in relation to modelling errors.

The classification of errors generated from interviews with HTA modellers demonstrates that we are in the very early stages of development in this area. The purpose of the HTA interview classification is to set down a starting point for developing our understanding of errors and to facilitate discussion. To take forward any strategy for reducing errors and improving the robustness of decision support is likely to require further development of this classification of errors. Specifically the lessons drawn from the Purser and Chadwick study<sup>51</sup> indicate that a close attention to describing the symptoms and characteristics of errors may be beneficial as a basis for improving error detection techniques.

### Criteria for taxonomy

Two important criteria of a taxonomy are (1) it should be complete, that is, any item being considered should be able to be located in the taxonomy, and (2) the classes should be mutually exclusive, so an item should fall in only one class within the taxonomy.

The spreadsheet error classification systems identified in the literature and indeed the HTA interview error classification, struggle with these criteria. In the first Panko and Halverson<sup>45</sup> system it is unclear why some omission errors might not fall into either the mechanical or the logical errors categories, similarly in the second Rajalingham system<sup>49</sup> it is unclear that the division of qualitative errors into structural and temporal is exhaustive.

The classification derived from the HTA modeller interviews exhibits similar difficulties to those found in the literature, a number of reasons underpin the difficulties in this regard. First, the absence of a common description of the modelling process means that location of errors in the modelling life cycle is problematic. Second, the interviewees do not have a common perspective on differentiating between root causes of errors, model errors and the impact of errors, as a result, differentiating between these characteristics is problematic. For these reasons the interview

taxonomy may be better thought of as an error classification rather than a formal taxonomy.

### The interview classification of HTA modelling errors

Panko and Halverson<sup>45</sup> and Ko and Myers<sup>48</sup> recognise the importance of the life cycle dimension of errors. The use of the modelling process description as a mechanism for discussing errors in the HTA modeller interviews, and the subsequent use of the modelling process to structure the interview classification, echoes this life cycle dimension.

The second classification system of Panko<sup>52</sup> provides perhaps the best match in terms of purpose and level of abstraction to the interview classification of modelling errors. The interview classification is mapped onto this second Panko classification in *Table 6* and this is used here to further discuss the details of the interview classification system.

The concept of a ‘violation error’ proposed by Panko is recognised explicitly by the HTA modelling community in three ways: first, in the deliberate structuring of models to produce given outcomes; second, in the overoptimistic interpretation of evidence; and third, in the overoptimistic interpretation of results and conclusions. All the relevant points of discussion here arose in relation to discussions about critically appraising models.

I suppose in terms of when you’re critically appraising something...the model’s been structured deliberately in a way to produce a given outcome.

[I9]

However, Panko also considers non-compliance with procedures and policies also to constitute a violation error. This issue, although not explicitly discussed by interviewees, is potentially implied where errors are associated with inadequacies of process or methods, especially where those processes and methods are subject to guidelines or process and methods documents. The definition of failure to follow guidance for process and methods depends significantly on the soundness and validity of that guidance.

The location of ‘violation errors’ at the top of the Panko hierarchy of errors is potentially misleading because each element of the subsequent error

TABLE 6 HTA modelling errors classified according to Panko<sup>52</sup>

Error type by HTA modelling
<b>Error in description of the decision problem</b>
Context error
<b>Error in model structure and methodology used</b>
Violation
Qualitative error
Context error: spreadsheet, module, section design
Context error: formula errors design
<b>Error in the use of evidence</b>
Violation
Context error: spreadsheet, module, section design
<b>Error in implementation of the model</b>
Slips
Context error: spreadsheet, module, section design
Context error: formula errors design
<b>Error in operation of the model</b>
Slips
Lapses
<b>Error in the presentation and understanding of results</b>
Violation
Context error: spreadsheet, module, section design
Slips

classification tree might also be the subject of a violation error. For instance, if there is a point at which a misjudgement becomes a negligent misjudgement then this would constitute a violation. Alternatively, where errors are the result of a breakdown in domain or programming knowledge, this may reflect a violation in the responsibility of the organisation or individuals within the organisation for recruitment of appropriately skilled people and adequate staff development. Hence, rather than being a separate class at the head of the hierarchy, violations may potentially be better considered as more a separate dimension of errors that run throughout the error classification tree and may themselves be subject to further grades in much the same way as sins are classed as venial or mortal.

The concept of qualitative errors in spreadsheet design was recognised by the interviewees in several ways. Interviewees referred to the vicious circle of poorly designed or executed model

implementation leading to excessive updating and debugging, leading in turn to further errors. Note that this description in the interviews reflects the chains of errors, faults and failures proposed by Ko and Myers.<sup>48</sup> Furthermore interviewees identified errors associated with the premature selection of software platforms or model structures and described these as leading to cumbersome and unwieldy models, prone to errors in programming and operation, this description precisely reflects the type of qualitative errors described by Panko.<sup>52</sup> ‘Jamming’, as defined by Panko as the mixing of values within formulae, was explicitly cited by several interviewees as bad practice leading to the potential for errors to arise.

It is worth noting that not all error types occur in all domains. Whereas the distribution of error types over the different domains appears at first sight intuitively sensible, for example errors in model structure are associated with violations, qualitative and context errors, it is not clear whether the

complete absence of slips and lapses in this domain is correct or represents a shortcoming in the perception of the modelling community.

## Understanding the decision problem

Walia *et al.*<sup>50</sup> focus on errors in the requirements specification phase of the programming cycle, this is the equivalent to the understanding the decision problem phase of the modelling cycle and is informative when compared with the classification arising from the HTA modeller interviews. The comments in the HTA modeller interviews focused primarily on failures in generating a common understanding between stakeholders or failures of communication. In the Panko classification this is captured as a context error, a broad grouping that is unhelpful in considering error avoidance and identification strategies.

The study by Walia *et al.*<sup>50</sup> describes three categories of error types, people errors, process errors and documentation errors. Whereas the central theme of communication is identified within 'people errors', Walia *et al.* also highlight errors in participation. Although not raised as errors by the HTA interviewers themselves, views such as 'It's not an error in the model because we did what they asked us to, but they could have asked us a better question,' [I2] might indicate errors of participation on behalf of the interviewee.

In terms of process errors, it should be noted that the interviews included only representatives of the HTA modelling community whereas the responsibility for defining and managing key elements of the process of defining the decision problem frequently does not lie directly in the hands of the HTA modellers.

...they didn't realise I'd have to build a treatment model alongside it and then they hadn't invited the right people to come to it... [they] dropped the ball if they believed you could do a screening [model] without having a treatment model, but that wasn't our error. [I2]

Addressing these issues may require development of process as well as modelling methods.

## Conclusion

The interviewees collectively demonstrated examples of all the major error types identified in the literature on errors in end-user developer spreadsheet systems. Taken together, the interview classification of modelling errors provides a basis for developing our understanding of errors and facilitating discussion of errors.

To take forward any strategy for reducing errors and improving the robustness of decision support, it is likely that further development of this classification of errors will be required. Particular attention should be shown in this regard when considering methods for improving the identification of errors.

The literature detailing classifications of spreadsheet errors, is based around the work of two key authors, Panko and Rajalingham. Six different versions of classifications have been identified, that all struggle in varying degrees to present classifications that are complete and mutually exclusive, whereas being rich enough to be useful. Ko and Myers<sup>48</sup> identified explicitly the often complex chains or networks of errors and faults that can lead to a failure and related these to research on the cognitive background to human error. The principle of complex chains underpinning errors is described in the interviews but the relation of modelling errors to human cognitive processes is poorly understood. This layer of complexity is poorly developed in the spreadsheet error literature and not explicitly discussed in the HTA modelling interviews.

The error category termed by Panko 'context errors' hides a multitude of different error causes. For example this category contains errors associated with breakdowns in domain/disease knowledge, modelling skills/knowledge, programming skills/knowledge, mathematical logic and statistical methods knowledge. All of these represent areas of judgement where the identification of and definition of error are problematic and may include at their most extreme, violations of methods and processes. This area requires specific attention when moving towards strategies to reduce errors in HTA modelling.



# Chapter 6

## Strategies for avoiding errors

### Overview

This chapter explores the procedures and techniques for preventing the introduction of errors in health economic models discussed by the interview respondents. As with the other themes that emerged from interrogation of the interview data, a descriptive account of discussions surrounding the avoidance of model errors was developed through the categorisation and classification of the charted transcript data. The interview data were analysed in terms of the methods and procedures for error avoidance as well as the experiences, views and perceived importance of these methods. Within the qualitative analysis, five classifications emerged from the analysis:

1. Mutual understanding, which relates to joint understanding and communication between stakeholders to the processes of model development and use. This includes decision-makers and clients, clinical advisors and other stakeholders, health economic modellers and other members of the research team such as systematic reviewers and information specialists.
2. Model complexity and its perceived relationship to the generation of errors.
3. Housekeeping processes, which relate to features and techniques that can be incorporated into implemented models to alert the modeller to possible errors and ensure the analyst is aware of the current state of the model.
4. Guidelines and interviewees' perceptions of their necessary content, role and limitations.
5. Skills and training, which explores the skills, abilities and knowledge that modellers working within the HTA process should possess.

The techniques and processes identified within the descriptive analysis were related to the model taxonomy to examine gaps between types of model error and current approaches currently adopted by the HTA community to ensure their avoidance (see *Table 7*).

### Mutual understanding

#### Clinician input to the model development process

The interview data suggested that respondents agreed that developing an understanding of the clinical situation or disease process being investigated is paramount in ensuring model credibility, highlighting the importance of clinical input within the model development process. Although interviewees' definitions of what constitutes a model error (see Chapter 4) did not consistently include 'softer' issues around model structuring, much of the discussion around avoiding errors focused on these. The primary area in which clinician involvement was indicated concerned the development of mutual understanding of the disease process under consideration:

...make sure that there is always some kind of oversight by someone with the appropriate clinical background to put you right... [I9]

...what we would have on every project is at least one member of the team with research training but also with a clinical background to help us with the day-to-day understanding... [I8]

All respondents either implicitly or explicitly suggested that developing an ongoing relationship with clinicians who have a specialism in the disease area under consideration is essential for the development of health economic models. A preference for involvement with multiple clinical advisors throughout the model development process was suggested by several respondents.

...people do have vested interests and sometimes people may only know part of the care pathway, they may be thinking I know the whole lot but they are not involved in the whole of the care pathway they are just saying what they think is happening and that means you need to speak to various people... [I8]

One interviewee working within an Outcomes Research organisation noted that while clinical advice may be internally available from within the client's organisation, the use of independent clinical advisors may also be preferable. The same respondent stated, however, that an independent clinical advisor may not be aware of important aspects of a clinical trial used to inform the model, hence independent and company-employed advisors may both be valuable in developing models.

...you can have clinical involvement of a clinician who is removed from the client and can offer an independent voice, not always possible due to time constraints...and then you are relying on clinical advice from within the company...sometimes that can be better than advice you get outside because sometimes they can be more aware of important aspects of the trial...

[I12]

There was some disagreement among respondents concerning the preferred nature of the relationship with clinical advisors. Some interviewees suggested that the research team should meet with clinical advisors early in the model development process to develop an understanding of the disease process, the health states that will need to be considered and their relationships, and to understand current clinical practice before attempting to implement the model on a computer platform.

...there will be a series of meetings where actually people sit down and try to work out and start to draw what they think the care pathways are going to be...It would be myself or one of the other economists on the project, sitting there hand drawing it on a big sheet of A3...

[I8]

However, an alternative view (see Chapter 3) was that the analyst should attempt to implement an initial model on a computer platform before meeting with advisors. The majority of statements relating to mutual understanding between modellers and clinicians concerned ensuring that the views of the clinician are adequately captured by the research team. One respondent discussed the importance of ensuring that clinicians understand the modelling process to ensure that they are able to fully contribute to model development:

...part of my job is actually to get the clinicians aware of what is possible to be modelled and sometimes that is a question, it depends on their previous experience of modelling in some cases it is trying to expand the horizons and make them realise that things are feasible that they might have been told in the past might not be on other occasion it might be...to say what can't be done. The clinicians that I deal with it's mostly a case of trying to expand their horizons out...

[I5]

One respondent highlighted that clinical advisors serve a role not only in developing an understanding of current practice (what is), but also the impact of new interventions on that practice (what will/may be).

...we are interested in the current practice and then we are starting to think about how that practice may change if we put our intervention in at a particular point...

[I8]

Other respondents indicated that they would seek advice from clinicians regarding the best sources of data with which to populate the model:

We worked closely with one clinician in developing the original concepts and he pointed us in the right direction for data for populating particular links in the model...

[I3]

Respondents indicated that clinical advice is also important in generating parameter estimates in instances whereby no evidence is available.

...we would go back to the clinician or other experts just to see whether the value you're choosing has any kind of validity...

[I9]

### **Using clinical input to ensure face validity**

Respondents have indicated that clinicians may be asked to contribute to the validation of the model in a number of ways. Their input may be sought as to whether they consider that the model adequately represents the disease process and disease management pathways.

...we have a lot of checks with various people who may be commentators or people who we



think can question what we are doing; those experts will be the people who have got an interest in the disease area or condition area and have got some understanding of the process...

[I8]

...you've got connections between health states or whatever and you're saying that this is how people move around in there and they [the clinicians] can say well that never happens and so it's partly I guess a kind of validation of your understanding...

[I9]

With regards to this point, one respondent emphasised the benefits of discrete event simulation software applications that involve visual interactive modelling, allowing clinical advisors to watch patients moving between various health states on screen.

...it's so easy to do, when you actually see people move through a model you would instinctively walk someone through a SIMUL8 model because that actually visualises what they are expecting. In EXCEL it becomes a lot harder...

[I2]

Respondents also stated that they invite clinical advisors to provide feedback on interim and final model results and to examine whether the model results are in agreement with their expectations:

...we will try to explain, well we will either find out where we've gone wrong or alternatively explain to them [the clinicians] why it shouldn't be a surprise because sometimes it shouldn't be or, you know, sometimes it is their intuition or the data or the modelling is wrong...but I think you tend to assume that once they're [the clinicians] not surprised by the thing then that means you have got it right...

[I5]

### **Mutual understanding between the decision-maker and other stakeholders**

The respondents suggested that communication between NICE and other stakeholders involved in the HTA process had improved over time.

...but you can see that in some of the NICE reviews, some of the older ones where the

question was asked incorrectly, or the question was framed early on in a way that was not coherent with the clinicians' understanding of the disease. And so the whole process flowed through and generated junk at the end...

[I6]

Furthermore, one respondent raised an issue concerning a mismatch in expectations between the research team and the decision-maker.

So now NICE wants in the protocol to appear what modelling technique and what software you will be using...but, from my point of view it's too early...

[I1]

The same respondent suggested that the opportunity for negotiation between the decision-maker and the research team was limited. Two further respondents expressed a similar view, suggesting that the specification of the scoping document was rigid.

The evaluation should have included all valid options but when the model is set up what is the difference in the evaluation of the model, the model itself could be fine in terms of what it's done but it's not really addressing the question which it should have been addressing...to miss out something that should have been there could potentially lead you to come to an incorrect conclusion...

[I8]

...so I think you can start with what you think the ideal is; I think you can then see how that then maps into what the decision-maker specified as what they think the decision problem is, and if there is some kind of mismatch, and the question is where do you go from there really with someone like NICE, who will be quite sort of explicit about 'we want you to look at this particular intervention,' 'we think these are potentially relevant comparators'. But, then we're pretty restricted by licence issues etc...

[I7]

One respondent raised a point concerning the Final Appraisal Determination document, highlighting the need for consensus of opinion where evidence was lacking or absent.

...in other places it's a real unknown and there is no way of satisfactorily filling that gap and

you will find often then e.g. a NICE Final Appraisal Determination, a lot of that report will be discussing that weakness in terms of the evidence then it just comes down to one person's judgment potentially over another's although you would hope more that there would be consensus around it...

[I12]

The same respondent further suggested the need for mutual understanding in instances whereby the clinical pathway was not well understood or subject to variation.

So in certain disease areas it might be useful to have some consensus around what we should do when we can no longer be explicit...

[I12]

### **Mutual understanding between the modeller and the client (Outcomes Research organisations).**

As noted earlier (see Chapter 3), the development of models within the consultancy field differs slightly to standard processes within the academic groups, particularly in terms of understanding the decision problem. The interview data suggested a somewhat iterative process of negotiation between the research team and the client, primarily at the RFP and proposal stage. One interpretation of this aspect of the process is to ensure mutual understanding between both parties and to reduce the possibility of a mismatch of expectations. One respondent stated that this negotiation often starts when the analyst is bidding for the work and specifying the disease area, treatment, outcomes and model purpose. The respondent also suggested that one approach to generating this shared understanding was through the production of a report summarising the evidence relating to the decision problem to be addressed.

...but so then we go through that process and we normally would put that together in some kind of report or tabular summary, as a draft that will get discussed with clients, that will come back and we'll kind of take on board comments and finalise that data set. So we try and get to a point where we've got an evidence summary that is a shared understanding of the evidence between us and the client and the in effect signed up...

[I10]

As further noted (see Chapter 3), clinical advice may also be sought through the use of clinical advisory boards with the clients to generate consensus and mutual understanding (often using external clinical experts). The same respondent suggested that a key role of clinical input concerned refining specific aspects of the model where no predetermined approach had been agreed, as well as providing iterative feedback on the model during its development. Indeed, rather than using clinical input at a single point in time, it was suggested that the clinician-modeller relationship should involve the modeller managing the expectations of the advisors, but that the entire process should be 'a shared journey':

...I think if we start hitting a specific issue that we want to really detail and nail down I think our tendency...and we don't have any fixed approach...I think what we end up probably doing more often than not...is producing something separate that looks at that particular aspect, either a short memo or a few slides to explain it, or we may mock something up in EXCEL to demonstrate what the key issue is and I think that's the way we do it...

[I10]

...it's about managing expectation. And it's back to you don't want them reaching here and going. This is not a 'da-da' moment at the final results, it's not, it should be like a shared journey. You should be finding these things out together and managing them in true collaborative style...

[I10]

### **Mutual understanding within the analytical team**

It was noted by the interviewees that research teams typically consist of a number of researchers and analysts with various backgrounds and expertise, e.g. modellers, reviewers, statisticians and information specialists. Researchers may be working individually on their own particular part of the project, or in small groups; several of the respondents indicated the importance of ensuring effective communication between the individual members of the team (and indeed with wider external stakeholders involved in the process). The qualitative data suggested that this was commonly done through meetings and by presenting the model to people outside the research team. One respondent described how they would present the model to individuals within the research group

who had not been involved in the design or implementation of the model.

...there's also a stage of exposing the model at that stage to I suppose to people within the health economics group but also outside of that, so I think there is a, well, I think there's a benefit in terms of, again, just going over the way the model's been put together...if it's a Markov model it would be a presentational state transition diagram. It would be back to then explaining the links between states, beginning to clarify where the data were coming from, I think I had a suggestion as to what type of outputs are coming from it...

[I9]

Another respondent stated that owing to the large number of people in the research team, considerable co-ordination is required to ensure that information sourced by one member of the team is coherently communicated to other team members; the failure of this process has the capacity to introduce errors into the model development process.

...what you have got is effectively quite a large research team, it's not one person doing it so a lot of co-ordination of effort is required. Things can get lost, not reported, lost along the way, or it might just seem that well actually for whatever reason I was probably having problems building the model that was going to reflect this so I thought about an assumption which seemed quite sensible to me but when you have to present it to someone else people say hold on and that means you cannot do *x*, *y* and *z* and that will imply something in terms of results and that may cause a problem later on...

[I8]

One interviewee suggested that many of the errors that are introduced into models are a result of miscommunication, the implication of which is that modellers need to ensure not only that they understand the decision problem and how the model relates to it but that all other members of the research group share this understanding. As noted above, one way of doing so was the exposition of the model at meetings throughout its development.

...I think that very often the source of a lot of errors is miscommunications so finding ways in which you can make sure communication

between different people can be made less error prone...

[I4]

...virtually everything we do will be presented either to the public health researchers at a team meeting...

[I5]

## Model complexity

One key issue that modellers are regularly faced with when designing or developing a model concerns its complexity. The interview data suggested that modellers do not want to spend a substantial amount of time designing and implementing a highly complex model when a simpler design would have sufficed. Nor, however, do they want to simplify the implementation of the model to such an extent that clinically relevant factors are lost or omitted. It should be noted that the concept of model complexity is in itself subjective (what is considered a complex model by one individual may be simplistic to another).

- model complexity and the importance of transparency
- arguments in favour of model simplicity
- arguments in favour of model complexity.

## Model complexity and the importance of transparency

The interview data suggested that modellers' preferences concerning the degree of model complexity was related to their perceived need for the model to be transparent. However, several interviewees suggested that the model was required to be sufficiently complex to answer the question at the level of detail required.

...I have a preference for the simplest model structure that will answer the question...what we do has to be transparent and verifiable by an external reviewer...

[I6]

In a related discussion, another respondent raised the issue of transparency with regards to debugging the model at the end of implementation, highlighting a trade-off between having a model that is sufficiently complex to answer the question but sufficiently simple to ensure transparency and understanding.

...transparency can be an interesting one, the worst kind of error is the one that is never labelled an error but has resulted in the wrong decision so transparency is key in allowing errors to be seen...

[I4]

### Arguments in favour of model simplicity

The qualitative analysis revealed a degree of agreement between the respondents that models should be as simple as possible provided that they answer the question to the required degree of accuracy and do not omit or neglect clinically relevant features of the decision problem. As indicated earlier (see Chapter 3) there was no clear set of principles guiding the appropriate breadth, depth and level of granularity of models. One respondent highlighted a trade-off between spending time understanding the evidence base and spending time adding complexity into models:

...my gut feel is that normally the bells and whistles don't add very much...in terms of the accuracy of the final product, your time is better spent really understanding the data that you've got than increasing the sophistication of the relationships that you create, or the parameters that you pull out of that data...

[I6]

The same respondent referred to an example drawn from his own experience in which the results of a complex model and a simple model relating to the same decision problem were compared; the interviewee noted very little difference in results so he expressed an opinion that the additional accuracy did not justify the extra commitment of time and resources.

...looking at the findings of quite detailed patient simulation models with a lot of flexibility and description of the individuals and sort of progressive process, and progressive degenerative diseases and comparing that with a much simpler Markov state process where you're grouping the patients much more rigidly and imposing a structure on what can happen to them and actually looking at similar inputs to the two, we got within I think two or three per cent, near enough, the final output, with the capacity to do the same sensitivity analyses. I used to wonder what you were gaining from the extra run-time and sophistication and it seemed that the extra flexibility in terms of

the timing of events or the patient history. I thought two to three per cent was not enough of a difference in your final output to justify the extra complexity...

[I6]

A similar viewpoint was expressed by other interview participants.

...unless there is a clinical imperative for it to be modelled in the more complex manner, I would try and keep it as simple as possible...

[I12]

...I think one simple rule might be to keep it as simple as you can. There is a lovely quote from Einstein which is 'things should be kept as simple as possible but made no simpler than that'...

[I4]

However, one respondent claimed that models typically become more complex as the model development process continues, although this point was shared by all (one respondent suggested that sometimes models may become less complex over time).

...a natural tendency for things to become more complex in a model over time because people say well what about this and what about that and you have to start building other states in or other transitions...

[I4]

Importantly, one respondent perceived a direct relationship between model complexity and the chance of creating technical model errors.

...I think it's a very interesting, fundamental thing which I haven't mentioned yet which is the extent to which you keep a model simple in order to avoid errors as well because I think there is an understanding that building more functionality and complexity into a model comes at the overhead of a greater probability that you make a mistake in the model...

[I4]

A similar view was expressed by another respondent, who suggested that there was no point in building an individual patient-level model if it is not needed, as it adds complexity, hinders peer review and increases the probability of an error being introduced into the model.

...there is no point using [an] individual [patient level] model if you don't need to, that's just a waste of effort and also adds in complexity that clinicians might not like, peer reviewers might not like and also adds in a chance of more errors...

[I2]

## Arguments in favour of model complexity

Several respondents noted that there are occasions when, given the level of understanding of the underlying disease or the presence of dynamic interactions between model entities, a greater level of model complexity may be required. One respondent noted that if implemented without technical error, the more complex model would be his preference.

...apart from that everything else as I see it at the minute it's all in the same box with different levels of how hard it is to do...I would tend to go with the more complicated one provided they're confident, in its ability to, that it's error free...

[I2]

However, respondents highlighted a cost associated with developing more complex models:

...the danger with the NICE process when you're up against hard deadlines is that if a particular set of data doesn't come you may forget that it hasn't come...so in that context one does try and flag up potential for leaving in arbitrary data. The simpler the model the more important it is that one particular thing is right...

[I5]

The analysis of responses concerning model complexity and transparency highlights the pragmatic difficulty in defining the appropriate level of complexity of a model. Inevitably this is a subjective matter of judgement that appeared to differ between interviewees.

## Housekeeping

Housekeeping concerns features and techniques employed within an implementation model (and its development) to facilitate error checking. Four approaches were discussed by respondents:

- use of automatic checks and flags
- use of model settings which are visible on the screen
- standard layout conventions
- naming of cells and cell ranges.

## Automatic checks and flags

One housekeeping approach raised by respondents involved the use of automatic checks and 'flags' incorporated into the model to draw the attention of the modeller to potential errors. They can be used to flag up errors within the calculations performed within the model, for example sum-checks to ensure that the sum of rows of transition probabilities add up to one. In particular, respondents suggested the use of automatic checks to ensure that no obvious errors have occurred in the data entry (e.g. typographical errors) and programming logic of the model; importantly, the use of such checks may raise the awareness not only of the modeller, but any other model user.

...you check that your columns and your rows add up to the same number, you check that your probabilities always come to 1, all those sort of things, they should be on, you know, if it's a spreadsheet, they should be on the spreadsheet. It should be obvious to anybody if there's a problem anywhere. And you know if there is a number which is always positive, make sure you have a test to flag this up if anything goes negative, even if it's 30 years into the future...

[I3]

...there are all sorts of checks that you can build into the model or apply to it, so you know there may be basic checks like making sure your transition probabilities add up to one. So you can build that in as a development model and have a check, or make sure you are accounting for everyone in the population...

[I11]

## Summary of model settings visible on screen

A further housekeeping technique suggested by some interview respondents involved the use of visible switches (potentially incorporating a master switch or monitor) to alert the model developer or user to the current status of the model. For example, this approach could be used to highlight whether a certain scenario was selected, e.g. base-case values, or to toggle between deterministic and

probabilistic analyses. One respondent suggested using a monitoring tool (perhaps a cell which displays a certain colour when all switches or flags are set at default values or settings). Although the incorporation of such techniques may involve some programming time, it was suggested that there is a payoff in terms of avoiding the presentation of unintended analysis results.

...to have a summary of what things are set to so that tends to keep reporting errors minimised to some extent... [I12]

...very often we have a model which has half a dozen switches in it, very easy thing to put a master monitoring cell that goes the right colour when you have any of those switches set. The base case will probably be with them unset, so then when you see the output you know if the big box is red you know that that output represents something with the switch on and it tells you the switch is on... [I4]

### Use of a standard model layout

At one extreme, one of the respondents stated that their institution had adopted a formalised standard model layout with colour coding of certain common elements of the model (e.g. parameters). However, many respondents stated that they adopted at least some degree of standardisation in terms of model layout; most commonly this involved retaining all model parameters within the same worksheet.

...try and keep all the inputs in one place and have them referenced so that someone can go through and say yes that unit cost was £43, tick that sort of check... [I11]

A further suggestion concerned the structured programming of formulae within the model by copying cell formulae rather than 'hard coding' numbers:

...I think you can design the spreadsheets to minimise that by keeping the inputs in one place by keeping the calculations in a nice ordered way...the calculation phases across several sheets in a spreadsheet it's more difficult to identify where it's all going wrong... it's possible just to copy formulae down rather

than having hard-coded things and stuff that can go wrong... [I11]

Several respondents expressed a belief that such housekeeping procedures facilitate internal peer review by other modellers who have not been directly involved in the design or implementation of the model:

...we use a standard set of colours for the different sorts of cell in our spreadsheet. So we use one colour for the check sum, we use one colour for the data entry points, we use one colour for the derived data entry points, use certain colours for each output, we arrange our sheets in a particular way and we have a separate work sheet for specific elements of the models. For example, each comparator arm will have its own sheet and we always do that... having those standards allows each modeller to much more quickly understand the model and to build it in a way that potentially avoids some of the sources of error... [I4]

It differs because if you want the model to be transparent within EXCEL...you need it to be in a nice way so that all your variables that are going to change within the model are on one spreadsheet, they're all in a nice line... [I2]

I would have all the parameters on the same page, all the input parameters on the same page... [I2]

### Naming of spreadsheet cells

A further housekeeping procedure suggested by the interviewees concerned the use of names for input cells (and cell ranges) within spreadsheets. It is clear that this approach may avoid some programming errors; whereas mistyping a cell reference in EXCEL may go unnoticed during model development, mistyping a cell name will return an error value (a '#NAME?' value). Such approaches may also enhance the transparency of relationships within the model if cell ranges are named clearly.

...it's a very different game when you're building a model in EXCEL as opposed to writing a document, the consequences of making a small mistake can be quite high you

have to be very careful about how you build things and how you then check them and so. Using named references for example, rather than saying A3 using a named reference you probably would not have made that mistake...

[I4]

If we use range names we try and label them in the actual thing as well so you can see the range names on cells...

[I10]

## Guidelines

Eleven of twelve respondents discussed the potential role of modelling guidelines. However, the term 'guidelines' may have been interpreted differently; at one extreme some may view guidelines as a standardised model development map that prescribes good practice under any eventuality, whereas others may have more relaxed views of what constitutes a guideline, for example, reference cases, critical appraisal checklists and good practice suggestions.

### Guidelines define good practice

The qualitative synthesis suggested some agreement between the respondents that guidelines which define either good practice or a 'minimum standard' should be attained. However, it was suggested that such modelling guidelines should not be restrictive or inhibit the exploration of potentially innovative techniques and solutions beyond their remit.

...[there is] no scientific reason of why [we] use EXCEL [it's] just a practical thing...other software that can be much more efficient and effective...So if you are going to see guidelines like that I don't want the guidelines but if you see it as a quality standard then it's another thing...

[I1]

...guidelines that are too prescriptive might inhibit the scope to explore things in the way they should be explored by using the appropriate methods and might inhibit innovation as well because things are always done in the same way so I think you need guidelines that are flexible but I think there are a lot of very sensible things that can be done in any situation that contribute greatly

to the prevention of errors in the models just adopting simple standards for writing models...

[I4]

Another respondent expressed a similar viewpoint but raised concerns about the potential to use guidelines as a means of reducing accountability for model errors.

...they would be 'best practice in modelling' type guidelines, I can't see any harm...you have got to be careful not to take those sorts of guidelines too prescriptively, you have got to be responsible for your own work and own errors, you can't disclaim it...

[I12]

One respondent further suggested a potential benefit of the development of guidelines in terms of defining a point of reference for models.

...in that sense they [guidelines] are useful because they allow anybody the same point of reference...

[I8]

...I think the principle of having an overriding set of standards or guidelines in a wider sense, is kind of necessary; and that's what's led to the reference case I guess and various things isn't it?...

[I10]

Not all respondents were in favour of guidelines in any form; one respondent suggested that guidelines were unnecessary in any form; that they are not a sufficient replacement of scrutiny.

...I'm not a great fan of guidelines and standard operating procedures...I'm a great believer in scrutiny and a great believer in... fitness for purpose...

[I5]

Similarly, a different respondent suggested that the value of guidelines was at best limited, serving a useful role only in terms of teaching purposes. In particular, several respondents expressed negative views concerning the value of 'ticking boxes' in checklists.

...guidelines for, you know, for people to operate and to use for checking and ticking boxes and things. I mean, I don't know that

it would do too much harm, but it wouldn't do much good either, I don't think...to be perfectly honest. I mean, the only useful place for using that sort of thing, I think, is in teaching...

[I3]

...We actually took the Drummond checklist for economic evaluations and applied it to the industry model and we got all the way down saying yes, yes, yes, yes until we got to I think it was number 34, do the results follow from the data at which point we said no in very large capitals...

[I5]

This concern was summarised by one of the respondents:

...It's all so easy for you to get a lot of ticks on guidelines and checklists when actually you've done something fundamental that is so special to the particular problem...but the guideline, and unless you are getting to questions like does the model reflect the clinical reality, that is the whole question from the start. It's not really a question on the checklist it is the whole objective of the checklist itself...

[I5]

### Remit of guidelines

As alluded to above, the interview data did suggest some consensus surrounding the value of modelling guidelines, indicating that although technical aspects could be easily covered, issues surrounding understanding the decision problem and conceptual modelling were more difficult to capture.

...so I would perceive a risk that if we go down the guidelines path, we end up with something which is a technical guide for software developers that doesn't address the basic the earlier question in the process, which is: 'is this model actually designed appropriately to answer the question for which it was intended?' So I'd be much more interested in guidelines or in ways of working that helped address that fit-for-purpose question, but I appreciate that's a much harder thing to do...

[I6]

...again it's easy to write a guideline that has a wish list of all our favourite distributions that

we would like to use for the different kinds of parameters...

[I6]

### Skills and training

The final subtheme relating to the prevention of errors within models concerned the skills and training of the modeller and the broader research team. Interview discussions covered a range of skills; the qualitative data were interrogated with respect to four classifications:

- skills related to the design and development of models
- skills related to locating errors in models
- skills in statistical analysis techniques
- background and other abilities of the modeller.

### Skills related to the design and development of models

One of the points that the qualitative analysis exposed was that analysts are concerned more about the ability to design and develop the correct model than the ability to actually implement the model correctly on the chosen computer platform:

...in the design stage a bit more...not to do with how to use the software, like...whether I learnt to use...EXCEL or whatever, but rather more on how to develop programs...

[I1]

One respondent stated that there were some modelling approaches that he would not attempt to implement because he was not confident in his ability to complete the investigation within agreed timescales.

...there will be some things you yourself will do particular modelling approaches that I will probably not want to do simply because I am not confident in my ability to complete them to my satisfaction within a time scale...

[I8]

### Skills related to locating errors in models

Some discussion was held concerning modellers' skills in identifying model errors. One respondent emphasised that the ability to scrutinise models is a key skill.



...well it is mostly appropriate training and I think again it is scrutiny...it all comes back to it's scrutiny...

[I5]

A separate respondent defined the skills to locate a model error as separate to skills concerning developing the correct model or implementing the model on a computer platform.

...it's nothing to do with how good you are at EXCEL and I don't think it's anything to do with how intellectual you are at conceptualising the problem, I just think it's. I think it's separate, I think it's, I don't know, it's just the ability to be prepared to see a problem as something that you break down into bits and you track it down in a logical fashion and you try and narrow down to where the problem is. There's the problem, cut it in half and the problem's in this half, right I'm going to cut it in thirds, the problems in this third. And within 5 minutes you've nailed it down to the five lines of code where it has to be otherwise this thing wouldn't be happening...

[I10]

### Skills in statistical analysis techniques

A minority of interviewees discussed training in the use of statistics techniques and software as a means of avoiding errors. One of the respondents raised an important point regarding the use of statistics: that user-friendly software may result in errors, not as a fault of the software itself, but rather through the misunderstanding of the user implementing the software (for example, WINBUGS software comes with a user 'health warning'; [www.mrc-bsu.cam.ac.uk/bugs](http://www.mrc-bsu.cam.ac.uk/bugs)).

...it's a matter of knowing what the tool does as you learn to use it and really trying to teach limitations alongside teaching use of tools...sometimes user friendly software isn't necessarily a good thing because there is a distinction between software that is easy to use well and software that is easy to use badly...

[I5]

...there's two levels of training. One which is training in sort of statistical approaches and... methods sort of research, and then there's a sort of actually...you know, the actual using software to then do that...

[I7]

### Background and other abilities of the modeller

The final classification concerned the impact of the background and abilities of the modeller in avoiding errors. Limited discussions were had on this issue. One respondent had a particularly strong view about the characteristics of 'good' modellers.

...if you are intelligent you should spot a hell of a load of them [errors] straight off...your best bet for stopping errors is to employ an intelligent person...

[I2]

...you will have people who are naturally better at estimating an answer from just looking at it on paper, than those that aren't, but I think those are the people who make the best modellers and then the best modellers will get it more right because they know what they are doing...

[I2]

The same respondent discussed an 'instinctive' characteristic of modellers and the necessary presence of 'gut feeling' with respect to the identification and avoidance of errors.

...you instinctively know what model to build, you instinctively know whether the answer looks wrong and if they do look wrong that's when your alarm bell goes off. Obviously, if I miss some what I'm calling non-significant errors, they might be in there and I've missed them because my gut feeling hasn't said they were wrong, anybody who never has a gut feeling telling them it's wrong will never spot any of those...

[I2]

Furthermore, a different respondent suggested that the ideal background for health economic modellers may not be health economics, but rather some associated modelling discipline.

...maybe the people that do that [design and develop models] best are the people who've not come to this from a pure health economist context but have come to it from a more professional modelling context and I think there's something in that kind of mathematical modelling mindset that we've already got before we've got into health economics and knew what a QALY was that says that's how you deal with that kind of problem. You either

come at it from another route, break it down into bits or you stick these numbers through, and I think to an extent you only get that through having that background of doing enough of these that you find your way with it...

[I10]

## Explanatory analysis concerning methods for preventing errors

Table 7 presents an explanatory analysis which attempts to link the current methods for preventing model errors to the taxonomy of model error presented earlier (see Chapter 5). The leftmost column details ‘methods’ currently used to avoid error raised by the interview respondents. These 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. These methods have been mapped against the types of model error

they are likely (or intended) to avoid, based on the interview data and through detailed discussion among the authors. As far as possible these approaches for avoiding errors have been mapped to the root cause of the model error; however, in several instances the primary source or cause of error was unclear. The mapping presented in Table 7 may not exhaustively capture every type of error that these methods address. In addition, given the variation in the elicited model development processes of individual respondents (see Chapter 3), the use of the full range of methods for avoiding errors is not necessarily typical.

A number of emergent issues arose from the interpretation of the mapping presented in Table 7. A simple crosstabulation of the number of avoidance strategies for each error domain gives an indication of the focus of the modelling community on different areas of model validity. Clearly, the range of methods and approaches to avoiding model errors is considerably broader than the housekeeping techniques focused on avoiding errors in the implementation of the model; only

TABLE 7 Relationship between methods for avoiding errors and error types

Interventions discussed by interviewees	Relationship with error taxonomy		
	Intervention type	Relevant error domains	Primary error targets
<b>Clinician input to the model development process and use of clinical input to ensure face validity</b>			
Establish ongoing long-term involvement with clinicians who know about disease	Process	Description of the decision problem/structure and methodology used/use of evidence	10, 40, 150
Use internal and external clinical input to elicit information on efficacy and effectiveness	Process	Use of evidence	150, 225
Ask clinicians to provide feedback on how results meet their expectations	Process	Presentation and understanding of results/potentially all error domains	290 (may relate to any point on taxonomy)
Discuss what is possible with clinicians	Process	Model structure and methodology used	40, 50
Discuss assumptions with clinicians throughout process	Process	Model structure and methodology used	20, 40, 50
Discuss likely impact of intervention with clinicians	Process	Model structure and methodology used	40
Discuss data sources with clinician (directed by clinician)	Process	Use of evidence	100, 180
Ask experts who know about the disease to comment on the model	Process	Model structure and methodology used	20, 40, 50
Ask experts who know about the disease to comment on the clinical pathways	Process	Model structure and methodology used	40

**TABLE 7** Relationship between methods for avoiding errors and error types (continued)

Interventions discussed by interviewees	Relationship with error taxonomy		
	Intervention type	Relevant error domains	Primary error targets
Step through model pathways with clinicians	Process	Model structure and methodology used	40
Build initial model before meeting clinicians	Process/technique	Model structure and methodology used	40, 50, 60
Draw out pathways on paper with clinicians	Technique	Model structure and methodology used	40
Developing written documentation of the proposed model structure	Technique	Model structure and methodology used	20, 40, 60
Use of diagrams or sketches of model designs and/or clinical/disease pathways	Technique	Model structure and methodology used	40
Memos	Technique	Model structure and methodology used	20, 40, 60
Representative mock-ups to illustrate specific issues in the proposed implementation model	Technique	Model structure and methodology used	40, 50, 60
Written interpretations of evidence	Technique	Model structure and methodology used	20, 30, 40, 50
<b>Mutual understanding between the researcher and decision-maker/client</b>			
Iterative negotiation and communication between the modeller and the client (managing expectations)	Process	Description of the decision problem	10
Advisory committees to generate consensus of best approach	Process	Model structure and methodology used	20, 30, 40, 50
Mutual understanding within the team			
Meetings with research team through project	Process	Model structure and methodology used/use of evidence	70, 150, 225
Presentations to internal team and wider team throughout project	Process	Model structure and methodology used/use of evidence	70, 150, 225
Exposition of model during research team meetings	Process	Model structure and methodology used/use of evidence	70, 150, 225
Handling model complexity			
Ensuring transparency of model	Technique	Model structure and methodology used	30
<b>Housekeeping techniques</b>			
Automatic checks	Technique	Implementation of the model	230, 240, 250
Automatic flags (including master flag)	Technique	Operation of the model	270
Summary of model settings visible on screen	Technique	Operation of the model	270
Use of a standard model layout	Technique	Implementation of the model/operation of the model	250, 270
Naming of spreadsheet cells and ranges	Technique	Implementation of the model	230, 240, 250
			<i>continued</i>

TABLE 7 Relationship between methods for avoiding errors and error types (continued)

Interventions discussed by interviewees	Relationship with error taxonomy		
	Intervention type	Relevant error domains	Primary error targets
<b>Establishment and dissemination of good practice</b>			
Development of guidelines	Process	Model structure and methodology used/use of evidence/presentation and understanding of results	30, 50, 110, 170, 190, 310
<b>Skills and training</b>			
Develop skills in understanding software	Process	Model structure and methodology used/implementation of the model	60, 70, 250
Develop skills in developing programs	Process	Model structure and methodology used	70, 80, 90
Develop skills in scrutinising models	Process	All error domains	Any point on taxonomy
Develop skills for analysing problems	Process	All error domains (primary communication and judgement errors)	Any point on taxonomy
Developing skills besides health economics	Process	All error domains	Any point on taxonomy

one-fifth of the methods target the implementation and operation of the model. The remaining four-fifths target the conceptual validity of the model. Very little emphasis is placed on presenting and communicating results and conclusions. An emergent issue that is evident in the synthesis concerns the nature, and balance, of processes and techniques for avoiding model errors. In particular, the techniques detailed are explicit, and can be interpreted as relating to *how* something should be done, for example, implementing a specific model layout. On the other hand, the processes for avoiding errors recognise that there is a need to be addressed and acknowledge that something should be done as part of model development; however, in many cases this is not accompanied by a clear strategy for achieving the required goal. This issue particularly affects those aspects relating to the definition of the decision problem and conceptual modelling. A number of requirements are identified to achieve clarity and mutual understanding; these are expressed as process requirements (e.g. establishment of long-term clinical input) to avoid errors in the description of the decision problem. Whereas a number of

techniques are suggested by the interviewees, for example, sketching out clinical pathways, these are not framed within an overall strategy for structuring complex problems.

### Placing the findings of the interviews in the context of the literature on avoiding errors

Fostering mutual understanding between the modelling team and all stakeholders to the decision-making process is targeted at achieving model credibility. The issue of defining appropriate level of complexity is well discussed within the literature; however, no objective methods are suggested for dealing with this issue. Literature on problem-structuring methods suggests that the focus should be on developing adequate models, that are 'owned' by the decision-maker. This in turn, highlights the importance of transparency of reporting, implementation and interpretation of models.

# Chapter 7

## Strategies for identifying errors

### Overview

This chapter presents a qualitative analysis of discussions concerning procedures and techniques that are employed to identify whether, where and why errors have been introduced into models. As part of the Framework analysis, a number of classifications emerged from the qualitative data.

- Check face validity with clinicians.
- Do the results appear reasonable?
- Test model behaviour.
- Is the model able to reproduce its inputs?
- Can the model replicate other data not used in its construction?
- Compare answers with the answers generated by alternative models.
- Peer review of models.
- Double checking of input values.
- Double-programming.

### Check face validity with experts

A commonly cited approach used in the identification of model errors concerned checking face validity of interim and final model results with people who know about the disease and treatment(s) under consideration. This has relevance both in the avoidance and the identification of errors; in this case the primary differences in its role concerning whether the error is found before or after the decision has been made, or whether a model is being scrutinised by clinical experts acting on behalf of the decision-maker (e.g. reviewing models submitted by manufacturers within the Single Technology Appraisal process). Key discussions surrounding this issue have therefore been presented (see Chapter 6).

### Do the results appear reasonable?

A common issue concerning the identification of model errors involved whether the results appeared to be 'reasonable'. As noted earlier (see Chapter 4),

the 'reasonableness' of model results, and whether they match preconceived expectations, were raised as dimensions of what constitutes a model error.

...there's a level of validation which can happen at [the] end of any model, because, you know, you want to make sure it's actually giving you numbers which aren't ridiculous...

[I7]

...I suppose, that there's a kind of issue of judgement in terms of whether it looks, whether the model itself looks reasonable...

[I9]

The ability of a modeller to develop an expectation of the answer has been discussed previously (see Chapter 6, Skills and training). As highlighted through these previous discussions, the source of these expectations differed between respondents. For example, one respondent suggested that an expectation of the answer can be formed by the modeller through an examination of the key features of the disease and the evidence used to inform the model.

...you'll look at the key things which... are going on, but you want this right at the beginning of the process. We call it the sniff test which is whether the result smells right...

[I6]

In practice, this approach echoes the idea of 'skeleton models'; 'back-of-the-envelope' models, which adopt a high level of abstraction to generate an expectation of the model result. One respondent highlighted how this might work in practice; however, as already noted (see Chapter 4), although a model result may match one's expectations, this does not provide a guarantee that the model is error-free.

...my back of the envelope said this was going to be about forty grand per QALY and there were some subgroups that might be interesting. If you come to the end, or you come to a draft model of the results, and you have something radically different from that, then there should be a reason. If it comes out and it's four grand

per QALY then it just smells wrong, and something has changed between the decision problem you thought you were looking at in the beginning and the data that's actually been implemented...

[I6]

...do the results, if it [the model] says the ICER is £50,000...does that seem about right using the back of the envelope sort of techniques... you're never going to be able to confirm that exactly without the full model full calculations but you should be able to at least justify [it] or not within a fairly...certain range...

[I11]

The same respondent described the action which he would take if there were a significant difference between the expectation and the answer from the model:

That's the sort of detective phase I think, identifying which bit is different, so your fag packet might have got differences in costs divided by differences in QALYs...so firstly which of those two or both is different from what the model is generating, and if it's say the difference in costs, is there a cost of intervention or the cost of the comparator, which one's wrong?

[I11]

Similarly, another respondent stated that he estimates what the model output should be using pen and paper at the start of the project; if this approximated result is not within a reasonable tolerance of the results from the final model, he then attempts to establish which is wrong and why.

...I regularly work out the answer before I do the project and see if I am right or not, and if I don't end up with the answer I think I have, work out why I was mistaken in the beginning, why my original calculation was wrong or why the new calculation is wrong...

[I2]

That's why it's interesting, when it doesn't match your back of the fag packet answer, you've got to be damn sure you are right...

[I2]

A different respondent stated that when a model is operating in an unexpected fashion then one of the approaches is to build up the same relationships and logic and try to track down which part of the model is responsible for the error.

...if issues are identified in the model during construction or in quality control and it's one of those really annoying, why the hell is this doing this? Then one of the approaches is, and again this is experience and not written down. It's out the model and I'm going to go down to back of the fag packet, start again and just build up the same thing, in essence, not replicate it, but just build up the same kind of relationship and logic to see whether that's at all possible or try and track down which part of that process must be going wrong to hit those kind of numbers...

[I10]

The same respondent suggested that such approaches should be a standard part of the model development process.

All the good modellers do and the bad ones don't...

[I10]

## Testing model behaviour

Testing model behaviour relates to the process of changing the input values in a way which will result in a change in the results, and examining whether the change was expected (note there is a direct link to the methods for developing expectations described previously). One respondent suggested a series of logical tests involving the efficacy of interventions as a means of identifying potential errors within a model.

...if I increase the efficacy of an intervention, the effectiveness ought to go up and the cost-effectiveness ought to go down. If I switch the inputs for the two arms of the model, I should get the opposite result out of it at the end. If I put the same inputs into both arms, I should get a null...

[I6]

...as you start to think through the model you start to think of an expectation of what the results are going to be and given a certain change how you think the results should change and the more specific you get in terms of the structure of a model and the data inputs the more specific you can be about how you expect the change to be so if it isn't in the direction or the magnitude that you are expecting, you are expecting that you have probably made a mistake somewhere...

[I8]

Another respondent states that it is possible to prospectively define a set of input parameters and develop an expectation of their impact on the model output.

...well...one thing you can do is you'd define a set of input[s], and you know what will be the set of output[s]. And with that, you know exactly...whether the results are right or wrong...

[I1]

...you see testing as a black box where you define input and you know what output you are expecting. So you change the inputs, and you see how the outputs change...

[I1]

The same respondent highlighted that techniques exist that can facilitate this process of design and analysis; however these are not currently used within HTA modelling.

Imagine you have five inputs that can have different values, each of them, and you know what the outputs should be for the values. What would be the minimum number of combinations that you can do with your five inputs in order to get to be sure that the outputs are correct...so there are techniques that tell you how to define your testing inputs, then, in order to get the outputs...

[I1]

Another suggestion concerned the use of model functionality tests such as altering assumptions concerning the initial distribution of patients in the model.

...model functionality tests...what happens if you set the starting population to be all in one state rather than another...

[I11]

Another approach was to test extreme input values in terms of stability and robustness of model outputs.

...taking extreme settings as inputs as ways of checking that the model is resilient to use its input...

[I11]

Whereas many of the respondents discussed the use of model testing to identify errors, it was suggested

that the use of such techniques did not guarantee that errors would be identified (a similar point was made concerning developing expectations; see Do the results appear reasonable?). One respondent suggested that the model scenario may match previously determined expectations as a result of luck rather than as a consequence of internal consistency within the model.

...if you can test this by a kind of univariate type sensitive analysis approach where you're saying I think this is going to have a huge impact or at least if I increase this value, the result, you know, that value should increase rather than decrease...you may have the right direction, seeing the right magnitude of change, but actually purely by luck...It's just pure luck that there happens to be the case and there's some error in there.

[I9]

Other examples of model tests included setting all utility parameters to zero, adjusting transition probabilities and assessing the impact, and setting the parameters in each arm of the model to the same value and checking that the results were identical.

Simple things like if you set utility values in the model to zero is the output of the utility zero in each arm? If you set the data points in each arm equal do you get an ICER of zero or not...

[I4]

...when you increase one of the transitions which you expect to reduce the utility then the utility goes downward. So you play around with it and make sure its pretty much behaving like you want...

[I4]

Although the majority of interviewees discussed the techniques they adopted, one respondent adopted a more normative stance, suggesting that the testing model behaviour should be a standard part of the model development process.

...every time you change something, you should as a matter of course just check that the ICER goes in the right direction...A new parameter, a parameter changes or a new variable or a new state or any of these things get added you rerun the model, check the ICER, did it go in the right direction...

[I2]

...what we have at the moment is a model prior to final or full analysis being run that we are happy with and at that point it is only really when we are starting the final stages of the model when you are producing results, some of the results, especially those exploring extreme values, that your problems become fully apparent and then the question becomes do we need to refine the model or is our model giving reasonably good predictions or any biases that we have within our model reasonable over the plausible range of those values and only becoming uncertain over implausible values. So we are actually maybe trying to specify at what points our model is going to fall over...

[I8]

...you'd hope that that sort of iterative process, the sort of extreme values would kind of dig up the major ones...

[I7]

## Is the model able to reproduce its inputs?

Testing internal validity, that is, checking that the model can generate the inputs used to populate it, was also discussed by participants. This corresponds to Eddy's second order of model validation.<sup>59</sup> Several respondents suggested that they would typically use this approach as part of the model checking process; examples included testing incidence data and survival analysis used to inform model parameters.

...I suppose it does depend on you know if you have taken for example, national statistics, cancer incidence data to go into your model then you very much expect the incidence coming out of it...

[I11]

...we might do some kind of level of validation in terms of...model survival curves [do they] look anything like the data itself? And so there would be some kind of, where it's possible, internal sort of validation on data inputs, and I think that would be done on a sort of statistical basis...

[I7]

However, some participants' views of the utility of this method were negative, indicating that it may be a necessary but not sufficient condition of any model.

...internal validity means that you get out what you put in, big deal, fine...you have to do that, but it doesn't get you anywhere. It doesn't tell you that your model is right. It only means that what you put in – if you put rubbish in it, then you'll get rubbish out of it. So all it, all it means is that, you know, you've got the mathematics... right...

[I3]

Another respondent highlighted that there may be instances whereby one would not expect the model outputs to match the input data used to inform it; however, this respondent did express a preference for demonstrating internal consistency where possible.

...even if you have good reasons your numbers won't come out exactly the same and some of those might be overriding reasons, because of the way you are trying to use the trial for a particular model in a particular setting. If it's possible to build a model in such a way that you can show consistency with a clinical trial then I think you should do that. That is a useful check...

[I12]

Interestingly, one respondent suggested that testing the internal consistency of the model did not necessarily take place within the model itself, but may be undertaken as part of a premodel analysis. The perceived benefit of this approach was that it may ensure that the data have been interpreted correctly before their incorporation into the model.

...I would just make sure that my survival models and other data that I have pulled from the trial is making sense, quite separately from any economic decision model. Can I reproduce these things in a way that looks right? Just to make sure I have done the analysis correctly and understood the results...

[I12]

## Can the model replicate other data not used in its construction?

An additional aspect of model validation discussed by interview participants concerned the ability of the model to replicate data that have not been used in its construction; this corresponds to Eddy's third order of model validation.<sup>59</sup> One respondent stated that this level of model validation is rarely possible



given issues of data scarcity and the general principle of using all relevant evidence to inform the model.

What does that mean – external validity? Well, if it means accurately predicting something from a different source, well, that’s all right, except that most of the modelling that we do, that I’ve ever done, is obliged to use every source that’s available! Because of data scarcity. So basically there you know, you cannot do external validation...

[I3]

The same respondent highlighted other problems concerning the relationship between the data used to populate the model and its relationship to evidence collected in the future.

...when you’re on a chronic model, by the time you get 10 years into the future, the whole circumstances have changed, anyway. The epidemiology’s changed, the interventions have changed, the population’s changed, you know, everything’s gone...if somebody predicted it and put it in a sealed envelope, it’s bound to be wrong. Does that mean that the model was wrong? No, because it was talking on the basis of what was known at the time and the circumstances at the time...no, predictive validation is impossible anyway. So the whole notion of model validation is either trivial or impossible...

[I3]

One respondent highlighted that even if such analyses were possible, time constraints may represent a barrier for this level of validation.

...we talked about [whether] our model had predicted the trial results but by that point you are on another project so you’ve got too many things going round your head and we’ve never actually done it, and then it’s probably too late to say you know that model three years ago, it’s wrong...

[I2]

## Compare answers with the answers generated by alternative models

Interviewees also discussed comparing the de novo model results against results from previously published studies. However, this is likely to be

problematic as a method of validation because models which attempt to address similar decision problems may have employed different structures, assumptions and parameters. As noted by one respondent, it may be necessary to accept that some difference in results is to be expected; however, the underlying cause of the difference may not be clear.

...it’s a judgement, but you’re not expecting to get absolutely the same result, but, you know, the convergent validity, let’s call it that...the differences between these models, do they, come down to fundamental differences in structure, differences in conceptualisation, differences in scope, or is it the parameter inputs? And what would obviously be reassuring in that situation is change the parameter inputs to suit those of the existing models produced, you know, similar results...

[I9]

...there are two levels I guess of validity, one of which is does...your model...look kind of reasonable in terms of what other people have done...

[I7]

One respondent suggested that this aspect of model checking should run alongside model development; however, there may be a danger of becoming reliant on the structure and assumptions of previous models (see Chapter 3, Discussion of model development process). Indeed, the presence of differences between models does not guarantee that *either* model is ‘wrong’, rather such differences may be explained by the modellers’ preference for a given set of assumptions and methods.

...if somebody else has got a model of a different answer to it, we wouldn’t automatically say it’s an error. Because, you know, we’ve got our own preferred set of assumptions and ways of dealing with it...

[I7]

Nonetheless, it was suggested by one respondent that certain aspects of the model should always follow accepted conventions and that failure to do so limits the comparability of model results.

So the accepted convention for discounting is that you discount after the first year in whole numbers...so this 12 months is not discounted, the next is discounted at 1 year and 2 years and so on...well, you know, people will

correctly use the correct discounting formula from point zero, so they are discounting every day from then onwards...it's not technically incorrect, but it is conventionally incorrect and inappropriate if you are then going to make a comparison with another study in the same area that's discounted in a different way. So then your answers are then not comparable...

[I3]

## Peer review of models

The majority of respondents discussed internal peer review as a common element of the model development process representing a useful means of identifying errors. Perhaps the principal reason for peer review was that 'your own mistakes are the hardest to see' [I4]. For example, one respondent stated that completed models are passed to a senior member of staff who is charged with the task of 'breaking' the model.

We have a process here where we have a sort of a peer review, so we have enough experienced staff that we could give the the draft model to one of the other project leaders and ask them to spend a few days kicking it about to see if there are errors or mistakes in it...

[I6]

Often, but not always, this was suggested to involve a modeller who has not been involved in the design or implementation of the model; however, some respondents suggested that meetings were held and model results presented throughout the development process (for example, in the supervision of junior staff).

...you might need someone external to the modelling to find those kind of errors...

[I1]

A key issue in the internal model peer review was that the reviewer must be capable of understanding the complexities of the model.

...you would like to think that the people checking it are reputable and able so they will understand the complexities that they have put into the model whereas others might just dismiss them...

[I12]

Similarly, one respondent suggested that external scrutiny of the model may also be seen as an ideal;

however, this may be constrained by time and concerns regarding intellectual property and a lack of willingness to share ideas. As a consequence, it was suggested that internal peer review is likely to be a viable second-best alternative:

...if there were time, you would want to send it out to some independent expert who would go through it all with a fine-tooth comb and tell you everything that's wrong with it. In practice that doesn't happen, it doesn't happen because there isn't the time to do it, and secondly, it doesn't happen, because if you're doing anything that's genuinely original, you know, you don't want to run the risk of anybody else picking up all your ideas. So, you know, there are problems on that front. The best that you can do is to try and expose it to somebody probably internal, who hasn't been involved in the development, and let them go through it and say, 'Well, can you find any obvious flaws here?' And that's probably the best that, in reality, we can usually do...

[I3]

It was suggested that the benefits of peer review did not lie solely in terms of identifying errors, but also in consolidating the modeller's own understanding of the decision problem and the justification for the modelling approach.

...there's also a stage of exposing the model at that stage to I suppose to people within the health economics group but also outside of that, so I think there is a, well, I think there's a benefit in terms of, again, just going over the way the model's been put together...

[I9]

Several respondents highlighted the importance of model scrutiny through peer review because of the model's political importance, as well as a feeling that if the peer review failed to identify substantial errors the modeller could be more confident that the model was error-free.

We have had people check my...model because that's a huge, politically important piece of work...

[I2]

...these things are subject to severe scrutiny and if they find something tiny well they've looked through a lot...it was a pure transcription error, I put the wrong number in and it was picked up by the technical lead...the

thought was if that had been checked with that much scrutiny and that was all they could find then there is actually a good chance that the rest of it was correct...

[I5]

However, despite the potential benefits of model peer review, some respondents raised concerns about the sensitivity of peer review (before publication) in detecting model errors:

...I remain baffled as to how general reviewers assess the quality of models without looking at them...but how a journal can perform a meaningful review on the basis of a three and a half thousand word manuscript describing a highly technical model...

[I6]

We try to publish in journals with the highest citation rating. Those are almost inevitably clinical journals. Clinical journals use clinicians and generally speaking the reviews for clinical journals are the older clinicians who've had very little direct exposure to economic evaluation, modelling or anything. They are pure clinicians. They don't know what they are reviewing...

[I3]

## Double checking of input values

One commonly cited process for identifying model errors involved double-checking model input parameters in terms of its incorporation within the model and cross-checking against the original sources from which parameter value(s) were derived, e.g. clinical trial publications. In relation to definitions of model error, these processes primarily concern ensuring that data have been transcribed and linked within the model as the modeller intended.

...we ask them to look at the accuracy of the input data, so that's been incorporated into the model appropriately...

[I6]

...there is a lot of checking to make sure the data is right not just so you talked about what the meaning of the data points are but also make sure that the actual values are what they should be so we step through the model and

check, go through each data point and make sure it's right...

[I4]

One respondent suggested that the effectiveness of such checking activities may be constrained by time availability and the tedium of the task.

...I mean, checking and double-checking, really. The only way really is having the time and the discipline to go through, crosscheck everything you do several times...

[I3]

One respondent stipulated that such checks must involve checking confidence intervals as well as mean input values.

It wasn't the point estimate that had been entered wrong it was the confidence interval and things like that are difficult to spot...

[I12]

## Double-programming

The issue of 'double-programming' was discussed implicitly or explicitly as a potential approach to identifying model errors. For the purposes of this analysis, double-programming relates to the development of the *same* mathematical model in two different platforms; a distinction should therefore be drawn between double-programming and the development of 'skeleton' models discussed earlier (see Chapter 3). Although respondents generally agreed the double-programming was a valuable activity, there was variation surrounding the extent to which double-programming was undertaken as a standard component of the model development process.

## Perceived value of double-programming

The potential value of double-programming primarily concerned developing confidence in the results of the original model; that the results should be comparable.

...I think that is about rebuilding models. It's about you know, rebuilding it in a different software package and making sure that you get similar answers...

[I7]

Clearly, the focus here is on technical and logical errors in the model implementation, rather than problems in the structure of the original model; as such the value of double-programming in identifying errors may depend on how the second model-build is implemented. However, one respondent noted that if the same analyst rebuilds the model in two different software platforms, they are less likely to identify any errors because they may not re-examine the structure of the model.

...building it in two platforms but that's effectively your double-build but the same person building it in two platforms but I would guess that is more prone to just using the same data and not re-examining your logic to see whether your logic was correct you'd just do exactly the same, so I think that's less likely to solve errors...

[I2]

However, one respondent highlighted that if the second model-build was undertaken by an individual who was not involved in the development of the original model, then this may highlight problems in interpretation of the decision problem.

...[It] depends on the circumstances but one of the things that might happen is there is an independent rebuild either by someone you know or outside...it's just as a check principally on whether you find any errors or in interpretation as the idea is that a reasonable person would come to a similar conclusion...

[I12]

For one respondent, double-programming was considered a standard part of model development, whereby the model is developed using two platforms in parallel:

...certainly at the stage of you know, developing it on the alternative platform... those are done with a model that you feel reasonably confident with...

[I9]

...I would normally do a quick development in TREEAGE. If we're thinking of sending it off to NICE, then we'd do it in EXCEL and, and we'd make sure that those two versions are kept in tandem...

[I9]

## Aspiration or realism?

Several respondents suggested that double-programming was an aspirational 'gold standard' rather than an attainable goal.

...one of the things we don't typically do is replicate our models using another piece of software and often that would be useful because that would be another way of checking things, so we don't do double-programming...

[I8]

One respondent highlighted that their gold standard would involve multiple modellers programming the model independently; however, one may foresee problems in terms of comparability of the resulting models (see quote from respondent I4, Chapter 3, Use of information in model development).

...well the real ideal is to have two people building the model independently...well even that is short of the ideal. You can have as many people as you like you know...At least three people building the model independently would be good but...

[I5]

## Constraints on double-programming

Two respondents discussed potential constraints surrounding double-programming. In particular, the ability to undertake such model checking activities may be limited by cost and time. However, one respondent suggested a relationship between double-programming and model complexity; that double-programming may be a viable option for simpler models.

...I know in a lot of those guidelines they talk about, in this verification stage they talk about parallel development of models in other software. We've never done it and I can't see us ever, commercially being able to do that because you are, by definition, you're doing it twice and you're doing it twice for a reason, to identify whether doing it in a different software or a slightly different approach makes a big difference...

[I10]

...if it is a big complex model it needn't be an aspiration you could do it but that can take a lot of time...it's an aspiration in some cases

because time or complexity doesn't allow in other cases it's achievable if it's a reasonably simple model...

[I12]

## Explanatory analysis concerning methods for preventing errors

*Table 8* presents an explanatory synthesis which attempts to link the current procedures and techniques for identifying model errors to the taxonomy of model error presented in Chapter 6. The layout and interpretation of the table, and the methods of synthesis used in its construction, are essentially the same as that for *Table 7*.

## Placing the findings of the interviews in the context of the literature on identifying model error

The literature revealed a number of validation techniques which have been developed in the field of computer science. Many of the techniques described in the literature were discussed by the interviewees. These included: comparison of the model with previous modelling of the same situation, seeking face validity through asking clinicians who have specialised in the disease process or medical condition being investigated to comment on the model, retrospective and prospective predictive validation, degenerate testing in which the value of a single parameter is changed and the behaviour of the model is examined to determine if it is in line with the expected behaviour, and extreme condition testing in which the model parameters are assigned extreme values and the behaviour of the model is investigated. Another process that is described in the literature and the interviews is the process of assigning the model fixed values that will facilitate manual calculation of the results or interim results, analogous to the pen and paper approach.

A number of processes were listed that were either not discussed or discussed in different terms. These include animation, operational graphics and trace testing. In animation, for example the number of patients in each health state across time might be displayed graphically across time to determine

whether the behaviour of the model is in line with the expected behaviour. An example of trace testing might involve tracking the proportion of patients in each health state numerically across time; the exhibited behaviour can then be compared to determine agreement with the expected behaviour.

The literature also describes Turing testing, in which the ability of an expert to distinguish between observed phenomena and output from a simulation model is used as a test of the model's validity. This process was not described by the interviewees and may have some potential for applications in HTA modelling.

Given the overlap between the identification and the prevention of model errors, the processes and techniques relating to clinical involvement in model development and model scrutiny are not repeated here. One key point is immediately noticeable from the interpretation of *Table 8*; many of the methods used to identify errors in the model involve specific techniques (rather than processes) that may be applied once the model is complete, in progress or under scrutiny by a third-party. However, the specific target of the techniques, i.e. the types of error that the technique is intended to identify is not always clear. Indeed, the majority of methods do not appear to map directly to specific error types in the taxonomy, but rather they may be used to identify *symptoms* of errors, the true source of which may relate to any point in the taxonomy of model errors. For example, the identification of a mismatch between actual model results and the derivation of prior expectations from a previous model may be suggestive of the presence of an error within the model, yet its root cause may be entirely unclear (there may even be legitimate reasons for such differences). This represents a considerable challenge in the peer review of models. The same may be true of certain black-box validation techniques; only the methods which test the underlying logic of the model can guarantee the presence of an error. In this sense, the methods outlined in *Table 8* that can be mapped directly to specific aspects of the taxonomy (e.g. switching model inputs between arms) are diagnostic in nature; mismatches in model results and expectations are indicative of the presence of model error. Conversely, those techniques and procedures which map to any or all points in the taxonomy are effectively non-specific model screening methods.

**TABLE 8** Relationship between methods for identifying errors and error types

Interventions discussed by interviewees	Relationship with error taxonomy		
	Intervention type	Relevant error domains	Primary error targets
<b>Checking face validity of the model</b>			
Clinical review of model structure, evidence and results	See Table 7		
<b>Evaluating whether model results appear reasonable</b>			
Compare interim or final model results against predetermined expectation (from previous models, from skeleton/back of the envelope model, from earlier version of model)	Technique	Any/all error domains	
<b>Testing model behaviour (black-box validation)</b>			
Increase efficacy of intervention – effectiveness up, ICER down	Technique	Any/all error domains	Any point on taxonomy
Switch inputs for two model arms – results should be opposite	Technique	Implementation of the model	230, 240, 250
Functionality tests, e.g. set initial distribution all to single state	Technique	Any/all error domains	Any point on taxonomy
Test how model behaves at extreme values	Technique	Any/all error domains	Any point on taxonomy
Set utility values to zero – no QALYs	Technique	Implementation of the model	230, 240, 250
Adjust transition probabilities	Technique	Any/all error domains	Any point on taxonomy
Set same model parameters for each arm	Technique	Implementation of the model	230, 240, 250
<b>Testing the internal consistency of the model</b>			
Testing the internal validity of the model – can it reproduce data used as inputs, e.g. incidence and survival data?	Technique	Structure and methodology used/ use of evidence/implementation of the model	70, 80, 90, 200, 230, 240, 250
<b>Testing the predictive validity of the model</b>			
Testing the ability of the model to predict data not used in its construction	Technique	Any/all error domains	
<b>Peer review of models</b>			
Internal peer review by modeller responsible for building model (e.g. supervisor of junior staff)	Process	Any/all error domains	
Internal peer review by modeller not external to the process	Process	Any/all error domains	
External peer review	Process	Any/all error domains	
<b>Checking model input values</b>			
Check model input values from source material	Technique	Use of evidence/implementation of model/presentation and understanding of results	130, 230, 240, 280
<b>Double-programming</b>			
Double-programming by same modeller in two platforms	Technique	Structure and methodology used/ implementation of the model	70, 80, 90, 230, 240, 250
Double/triple-programming by independent modelling groups	Technique	Any/all error domains	Any point on taxonomy

## Chapter 8

# Barriers and facilitators

The final theme concerned barriers and facilitators to model checking. Interviewees were asked about their views concerning barriers and facilitators to model checking and whether model checking is currently given sufficient priority. Respondents engaged in broader, more detailed discussion regarding barriers to model checking rather than facilitators.

### Barriers

Barriers to model checking activities discussed by interviewees are summarised in *Box 1*. The interviewees commonly focused on resource constraints (time constraints, people, money) and organisational/process constraints. Two further barriers included the perceived importance and priority of model checking activities and modellers' confidence in current model checking procedures.

Views concerning the nature of current barriers to model checking varied between respondents. Two respondents highlighted a trade-off between time available for developing the model and time available for checking the model:

...if you spend all of your time doing quality control you don't have as much to actually build a model.

[I6]

Time was highlighted as a barrier to some of the more aspirational model checking activities such

as double-programming, particularly for more complex models. Two respondents also suggested a relationship between time constraints and model validity issues in terms of the use of an appropriate methodology and plausible assumptions.

I've gone back and gone well why did I make that assumption? Actually a better assumption was this, but I've gone with an assumption because it was easier, quicker and they need the report tomorrow.

[I2]

A second respondent also suggested that while one may wish to explore whether more sophisticated models would have made a difference, this is constrained by time. Human resource constraints were also raised by several respondents, including availability of modellers as well as other stakeholders within the model development process, e.g. clinical experts providing input on clinical practice.

You can have clinician involvement in terms of the clinician who is removed from the client and can offer an independent voice, not always possible due to time constraints, it really needs to be done quite quickly to organise the necessary meetings, preparatory literature and so on...

[I12]

Alongside physical human resources, the dimensions of human constraints also included

#### **BOX 1** Summary of barriers to model checking

- Lack of time for checking models, timing of the process and time management
- Financial resources
- Availability of clinical input
- Tedium of model checking process
- Lack of confidence in usefulness of current model checking processes
- Lack of awareness of the need to check models
- Lack of awareness of the need for good housekeeping
- Lack of accountability for model errors
- Low priority accorded to model checking activities

the availability of sufficient levels of expertise in scrutinising HTA models. Financial constraints were noted explicitly by some respondents, and implicitly by others. However, even with sufficient financial resources, it was suggested that more extensive model checking activities may not identify and avoid all model errors potentially because of the tedious nature of model checking activities and a lack of confidence therein.

...we don't get paid enough, the cheaper option is to get someone else to go through it but after they've been through a thousand lines of coding and not spotted an error, do they give up and not check the next thousand.

[I4]

...it is a matter of painstakingly going through and seeing 'yes, this number and this model and where does this correspond to in the report', but that is not something you know – it is not something that is not guaranteed to work just because it is a tedious process...you are not going to have time to go through and find every single number.

[I5]

The above statements serve to highlight some of the gaps between the breadth of error types presented in the taxonomy and the effectiveness of current model checking activities to identify and avoid these (see *Tables 7 and 8*). Other barriers included developing models 'too quickly' and organisational barriers such as working within groups in which a shared understanding is not maintained. One respondent raised a number of less tangible barriers to model checking in which the onus is placed on the modeller rather than the constraints of the decision-making process. Barriers included a lack of proper time management, a lack of appreciation of the need to check models to identify errors, failure to adopt housekeeping approaches and a lack of accountability for model errors.

To explore the priority allocated to model checking activities interviewees were asked 'if you had to increase checking which other activities would you drop?' and/or 'where in the development process would additional monies/time be allocated?'. Some of the comments indicated that a low priority was

attached to these activities, indicating that simple resource constraints were not necessarily the key barrier.

I don't think we'd drop anything. I'd spend less time on validation...well, most of these projects are all underfunded anyway...so, we'd probably again apportion that money into those other activities, and again...put less on validation.

[I7]

...if we are under severe time pressure we try to cut back on what the outputs delivered will be, so if we were hoping to do a value of information analysis we will chop that, if we were hoping to do sophisticated sensitivity analysis we would chop that, just drop blocks of work rather than compromising the quality of key work which is the fundamental outputs in the model.

[I4]

## Facilitators

Discussion of potential facilitators to model checking was typically brief, with few respondents suggesting how any of the above barriers could be ameliorated. Interviews suggested that setting up internal processes to ensure that things are not done in a hurry, that projects are adequately staffed and that work is divided up appropriately might facilitate improved error checking activities. In contrast to the barriers to model checking, one respondent from an Outcomes Research organisation suggested that in his experience clients would provide a budget for quality control; this view was not reflected in any respondents working within an academic setting. One interviewee suggested the use of a time delay within the model development process to allow improved internal review:

If we had a 3-month delay in every project where you handed it in...and then you came back in 3 months to read your report and re-examine your assumptions, your logic and what you had written in the report, your report would be a hell of a lot better, which is why the peer review.

[I2]



# Chapter 9

## Discussion and conclusions

### Strengths and weaknesses of the study

The structure for considering the validity of qualitative research described by Mays and Pope<sup>24</sup> is used here to examine the strengths and weaknesses of this study. Mays and Pope identify the following key issues of validity: triangulation of methods, respondent validation, transparency of methods, reflexivity and fair dealing.

This study uses two primary sources of evidence, interviews with HTA modellers and reviews of the literature, this twin approach has the strength of providing a triangulation of different research methods, helping to ensure a complete coverage of issues and allowing a convergent validity of the results and conclusions to be examined.

Respondent validation was undertaken with the HTA modellers interviewed to ensure that the extracts from the transcripts used within the analysis were taken within context and did not misrepresent the views of the interviewees. This validation was undertaken by sending the interviewees a near final draft of the report, including the complete interview analyses and principle conclusions. The communication invited both general comment on the content of the report and specific comment on the above validity issues. All interviewees were happy that their views had not been misrepresented.

To ensure clarity in reporting of methods the report has been subjected to internal peer review from a qualitative researcher, mixed methodologist, modellers not involved in the interviews and a systematic reviewer.

Attention was paid in the research design to ensure as far as possible that the research methods and particular perspectives of the researchers did not bias the results. Three individuals were involved in the interviews to avoid bias associated with having a single interviewer. Ten of the twelve interviews were undertaken in pairs, with one lead interviewer and one interviewer responsible for ensuring complete coverage of the questioning and capturing of

loose ends raised during the interview. To ensure congruence between the lead interviewers the interview pilots and the first interview were undertaken by the two lead interviewers together. Some of the interviewees were known personally by the interviewers, this may have affected the nature of some of the responses. These effects may have been positive in aiding an open communication during the interviews, and negative in leading to interviewers assuming the nature of some of the interview responses.

Regarding the definition of errors, the preliminary questioning about the full breadth of the modelling process may have biased interviewees towards a definition that encompassed all stages of the process. An alternative method of questioning would have been to commence discussion on technical errors of implementation and let interviewees explore the definition outward until they defined their own boundary. However, this approach may equally have led to an artificially constrained definition of errors.

In the analysis the importance ascribed to individual accounts is only partially related to their frequency. The importance ascribed to comments is at least potentially subject to bias arising from the prior beliefs and prejudices of the analysts, this was guarded against by attention to deviant cases and by undertaking elements of the analyses in teams.

The interview sample was designed to ensure fair representation of organisational perspectives in the analysis. After the first draft of the descriptive interview analysis, the authors checked the distribution of comments used in the report across the different interviewees. Although there was some variation in the number of comments ascribed to different individuals, all interviewees were represented in the analysis. No rebalancing of the report was therefore deemed necessary.

With regard to the review of literature on error classifications, one potential weakness was that the literature focused on spreadsheet and program development rather than specifically on model development. It is assumed that this evidence

is transferable to the modelling domain. It should also be noted that the identified literature concerning programming errors was very narrow despite broad searches being undertaken.

The interviews focused on the process of modelling for HTA decision-making; however, much of this discussion focuses on generic modelling issues and one strength of this work is the potential validity outside its immediate domain. Perhaps the main impact of the focus on HTA modelling is in the area of understanding the decision problem. The HTA modellers interviewed moved directly to a discussion of the PICO definition of a problem scope. When moving to different decision-making areas, this process of defining the problem would have to be more broadly based with an initial focus on defining the key characteristics of the decision problem. Thereafter, much of the modelling discussion appears to be generic.

## **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 philosophical approach is generally taken across the HTA modelling community. Due to the nature of the interview process it cannot be assumed that the absence of an explicit discussion of such a theoretical underpinning means that individuals do not use such a framework in informing their modelling activities.

Although the interview discussions all implied this common pragmatic outlook, there was no common language used in the discussion of modelling errors and furthermore there was a somewhat inconsistent approach to the boundaries of what constitutes an error. For instance, when asked explicitly about the definition of model error, there was a tendency for participants to exclude matters of judgement from being errors and focus on what have been termed 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, the 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.

The original focus of this research, and for consistency the terminology throughout this report, has been in terms of errors in models. However, in the light of the previous discussion of error boundaries and when considering methods of improving modelling practice and processes and examining skills and training requirements, it may be more useful to refer to modelling risks rather than the more black and white term modelling errors.

During the interviews, a number of respondents discussed the concepts of validation and verification. Where validation and verification were discussed, there was a consistency among the participants in their interpretation. Verification was taken to mean the process of ensuring that the computer model correctly implemented the intended model, that is the implementation of the model is free of errors. Validation meant the process of ensuring that a model is fit for purpose, some interviewees explicitly noting that this concept might subsume verification. There was some discussion that related to the concept of credibility; however, this was not developed in any formal sense. As a consequence, issues of validation and verification are central to the discussion of errors. Sargent<sup>41</sup> provides a definition of overall validity, encapsulated within *Figure 2*, as being comprised of conceptual model validity, verification of the computer model, and operational validity of the use of the model in addressing the real world problem. The definitions of verification and validation provided by the interviewees are compatible with these definitions. These definitions are therefore recommended as the basis for further discussion of these topics.

The qualitative analysis highlights considerable variation in modelling practice across the HTA modelling community. This may be in part explained by the apparent variations in expertise, background and experience between the respondents. One particular area of variation concerned the explicit demonstration of conceptual modelling.

Although there was consensus surrounding the definition of verification, to make this a measurable and useful concept there needs to be an explicit and complete description (or design specification) of the intended or conceptual model. A common theme in the modelling process discussion was the absence of a fully transparent description of the intended model (indeed for some interviewees this

barely exists at all). So even the distinction between verification and validation breaks down in any practically useful sense.

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. This literature supports the proposition 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.<sup>35</sup> 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

The interviewees collectively demonstrated examples of all the major error types identified in the literature on errors in end-user developer spreadsheet systems. Taken together, the interview classification of modelling errors provides a basis for developing our understanding of errors and facilitating discussion of errors. The literature detailing classifications of spreadsheet errors is based around the work of two key authors, Panko and Rajalingham. Six different versions of classifications have been identified, each of which struggle to present classifications that are complete and mutually exclusive while being sufficiently rich to be useful, the HTA error classifications similarly struggle with these issues.

Ko explicitly identifies complex chains or networks of errors and faults that can lead to a failure. This layer of complexity is not explicitly discussed either in the spreadsheet error literature or in the HTA modelling interviews. Further development of model errors and risks should reflect this complexity in the relationship between errors in models and failures in decision-making. Existing research on the cognitive basis of human error should be brought into the examination of modelling errors.

To take forward any strategy for reducing errors and improving the robustness of decision support,

it is likely that further development of this classification of errors will be required, specifically with regard to considering methods for improving the retrospective identification of errors. This is because the taxonomy does not seek to classify errors by symptoms.

The error category termed by Panko as 'context errors' hides a multitude of different error causes. For example, this category contains errors associated with breakdowns in domain/disease knowledge, modelling skills/knowledge, programming skills/knowledge, mathematical logic and statistical methods knowledge. All of these represent areas of judgement where the identification and definition of error are problematic and may include at their most extreme, violations of methods and processes. This area requires specific attention when moving towards strategies to avoid errors in HTA modelling.

## 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 degree of overlap between methods used to identify errors and methods used to avoid errors in models; the distinction between avoidance and identification methods is determined by the point at which the strategy is applied. These 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 model development, 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, engaging with the client or decision-maker to ensure mutual understanding, developing 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 training in relevant disciplines (e.g. health economics, statistics, systematic reviewing), as well as ensuring appropriate skills in model programming, model scrutiny and general problem structuring. Evidently, the range of strategies and approaches for avoiding model errors is considerably broader than housekeeping techniques alone; only one-fifth of the strategies target the implementation and operation of the model. The remaining four-fifths target the conceptual validity of the model. Very little emphasis is placed on how model results are conveyed to users.

Clarity and mutual understanding, specifically in the definition of the decision problem and conceptual modelling, are identified as key issues. Currently adopted strategies supporting these aspects of model development are expressed as process requirements, e.g. 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 example, sketching out clinical pathways, their use appears to be partial, and the extent of their use appears to vary considerably between individual modellers. Hence, while there is 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

Current 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. Indeed, the majority of methods do not appear to map directly to specific error types in the taxonomy, but rather they may be used to identify *symptoms* of errors, the true source of which may relate to any point in the taxonomy of model errors. For example, the identification of

a mismatch between actual model results and the derivation of prior expectations from a previous model may be suggestive of the presence of an error, yet its 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. In this sense, the methods outlined in *Table 8* 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

There has been remarkably little development in the theoretical underpinning of model verification and validation since a *European Journal of Operational Research* special issue<sup>31</sup> focused on the topic in 1993. It was then noted that most modellers instinctively took a pragmatic approach to developing model credibility, operating in the middle ground between objectivism and relativism. This description accurately portrays the current position in HTA modelling. 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) the processes for making evidence-based policy and guidance. This research would inform how modelling is best used in the decision-making process and how credibility, the warranty for action, is best developed.

### 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

Despite a large literature on modelling best practice there is little evidence to suggest that models are improving in reliability. Further research is required to define, implement and importantly 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.





## Acknowledgements

Professor Alan Brennan and Dr Alicia O’Cathain commented on the draft report. We would like to thank Mrs Clare Watson, Ms Elizabeth Marsden, Ms Andrea Shippam and Miss Sarah McEvoy for transcribing the interview recordings. Ms Andrea Shippam organised the retrieval of papers and helped in preparing and formatting the report.

The authors also wish to thank Mr Dominic Muston, Mr Steve Beard, Dr Martin Pitt, Professor Adrian Bagust, Dr Matt Stevenson, Dr Alejandra Duenas, Dr Jeremy Jones, Mr Andrew Davies, Mr Pelham Barton and Mr Adam Lloyd for partaking

in the interviews. Two interviewees remained anonymous.

### Contributions of authors

Jim Chilcott led the project. Jim Chilcott, Paul Tappenden and Suzy Paisley were responsible for the design of the study. Jim Chilcott, Paul Tappenden and Andrew Rawdin undertook all in-depth interviews. All authors contributed towards the qualitative synthesis methods and implementation. Diana Papaioannou carried out the literature searches.







## References

1. Buxton M, Drummond M, Van Hout B, Prince R, Sheldon T, Szucs T, *et al.* Modelling in economic evaluation: an unavoidable fact of life. *Health Econ* 1997;**6**:217–27.
2. Cragg PB, King M. Spreadsheet modelling abuse: An opportunity for OR. *J Op Res Soc* 1993;**44**:743–52.
3. Panko RR. What we know about spreadsheet errors. *J End User Comp Special issue on Scaling Up End User Development* 2008;**10**:15–21.
4. Neuhauser D, Lewicki A. What do we gain from the sixth stool guaiac? *N Engl J Med* 1975;**293**:226–8.
5. Brown K, Burrows C. The sixth stool guaiac test. *J Health Econ* 1990;**9**:429–55.
6. Culyer A. *Economics*. Oxford: Blackwell Scientific Publications; 1985.
7. Drummond M, Stoddart G, Torrance G. *Methods for the Economic Evaluation of Health Programmes*. Oxford: Oxford University Press; 1987.
8. Drummond M, Sculpher M, Torrance G, O'Brien B, Stoddart G. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford: Oxford University Press; 2005.
9. Hill S, Mitchell A, Henry D. Problems with the interpretation of pharmacoeconomic analyses: a review of submissions to the Australian Pharmaceuticals Benefit Scheme. *J Am Med Assoc* 2000;**283**:2116–21.
10. Li J, Harris A, Chin G. Assessing the quality of evidence for decision-making (QED): a new instrument. Abstract 104. 2008. Health Priorities 2008. Conference of the International Society for Priorities in Health Care, Gateshead, UK 28–31 October 2008.
11. Holcombe M, Ipate F. *Correct Systems – Building a Business Process Solution. Applied Computing Series*. Berlin: Springer-Verlag; 1998.
12. 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 since 1997. *J Health Service Res Policy* 2005;**10**:245–50.
13. Gold M, Siegel J, Russell L, Weinstein M, editors. *Cost-effectiveness in Health and Medicine*. Oxford: Oxford University Press; 1996.
14. Canadian Coordinating Office for Health Technology Assessment. *Guidelines for Economic Evaluation of Pharmaceuticals*. 2nd edn. Ottawa, Canada: CCOHTA; 1997.
15. Garrison LJ. The ISPOR Good Practice Modeling Principles – a sensible approach: be transparent, be reasonable. *Value in Health* 2003;**6**(1):6–8.
16. McCabe C, Dixon S. Testing the validity of cost-effectiveness models. *Pharmacoeconomics* 2000;**5**:501–13.
17. Sculpher M, Fenwick E, Claxton C. Assessing quality in decision analytic cost-effectiveness models: a suggested framework and example of application. *Pharmacoeconomics* 2000;**17**:461–77.
18. Panko RR. A rant on the lousy use of science in best practice recommendations for spreadsheet development, testing and inspection. 2007. URL: <http://panko.cba.hawaii.edu>
19. Britten N. Qualitative interviews. In Pope C, Mays N, editors. *Qualitative Research in Health Care*. 3rd edn. Oxford: BMJ Blackwell Publishing; 2006. pp. 12–20.
20. Mason J. *Qualitative Researching*. London, UK: Sage Publications Ltd; 1998.
21. Ritchie J, Lewis J. *Qualitative Research Practice. A Guide for Social Science Students and Researchers*. London, UK: Sage Publications Ltd; 2003.
22. Black N, Brazier J, Fitzpatrick R, Reeves B. *Health services research methods: a guide to best practices*. London, UK: BMJ Books; 1998. pp. 187–98.
23. Fisher K, Erdelez S, McKechnie L. Theories of information behaviour. ASIST monograph series 2006. Medford, NJ: American Society for Information Science and Technology; 2006.
24. Mays N, Pope C. Quality in qualitative health research. In Pope C, Mays N, editors. *Qualitative Research in Health Care*. 3rd edn. Oxford: BMJ Blackwell Publishing; 2006. pp. 82–101.

25. Rosenhead J, Mingers J. Rational Analysis for a Problematic World Revisited. Problem Structuring Methods for Complexity, Uncertainty and Conflict. 2nd edn. Chichester: John Wiley & Sons Ltd; 2001.
26. Pidd M. *Computer Simulation in Management Science*. Chichester: John Wiley and Sons; 2004.
27. Schlesinger S, Crosbie RE, Gagne RE, Innis GS, Lalwani CS, Loch J, et al. Terminology for model credibility. *Simulation* 1979;**32**:103–4.
28. Naylor TH, Finger JM, McKenney JL, Schrank WE, Holt CC. Verification of computer simulation models. *Management Sci* 1967;**14**:B92–B106.
29. Luis SJ, McLaughlin D. A stochastic approach to model validation. *Adv Water Resour* 1992;**15**:15–32.
30. Kleijnen J. Verification and validation of simulation models. *Eur J Op Res* 1995;**82**:145–62.
31. Special issue on model validation. *Eur J Op Res* 2009;**66**(2).
32. Déry R, Landry M, Banville C. Revisiting the issue of model validation in OR: an epistemological view. *Eur J Op Res* 1993;**66**:168–83.
33. Raitt R. OR and science. *J Op Res Soc* 1974;**20**:835–6.
34. Rorty R. *Philosophy and the Mirror of Nature*. Princeton, NJ: Princeton University Press; 1979.
35. Kleindorfer GB, O'Neill L, Ganesahn R. Validation in simulation, various position in the philosophy of science. *Management Sci* 1998;**44**:1087–99.
36. Gadamer H. *Truth and Method*. 4th edn. New York, NY: Seabury Press; 1975.
37. Bernstein R. *Beyond Objectivism and Relativism: Science, Hermeneutics and Praxis*. Philadelphia, PA: University of Pennsylvania Press; 1983.
38. Howson C, Urbach P. *Scientific Reasoning: The Bayesian Approach*. 2nd edn. Chicago, IL: Open Court Publishing Company; 1996.
39. Sargent RG. An exposition on verification and validation of simulation models. In Gantz DT, Blais GC, Solomon SL, editors. *Proceedings of the 17th Conference on Winter Simulation*. New Jersey, NJ: IEEE Press; 1985; pp. 15–22.
40. Sargent RG. Validation and verification of simulation models. In Farrington PA, Nembhard HB, Sturrock DT, Evans GW, editors. *Proceedings of the 31st Conference on Winter Simulation: Simulation – a bridge to the future*, vol. 1, New Jersey, NJ: IEEE Press; 1999. pp. 39–48.
41. Sargent RG. Validation and verification of simulation models. In Ingall RG, Rossetti MD, Smith TS, Peters BA, editors. *Proceedings of the 36th Conference on Winter Simulation*. New Jersey, NJ: J Wiley & Sons, IEEE Press 2004; pp. 17–28.
42. Oreskes N, Shrader-Frechette K, Belitz K. Verification, validation and confirmation of numerical models in earth sciences. *Science* 1994;**263**(5147):641–6.
43. Kleijnen J. Verification and validation of simulation models. *Eur J Op Res* 1995;**82**:145–62.
44. Chwif L, Shimada LM. A prescriptive technique for V&V of simulation models when no real-life data are available. In Perrone LF, Lawson BG, Liu J, Wieland FP, editors. *Proceedings of the 38th Winter Simulation Conference*. New Jersey, NJ: J Wiley & Sons, IEEE Press; 2006; pp. 911–18.
45. Panko RR, Halverson Jr. RP. Spreadsheets on trial: a survey of research on spreadsheet risks. In *Proceedings of the 29th Hawaii International Conference on System Sciences*. New Jersey, NJ: IEEE Press; 1996;**2**:326.
46. Teo TSH, Tan M. Quantitative and qualitative errors in spreadsheet development. *Proceedings of the 30th Hawaii International Conference on System Sciences*. New Jersey, NJ: IEEE Press; 1997;**3**:149.
47. Rajalingham K, Chadwick DR, Knight B. Classification of spreadsheet errors. *Conference of the European Spreadsheet Risks Interest Group (EuSpRIG)* 2000.
48. Ko AJ, Myers BA. Development and evaluation of a model of programming errors. In *Proceedings of IEEE Symposium on Human-Centric Computing Languages and Environments*, Auckland, New Zealand, 28–31 October 2003. pp. 7–14.
49. Rajalingham K. A revised classification of spreadsheet errors. *Conference of the European Spreadsheet Risks Interest Group (EuSpRIG)* 2005.
50. Walia GS, Carver J, Philip T. Requirement error abstraction and classification: an empirical study. *Proceedings of the 2006 ACM/IEEE International Symposium on Empirical Software Engineering*; 21–22 September 2006.
51. Purser M, Chadwick D. Does an awareness of differing types of spreadsheet errors aid end-users in identifying spreadsheets errors? *Conference of the European Spreadsheet Risks Interest Group (EuSpRIG)* 2006; pp. 185–204.
52. Panko RR. Revising the Panko–Halverson taxonomy of spreadsheet risks. *42nd Hawaii International Conference on System Sciences*, 2009.

53. Powell SG, Baker KR, Lawson B. A critical review of the literature on spreadsheet errors. *Decision Support Systems* 2008;**46**:128–38.
54. Panko RR. *End User Computing: Management, Applications and Technology*. New York, NY: John Wiley; 1988.
55. Allwood C. Error detection processes in statistical problem solving. *Cogn Sci* 1984;**8**:413–37.
56. Flower L, Hayes J. The dynamics of composing: making plans and juggling constraints. In Greg LW, Steinberg ER. editors. *Cognitive Processes in Writing*. Hillsdale, NJ: Erlbaum; 1980. pp. 31–50.
57. Ko A, Myers B. A framework and methodology for studying the causes of errors in programming systems. *J Visual Lang Comp* 2005;**16**:41–84.
58. Reason J. *Human Error*. Cambridge: Cambridge University Press; 1990.
59. Eddy D. Technology assessment: the role of mathematical modeling. In *Assessing Medical Technologies*. Washington D.C.: National Academy Press; 1985. pp. 144–75.



# Appendix I

## Avoiding and identifying errors in health technology assessment models

### Interview topic guide

Research aims to explore:

- Definitions of model error
- Experience of model errors
- Reasons for/causes of model errors
- Perceived importance of model errors
- Interviewees' approaches to model checking
- Perceived facilitators and barriers to model checking
- Potential ways of reducing errors in the future – what can others learn from their experience

#### 1. Introduction

AIM: TO ENSURE THAT THE INTERVIEWEE IS AWARE OF THE PURPOSE OF THE RESEARCH AND THE NATURE OF THE INTERVIEW.

- Introduce self and School of Health and Related Research (SchARR).
- Explain:
  - nature and purpose of research
  - who the research is for
  - how the interview results will be used
  - interested in views and experiences – no right or wrong answers.
- Stress confidentiality and anonymity.
- Introduce tape recorder.
- Any other issues?

#### 2. Background

AIM: TO OBTAIN PRELIMINARY INFORMATION ABOUT THE INTERVIEWEE BUT ALSO TO GET THEM COMFORTABLE WITH TALKING. QUESTIONS SHOULD NOT BE INTRUSIVE.

- Context and composition of research organisation.
- Their role – how do they fit it?
- What types of modelling activities have they been involved in?
  - experiences
  - case studies
  - methods
  - software.

#### 3. The model development process

AIM: TO ELICIT THE MODEL DEVELOPMENT PROCESS FROM THE PERSPECTIVE OF THE INTERVIEWEE. THE SECOND INTERVIEWER SHOULD MAP VISUALLY THIS USING THE TERMS USED BY THE INTERVIEWEE.

- Elicitation of conceptual map:
  - Where does it start/stop?

- Types of activities at each point?
- Iterations?
- Recap of elicited map by second interviewer.

#### **4. Definition and experience of errors**

AIM: TO ELICIT WHAT THE INTERVIEWEE UNDERSTANDS BY THE TERM ERROR – TO DRAW A BOUNDARY AROUND THE TERM. WHAT ARE THE KEY FEATURES? WITH EXAMPLES. LOOK OUT FOR USE OF ‘VALIDATION’ AND ‘VERIFICATION’.

- How would they define what a model error is?
  - How would they explain what an error is to someone else?
  - What are the features or characteristics of a model error? Intended or not
  - What is it that makes an error an error?

#### **5 Types of model errors**

AIM: TO IDENTIFY DIFFERENT TYPES OF ERRORS AND WHERE AND WHY THEY MAY OCCUR WITHIN THE MODEL DEVELOPMENT PROCESS.

- Thinking back through the model development process you described earlier, what types of model errors might arise?
  - incorrect calculation (logic)
  - incorrect programming
  - poorly scoped problems
  - failure to meet the scope
  - failure to meet the reference case
  - inappropriate assumptions
  - inappropriate methodology
  - problems in evidence review
  - errors in reporting/poor reporting.
- For each type, ask:
  - What they mean by that type of error?
  - Why did they arise? Were they avoidable?
  - When and how were they discovered?
  - What was done about them? Why?
  - What impact did they have?
  - Did they influence your future practice?
  - What are the most important types of error; reasons.

#### **6. Current approaches to avoiding errors**

AIM: TO EXPLORE INTERVIEWEE’S CURRENT APPROACHES TO AVOIDING ERRORS.

[Interviewer to define ‘checking’ to avoid framing assumptions – didn’t want to use the word validation or verification. ‘Any activity that you embark on to ensure that the model is free from errors’]

- You’ve talked about  $x$  errors and how they were discovered – I’d like to explore that a little more systematically. What types of activities do you embark on to ensure that the model is free from errors?
  - conceptual model validation, e.g. with experts
  - computerised model validation
  - good modelling practice – e.g. common approach to tabulation, including all parameters on same sheet, not using values in formulae
  - data validation
  - external validation against other non-input data
  - reference to other models.

- At what point do you embark on these checking activities?
  - Throughout the model development process?
  - At specific stages? Which ones? By who?
  - At the end of model development?
  - When do you start/stop?
- Why do you undertake this checking?
  - confidence in results
  - does it matter – impact on decision
  - requirement of commissioner
  - avoid embarrassment later.
- If you find and correct an error in your model, does that make you more or less confident that the model is now error-free?

### **7. Future initiatives in model checking**

- You've talked about what you currently do, what might you do in terms of checking models? Are there any other initiatives that you've not covered?
  - guidelines
  - double-programming.
- What do you think are the pros and cons of *x*?
  - all models are different
  - use of different software packages.

### **8. Barriers and facilitators to model checking**

AIMS: TO EXPLORE INTERVIEWEE'S PERCEPTION OF BARRIERS AND FACILITATORS TO MODEL CHECKING.

- What factors affect the amount of testing and validation you are able to perform?
  - time constraints
  - knowledge of formal testing and validation procedures
  - idiosyncrasies of model/decision problem
  - lack of framework/guidelines on checking
  - lack of training/skills
  - what's possible
  - employing the right people.

### **9. Interview close**

- Are there any other issues regarding errors in HTA models that we haven't talked about that you would like to raise but haven't had the opportunity?
- Offer report prior to submission to ensure no quotes taken out of context.
- Anticipated timelines.
- Thanks.





## Appendix 2

### Search strategies

#### Scoping search

Database	Search query	Results
Computer and Information Systems Abstracts	{[DE <sup>a</sup> = ('modelling')] or [TI <sup>b</sup> = (model or models)]} and (DE= 'errors') and [TI= (error <sup>c</sup> *)]	115
Water Resources Abstracts	(TI= error*) and [(DE= 'errors') and (DE= 'model studies')]	49
Water Resources Abstracts	(Au <sup>d</sup> = Bevan)	160

- a DE, database index term.  
 b TI, title field.  
 c \*, truncation.  
 d Au, author field.

#### Search strategies to locate error avoidance and identification – February 2008

##### Water Resources Abstracts via CSA

	Results	Selected
1. (DE <sup>a</sup> = 'model studies') and (DE= 'model testing') and (DE= 'errors')	96	19
2. AB <sup>b</sup> = ( <b>model structure error</b> )	5	5
3. AB= ( <b>model structural error</b> )	7	5 (plus one included in step 2)
4. TI <sup>c</sup> = ( <b>GLUE<sup>d</sup> methodology</b> )	8	5 (2 previously included)
5. AB= (error <sup>e</sup> * in modelling)	3	1
6. AB= (error* in modelling)	2	0
7. TI= ( <b>model error*</b> )	9	3 (2 previously marked)
8. TI= <b>error*</b> and AB= ( <b>artificial neural network*</b> )	3	1 (1 previously marked)
9. TI= <b>error*</b> and AB= ( <b>artificial neural network*</b> )	2	1
10. TI=(error analysis)	56	13
11. TI=(model testing)	25	7
12. TI=(model verification)	26	8
13. refsgaard.au <sup>f</sup>	46	7

- a DE, database index term.  
 b AB, abstract field.  
 c TI, title field.  
 d GLUE, Generalized Likelihood Uncertainty Estimation.  
 e \*, truncation.  
 f au, author field.

**ACM Digital Library**

	Results	Selected
1. TI <sup>a</sup> =model testing	72	13
2. TI=model verification	90	5
3. TI=model error	49	0
a TI, title field		

**Computer and information systems abstracts via CSA**

	Results	Selected
1. DE <sup>a</sup> =model studies	32	2
2. TI <sup>b</sup> =model error <sup>*c</sup>	26	15
3. TI= <b>error</b> * and TI=( <b>modelling</b> or <b>modelling</b> )	173	0 scanned first 50
4. TI=( <b>error analysis</b> ) and TI= <b>model</b> *	31	20
5. TI=(model testing)	13	8
6. TI=(model verification)	27	10
a DE, database index term.		
b TI, title field.		
c *, truncation.		

**IEEE/IET Electronic Library (IEL)**

	Results	Selected
1. Model testing	58	0
2. {[[(model error <sup>*a</sup> )<in>ti <sup>b</sup> ] <and> [(assess or verif* or assurance or valid*)<in>ab <sup>c</sup> ]}	17	14
3. {[[(model error <sup>*</sup> )<in>ti] <and> [(identif* or analys <sup>*</sup> )<in>ab]}	41	19
4. (model verification).ti	46	7
a *, truncation.		
b ti, title field.		
c ab, abstract field.		

**ANTE: abstracts in new technologies and engineering via CSA**

	Results	Selected
1. Model verification.ti. <sup>a</sup>	10	2
2. Model validation.ti.	29	1
3. Model <sup>*b</sup> error*.ti.	5	5
4. model* error*.ab. <sup>c</sup>	109	0
a ti, title field.		
b *, truncation.		
c ab, abstract field.		

**Google Scholar**

	Results	Selected
1. Spreadsheet error* <sup>a</sup>	87	28
a *, truncation.		

**Web of Knowledge**

	Results	Selected
1. Refsgaard.au. <sup>a</sup>	21	5
2. (Model verification and error* <sup>b</sup> ).ti. <sup>c</sup>	9	4
3. model error correction.ti.	129	17
4. (model validation and error*).ti.	21	6
5. (model error* and uncertainty).ti.	23	7
6. GLUE methodology.ti.	9	2
7. (model testing and error*).ti.	43	2
a au, author field. b *, truncation. c ti, title field.		

**Search strategies to locate error classification systems – January 2009****Web of Knowledge**

Topic=(program\*<sup>a</sup> or spreadsheet\*) AND Topic=(error\*) AND Topic=(taxonom\*):62

**Abstracts in new technologies and engineering (ANTE)**

(program\* or spreadsheet\*) and error\* and taxonom\*): 10

**Computer and Information Systems Abstracts**

(program\* or spreadsheet\*) and error\* and taxonom\*: 68 results

**IEEE/IET Electronic Library**

program\* or spreadsheet\* in ti<sup>b</sup> and (error\*) in all fields and [(taxonom\* or classif\*)<in>all fields]: 40

**Google Scholar**

Spreadsheet error taxonom\* (JC sifted results and important relevant citations directly): 4630 results

Note: a \*, truncation; b ti, title field.

**Citation searches February 2009 undertaken on Google Scholar and Web of Knowledge**

Janvrin D, Morrison J. Using a structured design approach to reduce risks in end user spreadsheet development. *Inform Manage* 2000;**37**(1):1–12.

Panko R. What we know about spreadsheet errors. *J End User Comp* 1998;**10**(2):15–21.

Powell SG, Baker KR, Lawson B. A critical review of the literature on spreadsheet errors. *Decision Support Systems* 2008;**46**(1):128–38.

Purser M, Chadwick D. Does an awareness of differing types of spreadsheet errors aid end-users in identifying spreadsheets errors? *Proceedings of European Spreadsheet Risks International Group (EuSpRIG)* 2006; pp. 185–204.

Rajalingham K, Chadwick DR, Knight B. 2001, Classification of Spreadsheet Errors. *Proceedings of European Spreadsheet Risks International Group (EuSpRIG)* 2001.

Web of Knowledge: 0 results

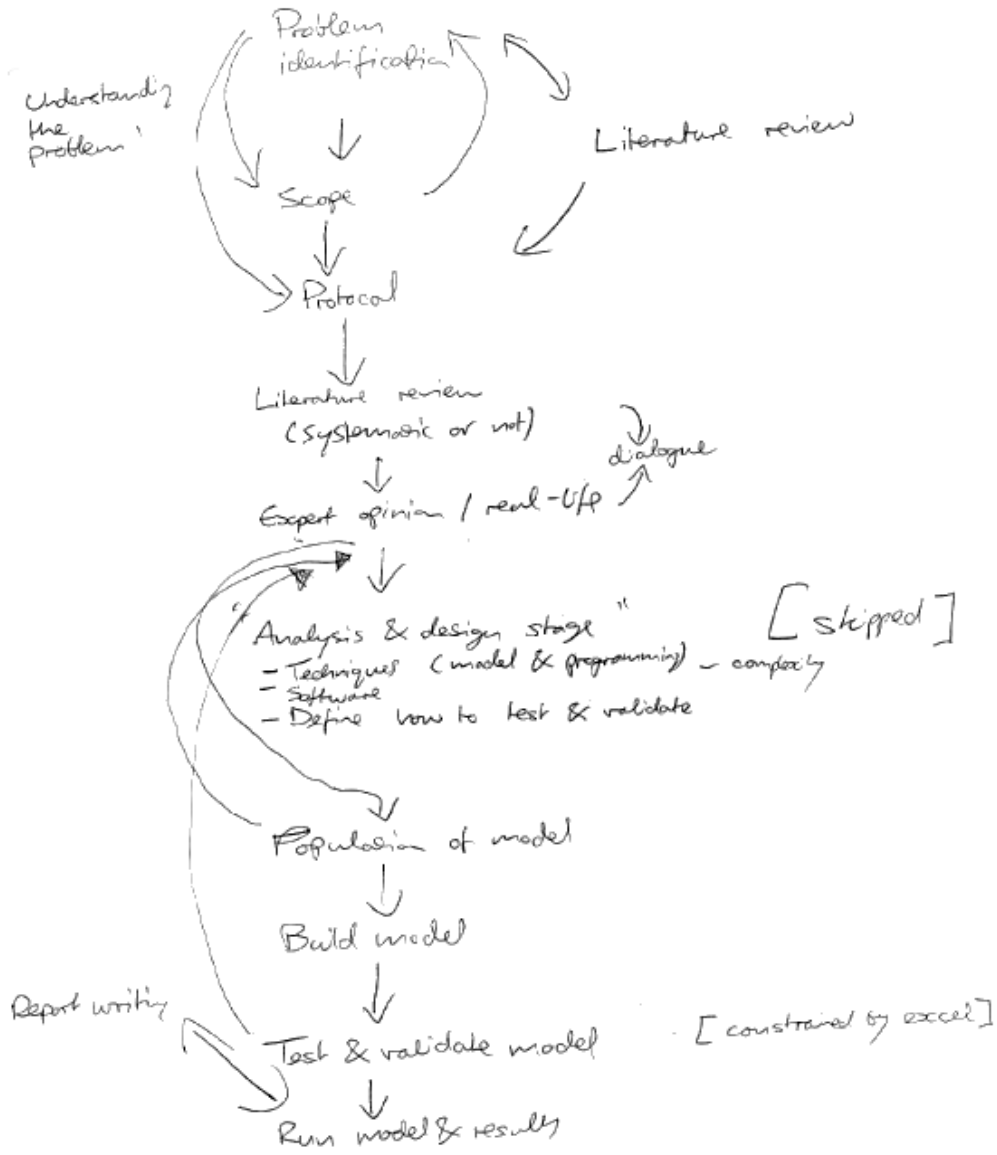
Google Scholar: JC sifted results and important relevant citations directly



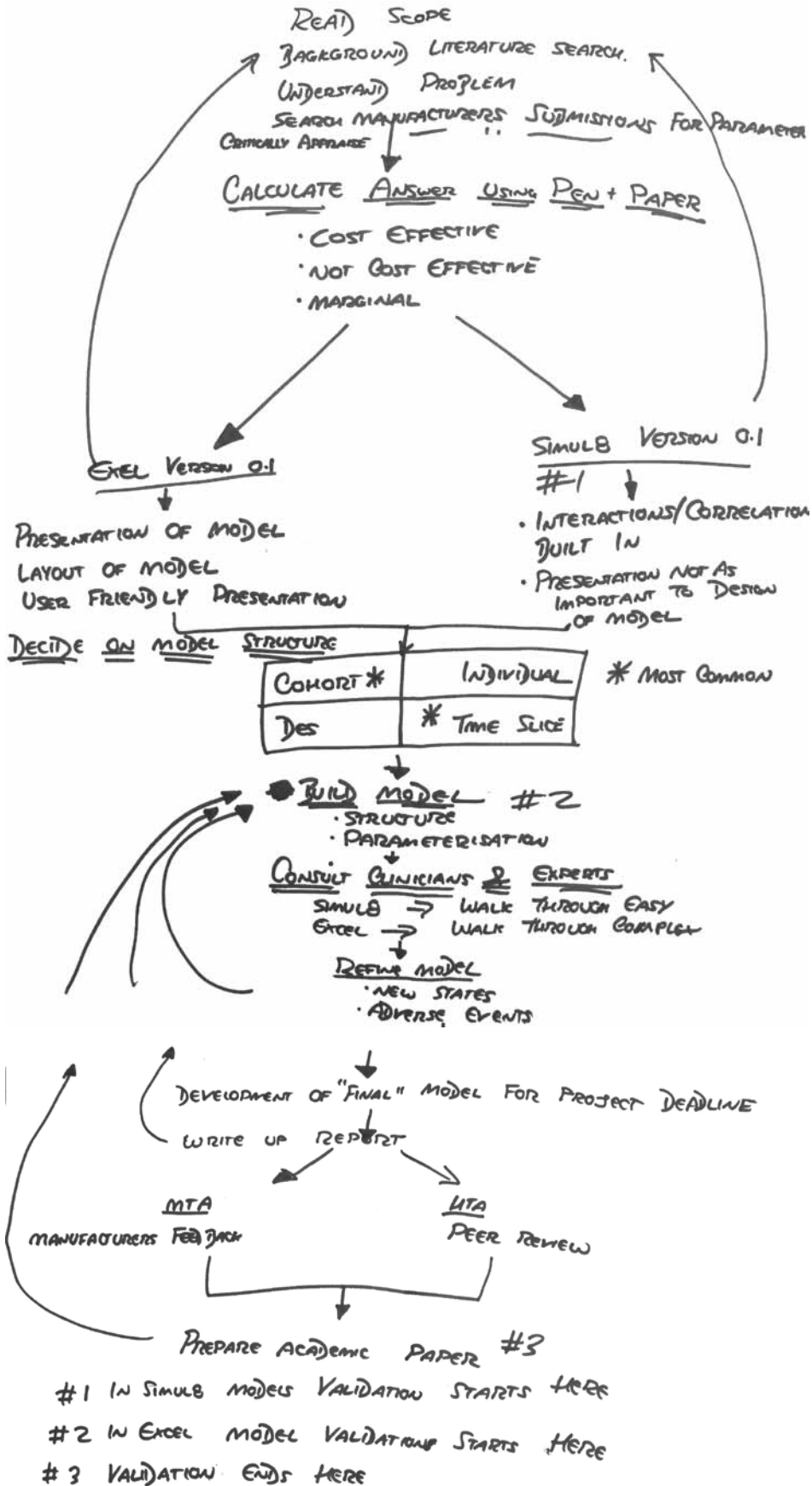
# Appendix 3

## Charts

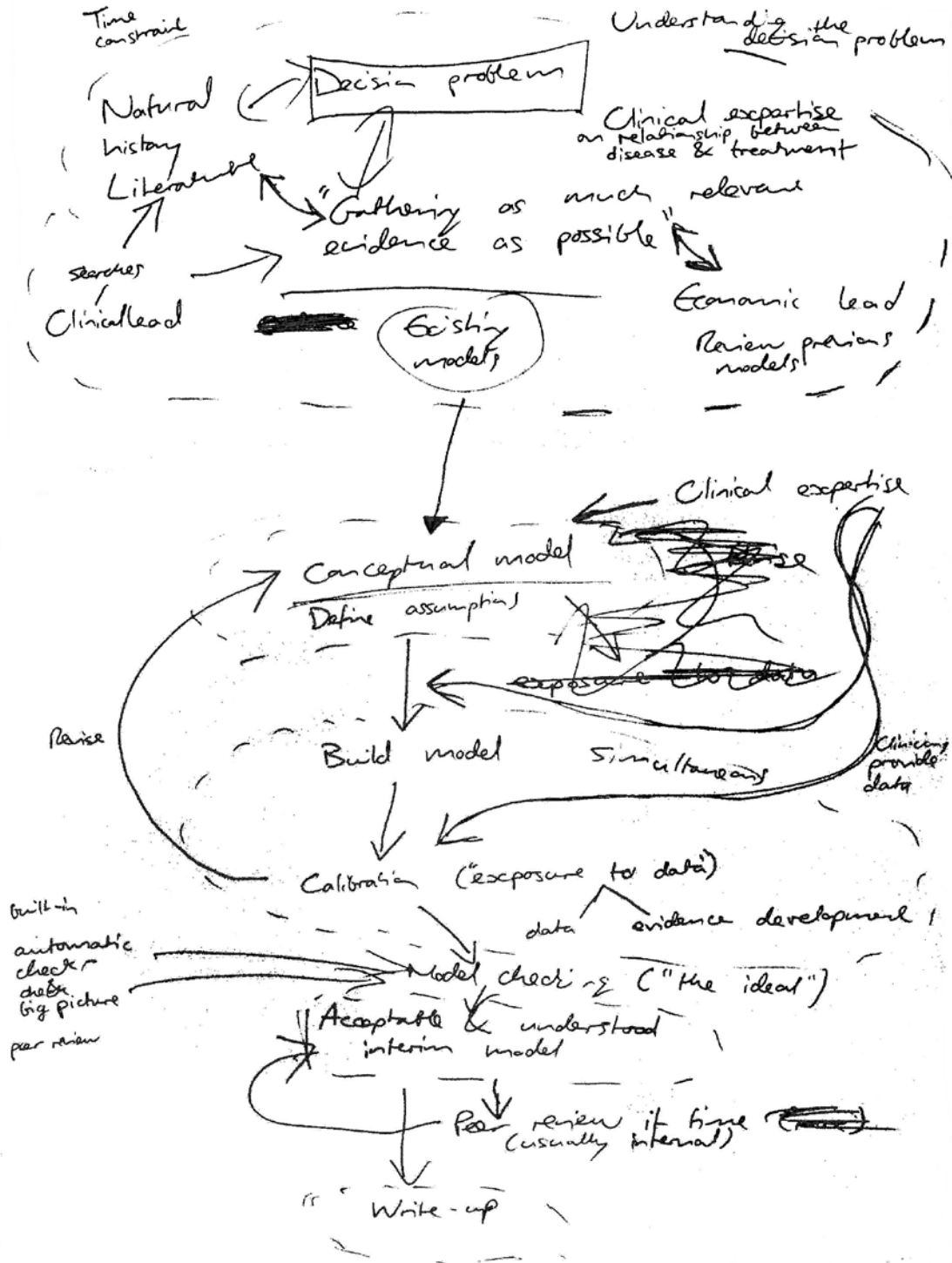
II



12

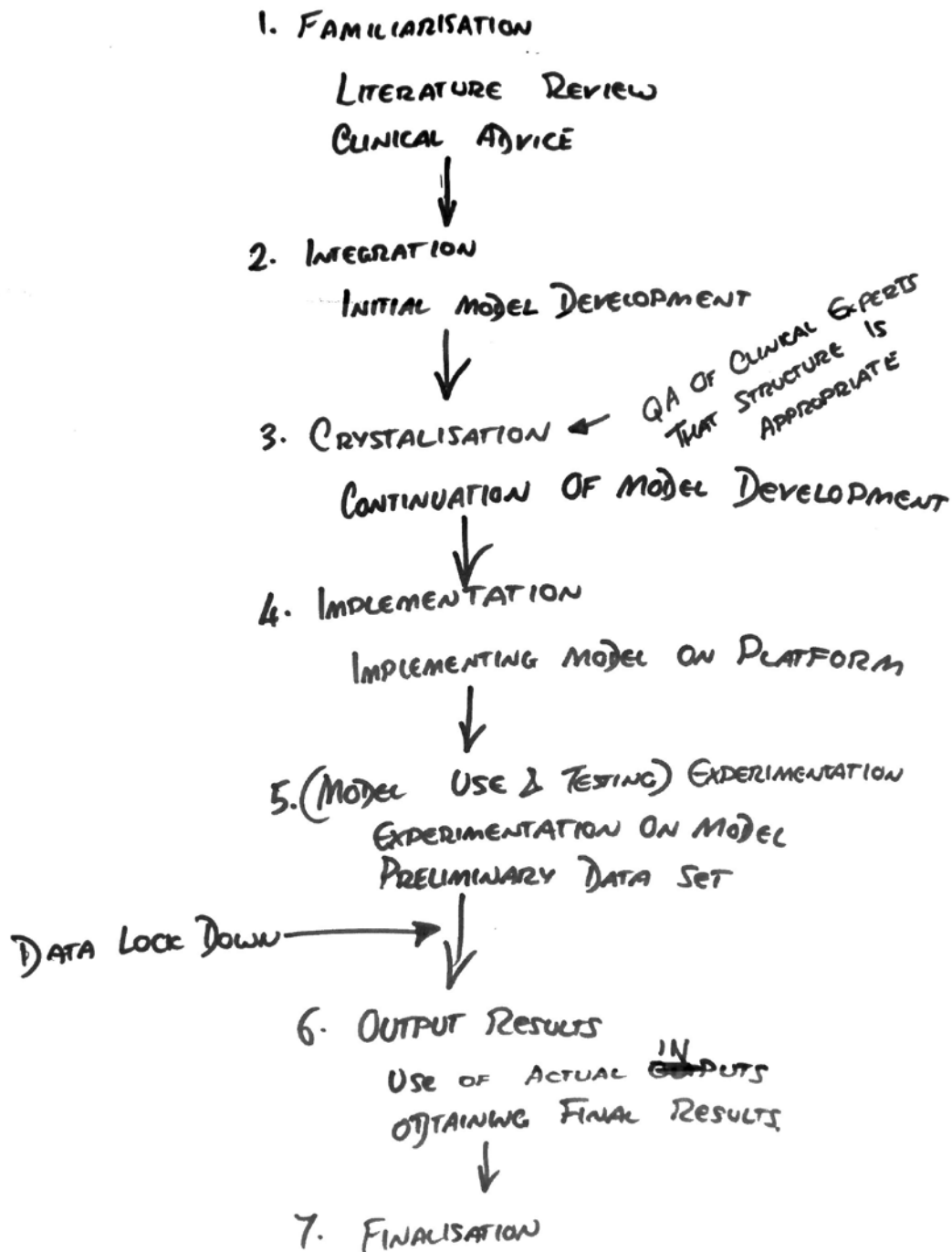


I3

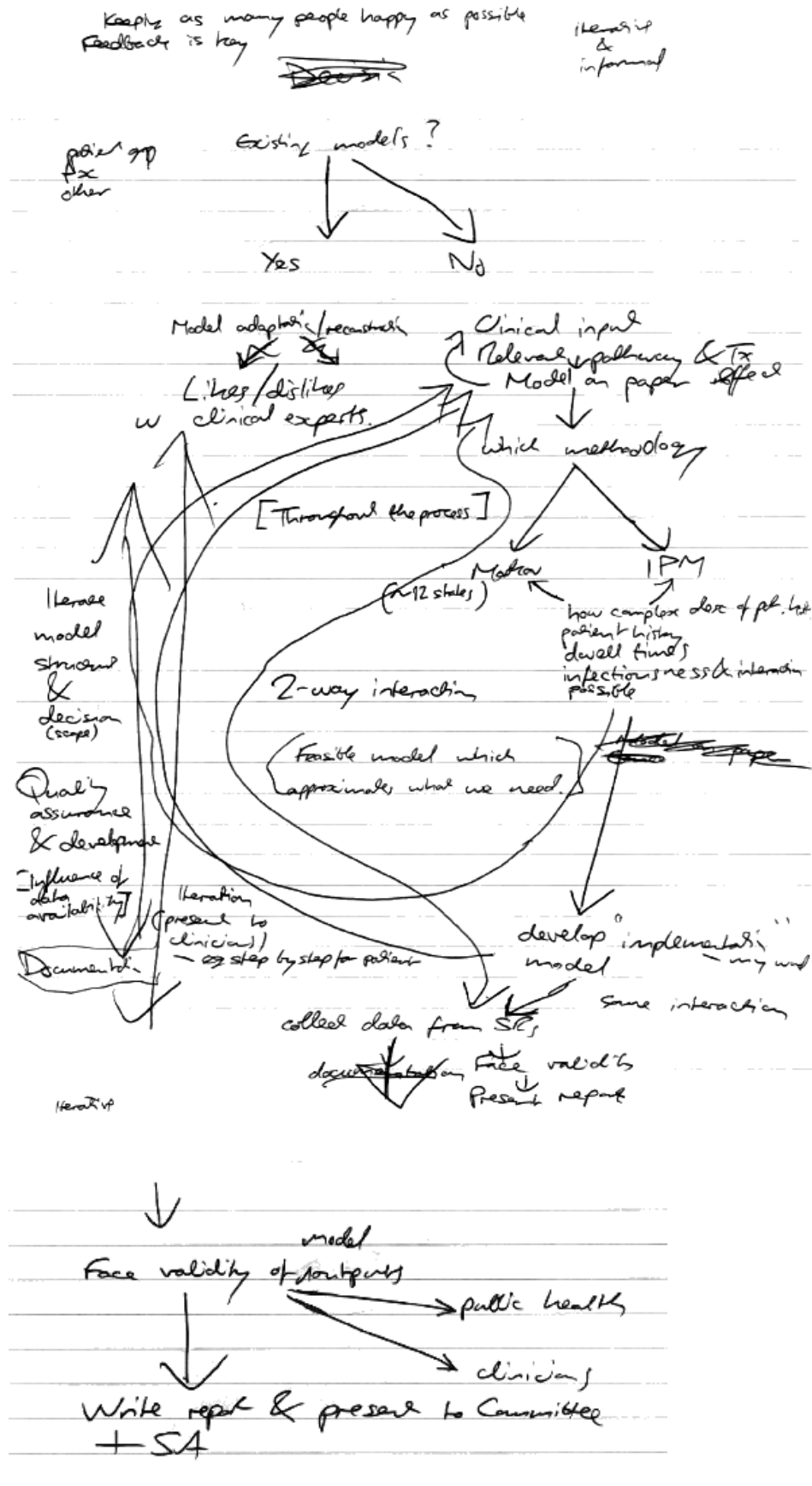


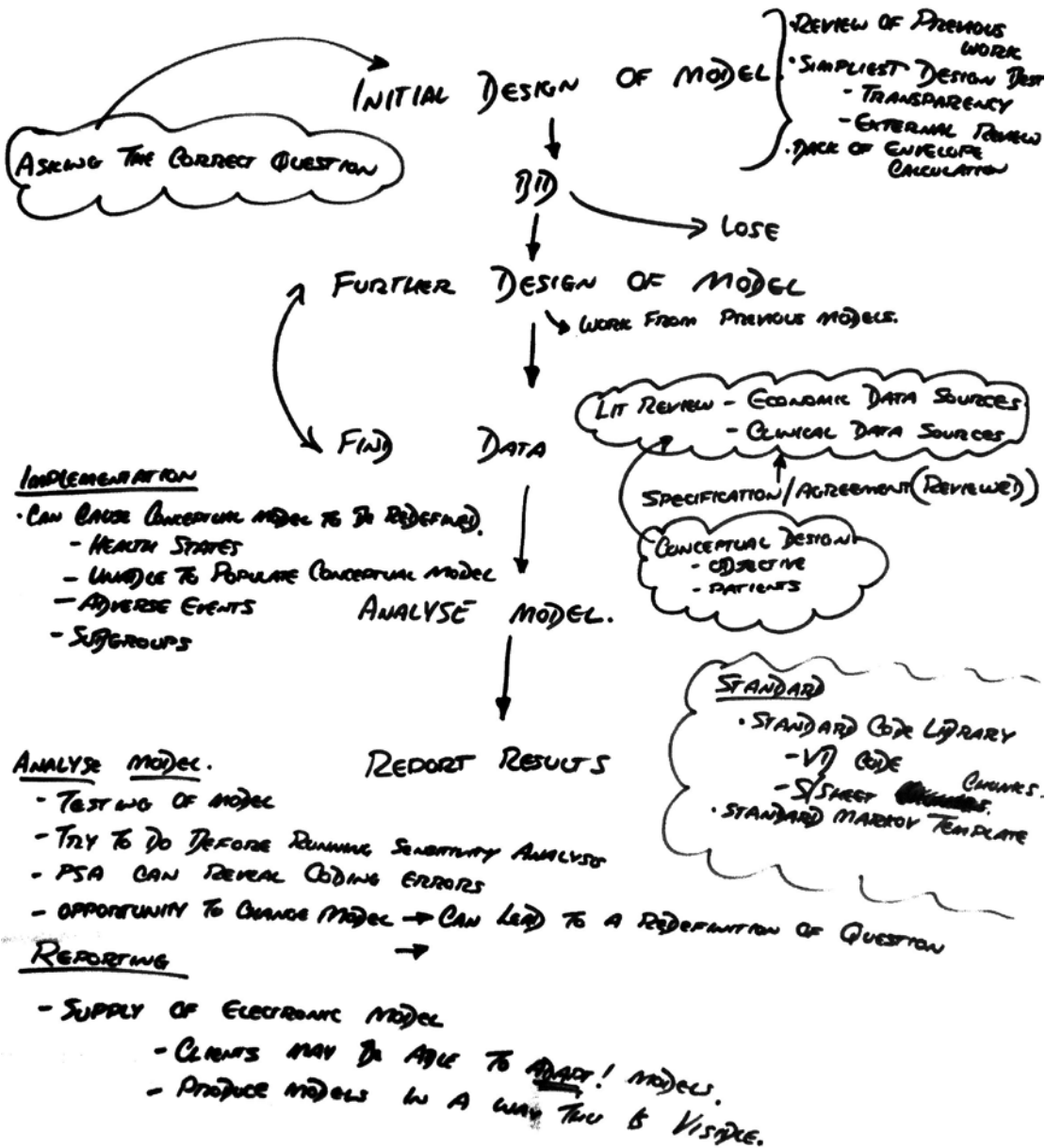


14

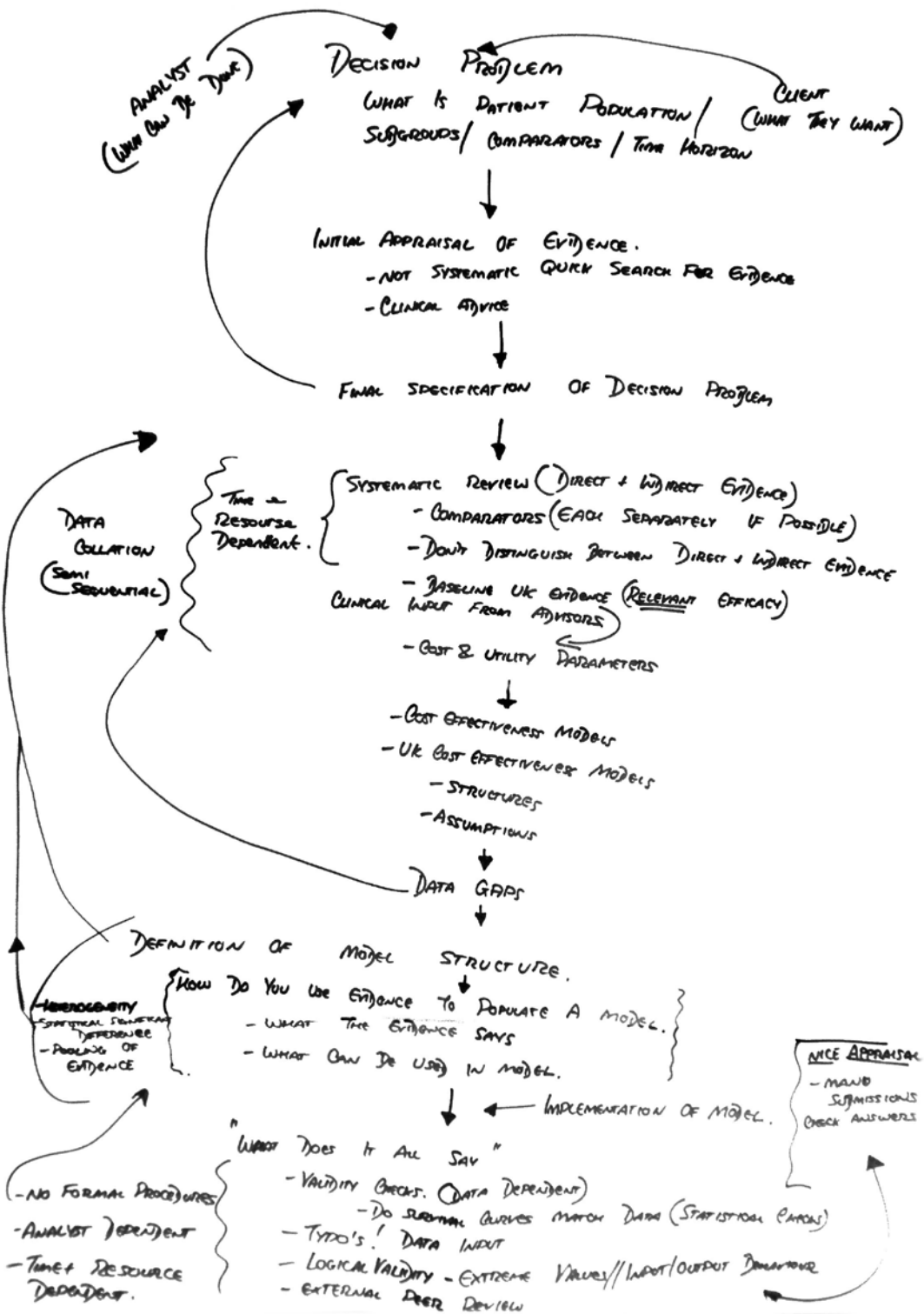


15

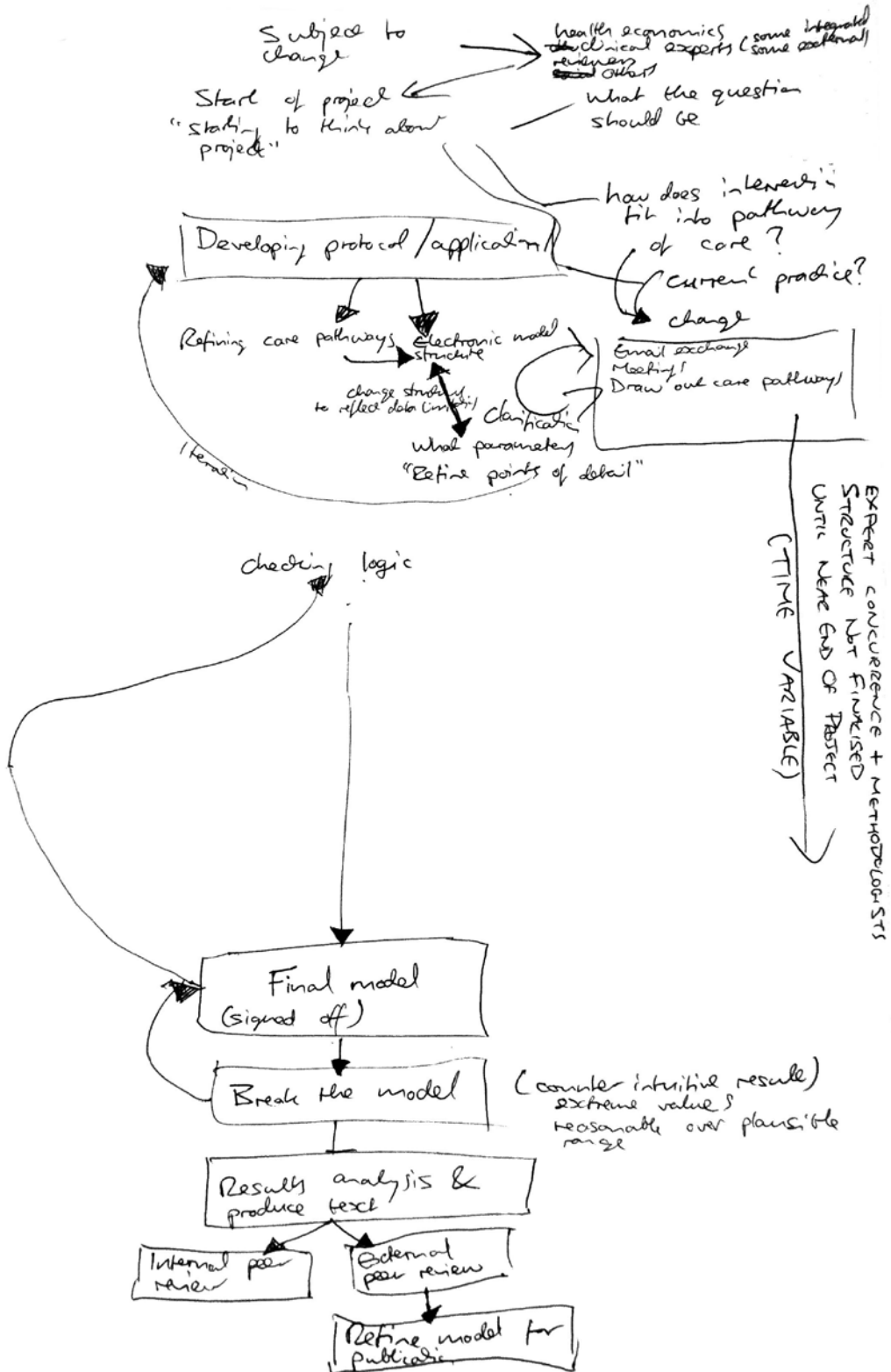




17



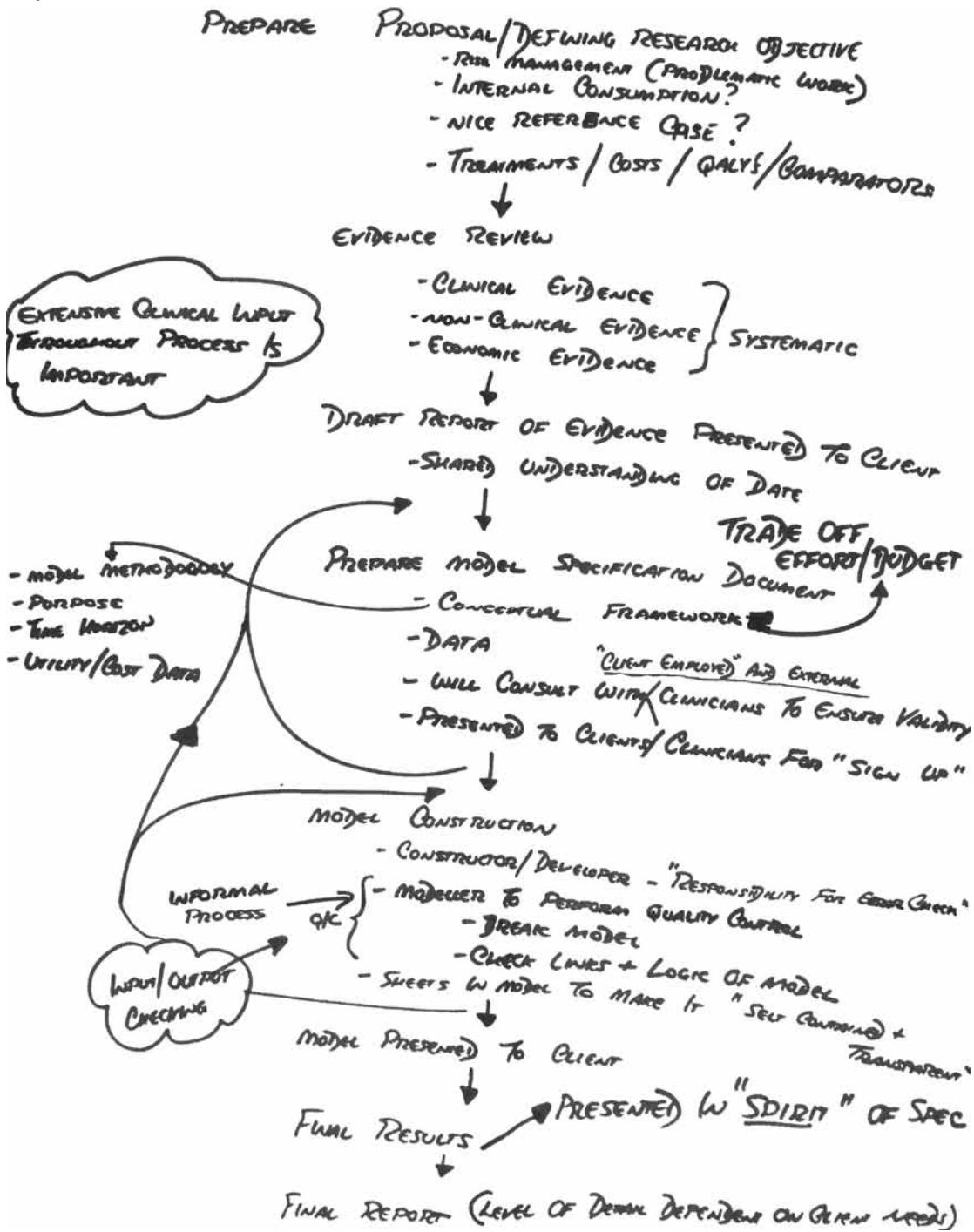
18

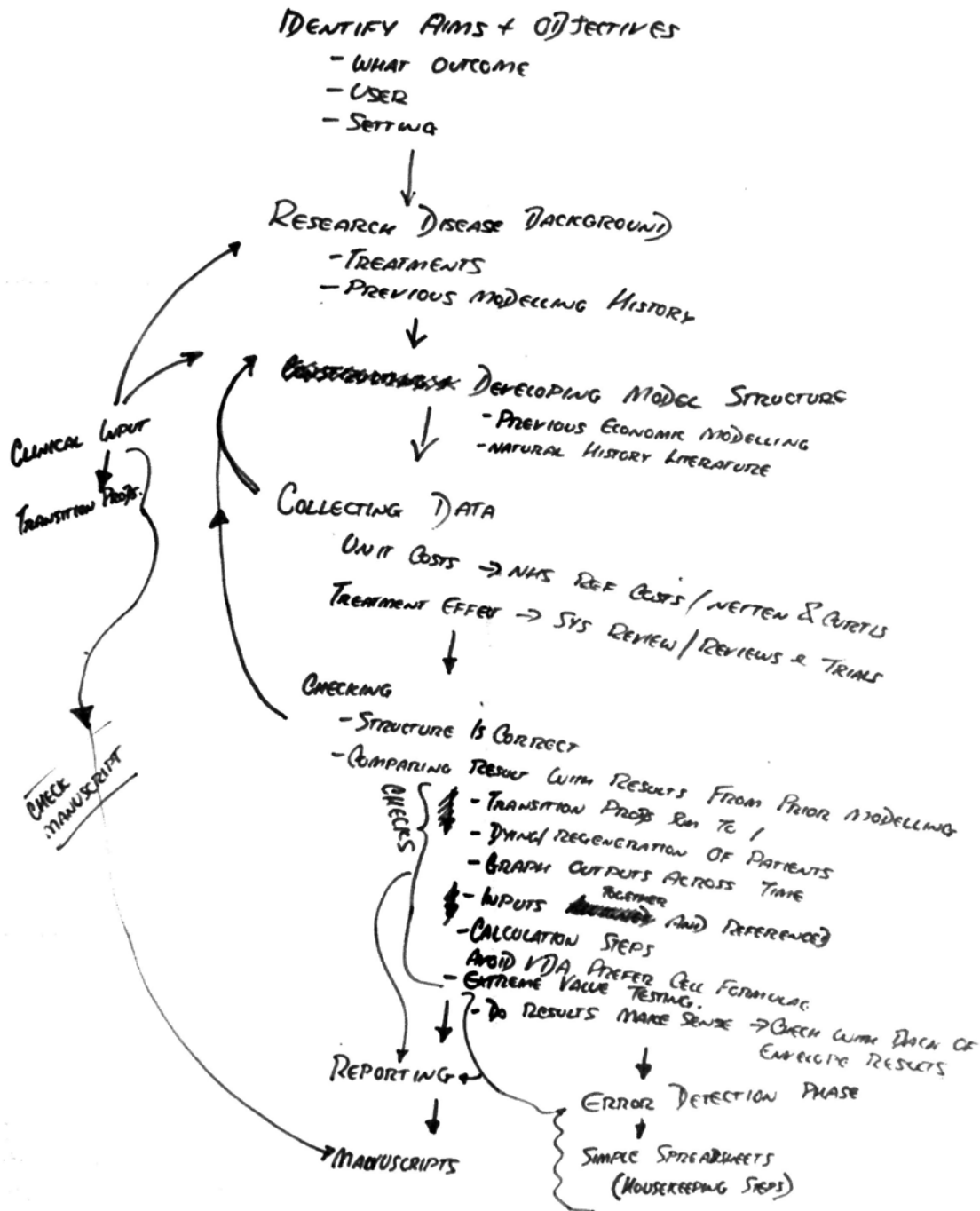


19



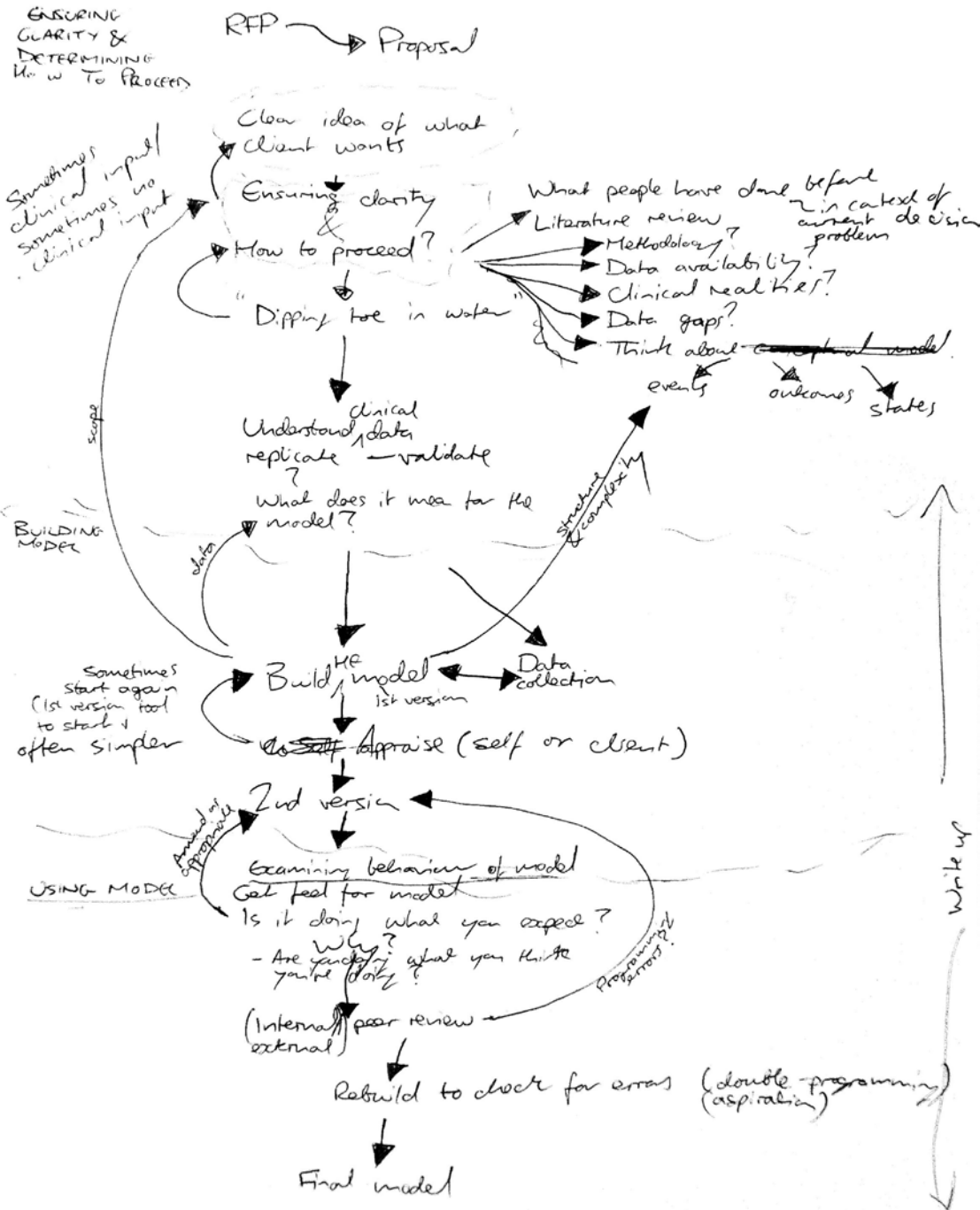
110







112





# Health Technology Assessment reports published to date

## Volume 1, 1997

### No. 1

Home parenteral nutrition: a systematic review.

By Richards DM, Deeks JJ, Sheldon TA, Shaffer JL.

### No. 2

Diagnosis, management and screening of early localised prostate cancer.

A review by Selley S, Donovan J, Faulkner A, Coast J, Gillatt D.

### No. 3

The diagnosis, management, treatment and costs of prostate cancer in England and Wales.

A review by Chamberlain J, Melia J, Moss S, Brown J.

### No. 4

Screening for fragile X syndrome.

A review by Murray J, Cuckle H, Taylor G, Hewison J.

### No. 5

A review of near patient testing in primary care.

By Hobbs FDR, Delaney BC, Fitzmaurice DA, Wilson S, Hyde CJ, Thorpe GH, *et al.*

### No. 6

Systematic review of outpatient services for chronic pain control.

By McQuay HJ, Moore RA, Eccleston C, Morley S, de C Williams AC.

### No. 7

Neonatal screening for inborn errors of metabolism: cost, yield and outcome.

A review by Pollitt RJ, Green A, McCabe CJ, Booth A, Cooper NJ, Leonard JV, *et al.*

### No. 8

Preschool vision screening.

A review by Snowdon SK, Stewart-Brown SL.

### No. 9

Implications of socio-cultural contexts for the ethics of clinical trials.

A review by Ashcroft RE, Chadwick DW, Clark SRL, Edwards RHT, Frith L, Hutton JL.

### No. 10

A critical review of the role of neonatal hearing screening in the detection of congenital hearing impairment.

By Davis A, Bamford J, Wilson I, Ramkalawan T, Forshaw M, Wright S.

### No. 11

Newborn screening for inborn errors of metabolism: a systematic review.

By Seymour CA, Thomason MJ, Chalmers RA, Addison GM, Bain MD, Cockburn F, *et al.*

### No. 12

Routine preoperative testing: a systematic review of the evidence.

By Munro J, Booth A, Nicholl J.

### No. 13

Systematic review of the effectiveness of laxatives in the elderly.

By Petticrew M, Watt I, Sheldon T.

### No. 14

When and how to assess fast-changing technologies: a comparative study of medical applications of four generic technologies.

A review by Mowatt G, Bower DJ, Brebner JA, Cairns JA, Grant AM, McKee L.

## Volume 2, 1998

### No. 1

Antenatal screening for Down's syndrome.

A review by Wald NJ, Kennard A, Hackshaw A, McGuire A.

### No. 2

Screening for ovarian cancer: a systematic review.

By Bell R, Petticrew M, Luengo S, Sheldon TA.

### No. 3

Consensus development methods, and their use in clinical guideline development.

A review by Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CFB, Askham J, *et al.*

### No. 4

A cost-utility analysis of interferon beta for multiple sclerosis.

By Parkin D, McNamee P, Jacoby A, Miller P, Thomas S, Bates D.

### No. 5

Effectiveness and efficiency of methods of dialysis therapy for end-stage renal disease: systematic reviews.

By MacLeod A, Grant A, Donaldson C, Khan I, Campbell M, Daly C, *et al.*

### No. 6

Effectiveness of hip prostheses in primary total hip replacement: a critical review of evidence and an economic model.

By Faulkner A, Kennedy LG, Baxter K, Donovan J, Wilkinson M, Bevan G.

### No. 7

Antimicrobial prophylaxis in colorectal surgery: a systematic review of randomised controlled trials.

By Song F, Glenny AM.

### No. 8

Bone marrow and peripheral blood stem cell transplantation for malignancy.

A review by Johnson PWM, Simnett SJ, Sweetenham JW, Morgan GJ, Stewart LA.

### No. 9

Screening for speech and language delay: a systematic review of the literature.

By Law J, Boyle J, Harris F, Harkness A, Nye C.

### No. 10

Resource allocation for chronic stable angina: a systematic review of effectiveness, costs and cost-effectiveness of alternative interventions.

By Sculpher MJ, Petticrew M, Kelland JL, Elliott RA, Holdright DR, Buxton MJ.

### No. 11

Detection, adherence and control of hypertension for the prevention of stroke: a systematic review.

By Ebrahim S.

### No. 12

Postoperative analgesia and vomiting, with special reference to day-case surgery: a systematic review.

By McQuay HJ, Moore RA.

### No. 13

Choosing between randomised and nonrandomised studies: a systematic review.

By Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C.

**No. 14**

Evaluating patient-based outcome measures for use in clinical trials.  
A review by Fitzpatrick R, Davey C, Buxton MJ, Jones DR.

**No. 15**

Ethical issues in the design and conduct of randomised controlled trials.  
A review by Edwards SJL, Lilford RJ, Braunholtz DA, Jackson JC, Hewison J, Thornton J.

**No. 16**

Qualitative research methods in health technology assessment: a review of the literature.  
By Murphy E, Dingwall R, Greatbatch D, Parker S, Watson P.

**No. 17**

The costs and benefits of paramedic skills in pre-hospital trauma care.  
By Nicholl J, Hughes S, Dixon S, Turner J, Yates D.

**No. 18**

Systematic review of endoscopic ultrasound in gastro-oesophageal cancer.  
By Harris KM, Kelly S, Berry E, Hutton J, Roderick P, Cullingworth J, *et al.*

**No. 19**

Systematic reviews of trials and other studies.  
By Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F.

**No. 20**

Primary total hip replacement surgery: a systematic review of outcomes and modelling of cost-effectiveness associated with different prostheses.  
A review by Fitzpatrick R, Shortall E, Sculpher M, Murray D, Morris R, Lodge M, *et al.*

**Volume 3, 1999**

**No. 1**

Informed decision making: an annotated bibliography and systematic review.  
By Bekker H, Thornton JG, Airey CM, Connelly JB, Hewison J, Robinson MB, *et al.*

**No. 2**

Handling uncertainty when performing economic evaluation of healthcare interventions.  
A review by Briggs AH, Gray AM.

**No. 3**

The role of expectancies in the placebo effect and their use in the delivery of health care: a systematic review.  
By Crow R, Gage H, Hampson S, Hart J, Kimber A, Thomas H.

**No. 4**

A randomised controlled trial of different approaches to universal antenatal HIV testing: uptake and acceptability. Annex: Antenatal HIV testing – assessment of a routine voluntary approach.  
By Simpson WM, Johnstone FD, Boyd FM, Goldberg DJ, Hart GJ, Gormley SM, *et al.*

**No. 5**

Methods for evaluating area-wide and organisation-based interventions in health and health care: a systematic review.  
By Ukoumunne OC, Gulliford MC, Chinn S, Sterne JAC, Burney PGJ.

**No. 6**

Assessing the costs of healthcare technologies in clinical trials.  
A review by Johnston K, Buxton MJ, Jones DR, Fitzpatrick R.

**No. 7**

Cooperatives and their primary care emergency centres: organisation and impact.  
By Hallam L, Henthorne K.

**No. 8**

Screening for cystic fibrosis.  
A review by Murray J, Cuckle H, Taylor G, Littlewood J, Hewison J.

**No. 9**

A review of the use of health status measures in economic evaluation.  
By Brazier J, Deverill M, Green C, Harper R, Booth A.

**No. 10**

Methods for the analysis of quality-of-life and survival data in health technology assessment.  
A review by Billingham LJ, Abrams KR, Jones DR.

**No. 11**

Antenatal and neonatal haemoglobinopathy screening in the UK: review and economic analysis.  
By Zeuner D, Ades AE, Karnon J, Brown J, Dezateux C, Anionwu EN.

**No. 12**

Assessing the quality of reports of randomised trials: implications for the conduct of meta-analyses.  
A review by Moher D, Cook DJ, Jadad AR, Tugwell P, Moher M, Jones A, *et al.*

**No. 13**

'Early warning systems' for identifying new healthcare technologies.  
By Robert G, Stevens A, Gabbay J.

**No. 14**

A systematic review of the role of human papillomavirus testing within a cervical screening programme.  
By Cuzick J, Sasieni P, Davies P, Adams J, Normand C, Frater A, *et al.*

**No. 15**

Near patient testing in diabetes clinics: appraising the costs and outcomes.  
By Grieve R, Beech R, Vincent J, Mazurkiewicz J.

**No. 16**

Positron emission tomography: establishing priorities for health technology assessment.  
A review by Robert G, Milne R.

**No. 17 (Pt 1)**

The debridement of chronic wounds: a systematic review.  
By Bradley M, Cullum N, Sheldon T.

**No. 17 (Pt 2)**

Systematic reviews of wound care management: (2) Dressings and topical agents used in the healing of chronic wounds.  
By Bradley M, Cullum N, Nelson EA, Petticrew M, Sheldon T, Torgerson D.

**No. 18**

A systematic literature review of spiral and electron beam computed tomography: with particular reference to clinical applications in hepatic lesions, pulmonary embolus and coronary artery disease.  
By Berry E, Kelly S, Hutton J, Harris KM, Roderick P, Boyce JC, *et al.*

**No. 19**

What role for statins? A review and economic model.  
By Ebrahim S, Davey Smith G, McCabe C, Payne N, Pickin M, Sheldon TA, *et al.*

**No. 20**

Factors that limit the quality, number and progress of randomised controlled trials.  
A review by Prescott RJ, Counsell CE, Gillespie WJ, Grant AM, Russell IT, Kiauka S, *et al.*

**No. 21**

Antimicrobial prophylaxis in total hip replacement: a systematic review.  
By Glenny AM, Song F.

**No. 22**

Health promoting schools and health promotion in schools: two systematic reviews.  
By Lister-Sharp D, Chapman S, Stewart-Brown S, Sowden A.

**No. 23**

Economic evaluation of a primary care-based education programme for patients with osteoarthritis of the knee.

A review by Lord J, Victor C, Littlejohns P, Ross FM, Axford JS.

**Volume 4, 2000****No. 1**

The estimation of marginal time preference in a UK-wide sample (TEMPUS) project.

A review by Cairns JA, van der Pol MM.

**No. 2**

Geriatric rehabilitation following fractures in older people: a systematic review.

By Cameron I, Crotty M, Currie C, Finnegan T, Gillespie L, Gillespie W, *et al.*

**No. 3**

Screening for sickle cell disease and thalassaemia: a systematic review with supplementary research.

By Davies SC, Cronin E, Gill M, Greengross P, Hickman M, Normand C.

**No. 4**

Community provision of hearing aids and related audiology services.

A review by Reeves DJ, Alborz A, Hickson FS, Bamford JM.

**No. 5**

False-negative results in screening programmes: systematic review of impact and implications.

By Petticrew MP, Sowden AJ, Lister-Sharp D, Wright K.

**No. 6**

Costs and benefits of community postnatal support workers: a randomised controlled trial.

By Morrell CJ, Spiby H, Stewart P, Walters S, Morgan A.

**No. 7**

Implantable contraceptives (subdermal implants and hormonally impregnated intrauterine systems) versus other forms of reversible contraceptives: two systematic reviews to assess relative effectiveness, acceptability, tolerability and cost-effectiveness.

By French RS, Cowan FM, Mansour DJA, Morris S, Procter T, Hughes D, *et al.*

**No. 8**

An introduction to statistical methods for health technology assessment.

A review by White SJ, Ashby D, Brown PJ.

**No. 9**

Disease-modifying drugs for multiple sclerosis: a rapid and systematic review.

By Clegg A, Bryant J, Milne R.

**No. 10**

Publication and related biases.

A review by Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ.

**No. 11**

Cost and outcome implications of the organisation of vascular services.

By Michaels J, Brazier J, Palfreyman S, Shackley P, Slack R.

**No. 12**

Monitoring blood glucose control in diabetes mellitus: a systematic review.

By Coster S, Gulliford MC, Seed PT, Powrie JK, Swaminathan R.

**No. 13**

The effectiveness of domiciliary health visiting: a systematic review of international studies and a selective review of the British literature.

By Elkan R, Kendrick D, Hewitt M, Robinson JJA, Tolley K, Blair M, *et al.*

**No. 14**

The determinants of screening uptake and interventions for increasing uptake: a systematic review.

By Jepson R, Clegg A, Forbes C, Lewis R, Sowden A, Kleijnen J.

**No. 15**

The effectiveness and cost-effectiveness of prophylactic removal of wisdom teeth. A rapid review by Song F, O'Meara S, Wilson P, Golder S, Kleijnen J.

**No. 16**

Ultrasound screening in pregnancy: a systematic review of the clinical effectiveness, cost-effectiveness and women's views.

By Bricker L, Garcia J, Henderson J, Mugford M, Neilson J, Roberts T, *et al.*

**No. 17**

A rapid and systematic review of the effectiveness and cost-effectiveness of the taxanes used in the treatment of advanced breast and ovarian cancer.

By Lister-Sharp D, McDonagh MS, Khan KS, Kleijnen J.

**No. 18**

Liquid-based cytology in cervical screening: a rapid and systematic review.

By Payne N, Chilcott J, McGoogan E.

**No. 19**

Randomised controlled trial of non-directive counselling, cognitive-behaviour therapy and usual general practitioner care in the management of depression as well as mixed anxiety and depression in primary care.

By King M, Sibbald B, Ward E, Bower P, Lloyd M, Gabbay M, *et al.*

**No. 20**

Routine referral for radiography of patients presenting with low back pain: is patients' outcome influenced by GPs' referral for plain radiography?

By Kerry S, Hilton S, Patel S, Dundas D, Rink E, Lord J.

**No. 21**

Systematic reviews of wound care management: (3) antimicrobial agents for chronic wounds; (4) diabetic foot ulceration.

By O'Meara S, Cullum N, Majid M, Sheldon T.

**No. 22**

Using routine data to complement and enhance the results of randomised controlled trials.

By Lewsey JD, Leyland AH, Murray GD, Boddy FA.

**No. 23**

Coronary artery stents in the treatment of ischaemic heart disease: a rapid and systematic review.

By Meads C, Cummins C, Jolly K, Stevens A, Burls A, Hyde C.

**No. 24**

Outcome measures for adult critical care: a systematic review.

By Hayes JA, Black NA, Jenkinson C, Young JD, Rowan KM, Daly K, *et al.*

**No. 25**

A systematic review to evaluate the effectiveness of interventions to promote the initiation of breastfeeding.

By Fairbank L, O'Meara S, Renfrew MJ, Woolridge M, Sowden AJ, Lister-Sharp D.

**No. 26**

Implantable cardioverter defibrillators: arrhythmias. A rapid and systematic review.

By Parkes J, Bryant J, Milne R.

**No. 27**

Treatments for fatigue in multiple sclerosis: a rapid and systematic review.

By Brañas P, Jordan R, Fry-Smith A, Burls A, Hyde C.

**No. 28**

Early asthma prophylaxis, natural history, skeletal development and economy (EASE): a pilot randomised controlled trial.

By Baxter-Jones ADG, Helms PJ, Russell G, Grant A, Ross S, Cairns JA, *et al.*

**No. 29**

Screening for hypercholesterolaemia versus case finding for familial hypercholesterolaemia: a systematic review and cost-effectiveness analysis.

By Marks D, Wonderling D, Thorogood M, Lambert H, Humphries SE, Neil HAW.

**No. 30**

A rapid and systematic review of the clinical effectiveness and cost-effectiveness of glycoprotein IIb/IIIa antagonists in the medical management of unstable angina.

By McDonagh MS, Bachmann LM, Golder S, Kleijnen J, ter Riet G.

**No. 31**

A randomised controlled trial of prehospital intravenous fluid replacement therapy in serious trauma.

By Turner J, Nicholl J, Webber L, Cox H, Dixon S, Yates D.

**No. 32**

Intrathecal pumps for giving opioids in chronic pain: a systematic review.

By Williams JE, Louw G, Towlerton G.

**No. 33**

Combination therapy (interferon alfa and ribavirin) in the treatment of chronic hepatitis C: a rapid and systematic review.

By Shepherd J, Waugh N, Hewitson P.

**No. 34**

A systematic review of comparisons of effect sizes derived from randomised and non-randomised studies.

By MacLehose RR, Reeves BC, Harvey IM, Sheldon TA, Russell IT, Black AMS.

**No. 35**

Intravascular ultrasound-guided interventions in coronary artery disease: a systematic literature review, with decision-analytic modelling, of outcomes and cost-effectiveness.

By Berry E, Kelly S, Hutton J, Lindsay HSJ, Blaxill JM, Evans JA, *et al.*

**No. 36**

A randomised controlled trial to evaluate the effectiveness and cost-effectiveness of counselling patients with chronic depression.

By Simpson S, Corney R, Fitzgerald P, Beecham J.

**No. 37**

Systematic review of treatments for atopic eczema.

By Hoare C, Li Wan Po A, Williams H.

**No. 38**

Bayesian methods in health technology assessment: a review.

By Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR.

**No. 39**

The management of dyspepsia: a systematic review.

By Delaney B, Moayyedi P, Deeks J, Innes M, Soo S, Barton P, *et al.*

**No. 40**

A systematic review of treatments for severe psoriasis.

By Griffiths CEM, Clark CM, Chalmers RJG, Li Wan Po A, Williams HC.

**Volume 5, 2001**

**No. 1**

Clinical and cost-effectiveness of donepezil, rivastigmine and galantamine for Alzheimer's disease: a rapid and systematic review.

By Clegg A, Bryant J, Nicholson T, McIntyre L, De Broe S, Gerard K, *et al.*

**No. 2**

The clinical effectiveness and cost-effectiveness of riluzole for motor neurone disease: a rapid and systematic review.

By Stewart A, Sandercock J, Bryan S, Hyde C, Barton PM, Fry-Smith A, *et al.*

**No. 3**

Equity and the economic evaluation of healthcare.

By Sassi F, Archard L, Le Grand J.

**No. 4**

Quality-of-life measures in chronic diseases of childhood.

By Eiser C, Morse R.

**No. 5**

Eliciting public preferences for healthcare: a systematic review of techniques.

By Ryan M, Scott DA, Reeves C, Bate A, van Teijlingen ER, Russell EM, *et al.*

**No. 6**

General health status measures for people with cognitive impairment: learning disability and acquired brain injury.

By Riemsma RP, Forbes CA, Glanville JM, Eastwood AJ, Kleijnen J.

**No. 7**

An assessment of screening strategies for fragile X syndrome in the UK.

By Pembrey ME, Barnicoat AJ, Carmichael B, Bobrow M, Turner G.

**No. 8**

Issues in methodological research: perspectives from researchers and commissioners.

By Lilford RJ, Richardson A, Stevens A, Fitzpatrick R, Edwards S, Rock F, *et al.*

**No. 9**

Systematic reviews of wound care management: (5) beds; (6) compression; (7) laser therapy, therapeutic ultrasound, electrotherapy and electromagnetic therapy.

By Cullum N, Nelson EA, Flemming K, Sheldon T.

**No. 10**

Effects of educational and psychosocial interventions for adolescents with diabetes mellitus: a systematic review.

By Hampson SE, Skinner TC, Hart J, Storey L, Gage H, Foxcroft D, *et al.*

**No. 11**

Effectiveness of autologous chondrocyte transplantation for hyaline cartilage defects in knees: a rapid and systematic review.

By Jobanputra P, Parry D, Fry-Smith A, Burls A.

**No. 12**

Statistical assessment of the learning curves of health technologies.

By Ramsay CR, Grant AM, Wallace SA, Garthwaite PH, Monk AF, Russell IT.

**No. 13**

The effectiveness and cost-effectiveness of temozolomide for the treatment of recurrent malignant glioma: a rapid and systematic review.

By Dinnes J, Cave C, Huang S, Major K, Milne R.

**No. 14**

A rapid and systematic review of the clinical effectiveness and cost-effectiveness of debriding agents in treating surgical wounds healing by secondary intention.

By Lewis R, Whiting P, ter Riet G, O'Meara S, Glanville J.

**No. 15**

Home treatment for mental health problems: a systematic review.

By Burns T, Knapp M, Catty J, Healey A, Henderson J, Watt H, *et al.*

**No. 16**

How to develop cost-conscious guidelines.

By Eccles M, Mason J.

**No. 17**

The role of specialist nurses in multiple sclerosis: a rapid and systematic review.

By De Broe S, Christopher F, Waugh N.

**No. 18**

A rapid and systematic review of the clinical effectiveness and cost-effectiveness of orlistat in the management of obesity.

By O'Meara S, Riemsma R, Shirran L, Mather L, ter Riet G.

**No. 19**

The clinical effectiveness and cost-effectiveness of pioglitazone for type 2 diabetes mellitus: a rapid and systematic review.

By Chilcott J, Wight J, Lloyd Jones M, Tappenden P.

**No. 20**

Extended scope of nursing practice: a multicentre randomised controlled trial of appropriately trained nurses and preregistration house officers in preoperative assessment in elective general surgery.

By Kinley H, Czoski-Murray C, George S, McCabe C, Primrose J, Reilly C, *et al.*

**No. 21**

Systematic reviews of the effectiveness of day care for people with severe mental disorders: (1) Acute day hospital versus admission; (2) Vocational rehabilitation; (3) Day hospital versus outpatient care.

By Marshall M, Crowther R, Almaraz-Serrano A, Creed F, Sledge W, Kluiter H, *et al.*

**No. 22**

The measurement and monitoring of surgical adverse events.

By Bruce J, Russell EM, Mollison J, Krukowski ZH.

**No. 23**

Action research: a systematic review and guidance for assessment.

By Waterman H, Tillen D, Dickson R, de Koning K.

**No. 24**

A rapid and systematic review of the clinical effectiveness and cost-effectiveness of gemcitabine for the treatment of pancreatic cancer.

By Ward S, Morris E, Bansback N, Calvert N, Crellin A, Forman D, *et al.*

**No. 25**

A rapid and systematic review of the evidence for the clinical effectiveness and cost-effectiveness of irinotecan, oxaliplatin and raltitrexed for the treatment of advanced colorectal cancer.

By Lloyd Jones M, Hummel S, Bansback N, Orr B, Seymour M.

**No. 26**

Comparison of the effectiveness of inhaler devices in asthma and chronic obstructive airways disease: a systematic review of the literature.

By Brocklebank D, Ram F, Wright J, Barry P, Cates C, Davies L, *et al.*

**No. 27**

The cost-effectiveness of magnetic resonance imaging for investigation of the knee joint.

By Bryan S, Weatherburn G, Bungay H, Hatrick C, Salas C, Parry D, *et al.*

**No. 28**

A rapid and systematic review of the clinical effectiveness and cost-effectiveness of topotecan for ovarian cancer.

By Forbes C, Shirran L, Bagnall A-M, Duffy S, ter Riet G.

**No. 29**

Superseded by a report published in a later volume.

**No. 30**

The role of radiography in primary care patients with low back pain of at least 6 weeks duration: a randomised (unblinded) controlled trial.

By Kendrick D, Fielding K, Bentley E, Miller P, Kerslake R, Pringle M.

**No. 31**

Design and use of questionnaires: a review of best practice applicable to surveys of health service staff and patients.

By McColl E, Jacoby A, Thomas L, Soutter J, Bamford C, Steen N, *et al.*

**No. 32**

A rapid and systematic review of the clinical effectiveness and cost-effectiveness of paclitaxel, docetaxel, gemcitabine and vinorelbine in non-small-cell lung cancer.

By Clegg A, Scott DA, Sidhu M, Hewitson P, Waugh N.

**No. 33**

Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives.

By Brookes ST, Whitley E, Peters TJ, Mulheran PA, Egger M, Davey Smith G.

**No. 34**

Depot antipsychotic medication in the treatment of patients with schizophrenia: (1) Meta-review; (2) Patient and nurse attitudes.

By David AS, Adams C.

**No. 35**

A systematic review of controlled trials of the effectiveness and cost-effectiveness of brief psychological treatments for depression.

By Churchill R, Hunot V, Corney R, Knapp M, McGuire H, Tylee A, *et al.*

**No. 36**

Cost analysis of child health surveillance.

By Sanderson D, Wright D, Acton C, Duree D.

**Volume 6, 2002****No. 1**

A study of the methods used to select review criteria for clinical audit.

By Hearnshaw H, Harker R, Cheater F, Baker R, Grimshaw G.

**No. 2**

Fludarabine as second-line therapy for B cell chronic lymphocytic leukaemia: a technology assessment.

By Hyde C, Wake B, Bryan S, Barton P, Fry-Smith A, Davenport C, *et al.*

**No. 3**

Rituximab as third-line treatment for refractory or recurrent Stage III or IV follicular non-Hodgkin's lymphoma: a systematic review and economic evaluation.

By Wake B, Hyde C, Bryan S, Barton P, Song F, Fry-Smith A, *et al.*

**No. 4**

A systematic review of discharge arrangements for older people.

By Parker SG, Peet SM, McPherson A, Cannaby AM, Baker R, Wilson A, *et al.*

**No. 5**

The clinical effectiveness and cost-effectiveness of inhaler devices used in the routine management of chronic asthma in older children: a systematic review and economic evaluation.

By Peters J, Stevenson M, Beverley C, Lim J, Smith S.

**No. 6**

The clinical effectiveness and cost-effectiveness of sibutramine in the management of obesity: a technology assessment.

By O'Meara S, Riemsma R, Shirran L, Mather L, ter Riet G.

**No. 7**

The cost-effectiveness of magnetic resonance angiography for carotid artery stenosis and peripheral vascular disease: a systematic review.

By Berry E, Kelly S, Westwood ME, Davies LM, Gough MJ, Bamford JM, *et al.*

**No. 8**

Promoting physical activity in South Asian Muslim women through 'exercise on prescription'.

By Carroll B, Ali N, Azam N.

**No. 9**

Zanamivir for the treatment of influenza in adults: a systematic review and economic evaluation.

By Burls A, Clark W, Stewart T, Preston C, Bryan S, Jefferson T, *et al.*

**No. 10**

A review of the natural history and epidemiology of multiple sclerosis: implications for resource allocation and health economic models.

By Richards RG, Sampson FC, Beard SM, Tappenden P.

**No. 11**

Screening for gestational diabetes: a systematic review and economic evaluation.

By Scott DA, Loveman E, McIntyre L, Waugh N.

**No. 12**

The clinical effectiveness and cost-effectiveness of surgery for people with morbid obesity: a systematic review and economic evaluation.

By Clegg AJ, Colquitt J, Sidhu MK, Royle P, Loveman E, Walker A.

**No. 13**

The clinical effectiveness of trastuzumab for breast cancer: a systematic review.

By Lewis R, Bagnall A-M, Forbes C, Shirran E, Duffy S, Kleijnen J, *et al.*

**No. 14**

The clinical effectiveness and cost-effectiveness of vinorelbine for breast cancer: a systematic review and economic evaluation.

By Lewis R, Bagnall A-M, King S, Woolacott N, Forbes C, Shirran L, *et al.*

**No. 15**

A systematic review of the effectiveness and cost-effectiveness of metal-on-metal hip resurfacing arthroplasty for treatment of hip disease.

By Vale L, Wyness L, McCormack K, McKenzie L, Brazzelli M, Stearns SC.

**No. 16**

The clinical effectiveness and cost-effectiveness of bupropion and nicotine replacement therapy for smoking cessation: a systematic review and economic evaluation.

By Woolcott NF, Jones L, Forbes CA, Mather LC, Sowden AJ, Song FJ, *et al.*

**No. 17**

A systematic review of effectiveness and economic evaluation of new drug treatments for juvenile idiopathic arthritis: etanercept.

By Cummins C, Connock M, Fry-Smith A, Burls A.

**No. 18**

Clinical effectiveness and cost-effectiveness of growth hormone in children: a systematic review and economic evaluation.

By Bryant J, Cave C, Mihaylova B, Chase D, McIntyre L, Gerard K, *et al.*

**No. 19**

Clinical effectiveness and cost-effectiveness of growth hormone in adults in relation to impact on quality of life: a systematic review and economic evaluation.

By Bryant J, Loveman E, Chase D, Mihaylova B, Cave C, Gerard K, *et al.*

**No. 20**

Clinical medication review by a pharmacist of patients on repeat prescriptions in general practice: a randomised controlled trial.

By Zermansky AG, Petty DR, Raynor DK, Lowe CJ, Freemantle N, Vail A.

**No. 21**

The effectiveness of infliximab and etanercept for the treatment of rheumatoid arthritis: a systematic review and economic evaluation.

By Jobanputra P, Barton P, Bryan S, Burls A.

**No. 22**

A systematic review and economic evaluation of computerised cognitive behaviour therapy for depression and anxiety.

By Kaltenthaler E, Shackley P, Stevens K, Beverley C, Parry G, Chilcott J.

**No. 23**

A systematic review and economic evaluation of pegylated liposomal doxorubicin hydrochloride for ovarian cancer.

By Forbes C, Wilby J, Richardson G, Sculpher M, Mather L, Riemsma R.

**No. 24**

A systematic review of the effectiveness of interventions based on a stages-of-change approach to promote individual behaviour change.

By Riemsma RP, Pattenden J, Bridle C, Sowden AJ, Mather L, Watt IS, *et al.*

**No. 25**

A systematic review update of the clinical effectiveness and cost-effectiveness of glycoprotein IIb/IIIa antagonists.

By Robinson M, Ginnelly L, Sculpher M, Jones L, Riemsma R, Palmer S, *et al.*

**No. 26**

A systematic review of the effectiveness, cost-effectiveness and barriers to implementation of thrombolytic and neuroprotective therapy for acute ischaemic stroke in the NHS.

By Sandercock P, Berge E, Dennis M, Forbes J, Hand P, Kwan J, *et al.*

**No. 27**

A randomised controlled crossover trial of nurse practitioner versus doctor-led outpatient care in a bronchiectasis clinic.

By Caine N, Sharples LD, Hollingworth W, French J, Keogan M, Exley A, *et al.*

**No. 28**

Clinical effectiveness and cost – consequences of selective serotonin reuptake inhibitors in the treatment of sex offenders.

By Adi Y, Ashcroft D, Browne K, Beech A, Fry-Smith A, Hyde C.

**No. 29**

Treatment of established osteoporosis: a systematic review and cost–utility analysis.

By Kanis JA, Brazier JE, Stevenson M, Calvert NW, Lloyd Jones M.

**No. 30**

Which anaesthetic agents are cost-effective in day surgery? Literature review, national survey of practice and randomised controlled trial.

By Elliott RA Payne K, Moore JK, Davies LM, Harper NJN, St Leger AS, *et al.*

**No. 31**

Screening for hepatitis C among injecting drug users and in genitourinary medicine clinics: systematic reviews of effectiveness, modelling study and national survey of current practice.

By Stein K, Dalziel K, Walker A, McIntyre L, Jenkins B, Horne J, *et al.*

**No. 32**

The measurement of satisfaction with healthcare: implications for practice from a systematic review of the literature.

By Crow R, Gage H, Hampson S, Hart J, Kimber A, Storey L, *et al.*

**No. 33**

The effectiveness and cost-effectiveness of imatinib in chronic myeloid leukaemia: a systematic review.

By Garside R, Round A, Dalziel K, Stein K, Royle R.

**No. 34**

A comparative study of hypertonic saline, daily and alternate-day rhDNase in children with cystic fibrosis.

By Suri R, Wallis C, Bush A, Thompson S, Normand C, Flather M, *et al.*

**No. 35**

A systematic review of the costs and effectiveness of different models of paediatric home care.

By Parker G, Bhakta P, Lovett CA, Paisley S, Olsen R, Turner D, *et al.*

**Volume 7, 2003**

**No. 1**

How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study.

By Egger M, Juni P, Bartlett C, Holenstein F, Sterne J.

**No. 2**

Systematic review of the effectiveness and cost-effectiveness, and economic evaluation, of home versus hospital or satellite unit haemodialysis for people with end-stage renal failure.

By Mowatt G, Vale L, Perez J, Wyness L, Fraser C, MacLeod A, *et al.*

**No. 3**

Systematic review and economic evaluation of the effectiveness of infliximab for the treatment of Crohn's disease.

By Clark W, Raftery J, Barton P, Song F, Fry-Smith A, Burls A.

**No. 4**

A review of the clinical effectiveness and cost-effectiveness of routine anti-D prophylaxis for pregnant women who are rhesus negative.

By Chilcott J, Lloyd Jones M, Wight J, Forman K, Wray J, Beverley C, *et al.*

**No. 5**

Systematic review and evaluation of the use of tumour markers in paediatric oncology: Ewing's sarcoma and neuroblastoma.

By Riley RD, Burchill SA, Abrams KR, Heney D, Lambert PC, Jones DR, *et al.*

**No. 6**

The cost-effectiveness of screening for *Helicobacter pylori* to reduce mortality and morbidity from gastric cancer and peptic ulcer disease: a discrete-event simulation model.

By Roderick P, Davies R, Raftery J, Crabbe D, Pearce R, Bhandari P, *et al.*



**No. 7**

The clinical effectiveness and cost-effectiveness of routine dental checks: a systematic review and economic evaluation.

By Davenport C, Elley K, Salas C, Taylor-Weetman CL, Fry-Smith A, Bryan S, *et al.*

**No. 8**

A multicentre randomised controlled trial assessing the costs and benefits of using structured information and analysis of women's preferences in the management of menorrhagia.

By Kennedy ADM, Sculpher MJ, Coulter A, Dwyer N, Rees M, Horsley S, *et al.*

**No. 9**

Clinical effectiveness and cost-utility of photodynamic therapy for wet age-related macular degeneration: a systematic review and economic evaluation.

By Meads C, Salas C, Roberts T, Moore D, Fry-Smith A, Hyde C.

**No. 10**

Evaluation of molecular tests for prenatal diagnosis of chromosome abnormalities.

By Grimshaw GM, Szczepura A, Hultén M, MacDonald F, Nevin NC, Sutton F, *et al.*

**No. 11**

First and second trimester antenatal screening for Down's syndrome: the results of the Serum, Urine and Ultrasound Screening Study (SURUSS).

By Wald NJ, Rodeck C, Hackshaw AK, Walters J, Chitty L, Mackinson AM.

**No. 12**

The effectiveness and cost-effectiveness of ultrasound locating devices for central venous access: a systematic review and economic evaluation.

By Calvert N, Hind D, McWilliams RG, Thomas SM, Beverley C, Davidson A.

**No. 13**

A systematic review of atypical antipsychotics in schizophrenia.

By Bagnall A-M, Jones L, Lewis R, Ginnelly L, Glanville J, Torgerson D, *et al.*

**No. 14**

Prostate Testing for Cancer and Treatment (ProtecT) feasibility study.

By Donovan J, Hamdy F, Neal D, Peters T, Oliver S, Brindle L, *et al.*

**No. 15**

Early thrombolysis for the treatment of acute myocardial infarction: a systematic review and economic evaluation.

By Boland A, Dundar Y, Bagust A, Haycox A, Hill R, Mujica Mota R, *et al.*

**No. 16**

Screening for fragile X syndrome: a literature review and modelling.

By Song FJ, Barton P, Sleightholme V, Yao GL, Fry-Smith A.

**No. 17**

Systematic review of endoscopic sinus surgery for nasal polyps.

By Dalziel K, Stein K, Round A, Garside R, Royle P.

**No. 18**

Towards efficient guidelines: how to monitor guideline use in primary care.

By Hutchinson A, McIntosh A, Cox S, Gilbert C.

**No. 19**

Effectiveness and cost-effectiveness of acute hospital-based spinal cord injuries services: systematic review.

By Bagnall A-M, Jones L, Richardson G, Duffy S, Riemsma R.

**No. 20**

Prioritisation of health technology assessment. The PATHS model: methods and case studies.

By Townsend J, Buxton M, Harper G.

**No. 21**

Systematic review of the clinical effectiveness and cost-effectiveness of tension-free vaginal tape for treatment of urinary stress incontinence.

By Cody J, Wyness L, Wallace S, Glazener C, Kilonzo M, Stearns S, *et al.*

**No. 22**

The clinical and cost-effectiveness of patient education models for diabetes: a systematic review and economic evaluation.

By Loveman E, Cave C, Green C, Royle P, Dunn N, Waugh N.

**No. 23**

The role of modelling in prioritising and planning clinical trials.

By Chilcott J, Brennan A, Booth A, Karnon J, Tappenden P.

**No. 24**

Cost-benefit evaluation of routine influenza immunisation in people 65-74 years of age.

By Allsup S, Gosney M, Haycox A, Regan M.

**No. 25**

The clinical and cost-effectiveness of pulsatile machine perfusion versus cold storage of kidneys for transplantation retrieved from heart-beating and non-heart-beating donors.

By Wight J, Chilcott J, Holmes M, Brewer N.

**No. 26**

Can randomised trials rely on existing electronic data? A feasibility study to explore the value of routine data in health technology assessment.

By Williams JG, Cheung WY, Cohen DR, Hutchings HA, Longo MF, Russell IT.

**No. 27**

Evaluating non-randomised intervention studies.

By Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovich C, Song F, *et al.*

**No. 28**

A randomised controlled trial to assess the impact of a package comprising a patient-orientated, evidence-based self-help guidebook and patient-centred consultations on disease management and satisfaction in inflammatory bowel disease.

By Kennedy A, Nelson E, Reeves D, Richardson G, Roberts C, Robinson A, *et al.*

**No. 29**

The effectiveness of diagnostic tests for the assessment of shoulder pain due to soft tissue disorders: a systematic review.

By Dinnes J, Loveman E, McIntyre L, Waugh N.

**No. 30**

The value of digital imaging in diabetic retinopathy.

By Sharp PF, Olson J, Strachan F, Hipwell J, Ludbrook A, O'Donnell M, *et al.*

**No. 31**

Lowering blood pressure to prevent myocardial infarction and stroke: a new preventive strategy.

By Law M, Wald N, Morris J.

**No. 32**

Clinical and cost-effectiveness of capecitabine and tegafur with uracil for the treatment of metastatic colorectal cancer: systematic review and economic evaluation.

By Ward S, Kaltenthaler E, Cowan J, Brewer N.

**No. 33**

Clinical and cost-effectiveness of new and emerging technologies for early localised prostate cancer: a systematic review.

By Hummel S, Paisley S, Morgan A, Currie E, Brewer N.

**No. 34**

Literature searching for clinical and cost-effectiveness studies used in health technology assessment reports carried out for the National Institute for Clinical Excellence appraisal system.

By Royle P, Waugh N.

**No. 35**

Systematic review and economic decision modelling for the prevention and treatment of influenza A and B.

By Turner D, Wailoo A, Nicholson K, Cooper N, Sutton A, Abrams K.

**No. 36**

A randomised controlled trial to evaluate the clinical and cost-effectiveness of Hickman line insertions in adult cancer patients by nurses.

By Boland A, Haycox A, Bagust A, Fitzsimmons L.

**No. 37**

Redesigning postnatal care: a randomised controlled trial of protocol-based midwifery-led care focused on individual women's physical and psychological health needs.

By MacArthur C, Winter HR, Bick DE, Lilford RJ, Lancashire RJ, Knowles H, *et al.*

**No. 38**

Estimating implied rates of discount in healthcare decision-making.

By West RR, McNabb R, Thompson AGH, Sheldon TA, Grimley Evans J.

**No. 39**

Systematic review of isolation policies in the hospital management of methicillin-resistant *Staphylococcus aureus*: a review of the literature with epidemiological and economic modelling.

By Cooper BS, Stone SP, Kibbler CC, Cookson BD, Roberts JA, Medley GF, *et al.*

**No. 40**

Treatments for spasticity and pain in multiple sclerosis: a systematic review.

By Beard S, Hunn A, Wight J.

**No. 41**

The inclusion of reports of randomised trials published in languages other than English in systematic reviews.

By Moher D, Pham B, Lawson ML, Klassen TP.

**No. 42**

The impact of screening on future health-promoting behaviours and health beliefs: a systematic review.

By Bankhead CR, Brett J, Bukach C, Webster P, Stewart-Brown S, Munafo M, *et al.*

**Volume 8, 2004**

**No. 1**

What is the best imaging strategy for acute stroke?

By Wardlaw JM, Keir SL, Seymour J, Lewis S, Sandercock PAG, Dennis MS, *et al.*

**No. 2**

Systematic review and modelling of the investigation of acute and chronic chest pain presenting in primary care.

By Mant J, McManus RJ, Oakes RAL, Delaney BC, Barton PM, Deeks JJ, *et al.*

**No. 3**

The effectiveness and cost-effectiveness of microwave and thermal balloon endometrial ablation for heavy menstrual bleeding: a systematic review and economic modelling.

By Garside R, Stein K, Wyatt K, Round A, Price A.

**No. 4**

A systematic review of the role of bisphosphonates in metastatic disease.

By Ross JR, Saunders Y, Edmonds PM, Patel S, Wonderling D, Normand C, *et al.*

**No. 5**

Systematic review of the clinical effectiveness and cost-effectiveness of capecitabine (Xeloda®) for locally advanced and/or metastatic breast cancer.

By Jones L, Hawkins N, Westwood M, Wright K, Richardson G, Riemsma R.

**No. 6**

Effectiveness and efficiency of guideline dissemination and implementation strategies.

By Grimshaw JM, Thomas RE, MacLennan G, Fraser C, Ramsay CR, Vale L, *et al.*

**No. 7**

Clinical effectiveness and costs of the Sugarbaker procedure for the treatment of pseudomyxoma peritonei.

By Bryant J, Clegg AJ, Sidhu MK, Brodin H, Royle P, Davidson P.

**No. 8**

Psychological treatment for insomnia in the regulation of long-term hypnotic drug use.

By Morgan K, Dixon S, Mathers N, Thompson J, Tomeny M.

**No. 9**

Improving the evaluation of therapeutic interventions in multiple sclerosis: development of a patient-based measure of outcome.

By Hobart JC, Riazi A, Lamping DL, Fitzpatrick R, Thompson AJ.

**No. 10**

A systematic review and economic evaluation of magnetic resonance cholangiopancreatography compared with diagnostic endoscopic retrograde cholangiopancreatography.

By Kaltenthaler E, Bravo Vergel Y, Chilcott J, Thomas S, Blakeborough T, Walters SJ, *et al.*

**No. 11**

The use of modelling to evaluate new drugs for patients with a chronic condition: the case of antibodies against tumour necrosis factor in rheumatoid arthritis.

By Barton P, Jobanputra P, Wilson J, Bryan S, Burls A.

**No. 12**

Clinical effectiveness and cost-effectiveness of neonatal screening for inborn errors of metabolism using tandem mass spectrometry: a systematic review.

By Pandor A, Eastham J, Beverley C, Chilcott J, Paisley S.

**No. 13**

Clinical effectiveness and cost-effectiveness of pioglitazone and rosiglitazone in the treatment of type 2 diabetes: a systematic review and economic evaluation.

By Czoski-Murray C, Warren E, Chilcott J, Beverley C, Psyllaki MA, Cowan J.

**No. 14**

Routine examination of the newborn: the EMREN study. Evaluation of an extension of the midwife role including a randomised controlled trial of appropriately trained midwives and paediatric senior house officers.

By Townsend J, Wolke D, Hayes J, Davé S, Rogers C, Bloomfield L, *et al.*

**No. 15**

Involving consumers in research and development agenda setting for the NHS: developing an evidence-based approach.

By Oliver S, Clarke-Jones L, Rees R, Milne R, Buchanan P, Gabbay J, *et al.*

**No. 16**

A multi-centre randomised controlled trial of minimally invasive direct coronary bypass grafting versus percutaneous transluminal coronary angioplasty with stenting for proximal stenosis of the left anterior descending coronary artery.

By Reeves BC, Angelini GD, Bryan AJ, Taylor FC, Cripps T, Spyt TJ, *et al.*

**No. 17**

Does early magnetic resonance imaging influence management or improve outcome in patients referred to secondary care with low back pain? A pragmatic randomised controlled trial.

By Gilbert FJ, Grant AM, Gillan MGC, Vale L, Scott NW, Campbell MK, *et al.*

**No. 18**

The clinical and cost-effectiveness of anakinra for the treatment of rheumatoid arthritis in adults: a systematic review and economic analysis.

By Clark W, Jobanputra P, Barton P, Burls A.

**No. 19**

A rapid and systematic review and economic evaluation of the clinical and cost-effectiveness of newer drugs for treatment of mania associated with bipolar affective disorder.

By Bridle C, Palmer S, Bagnall A-M, Darba J, Duffy S, Sculpher M, *et al.*

**No. 20**

Liquid-based cytology in cervical screening: an updated rapid and systematic review and economic analysis.

By Karnon J, Peters J, Platt J, Chilcott J, McGoogan E, Brewer N.

**No. 21**

Systematic review of the long-term effects and economic consequences of treatments for obesity and implications for health improvement.

By Avenell A, Broom J, Brown TJ, Poobalan A, Aucott L, Stearns SC, *et al.*

**No. 22**

Autoantibody testing in children with newly diagnosed type 1 diabetes mellitus.

By Dretzke J, Cummins C, Sandercock J, Fry-Smith A, Barrett T, Burls A.

**No. 23**

Clinical effectiveness and cost-effectiveness of prehospital intravenous fluids in trauma patients.

By Dretzke J, Sandercock J, Bayliss S, Burls A.

**No. 24**

Newer hypnotic drugs for the short-term management of insomnia: a systematic review and economic evaluation.

By Dündar Y, Boland A, Strobl J, Dodd S, Haycox A, Bagust A, *et al.*

**No. 25**

Development and validation of methods for assessing the quality of diagnostic accuracy studies.

By Whiting P, Rutjes AWS, Dinnes J, Reitsma JB, Bossuyt PMM, Kleijnen J.

**No. 26**

EVALUATE hysterectomy trial: a multicentre randomised trial comparing abdominal, vaginal and laparoscopic methods of hysterectomy.

By Garry R, Fountain J, Brown J, Manca A, Mason S, Sculpher M, *et al.*

**No. 27**

Methods for expected value of information analysis in complex health economic models: developments on the health economics of interferon- $\beta$  and glatiramer acetate for multiple sclerosis.

By Tappenden P, Chilcott JB, Eggington S, Oakley J, McCabe C.

**No. 28**

Effectiveness and cost-effectiveness of imatinib for first-line treatment of chronic myeloid leukaemia in chronic phase: a systematic review and economic analysis.

By Dalziel K, Round A, Stein K, Garside R, Price A.

**No. 29**

VenUS I: a randomised controlled trial of two types of bandage for treating venous leg ulcers.

By Iglesias C, Nelson EA, Cullum NA, Torgerson DJ, on behalf of the VenUS Team.

**No. 30**

Systematic review of the effectiveness and cost-effectiveness, and economic evaluation, of myocardial perfusion scintigraphy for the diagnosis and management of angina and myocardial infarction.

By Mowatt G, Vale L, Brazzelli M, Hernandez R, Murray A, Scott N, *et al.*

**No. 31**

A pilot study on the use of decision theory and value of information analysis as part of the NHS Health Technology Assessment programme.

By Claxton K, Ginnelly L, Sculpher M, Philips Z, Palmer S.

**No. 32**

The Social Support and Family Health Study: a randomised controlled trial and economic evaluation of two alternative forms of postnatal support for mothers living in disadvantaged inner-city areas.

By Wiggins M, Oakley A, Roberts I, Turner H, Rajan L, Austerberry H, *et al.*

**No. 33**

Psychosocial aspects of genetic screening of pregnant women and newborns: a systematic review.

By Green JM, Hewison J, Bekker HL, Bryant, Cuckle HS.

**No. 34**

Evaluation of abnormal uterine bleeding: comparison of three outpatient procedures within cohorts defined by age and menopausal status.

By Critchley HOD, Warner P, Lee AJ, Brechin S, Guise J, Graham B.

**No. 35**

Coronary artery stents: a rapid systematic review and economic evaluation.

By Hill R, Bagust A, Bakhai A, Dickson R, Dündar Y, Haycox A, *et al.*

**No. 36**

Review of guidelines for good practice in decision-analytic modelling in health technology assessment.

By Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R, *et al.*

**No. 37**

Rituximab (MabThera®) for aggressive non-Hodgkin's lymphoma: systematic review and economic evaluation.

By Knight C, Hind D, Brewer N, Abbott V.

**No. 38**

Clinical effectiveness and cost-effectiveness of clopidogrel and modified-release dipyridamole in the secondary prevention of occlusive vascular events: a systematic review and economic evaluation.

By Jones L, Griffin S, Palmer S, Main C, Orton V, Sculpher M, *et al.*

**No. 39**

Pegylated interferon  $\alpha$ -2a and -2b in combination with ribavirin in the treatment of chronic hepatitis C: a systematic review and economic evaluation.

By Shepherd J, Brodin H, Cave C, Waugh N, Price A, Gabbay J.

**No. 40**

Clopidogrel used in combination with aspirin compared with aspirin alone in the treatment of non-ST-segment-elevation acute coronary syndromes: a systematic review and economic evaluation.

By Main C, Palmer S, Griffin S, Jones L, Orton V, Sculpher M, *et al.*

**No. 41**

Provision, uptake and cost of cardiac rehabilitation programmes: improving services to under-represented groups.

By Beswick AD, Rees K, Griebisch I, Taylor FC, Burke M, West RR, *et al.*

**No. 42**

Involving South Asian patients in clinical trials.

By Hussain-Gambles M, Leese B, Atkin K, Brown J, Mason S, Tovey P.

**No. 43**

Clinical and cost-effectiveness of continuous subcutaneous insulin infusion for diabetes.

By Colquitt JL, Green C, Sidhu MK, Hartwell D, Waugh N.

**No. 44**

Identification and assessment of ongoing trials in health technology assessment reviews.

By Song FJ, Fry-Smith A, Davenport C, Bayliss S, Adi Y, Wilson JS, *et al.*

**No. 45**

Systematic review and economic evaluation of a long-acting insulin analogue, insulin glargine

By Warren E, Weatherley-Jones E, Chilcott J, Beverley C.

**No. 46**

Supplementation of a home-based exercise programme with a class-based programme for people with osteoarthritis of the knees: a randomised controlled trial and health economic analysis.

By McCarthy CJ, Mills PM, Pullen R, Richardson G, Hawkins N, Roberts CR, *et al.*

**No. 47**

Clinical and cost-effectiveness of once-daily versus more frequent use of same potency topical corticosteroids for atopic eczema: a systematic review and economic evaluation.

By Green C, Colquitt JL, Kirby J, Davidson P, Payne E.

**No. 48**

Acupuncture of chronic headache disorders in primary care: randomised controlled trial and economic analysis.

By Vickers AJ, Rees RW, Zollman CE, McCarney R, Smith CM, Ellis N, *et al.*

**No. 49**

Generalisability in economic evaluation studies in healthcare: a review and case studies.

By Sculpher MJ, Pang FS, Manca A, Drummond MF, Golder S, Urdahl H, *et al.*

**No. 50**

Virtual outreach: a randomised controlled trial and economic evaluation of joint teleconferenced medical consultations.

By Wallace P, Barber J, Clayton W, Currell R, Fleming K, Garner P, *et al.*

**Volume 9, 2005**

**No. 1**

Randomised controlled multiple treatment comparison to provide a cost-effectiveness rationale for the selection of antimicrobial therapy in acne.

By Ozolins M, Eady EA, Avery A, Cunliffe WJ, O'Neill C, Simpson NB, *et al.*

**No. 2**

Do the findings of case series studies vary significantly according to methodological characteristics?

By Dalziel K, Round A, Stein K, Garside R, Castelnovo E, Payne L.

**No. 3**

Improving the referral process for familial breast cancer genetic counselling: findings of three randomised controlled trials of two interventions.

By Wilson BJ, Torrance N, Mollison J, Wordsworth S, Gray JR, Haites NE, *et al.*

**No. 4**

Randomised evaluation of alternative electrosurgical modalities to treat bladder outflow obstruction in men with benign prostatic hyperplasia.

By Fowler C, McAllister W, Plail R, Karim O, Yang Q.

**No. 5**

A pragmatic randomised controlled trial of the cost-effectiveness of palliative therapies for patients with inoperable oesophageal cancer.

By Shenfine J, McNamee P, Steen N, Bond J, Griffin SM.

**No. 6**

Impact of computer-aided detection prompts on the sensitivity and specificity of screening mammography.

By Taylor P, Champness J, Given-Wilson R, Johnston K, Potts H.

**No. 7**

Issues in data monitoring and interim analysis of trials.

By Grant AM, Altman DG, Babiker AB, Campbell MK, Clemens FJ, Darbyshire JH, *et al.*

**No. 8**

Lay public's understanding of equipoise and randomisation in randomised controlled trials.

By Robinson EJ, Kerr CEP, Stevens AJ, Lilford RJ, Braunnholtz DA, Edwards SJ, *et al.*

**No. 9**

Clinical and cost-effectiveness of electroconvulsive therapy for depressive illness, schizophrenia, catatonia and mania: systematic reviews and economic modelling studies.

By Greenhalgh J, Knight C, Hind D, Beverley C, Walters S.

**No. 10**

Measurement of health-related quality of life for people with dementia: development of a new instrument (DEM-QOL) and an evaluation of current methodology.

By Smith SC, Lamping DL, Banerjee S, Harwood R, Foley B, Smith P, *et al.*

**No. 11**

Clinical effectiveness and cost-effectiveness of drotrecogin alfa (activated) (Xigris®) for the treatment of severe sepsis in adults: a systematic review and economic evaluation.

By Green C, Dinnes J, Takeda A, Shepherd J, Hartwell D, Cave C, *et al.*

**No. 12**

A methodological review of how heterogeneity has been examined in systematic reviews of diagnostic test accuracy.

By Dinnes J, Deeks J, Kirby J, Roderick P.

**No. 13**

Cervical screening programmes: can automation help? Evidence from systematic reviews, an economic analysis and a simulation modelling exercise applied to the UK.

By Willis BH, Barton P, Pearmain P, Bryan S, Hyde C.

**No. 14**

Laparoscopic surgery for inguinal hernia repair: systematic review of effectiveness and economic evaluation.

By McCormack K, Wake B, Perez J, Fraser C, Cook J, McIntosh E, *et al.*

**No. 15**

Clinical effectiveness, tolerability and cost-effectiveness of newer drugs for epilepsy in adults: a systematic review and economic evaluation.

By Wilby J, Kainth A, Hawkins N, Epstein D, McIntosh H, McDaid C, *et al.*

**No. 16**

A randomised controlled trial to compare the cost-effectiveness of tricyclic antidepressants, selective serotonin reuptake inhibitors and lofepramine.

By Peveler R, Kendrick T, Buxton M, Longworth L, Baldwin D, Moore M, *et al.*

**No. 17**

Clinical effectiveness and cost-effectiveness of immediate angioplasty for acute myocardial infarction: systematic review and economic evaluation.

By Hartwell D, Colquitt J, Loveman E, Clegg AJ, Brodin H, Waugh N, *et al.*

**No. 18**

A randomised controlled comparison of alternative strategies in stroke care.

By Kalra L, Evans A, Perez I, Knapp M, Swift C, Donaldson N.

**No. 19**

The investigation and analysis of critical incidents and adverse events in healthcare.

By Woloshynowych M, Rogers S, Taylor-Adams S, Vincent C.

**No. 20**

Potential use of routine databases in health technology assessment.

By Raftery J, Roderick P, Stevens A.

**No. 21**

Clinical and cost-effectiveness of newer immunosuppressive regimens in renal transplantation: a systematic review and modelling study.

By Woodroffe R, Yao GL, Meads C, Bayliss S, Ready A, Raftery J, *et al.*

**No. 22**

A systematic review and economic evaluation of alendronate, etidronate, risedronate, raloxifene and teriparatide for the prevention and treatment of postmenopausal osteoporosis.

By Stevenson M, Lloyd Jones M, De Nigris E, Brewer N, Davis S, Oakley J.

**No. 23**

A systematic review to examine the impact of psycho-educational interventions on health outcomes and costs in adults and children with difficult asthma.

By Smith JR, Mugford M, Holland R, Candy B, Noble MJ, Harrison BDW, *et al.*

**No. 24**

An evaluation of the costs, effectiveness and quality of renal replacement therapy provision in renal satellite units in England and Wales.

By Roderick P, Nicholson T, Armitage A, Mehta R, Mullee M, Gerard K, *et al.*

**No. 25**

Imatinib for the treatment of patients with unresectable and/or metastatic gastrointestinal stromal tumours: systematic review and economic evaluation.

By Wilson J, Connock M, Song F, Yao G, Fry-Smith A, Raftery J, *et al.*

**No. 26**

Indirect comparisons of competing interventions.

By Glenny AM, Altman DG, Song F, Sakarovich C, Deeks JJ, D'Amico R, *et al.*

**No. 27**

Cost-effectiveness of alternative strategies for the initial medical management of non-ST elevation acute coronary syndrome: systematic review and decision-analytical modelling.

By Robinson M, Palmer S, Sculpher M, Philips Z, Ginnelly L, Bowens A, *et al.*

**No. 28**

Outcomes of electrically stimulated gracilis neosphincter surgery.

By Tillin T, Chambers M, Feldman R.

**No. 29**

The effectiveness and cost-effectiveness of pimecrolimus and tacrolimus for atopic eczema: a systematic review and economic evaluation.

By Garside R, Stein K, Castelnovo E, Pitt M, Ashcroft D, Dimmock P, *et al.*

**No. 30**

Systematic review on urine albumin testing for early detection of diabetic complications.

By Newman DJ, Mattock MB, Dawnay ABS, Kerry S, McGuire A, Yaqoob M, *et al.*

**No. 31**

Randomised controlled trial of the cost-effectiveness of water-based therapy for lower limb osteoarthritis.

By Cochrane T, Davey RC, Matthes Edwards SM.

**No. 32**

Longer term clinical and economic benefits of offering acupuncture care to patients with chronic low back pain.

By Thomas KJ, MacPherson H, Ratcliffe J, Thorpe L, Brazier J, Campbell M, *et al.*

**No. 33**

Cost-effectiveness and safety of epidural steroids in the management of sciatica.

By Price C, Arden N, Coglán L, Rogers P.

**No. 34**

The British Rheumatoid Outcome Study Group (BROSG) randomised controlled trial to compare the effectiveness and cost-effectiveness of aggressive versus symptomatic therapy in established rheumatoid arthritis.

By Symmons D, Tricker K, Roberts C, Davies L, Dawes P, Scott DL.

**No. 35**

Conceptual framework and systematic review of the effects of participants' and professionals' preferences in randomised controlled trials.

By King M, Nazareth I, Lampe F, Bower P, Chandler M, Morou M, *et al.*

**No. 36**

The clinical and cost-effectiveness of implantable cardioverter defibrillators: a systematic review.

By Bryant J, Brodin H, Loveman E, Payne E, Clegg A.

**No. 37**

A trial of problem-solving by community mental health nurses for anxiety, depression and life difficulties among general practice patients. The CPN-GP study.

By Kendrick T, Simons L, Mynors-Wallis L, Gray A, Lathlean J, Pickering R, *et al.*

**No. 38**

The causes and effects of socio-demographic exclusions from clinical trials.

By Bartlett C, Doyal L, Ebrahim S, Davey P, Bachmann M, Egger M, *et al.*

**No. 39**

Is hydrotherapy cost-effective? A randomised controlled trial of combined hydrotherapy programmes compared with physiotherapy land techniques in children with juvenile idiopathic arthritis.

By Epps H, Ginnelly L, Utley M, Southwood T, Gallivan S, Sculpher M, *et al.*

**No. 40**

A randomised controlled trial and cost-effectiveness study of systematic screening (targeted and total population screening) versus routine practice for the detection of atrial fibrillation in people aged 65 and over. The SAFE study.

By Hobbs FDR, Fitzmaurice DA, Mant J, Murray E, Jowett S, Bryan S, *et al.*

**No. 41**

Displaced intracapsular hip fractures in fit, older people: a randomised comparison of reduction and fixation, bipolar hemiarthroplasty and total hip arthroplasty.

By Keating JF, Grant A, Masson M, Scott NW, Forbes JF.

**No. 42**

Long-term outcome of cognitive behaviour therapy clinical trials in central Scotland.

By Durham RC, Chambers JA, Power KG, Sharp DM, Macdonald RR, Major KA, *et al.*

**No. 43**

The effectiveness and cost-effectiveness of dual-chamber pacemakers compared with single-chamber pacemakers for bradycardia due to atrioventricular block or sick sinus syndrome: systematic review and economic evaluation.

By Castelnovo E, Stein K, Pitt M, Garside R, Payne E.

**No. 44**

Newborn screening for congenital heart defects: a systematic review and cost-effectiveness analysis.

By Knowles R, Griesch I, Dezateux C, Brown J, Bull C, Wren C.

**No. 45**

The clinical and cost-effectiveness of left ventricular assist devices for end-stage heart failure: a systematic review and economic evaluation.

By Clegg AJ, Scott DA, Loveman E, Colquitt J, Hutchinson J, Royle P, *et al.*

**No. 46**

The effectiveness of the Heidelberg Retina Tomograph and laser diagnostic glaucoma scanning system (GDx) in detecting and monitoring glaucoma.

By Kwartz AJ, Henson DB, Harper RA, Spencer AF, McLeod D.

**No. 47**

Clinical and cost-effectiveness of autologous chondrocyte implantation for cartilage defects in knee joints: systematic review and economic evaluation.

By Clar C, Cummins E, McIntyre L, Thomas S, Lamb J, Bain L, *et al.*

**No. 48**

Systematic review of effectiveness of different treatments for childhood retinoblastoma.

By McDaid C, Hartley S, Bagnall A-M, Ritchie G, Light K, Riemsma R.

**No. 49**

Towards evidence-based guidelines for the prevention of venous thromboembolism: systematic reviews of mechanical methods, oral anticoagulation, dextran and regional anaesthesia as thromboprophylaxis.

By Roderick P, Ferris G, Wilson K, Halls H, Jackson D, Collins R, *et al.*

**No. 50**

The effectiveness and cost-effectiveness of parent training/education programmes for the treatment of conduct disorder, including oppositional defiant disorder, in children.

By Dretzke J, Frew E, Davenport C, Barlow J, Stewart-Brown S, Sandercock J, *et al.*

**Volume 10, 2006**

**No. 1**

The clinical and cost-effectiveness of donepezil, rivastigmine, galantamine and memantine for Alzheimer's disease.

By Loveman E, Green C, Kirby J, Takeda A, Picot J, Payne E, *et al.*

**No. 2**

FOOD: a multicentre randomised trial evaluating feeding policies in patients admitted to hospital with a recent stroke.

By Dennis M, Lewis S, Cranswick G, Forbes J.

**No. 3**

The clinical effectiveness and cost-effectiveness of computed tomography screening for lung cancer: systematic reviews.

By Black C, Bagust A, Boland A, Walker S, McLeod C, De Verteuil R, *et al.*

**No. 4**

A systematic review of the effectiveness and cost-effectiveness of neuroimaging assessments used to visualise the seizure focus in people with refractory epilepsy being considered for surgery.

By Whiting P, Gupta R, Burch J, Mujica Mota RE, Wright K, Marson A, *et al.*

**No. 5**

Comparison of conference abstracts and presentations with full-text articles in the health technology assessments of rapidly evolving technologies.

By Dunder Y, Dodd S, Dickson R, Walley T, Haycox A, Williamson PR.

**No. 6**

Systematic review and evaluation of methods of assessing urinary incontinence.

By Martin JL, Williams KS, Abrams KR, Turner DA, Sutton AJ, Chapple C, *et al.*

**No. 7**

The clinical effectiveness and cost-effectiveness of newer drugs for children with epilepsy. A systematic review.

By Connock M, Frew E, Evans B-W, Bryan S, Cummins C, Fry-Smith A, *et al.*

**No. 8**

Surveillance of Barrett's oesophagus: exploring the uncertainty through systematic review, expert workshop and economic modelling.

By Garside R, Pitt M, Somerville M, Stein K, Price A, Gilbert N.

**No. 9**

Topotecan, pegylated liposomal doxorubicin hydrochloride and paclitaxel for second-line or subsequent treatment of advanced ovarian cancer: a systematic review and economic evaluation.

By Main C, Bojke L, Griffin S, Norman G, Barbieri M, Mather L, *et al.*

**No. 10**

Evaluation of molecular techniques in prediction and diagnosis of cytomegalovirus disease in immunocompromised patients.

By Szczepura A, Westmoreland D, Vinogradova Y, Fox J, Clark M.

**No. 11**

Screening for thrombophilia in high-risk situations: systematic review and cost-effectiveness analysis. The Thrombosis: Risk and Economic Assessment of Thrombophilia Screening (TREATS) study.

By Wu O, Robertson L, Twaddle S, Lowe GDO, Clark P, Greaves M, *et al.*

**No. 12**

A series of systematic reviews to inform a decision analysis for sampling and treating infected diabetic foot ulcers.

By Nelson EA, O'Meara S, Craig D, Iglesias C, Golder S, Dalton J, *et al.*

**No. 13**

Randomised clinical trial, observational study and assessment of cost-effectiveness of the treatment of varicose veins (REACTIV trial).

By Michaels JA, Campbell WB, Brazier JE, MacIntyre JB, Palfreyman SJ, Ratcliffe J, *et al.*

**No. 14**

The cost-effectiveness of screening for oral cancer in primary care.

By Speight PM, Palmer S, Moles DR, Downer MC, Smith DH, Henriksson M, *et al.*

**No. 15**

Measurement of the clinical and cost-effectiveness of non-invasive diagnostic testing strategies for deep vein thrombosis.

By Goodacre S, Sampson F, Stevenson M, Wailoo A, Sutton A, Thomas S, *et al.*

**No. 16**

Systematic review of the effectiveness and cost-effectiveness of HealOzone® for the treatment of occlusal pit/fissure caries and root caries.

By Brazzelli M, McKenzie L, Fielding S, Fraser C, Clarkson J, Kilonzo M, *et al.*

**No. 17**

Randomised controlled trials of conventional antipsychotic versus new atypical drugs, and new atypical drugs versus clozapine, in people with schizophrenia responding poorly to, or intolerant of, current drug treatment.

By Lewis SW, Davies L, Jones PB, Barnes TRE, Murray RM, Kerwin R, *et al.*

**No. 18**

Diagnostic tests and algorithms used in the investigation of haematuria: systematic reviews and economic evaluation.

By Rodgers M, Nixon J, Hempel S, Aho T, Kelly J, Neal D, *et al.*

**No. 19**

Cognitive behavioural therapy in addition to antispasmodic therapy for irritable bowel syndrome in primary care: randomised controlled trial.

By Kennedy TM, Chalder T, McCrone P, Darnley S, Knapp M, Jones RH, *et al.*

**No. 20**

A systematic review of the clinical effectiveness and cost-effectiveness of enzyme replacement therapies for Fabry's disease and mucopolysaccharidosis type 1.

By Connock M, Juarez-Garcia A, Frew E, Mans A, Dretzke J, Fry-Smith A, *et al.*

**No. 21**

Health benefits of antiviral therapy for mild chronic hepatitis C: randomised controlled trial and economic evaluation.

By Wright M, Grieve R, Roberts J, Main J, Thomas HC, on behalf of the UK Mild Hepatitis C Trial Investigators.

**No. 22**

Pressure relieving support surfaces: a randomised evaluation.

By Nixon J, Nelson EA, Cranny G, Iglesias CP, Hawkins K, Cullum NA, *et al.*

**No. 23**

A systematic review and economic model of the effectiveness and cost-effectiveness of methylphenidate, dexamfetamine and atomoxetine for the treatment of attention deficit hyperactivity disorder in children and adolescents.

By King S, Griffin S, Hodges Z, Weatherly H, Asseburg C, Richardson G, *et al.*

**No. 24**

The clinical effectiveness and cost-effectiveness of enzyme replacement therapy for Gaucher's disease: a systematic review.

By Connock M, Burls A, Frew E, Fry-Smith A, Juarez-Garcia A, McCabe C, *et al.*

**No. 25**

Effectiveness and cost-effectiveness of salicylic acid and cryotherapy for cutaneous warts. An economic decision model.

By Thomas KS, Keogh-Brown MR, Chalmers JR, Fordham RJ, Holland RC, Armstrong SJ, *et al.*

**No. 26**

A systematic literature review of the effectiveness of non-pharmacological interventions to prevent wandering in dementia and evaluation of the ethical implications and acceptability of their use.

By Robinson L, Hutchings D, Corner L, Beyer F, Dickinson H, Vanoli A, *et al.*

**No. 27**

A review of the evidence on the effects and costs of implantable cardioverter defibrillator therapy in different patient groups, and modelling of cost-effectiveness and cost-utility for these groups in a UK context.

By Buxton M, Caine N, Chase D, Connelly D, Grace A, Jackson C, *et al.*

**No. 28**

Adefovir dipivoxil and pegylated interferon alfa-2a for the treatment of chronic hepatitis B: a systematic review and economic evaluation.

By Shepherd J, Jones J, Takeda A, Davidson P, Price A.

**No. 29**

An evaluation of the clinical and cost-effectiveness of pulmonary artery catheters in patient management in intensive care: a systematic review and a randomised controlled trial.

By Harvey S, Stevens K, Harrison D, Young D, Brampton W, McCabe C, *et al.*

**No. 30**

Accurate, practical and cost-effective assessment of carotid stenosis in the UK.

By Wardlaw JM, Chappell FM, Stevenson M, De Nigris E, Thomas S, Gillard J, *et al.*

**No. 31**

Etanercept and infliximab for the treatment of psoriatic arthritis: a systematic review and economic evaluation.

By Woolacott N, Bravo Vergel Y, Hawkins N, Kainth A, Khadjesari Z, Misso K, *et al.*

**No. 32**

The cost-effectiveness of testing for hepatitis C in former injecting drug users.

By Castelnovo E, Thompson-Coon J, Pitt M, Cramp M, Siebert U, Price A, *et al.*

**No. 33**

Computerised cognitive behaviour therapy for depression and anxiety update: a systematic review and economic evaluation.

By Kaltenthaler E, Brazier J, De Nigris E, Tumor I, Ferriter M, Beverley C, *et al.*

**No. 34**

Cost-effectiveness of using prognostic information to select women with breast cancer for adjuvant systemic therapy.

By Williams C, Brunskill S, Altman D, Briggs A, Campbell H, Clarke M, *et al.*

**No. 35**

Psychological therapies including dialectical behaviour therapy for borderline personality disorder: a systematic review and preliminary economic evaluation.

By Brazier J, Tumor I, Holmes M, Ferriter M, Parry G, Dent-Brown K, *et al.*

**No. 36**

Clinical effectiveness and cost-effectiveness of tests for the diagnosis and investigation of urinary tract infection in children: a systematic review and economic model.

By Whiting P, Westwood M, Bojke L, Palmer S, Richardson G, Cooper J, *et al.*

**No. 37**

Cognitive behavioural therapy in chronic fatigue syndrome: a randomised controlled trial of an outpatient group programme.

By O'Dowd H, Gladwell P, Rogers CA, Hollinghurst S, Gregory A.

**No. 38**

A comparison of the cost-effectiveness of five strategies for the prevention of nonsteroidal anti-inflammatory drug-induced gastrointestinal toxicity: a systematic review with economic modelling.

By Brown TJ, Hooper L, Elliott RA, Payne K, Webb R, Roberts C, *et al.*

**No. 39**

The effectiveness and cost-effectiveness of computed tomography screening for coronary artery disease: systematic review.

By Waugh N, Black C, Walker S, McIntyre L, Cummins E, Hillis G.

**No. 40**

What are the clinical outcome and cost-effectiveness of endoscopy undertaken by nurses when compared with doctors? A Multi-Institution Nurse Endoscopy Trial (MINuET).

By Williams J, Russell I, Durai D, Cheung W-Y, Farrin A, Bloor K, *et al.*

**No. 41**

The clinical and cost-effectiveness of oxaliplatin and capecitabine for the adjuvant treatment of colon cancer: systematic review and economic evaluation.

By Pandor A, Eggington S, Paisley S, Tappenden P, Sutcliffe P.

**No. 42**

A systematic review of the effectiveness of adalimumab, etanercept and infliximab for the treatment of rheumatoid arthritis in adults and an economic evaluation of their cost-effectiveness.

By Chen Y-F, Jobanputra P, Barton P, Jowett S, Bryan S, Clark W, *et al.*

**No. 43**

Telemedicine in dermatology: a randomised controlled trial.

By Bowns IR, Collins K, Walters SJ, McDonagh AJG.

**No. 44**

Cost-effectiveness of cell salvage and alternative methods of minimising perioperative allogeneic blood transfusion: a systematic review and economic model.

By Davies L, Brown TJ, Haynes S, Payne K, Elliott RA, McCollum C.

**No. 45**

Clinical effectiveness and cost-effectiveness of laparoscopic surgery for colorectal cancer: systematic reviews and economic evaluation.

By Murray A, Lourenco T, de Verteuil R, Hernandez R, Fraser C, McKinley A, *et al.*

**No. 46**

Etanercept and efalizumab for the treatment of psoriasis: a systematic review.

By Woolacott N, Hawkins N, Mason A, Kainth A, Khadjesari Z, Bravo Vergel Y, *et al.*

**No. 47**

Systematic reviews of clinical decision tools for acute abdominal pain.

By Liu JLY, Wyatt JC, Deeks JJ, Clamp S, Keen J, Verde P, *et al.*

**No. 48**

Evaluation of the ventricular assist device programme in the UK.

By Sharples L, Buxton M, Caine N, Cafferty F, Demiris N, Dyer M, *et al.*

**No. 49**

A systematic review and economic model of the clinical and cost-effectiveness of immunosuppressive therapy for renal transplantation in children.

By Yao G, Albon E, Adi Y, Milford D, Bayliss S, Ready A, *et al.*

**No. 50**

Amniocentesis results: investigation of anxiety. The ARIA trial.

By Hewison J, Nixon J, Fountain J, Cocks K, Jones C, Mason G, *et al.*

**Volume 11, 2007**

**No. 1**

Pemetrexed disodium for the treatment of malignant pleural mesothelioma: a systematic review and economic evaluation.

By Dundar Y, Bagust A, Dickson R, Dodd S, Green J, Haycox A, *et al.*

**No. 2**

A systematic review and economic model of the clinical effectiveness and cost-effectiveness of docetaxel in combination with prednisone or prednisolone for the treatment of hormone-refractory metastatic prostate cancer.

By Collins R, Fenwick E, Trowman R, Perard R, Norman G, Light K, *et al.*

**No. 3**

A systematic review of rapid diagnostic tests for the detection of tuberculosis infection.

By Dinnes J, Deeks J, Kunst H, Gibson A, Cummins E, Waugh N, *et al.*

**No. 4**

The clinical effectiveness and cost-effectiveness of strontium ranelate for the prevention of osteoporotic fragility fractures in postmenopausal women.

By Stevenson M, Davis S, Lloyd-Jones M, Beverley C.

**No. 5**

A systematic review of quantitative and qualitative research on the role and effectiveness of written information available to patients about individual medicines.

By Raynor DK, Blenkinsopp A, Knapp P, Grime J, Nicolson DJ, Pollock K, *et al.*

**No. 6**

Oral naltrexone as a treatment for relapse prevention in formerly opioid-dependent drug users: a systematic review and economic evaluation.

By Adi Y, Juarez-Garcia A, Wang D, Jowett S, Frew E, Day E, *et al.*

**No. 7**

Glucocorticoid-induced osteoporosis: a systematic review and cost-utility analysis.

By Kanis JA, Stevenson M, McCloskey EV, Davis S, Lloyd-Jones M.

**No. 8**

Epidemiological, social, diagnostic and economic evaluation of population screening for genital chlamydial infection.

By Low N, McCarthy A, Macleod J, Salisbury C, Campbell R, Roberts TE, *et al.*

**No. 9**

Methadone and buprenorphine for the management of opioid dependence: a systematic review and economic evaluation.

By Connock M, Juarez-Garcia A, Jowett S, Frew E, Liu Z, Taylor RJ, *et al.*

**No. 10**

Exercise Evaluation Randomised Trial (EXERT): a randomised trial comparing GP referral for leisure centre-based exercise, community-based walking and advice only.

By Isaacs AJ, Critchley JA, See Tai S, Buckingham K, Westley D, Harridge SDR, *et al.*

**No. 11**

Interferon alfa (pegylated and non-pegylated) and ribavirin for the treatment of mild chronic hepatitis C: a systematic review and economic evaluation.

By Shepherd J, Jones J, Hartwell D, Davidson P, Price A, Waugh N.

**No. 12**

Systematic review and economic evaluation of bevacizumab and cetuximab for the treatment of metastatic colorectal cancer.

By Tappenden P, Jones R, Paisley S, Carroll C.

**No. 13**

A systematic review and economic evaluation of epoetin alfa, epoetin beta and darbepoetin alfa in anaemia associated with cancer, especially that attributable to cancer treatment.

By Wilson J, Yao GL, Raftery J, Bohlius J, Brunskill S, Sandercock J, *et al.*

**No. 14**

A systematic review and economic evaluation of statins for the prevention of coronary events.

By Ward S, Lloyd Jones M, Pandor A, Holmes M, Ara R, Ryan A, *et al.*

**No. 15**

A systematic review of the effectiveness and cost-effectiveness of different models of community-based respite care for frail older people and their carers.

By Mason A, Weatherly H, Spilsbury K, Arksey H, Golder S, Adamson J, *et al.*

**No. 16**

Additional therapy for young children with spastic cerebral palsy: a randomised controlled trial.

By Weindling AM, Cunningham CC, Glenn SM, Edwards RT, Reeves DJ.

**No. 17**

Screening for type 2 diabetes: literature review and economic modelling.

By Waugh N, Scotland G, McNamee P, Gillett M, Brennan A, Goyder E, *et al.*

**No. 18**

The effectiveness and cost-effectiveness of cinacalcet for secondary hyperparathyroidism in end-stage renal disease patients on dialysis: a systematic review and economic evaluation.

By Garside R, Pitt M, Anderson R, Mealing S, Roome C, Snaith A, *et al.*

**No. 19**

The clinical effectiveness and cost-effectiveness of gemcitabine for metastatic breast cancer: a systematic review and economic evaluation.

By Takeda AL, Jones J, Loveman E, Tan SC, Clegg AJ.

**No. 20**

A systematic review of duplex ultrasound, magnetic resonance angiography and computed tomography angiography for the diagnosis and assessment of symptomatic, lower limb peripheral arterial disease.

By Collins R, Cranny G, Burch J, Aguiar-Ibáñez R, Craig D, Wright K, *et al.*

**No. 21**

The clinical effectiveness and cost-effectiveness of treatments for children with idiopathic steroid-resistant nephrotic syndrome: a systematic review.

By Colquitt JL, Kirby J, Green C, Cooper K, Trompeter RS.

**No. 22**

A systematic review of the routine monitoring of growth in children of primary school age to identify growth-related conditions.

By Fayer D, Nixon J, Hartley S, Rithalia A, Butler G, Rudolf M, *et al.*

**No. 23**

Systematic review of the effectiveness of preventing and treating *Staphylococcus aureus* carriage in reducing peritoneal catheter-related infections.

By McCormack K, Rabindranath K, Kilonzo M, Vale L, Fraser C, McIntyre L, *et al.*

**No. 24**

The clinical effectiveness and cost of repetitive transcranial magnetic stimulation versus electroconvulsive therapy in severe depression: a multicentre pragmatic randomised controlled trial and economic analysis.

By McLoughlin DM, Mogg A, Eranti S, Pluck G, Purvis R, Edwards D, *et al.*

**No. 25**

A randomised controlled trial and economic evaluation of direct versus indirect and individual versus group modes of speech and language therapy for children with primary language impairment.

By Boyle J, McCartney E, Forbes J, O'Hare A.

**No. 26**

Hormonal therapies for early breast cancer: systematic review and economic evaluation.

By Hind D, Ward S, De Nigris E, Simpson E, Carroll C, Wyld L.



**No. 27**

Cardioprotection against the toxic effects of anthracyclines given to children with cancer: a systematic review.

By Bryant J, Picot J, Levitt G, Sullivan I, Baxter L, Clegg A.

**No. 28**

Adalimumab, etanercept and infliximab for the treatment of ankylosing spondylitis: a systematic review and economic evaluation.

By McLeod C, Bagust A, Boland A, Dagenais P, Dickson R, Dundar Y, *et al.*

**No. 29**

Prenatal screening and treatment strategies to prevent group B streptococcal and other bacterial infections in early infancy: cost-effectiveness and expected value of information analyses.

By Colbourn T, Asseburg C, Bojke L, Philips Z, Claxton K, Ades AE, *et al.*

**No. 30**

Clinical effectiveness and cost-effectiveness of bone morphogenetic proteins in the non-healing of fractures and spinal fusion: a systematic review.

By Garrison KR, Donell S, Ryder J, Shemilt I, Mugford M, Harvey I, *et al.*

**No. 31**

A randomised controlled trial of postoperative radiotherapy following breast-conserving surgery in a minimum-risk older population. The PRIME trial.

By Prescott RJ, Kunkler IH, Williams LJ, King CC, Jack W, van der Pol M, *et al.*

**No. 32**

Current practice, accuracy, effectiveness and cost-effectiveness of the school entry hearing screen.

By Bamford J, Fortnum H, Bristow K, Smith J, Vamvakas G, Davies L, *et al.*

**No. 33**

The clinical effectiveness and cost-effectiveness of inhaled insulin in diabetes mellitus: a systematic review and economic evaluation.

By Black C, Cummins E, Royle P, Phillip S, Waugh N.

**No. 34**

Surveillance of cirrhosis for hepatocellular carcinoma: systematic review and economic analysis.

By Thompson Coon J, Rogers G, Hewson P, Wright D, Anderson R, Cramp M, *et al.*

**No. 35**

The Birmingham Rehabilitation Uptake Maximisation Study (BRUM). Homebased compared with hospital-based cardiac rehabilitation in a multi-ethnic population: cost-effectiveness and patient adherence.

By Jolly K, Taylor R, Lip GYH, Greenfield S, Raftery J, Mant J, *et al.*

**No. 36**

A systematic review of the clinical, public health and cost-effectiveness of rapid diagnostic tests for the detection and identification of bacterial intestinal pathogens in faeces and food.

By Abubakar I, Irvine L, Aldus CF, Wyatt GM, Fordham R, Schelenz S, *et al.*

**No. 37**

A randomised controlled trial examining the longer-term outcomes of standard versus new antiepileptic drugs. The SANAD trial.

By Marson AG, Appleton R, Baker GA, Chadwick DW, Doughty J, Eaton B, *et al.*

**No. 38**

Clinical effectiveness and cost-effectiveness of different models of managing long-term oral anti-coagulation therapy: a systematic review and economic modelling.

By Connock M, Stevens C, Fry-Smith A, Jowett S, Fitzmaurice D, Moore D, *et al.*

**No. 39**

A systematic review and economic model of the clinical effectiveness and cost-effectiveness of interventions for preventing relapse in people with bipolar disorder.

By Soares-Weiser K, Bravo Vergel Y, Beynon S, Dunn G, Barbieri M, Duffy S, *et al.*

**No. 40**

Taxanes for the adjuvant treatment of early breast cancer: systematic review and economic evaluation.

By Ward S, Simpson E, Davis S, Hind D, Rees A, Wilkinson A.

**No. 41**

The clinical effectiveness and cost-effectiveness of screening for open angle glaucoma: a systematic review and economic evaluation.

By Burr JM, Mowatt G, Hernández R, Siddiqui MAR, Cook J, Lourenco T, *et al.*

**No. 42**

Acceptability, benefit and costs of early screening for hearing disability: a study of potential screening tests and models.

By Davis A, Smith P, Ferguson M, Stephens D, Gianopoulos I.

**No. 43**

Contamination in trials of educational interventions.

By Keogh-Brown MR, Bachmann MO, Shepstone L, Hewitt C, Howe A, Ramsay CR, *et al.*

**No. 44**

Overview of the clinical effectiveness of positron emission tomography imaging in selected cancers.

By Facey K, Bradbury I, Laking G, Payne E.

**No. 45**

The effectiveness and cost-effectiveness of carmustine implants and temozolomide for the treatment of newly diagnosed high-grade glioma: a systematic review and economic evaluation.

By Garside R, Pitt M, Anderson R, Rogers G, Dyer M, Mealing S, *et al.*

**No. 46**

Drug-eluting stents: a systematic review and economic evaluation.

By Hill RA, Boland A, Dickson R, Dündar Y, Haycox A, McLeod C, *et al.*

**No. 47**

The clinical effectiveness and cost-effectiveness of cardiac resynchronisation (biventricular pacing) for heart failure: systematic review and economic model.

By Fox M, Mealing S, Anderson R, Dean J, Stein K, Price A, *et al.*

**No. 48**

Recruitment to randomised trials: strategies for trial enrolment and participation study. The STEPS study.

By Campbell MK, Snowdon C, Francis D, Elbourne D, McDonald AM, Knight R, *et al.*

**No. 49**

Cost-effectiveness of functional cardiac testing in the diagnosis and management of coronary artery disease: a randomised controlled trial. The CECaT trial.

By Sharples L, Hughes V, Crean A, Dyer M, Buxton M, Goldsmith K, *et al.*

**No. 50**

Evaluation of diagnostic tests when there is no gold standard. A review of methods.

By Rutjes AWS, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PMM.

**No. 51**

Systematic reviews of the clinical effectiveness and cost-effectiveness of proton pump inhibitors in acute upper gastrointestinal bleeding.

By Leontiadis GI, Sreedharan A, Dorward S, Barton P, Delaney B, Howden CW, *et al.*

**No. 52**

A review and critique of modelling in prioritising and designing screening programmes.

By Karnon J, Goyder E, Tappenden P, McPhie S, Towers I, Brazier J, *et al.*

**No. 53**

An assessment of the impact of the NHS Health Technology Assessment Programme.

By Hanney S, Buxton M, Green C, Coulson D, Raftery J.

**Volume 12, 2008**

**No. 1**

A systematic review and economic model of switching from nonglycopeptide to glycopeptide antibiotic prophylaxis for surgery.

By Cranny G, Elliott R, Weatherly H, Chambers D, Hawkins N, Myers L, *et al.*

**No. 2**

'Cut down to quit' with nicotine replacement therapies in smoking cessation: a systematic review of effectiveness and economic analysis.

By Wang D, Connock M, Barton P, Fry-Smith A, Aveyard P, Moore D.

**No. 3**

A systematic review of the effectiveness of strategies for reducing fracture risk in children with juvenile idiopathic arthritis with additional data on long-term risk of fracture and cost of disease management.

By Thornton J, Ashcroft D, O'Neill T, Elliott R, Adams J, Roberts C, *et al.*

**No. 4**

Does befriending by trained lay workers improve psychological well-being and quality of life for carers of people with dementia, and at what cost? A randomised controlled trial.

By Charlesworth G, Shepstone L, Wilson E, Thalanany M, Mugford M, Poland F.

**No. 5**

A multi-centre retrospective cohort study comparing the efficacy, safety and cost-effectiveness of hysterectomy and uterine artery embolisation for the treatment of symptomatic uterine fibroids. The HOPEFUL study.

By Hirst A, Dutton S, Wu O, Briggs A, Edwards C, Waldenmaier L, *et al.*

**No. 6**

Methods of prediction and prevention of pre-eclampsia: systematic reviews of accuracy and effectiveness literature with economic modelling.

By Meads CA, Cnossen JS, Meher S, Juarez-Garcia A, ter Riet G, Duley L, *et al.*

**No. 7**

The use of economic evaluations in NHS decision-making: a review and empirical investigation.

By Williams I, McIver S, Moore D, Bryan S.

**No. 8**

Stapled haemorrhoidectomy (haemorrhoidopexy) for the treatment of haemorrhoids: a systematic review and economic evaluation.

By Burch J, Epstein D, Baba-Akbari A, Weatherly H, Fox D, Golder S, *et al.*

**No. 9**

The clinical effectiveness of diabetes education models for Type 2 diabetes: a systematic review.

By Loveman E, Frampton GK, Clegg AJ.

**No. 10**

Payment to healthcare professionals for patient recruitment to trials: systematic review and qualitative study.

By Raftery J, Bryant J, Powell J, Kerr C, Hawker S.

**No. 11**

Cyclooxygenase-2 selective non-steroidal anti-inflammatory drugs (etodolac, meloxicam, celecoxib, rofecoxib, etoricoxib, valdecoxib and lumiracoxib) for osteoarthritis and rheumatoid arthritis: a systematic review and economic evaluation.

By Chen Y-F, Jobanputra P, Barton P, Bryan S, Fry-Smith A, Harris G, *et al.*

**No. 12**

The clinical effectiveness and cost-effectiveness of central venous catheters treated with anti-infective agents in preventing bloodstream infections: a systematic review and economic evaluation.

By Hockenhull JC, Dwan K, Boland A, Smith G, Bagust A, Dundar Y, *et al.*

**No. 13**

Stepped treatment of older adults on laxatives. The STOOL trial.

By Mihaylov S, Stark C, McColl E, Steen N, Vanoli A, Rubin G, *et al.*

**No. 14**

A randomised controlled trial of cognitive behaviour therapy in adolescents with major depression treated by selective serotonin reuptake inhibitors. The ADAPT trial.

By Goodyer IM, Dubicka B, Wilkinson P, Kelvin R, Roberts C, Byford S, *et al.*

**No. 15**

The use of irinotecan, oxaliplatin and raltitrexed for the treatment of advanced colorectal cancer: systematic review and economic evaluation.

By Hind D, Tappenden P, Tumor I, Eggington E, Sutcliffe P, Ryan A.

**No. 16**

Ranibizumab and pegaptanib for the treatment of age-related macular degeneration: a systematic review and economic evaluation.

By Colquitt JL, Jones J, Tan SC, Takeda A, Clegg AJ, Price A.

**No. 17**

Systematic review of the clinical effectiveness and cost-effectiveness of 64-slice or higher computed tomography angiography as an alternative to invasive coronary angiography in the investigation of coronary artery disease.

By Mowatt G, Cummins E, Waugh N, Walker S, Cook J, Jia X, *et al.*

**No. 18**

Structural neuroimaging in psychosis: a systematic review and economic evaluation.

By Albon E, Tsourapas A, Frew E, Davenport C, Oyeboode F, Bayliss S, *et al.*

**No. 19**

Systematic review and economic analysis of the comparative effectiveness of different inhaled corticosteroids and their usage with long-acting beta<sub>2</sub> agonists for the treatment of chronic asthma in adults and children aged 12 years and over.

By Shepherd J, Rogers G, Anderson R, Main C, Thompson-Coon J, Hartwell D, *et al.*

**No. 20**

Systematic review and economic analysis of the comparative effectiveness of different inhaled corticosteroids and their usage with long-acting beta<sub>2</sub> agonists for the treatment of chronic asthma in children under the age of 12 years.

By Main C, Shepherd J, Anderson R, Rogers G, Thompson-Coon J, Liu Z, *et al.*

**No. 21**

Ezetimibe for the treatment of hypercholesterolaemia: a systematic review and economic evaluation.

By Ara R, Tumor I, Pandor A, Duenas A, Williams R, Wilkinson A, *et al.*

**No. 22**

Topical or oral ibuprofen for chronic knee pain in older people. The TOIB study.

By Underwood M, Ashby D, Carnes D, Castelnuovo E, Cross P, Harding G, *et al.*

**No. 23**

A prospective randomised comparison of minor surgery in primary and secondary care. The MiSTIC trial.

By George S, Pockney P, Primrose J, Smith H, Little P, Kinley H, *et al.*

**No. 24**

A review and critical appraisal of measures of therapist-patient interactions in mental health settings.

By Cahill J, Barkham M, Hardy G, Gilbody S, Richards D, Bower P, *et al.*

**No. 25**

The clinical effectiveness and cost-effectiveness of screening programmes for amblyopia and strabismus in children up to the age of 4–5 years: a systematic review and economic evaluation.

By Carlton J, Karnon J, Czoski-Murray C, Smith KJ, Marr J.

**No. 26**

A systematic review of the clinical effectiveness and cost-effectiveness and economic modelling of minimal incision total hip replacement approaches in the management of arthritic disease of the hip.

By de Verteuil R, Imamura M, Zhu S, Glazener C, Fraser C, Munro N, *et al.*

**No. 27**

A preliminary model-based assessment of the cost-utility of a screening programme for early age-related macular degeneration.

By Karnon J, Czoski-Murray C, Smith K, Brand C, Chakravarthy U, Davis S, *et al.*

**No. 28**

Intravenous magnesium sulphate and sotalol for prevention of atrial fibrillation after coronary artery bypass surgery: a systematic review and economic evaluation.

By Shepherd J, Jones J, Frampton GK, Tanajewski L, Turner D, Price A.

**No. 29**

Absorbent products for urinary/faecal incontinence: a comparative evaluation of key product categories.

By Fader M, Cottenden A, Getliffe K, Gage H, Clarke-O'Neill S, Jamieson K, *et al.*

**No. 30**

A systematic review of repetitive functional task practice with modelling of resource use, costs and effectiveness.

By French B, Leathley M, Sutton C, McAdam J, Thomas L, Forster A, *et al.*

**No. 31**

The effectiveness and cost-effectiveness of minimal access surgery amongst people with gastro-oesophageal reflux disease – a UK collaborative study. The REFLUX trial.

By Grant A, Wileman S, Ramsay C, Bojke L, Epstein D, Sculpher M, *et al.*

**No. 32**

Time to full publication of studies of anti-cancer medicines for breast cancer and the potential for publication bias: a short systematic review.

By Takeda A, Loveman E, Harris P, Hartwell D, Welch K.

**No. 33**

Performance of screening tests for child physical abuse in accident and emergency departments.

By Woodman J, Pitt M, Wentz R, Taylor B, Hodes D, Gilbert RE.

**No. 34**

Curative catheter ablation in atrial fibrillation and typical atrial flutter: systematic review and economic evaluation.

By Rodgers M, McKenna C, Palmer S, Chambers D, Van Hout S, Golder S, *et al.*

**No. 35**

Systematic review and economic modelling of effectiveness and cost utility of surgical treatments for men with benign prostatic enlargement.

By Lourenco T, Armstrong N, N'Dow J, Nabi G, Deverill M, Pickard R, *et al.*

**No. 36**

Immunoprophylaxis against respiratory syncytial virus (RSV) with palivizumab in children: a systematic review and economic evaluation.

By Wang D, Cummins C, Bayliss S, Sandercock J, Burls A.

**Volume 13, 2009****No. 1**

Deferasirox for the treatment of iron overload associated with regular blood transfusions (transfusional haemosiderosis) in patients suffering with chronic anaemia: a systematic review and economic evaluation.

By McLeod C, Fleeman N, Kirkham J, Bagust A, Boland A, Chu P, *et al.*

**No. 2**

Thrombophilia testing in people with venous thromboembolism: systematic review and cost-effectiveness analysis.

By Simpson EL, Stevenson MD, Rawdin A, Papaioannou D.

**No. 3**

Surgical procedures and non-surgical devices for the management of non-apnoeic snoring: a systematic review of clinical effects and associated treatment costs.

By Main C, Liu Z, Welch K, Weiner G, Quentin Jones S, Stein K.

**No. 4**

Continuous positive airway pressure devices for the treatment of obstructive sleep apnoea-hypopnoea syndrome: a systematic review and economic analysis.

By McDaid C, Griffin S, Weatherly H, Durée K, van der Burgt M, van Hout S, Akers J, *et al.*

**No. 5**

Use of classical and novel biomarkers as prognostic risk factors for localised prostate cancer: a systematic review.

By Sutcliffe P, Hummel S, Simpson E, Young T, Rees A, Wilkinson A, *et al.*

**No. 6**

The harmful health effects of recreational ecstasy: a systematic review of observational evidence.

By Rogers G, Elston J, Garside R, Roome C, Taylor R, Younger P, *et al.*

**No. 7**

Systematic review of the clinical effectiveness and cost-effectiveness of oesophageal Doppler monitoring in critically ill and high-risk surgical patients.

By Mowatt G, Houston G, Hernández R, de Verteuil R, Fraser C, Cuthbertson B, *et al.*

**No. 8**

The use of surrogate outcomes in model-based cost-effectiveness analyses: a survey of UK Health Technology Assessment reports.

By Taylor RS, Elston J.

**No. 9**

Controlling Hypertension and Hypotension Immediately Post Stroke (CHHIPS) – a randomised controlled trial.

By Potter J, Mistri A, Brodie F, Chernova J, Wilson E, Jagger C, *et al.*

**No. 10**

Routine antenatal anti-D prophylaxis for RhD-negative women: a systematic review and economic evaluation.

By Pilgrim H, Lloyd-Jones M, Rees A.

**No. 11**

Amantadine, oseltamivir and zanamivir for the prophylaxis of influenza (including a review of existing guidance no. 67): a systematic review and economic evaluation.

By Tappenden P, Jackson R, Cooper K, Rees A, Simpson E, Read R, *et al.*

**No. 12**

Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods.

By Hobart J, Cano S.

**No. 13**

Treatment of severe ankle sprain: a pragmatic randomised controlled trial comparing the clinical effectiveness and cost-effectiveness of three types of mechanical ankle support with tubular bandage. The CAST trial.

By Cooke MW, Marsh JL, Clark M, Nakash R, Jarvis RM, Hutton JL, *et al.*, on behalf of the CAST trial group.

**No. 14**

Non-occupational postexposure prophylaxis for HIV: a systematic review.

By Bryant J, Baxter L, Hird S.

**No. 15**

Blood glucose self-monitoring in type 2 diabetes: a randomised controlled trial.

By Farmer AJ, Wade AN, French DP, Simon J, Yudkin P, Gray A, *et al.*

**No. 16**

How far does screening women for domestic (partner) violence in different health-care settings meet criteria for a screening programme? Systematic reviews of nine UK National Screening Committee criteria.

By Feder G, Ramsay J, Dunne D, Rose M, Arsene C, Norman R, *et al.*

**No. 17**

Spinal cord stimulation for chronic pain of neuropathic or ischaemic origin: systematic review and economic evaluation.

By Simpson, EL, Duenas A, Holmes MW, Papaioannou D, Chilcott J.

**No. 18**

The role of magnetic resonance imaging in the identification of suspected acoustic neuroma: a systematic review of clinical and cost-effectiveness and natural history.

By Fortnum H, O'Neill C, Taylor R, Lenthall R, Nikolopoulos T, Lightfoot G, *et al.*

**No. 19**

Dipsticks and diagnostic algorithms in urinary tract infection: development and validation, randomised trial, economic analysis, observational cohort and qualitative study.

By Little P, Turner S, Rumsby K, Warner G, Moore M, Lowes JA, *et al.*

**No. 20**

Systematic review of respite care in the frail elderly.

By Shaw C, McNamara R, Abrams K, Cannings-John R, Hood K, Longo M, *et al.*

**No. 21**

Neuroleptics in the treatment of aggressive challenging behaviour for people with intellectual disabilities: a randomised controlled trial (NACHBID).

By Tyrer P, Oliver-Africano P, Romeo R, Knapp M, Dickens S, Bouras N, *et al.*

**No. 22**

Randomised controlled trial to determine the clinical effectiveness and cost-effectiveness of selective serotonin reuptake inhibitors plus supportive care, versus supportive care alone, for mild to moderate depression with somatic symptoms in primary care: the THREAD (THREShold for AntiDepressant response) study.

By Kendrick T, Chatwin J, Dowrick C, Tylee A, Morriss R, Peveler R, *et al.*

**No. 23**

Diagnostic strategies using DNA testing for hereditary haemochromatosis in at-risk populations: a systematic review and economic evaluation.

By Bryant J, Cooper K, Picot J, Clegg A, Roderick P, Rosenberg W, *et al.*

**No. 24**

Enhanced external counterpulsation for the treatment of stable angina and heart failure: a systematic review and economic analysis.

By McKenna C, McDaid C, Suekarran S, Hawkins N, Claxton K, Light K, *et al.*

**No. 25**

Development of a decision support tool for primary care management of patients with abnormal liver function tests without clinically apparent liver disease: a record-linkage population cohort study and decision analysis (ALFIE).

By Donnan PT, McLernon D, Dillon JF, Ryder S, Roderick P, Sullivan F, *et al.*

**No. 26**

A systematic review of presumed consent systems for deceased organ donation.

By Rithalia A, McDaid C, Suekarran S, Norman G, Myers L, Sowden A.

**No. 27**

Paracetamol and ibuprofen for the treatment of fever in children: the PITCH randomised controlled trial.

By Hay AD, Redmond NM, Costelloe C, Montgomery AA, Fletcher M, Hollinghurst S, *et al.*

**No. 28**

A randomised controlled trial to compare minimally invasive glucose monitoring devices with conventional monitoring in the management of insulin-treated diabetes mellitus (MITRE).

By Newman SP, Cooke D, Casbard A, Walker S, Meredith S, Nunn A, *et al.*

**No. 29**

Sensitivity analysis in economic evaluation: an audit of NICE current practice and a review of its use and value in decision-making.

By Andronis L, Barton P, Bryan S.

**Suppl. 1**

Trastuzumab for the treatment of primary breast cancer in HER2-positive women: a single technology appraisal.

By Ward S, Pilgrim H, Hind D.

Docetaxel for the adjuvant treatment of early node-positive breast cancer: a single technology appraisal.

By Chilcott J, Lloyd Jones M, Wilkinson A.

The use of paclitaxel in the management of early stage breast cancer.

By Griffin S, Dunn G, Palmer S, Macfarlane K, Brent S, Dyker A, *et al.*

Rituximab for the first-line treatment of stage III/IV follicular non-Hodgkin's lymphoma.

By Dundar Y, Bagust A, Hounsome J, McLeod C, Boland A, Davis H, *et al.*

Bortezomib for the treatment of multiple myeloma patients.

By Green C, Bryant J, Takeda A, Cooper K, Clegg A, Smith A, *et al.*

Fludarabine phosphate for the first-line treatment of chronic lymphocytic leukaemia.

By Walker S, Palmer S, Erhorn S, Brent S, Dyker A, Ferrie L, *et al.*

Erlotinib for the treatment of relapsed non-small cell lung cancer.

By McLeod C, Bagust A, Boland A, Hockenhull J, Dundar Y, Proudlove C, *et al.*

Cetuximab plus radiotherapy for the treatment of locally advanced squamous cell carcinoma of the head and neck.

By Griffin S, Walker S, Sculpher M, White S, Erhorn S, Brent S, *et al.*

Infliximab for the treatment of adults with psoriasis.

By Loveman E, Turner D, Hartwell D, Cooper K, Clegg A.

**No. 30**

Psychological interventions for postnatal depression: cluster randomised trial and economic evaluation. The PoNDER trial.

By Morrell CJ, Warner R, Slade P, Dixon S, Walters S, Paley G, *et al.*

**No. 31**

The effect of different treatment durations of clopidogrel in patients with non-ST-segment elevation acute coronary syndromes: a systematic review and value of information analysis.

By Rogowski R, Burch J, Palmer S, Craigs C, Golder S, Woolcott N.

**No. 32**

Systematic review and individual patient data meta-analysis of diagnosis of heart failure, with modelling of implications of different diagnostic strategies in primary care.

By Mant J, Doust J, Roalfe A, Barton P, Cowie MR, Glasziou P, *et al.*

**No. 33**

A multicentre randomised controlled trial of the use of continuous positive airway pressure and non-invasive positive pressure ventilation in the early treatment of patients presenting to the emergency department with severe acute cardiogenic pulmonary oedema: the 3CPO trial.

By Gray AJ, Goodacre S, Newby DE, Masson MA, Sampson F, Dixon S, *et al.*, on behalf of the 3CPO study investigators.

**No. 34**

Early high-dose lipid-lowering therapy to avoid cardiac events: a systematic review and economic evaluation.

By Ara R, Pandor A, Stevens J, Rees A, Rafia R.

**No. 35**

Adefovir dipivoxil and pegylated interferon alpha for the treatment of chronic hepatitis B: an updated systematic review and economic evaluation.

By Jones J, Shepherd J, Baxter L, Gospodarevskaya E, Hartwell D, Harris P, *et al.*

**No. 36**

Methods to identify postnatal depression in primary care: an integrated evidence synthesis and value of information analysis.

By Hewitt CE, Gilbody SM, Brealey S, Paulden M, Palmer S, Mann R, *et al.*

**No. 37**

A double-blind randomised placebo-controlled trial of topical intranasal corticosteroids in 4- to 11-year-old children with persistent bilateral otitis media with effusion in primary care.

By Williamson I, Benges S, Barton S, Petrou S, Letley L, Faisey N, *et al.*

**No. 38**

The effectiveness and cost-effectiveness of methods of storing donated kidneys from deceased donors: a systematic review and economic model.

By Bond M, Pitt M, Akoh J, Moxham T, Hoyle M, Anderson R.

**No. 39**

Rehabilitation of older patients: day hospital compared with rehabilitation at home. A randomised controlled trial.

By Parker SG, Oliver P, Pennington M, Bond J, Jagger C, Enderby PM, *et al.*

**No. 40**

Breastfeeding promotion for infants in neonatal units: a systematic review and economic analysis

By Renfrew MJ, Craig D, Dyson L, McCormick F, Rice S, King SE, *et al.*

**No. 41**

The clinical effectiveness and cost-effectiveness of bariatric (weight loss) surgery for obesity: a systematic review and economic evaluation.

By Picot J, Jones J, Colquitt JL, Gospodarevskaya E, Loveman E, Baxter L, *et al.*

**No. 42**

Rapid testing for group B streptococcus during labour: a test accuracy study with evaluation of acceptability and cost-effectiveness.

By Daniels J, Gray J, Pattison H, Roberts T, Edwards E, Milner P, *et al.*

**No. 43**

Screening to prevent spontaneous preterm birth: systematic reviews of accuracy and effectiveness literature with economic modelling.

By Honest H, Forbes CA, Durée KH, Norman G, Duffy SB, Tsourapas A, *et al.*

**No. 44**

The effectiveness and cost-effectiveness of cochlear implants for severe to profound deafness in children and adults: a systematic review and economic model.

By Bond M, Mealing S, Anderson R, Elston J, Weiner G, Taylor RS, *et al.*

**Suppl. 2**

Gemcitabine for the treatment of metastatic breast cancer.

By Jones J, Takeda A, Tan SC, Cooper K, Loveman E, Clegg A.

Varenicline in the management of smoking cessation: a single technology appraisal.

By Hind D, Tappenden P, Peters J, Kenjegalieva K.

Alteplase for the treatment of acute ischaemic stroke: a single technology appraisal.

By Lloyd Jones M, Holmes M.

Rituximab for the treatment of rheumatoid arthritis.

By Bagust A, Boland A, Hockenhull J, Fleeman N, Greenhalgh J, Dundar Y, *et al.*

Omalizumab for the treatment of severe persistent allergic asthma.

By Jones J, Shepherd J, Hartwell D, Harris P, Cooper K, Takeda A, *et al.*

Rituximab for the treatment of relapsed or refractory stage III or IV follicular non-Hodgkin's lymphoma.

By Boland A, Bagust A, Hockenhull J, Davis H, Chu P, Dickson R.

Adalimumab for the treatment of psoriasis.

By Turner D, Picot J, Cooper K, Loveman E.

Dabigatran etexilate for the prevention of venous thromboembolism in patients undergoing elective hip and knee surgery: a single technology appraisal.

By Holmes M, C Carroll C, Papaioannou D.

Romiplostim for the treatment of chronic immune or idiopathic thrombocytopenic purpura: a single technology appraisal.

By Mowatt G, Boachie C, Crowther M, Fraser C, Hernández R, Jia X, *et al.*

Sunitinib for the treatment of gastrointestinal stromal tumours: a critique of the submission from Pfizer.

By Bond M, Hoyle M, Moxham T, Napier M, Anderson R.

**No. 45**

Vitamin K to prevent fractures in older women: systematic review and economic evaluation.

By Stevenson M, Lloyd-Jones M, Papaioannou D.

**No. 46**

The effects of biofeedback for the treatment of essential hypertension: a systematic review.

By Greenhalgh J, Dickson R, Dundar Y.

**No. 47**

A randomised controlled trial of the use of aciclovir and/or prednisolone for the early treatment of Bell's palsy: the BELLS study.

By Sullivan FM, Swan IRC, Donnan PT, Morrison JM, Smith BH, McKinstry B, *et al.*

**Suppl. 3**

Lapatinib for the treatment of HER2-overexpressing breast cancer.

By Jones J, Takeda A, Picot J, von Keyserlingk C, Clegg A.

Infliximab for the treatment of ulcerative colitis.

By Hyde C, Bryan S, Juarez-Garcia A, Andronis L, Fry-Smith A.

Rimonabant for the treatment of overweight and obese people.

By Burch J, McKenna C, Palmer S, Norman G, Glanville J, Sculpher M, *et al.*

Telbivudine for the treatment of chronic hepatitis B infection.

By Hartwell D, Jones J, Harris P, Cooper K.

Entecavir for the treatment of chronic hepatitis B infection.

By Shepherd J, Gospodarevskaya E, Frampton G, Cooper K.

Febuxostat for the treatment of hyperuricaemia in people with gout: a single technology appraisal.

By Stevenson M, Pandor A.

Rivaroxaban for the prevention of venous thromboembolism: a single technology appraisal.

By Stevenson M, Scope A, Holmes M, Rees A, Kaltenthaler E.

Cetuximab for the treatment of recurrent and/or metastatic squamous cell carcinoma of the head and neck.

By Greenhalgh J, Bagust A, Boland A, Fleeman N, McLeod C, Dundar Y, *et al.*

Mifamurtide for the treatment of osteosarcoma: a single technology appraisal.

By Pandor A, Fitzgerald P, Stevenson M, Papaioannou D.

Ustekinumab for the treatment of moderate to severe psoriasis.

By Gospodarevskaya E, Picot J, Cooper K, Loveman E, Takeda A.

**No. 48**

Endovascular stents for abdominal aortic aneurysms: a systematic review and economic model.

By Chambers D, Epstein D, Walker S, Fayter D, Paton F, Wright K, *et al.*

**No. 49**

Clinical and cost-effectiveness of epoprostenol, iloprost, bosentan, sitaxentan and sildenafil for pulmonary arterial hypertension within their licensed indications: a systematic review and economic evaluation.

By Chen Y-F, Jowett S, Barton P, Malottki K, Hyde C, Gibbs JSR, *et al.*

**No. 50**

Cessation of attention deficit hyperactivity disorder drugs in the young (CADDY) – a pharmacoepidemiological and qualitative study.

By Wong ICK, Asherson P, Bilbow A, Clifford S, Coghill D, R DeSoysa R, *et al.*

**No. 51**

ARTISTIC: a randomised trial of human papillomavirus (HPV) testing in primary cervical screening.

By Kitchener HC, Almonte M, Gilham C, Dowie R, Stoykova B, Sargent A, *et al.*

**No. 52**

The clinical effectiveness of glucosamine and chondroitin supplements in slowing or arresting progression of osteoarthritis of the knee: a systematic review and economic evaluation.

By Black C, Clar C, Henderson R, MacEachern C, McNamee P, Quayyum Z, *et al.*

**No. 53**

Randomised preference trial of medical versus surgical termination of pregnancy less than 14 weeks' gestation (TOPS).

By Robson SC, Kelly T, Howel D, Deverill M, Hewison J, Lie MLS, *et al.*

**No. 54**

Randomised controlled trial of the use of three dressing preparations in the management of chronic ulceration of the foot in diabetes.

By Jeffcoate WJ, Price PE, Phillips CJ, Game FL, Mudge E, Davies S, *et al.*

**No. 55**

VenUS II: a randomised controlled trial of larval therapy in the management of leg ulcers.

By Dumville JC, Worthy G, Soares MO, Bland JM, Cullum N, Dowson C, *et al.*

**No. 56**

A prospective randomised controlled trial and economic modelling of antimicrobial silver dressings versus non-adherent control dressings for venous leg ulcers: the VULCAN trial

By Michaels JA, Campbell WB, King BM, MacIntyre J, Palfreyman SJ, Shackley P, *et al.*

**No. 57**

Communication of carrier status information following universal newborn screening for sickle cell disorders and cystic fibrosis: qualitative study of experience and practice.

By Kai J, Ulph F, Cullinan T, Qureshi N.

**No. 58**

Antiviral drugs for the treatment of influenza: a systematic review and economic evaluation.

By Burch J, Paulden M, Conti S, Stock C, Corbett M, Welton NJ, *et al.*

**No. 59**

Development of a toolkit and glossary to aid in the adaptation of health technology assessment (HTA) reports for use in different contexts.

By Chase D, Rosten C, Turner S, Hicks N, Milne R.

**No. 60**

Colour vision testing for diabetic retinopathy: a systematic review of diagnostic accuracy and economic evaluation.

By Rodgers M, Hodges R, Hawkins J, Hollingworth W, Duffy S, McKibbin M, *et al.*

**No. 61**

Systematic review of the effectiveness and cost-effectiveness of weight management schemes for the under fives: a short report.

By Bond M, Wyatt K, Lloyd J, Welch K, Taylor R.

**No. 62**

Are adverse effects incorporated in economic models? An initial review of current practice.

By Craig D, McDaid C, Fonseca T, Stock C, Duffy S, Woolacott N.

**Volume 14, 2010**

**No. 1**

Multicentre randomised controlled trial examining the cost-effectiveness of contrast-enhanced high field magnetic resonance imaging in women with primary breast cancer scheduled for wide local excision (COMICE).

By Turnbull LW, Brown SR, Olivier C, Harvey I, Brown J, Drew P, *et al.*

**No. 2**

Bevacizumab, sorafenib tosylate, sunitinib and temsirolimus for renal cell carcinoma: a systematic review and economic evaluation.

By Thompson Coon J, Hoyle M, Green C, Liu Z, Welch K, Moxham T, *et al.*

**No. 3**

The clinical effectiveness and cost-effectiveness of testing for cytochrome P450 polymorphisms in patients with schizophrenia treated with antipsychotics: a systematic review and economic evaluation.

By Fleeman N, McLeod C, Bagust A, Beale S, Boland A, Dundar Y, *et al.*

**No. 4**

Systematic review of the clinical effectiveness and cost-effectiveness of photodynamic diagnosis and urine biomarkers (FISH, ImmunoCyt, NMP22) and cytology for the detection and follow-up of bladder cancer.

By Mowatt G, Zhu S, Kilonzo M, Boachie C, Fraser C, Griffiths TRL, *et al.*

**No. 5**

Effectiveness and cost-effectiveness of arthroscopic lavage in the treatment of osteoarthritis of the knee: a mixed methods study of the feasibility of conducting a surgical placebo-controlled trial (the KORAL study).

By Campbell MK, Skea ZC, Sutherland AG, Cuthbertson BH, Entwistle VA, McDonald AM, *et al.*

**No. 6**

A randomised 2 × 2 trial of community versus hospital pulmonary rehabilitation for chronic obstructive pulmonary disease followed by telephone or conventional follow-up.

By Waterhouse JC, Walters SJ, Oluboyede Y, Lawson RA.

**No. 7**

The effectiveness and cost-effectiveness of behavioural interventions for the prevention of sexually transmitted infections in young people aged 13–19: a systematic review and economic evaluation.

By Shepherd J, Kavanagh J, Picot J, Cooper K, Harden A, Barnett-Page E, *et al.*

**No. 8**

Dissemination and publication of research findings: an updated review of related biases.

By Song F, Parekh S, Hooper L, Loke YK, Ryder J, Sutton AJ, *et al.*

**No. 9**

The effectiveness and cost-effectiveness of biomarkers for the prioritisation of patients awaiting coronary revascularisation: a systematic review and decision model.

By Hemingway H, Henriksson M, Chen R, Damant J, Fitzpatrick N, Abrams K, *et al.*

**No. 10**

Comparison of case note review methods for evaluating quality and safety in health care.

By Hutchinson A, Coster JE, Cooper KL, McIntosh A, Walters SJ, Bath PA, *et al.*

**No. 11**

Clinical effectiveness and cost-effectiveness of continuous subcutaneous insulin infusion for diabetes: systematic review and economic evaluation.

By Cummins E, Royle P, Snaith A, Greene A, Robertson L, McIntyre L, *et al.*

**No. 12**

Self-monitoring of blood glucose in type 2 diabetes: systematic review.

By Clar C, Barnard K, Cummins E, Royle P, Waugh N.

**No. 13**

North of England and Scotland Study of Tonsillectomy and Adeno-tonsillectomy in Children (NESSTAC): a pragmatic randomised controlled trial with a parallel non-randomised preference study.

By Lock C, Wilson J, Steen N, Eccles M, Mason H, Carrie S, *et al.*

**No. 14**

Multicentre randomised controlled trial of the clinical and cost-effectiveness of a bypass-surgery-first versus a balloon-angioplasty-first revascularisation strategy for severe limb ischaemia due to infrainguinal disease. The Bypass versus Angioplasty in Severe Ischaemia of the Leg (BASIL) trial.

By Bradbury AW, Adam DJ, Bell J, Forbes JF, Fowkes FGR, Gillespie I, *et al.*

**No. 15**

A randomised controlled multicentre trial of treatments for adolescent anorexia nervosa including assessment of cost-effectiveness and patient acceptability – the TOuCAN trial.

By Gowers SG, Clark AF, Roberts C, Byford S, Barrett B, Griffiths A, *et al.*

**No. 16**

Randomised controlled trials for policy interventions: a review of reviews and meta-regression.

By Oliver S, Bagnall AM, Thomas J, Shepherd J, Sowden A, White I, *et al.*

**No. 17**

Paracetamol and selective and non-selective non-steroidal anti-inflammatory drugs (NSAIDs) for the reduction of morphine-related side effects after major surgery: a systematic review.

By McDaid C, Maund E, Rice S, Wright K, Jenkins B, Woolacott N.

**No. 18**

A systematic review of outcome measures used in forensic mental health research with consensus panel opinion.

By Fitzpatrick R, Chambers J, Burns T, Doll H, Fazel S, Jenkinson C, *et al.*

**No. 19**

The clinical effectiveness and cost-effectiveness of topotecan for small cell lung cancer: a systematic review and economic evaluation.

By Loveman E, Jones J, Hartwell D, Bird A, Harris P, Welch K, *et al.*

**No. 20**

Antenatal screening for haemoglobinopathies in primary care: a cohort study and cluster randomised trial to inform a simulation model. The Screening for Haemoglobinopathies in First Trimester (SHIFT) trial.

By Dormandy E, Bryan S, Gulliford MC, Roberts T, Ades T, Calnan M, *et al.*

**No. 21**

Early referral strategies for management of people with markers of renal disease: a systematic review of the evidence of clinical effectiveness, cost-effectiveness and economic analysis.

By Black C, Sharma P, Scotland G, McCullough K, McGurn D, Robertson L, *et al.*

**No. 22**

A randomised controlled trial of cognitive behaviour therapy and motivational interviewing for people with Type 1 diabetes mellitus with persistent sub-optimal glycaemic control: A Diabetes and Psychological Therapies (ADaPT) study.

By Ismail K, Maissi E, Thomas S, Chalder T, Schmidt U, Bartlett J, *et al.*

**No. 23**

A randomised controlled equivalence trial to determine the effectiveness and cost-utility of manual chest physiotherapy techniques in the management of exacerbations of chronic obstructive pulmonary disease (MATREX).

By Cross J, Elender F, Barton G, Clark A, Shepstone L, Blyth A, *et al.*


**No. 24**

A systematic review and economic evaluation of the clinical effectiveness and cost-effectiveness of aldosterone antagonists for postmyocardial infarction heart failure.

By McKenna C, Burch J, Suekarran S, Walker S, Bakhai A, Witte K, *et al.*







# Health Technology Assessment programme

**Director,**  
**Professor Tom Walley,**  
Director, NIHR HTA  
programme, Professor of  
Clinical Pharmacology,  
University of Liverpool

**Deputy Director,**  
**Professor Jon Nicholl,**  
Director, Medical Care Research  
Unit, University of Sheffield

## Prioritisation Strategy Group

### Members

**Chair,**  
**Professor Tom Walley,**  
Director, NIHR HTA  
programme, Professor of  
Clinical Pharmacology,  
University of Liverpool

**Deputy Chair,**  
**Professor Jon Nicholl,**  
Director, Medical Care Research  
Unit, University of Sheffield

Dr Bob Coates,  
Consultant Advisor, NETSCC,  
HTA

Dr Andrew Cook,  
Consultant Advisor, NETSCC,  
HTA

Dr Peter Davidson,  
Director of Science Support,  
NETSCC, HTA

Professor Robin E Ferner,  
Consultant Physician and  
Director, West Midlands Centre  
for Adverse Drug Reactions,  
City Hospital NHS Trust,  
Birmingham

Professor Paul Glasziou,  
Professor of Evidence-Based  
Medicine, University of Oxford

Dr Nick Hicks,  
Director of NHS Support,  
NETSCC, HTA

Dr Edmund Jessop,  
Medical Adviser, National  
Specialist, National  
Commissioning Group (NCG),  
Department of Health, London

Ms Lynn Kerridge,  
Chief Executive Officer,  
NETSCC and NETSCC, HTA

Dr Ruairidh Milne,  
Director of Strategy and  
Development, NETSCC

Ms Kay Pattison,  
Section Head, NHS R&D  
Programme, Department of  
Health

Ms Pamela Young,  
Specialist Programme Manager,  
NETSCC, HTA

## HTA Commissioning Board

### Members

**Programme Director,**  
**Professor Tom Walley,**  
Director, NIHR HTA  
programme, Professor of  
Clinical Pharmacology,  
University of Liverpool

**Chair,**  
**Professor Jon Nicholl,**  
Director, Medical Care Research  
Unit, University of Sheffield

**Deputy Chair,**  
**Dr Andrew Farmer,**  
Senior Lecturer in General  
Practice, Department of  
Primary Health Care,  
University of Oxford

Professor Ann Ashburn,  
Professor of Rehabilitation  
and Head of Research,  
Southampton General Hospital

Professor Deborah Ashby,  
Professor of Medical Statistics,  
Queen Mary, University of  
London

Professor John Cairns,  
Professor of Health Economics,  
London School of Hygiene and  
Tropical Medicine

Professor Peter Croft,  
Director of Primary Care  
Sciences Research Centre, Keele  
University

Professor Nicky Cullum,  
Director of Centre for Evidence-  
Based Nursing, University of  
York

Professor Jenny Donovan,  
Professor of Social Medicine,  
University of Bristol

Professor Steve Halligan,  
Professor of Gastrointestinal  
Radiology, University College  
Hospital, London

Professor Freddie Hamdy,  
Professor of Urology,  
University of Sheffield

Professor Allan House,  
Professor of Liaison Psychiatry,  
University of Leeds

Dr Martin J Landray,  
Reader in Epidemiology,  
Honorary Consultant Physician,  
Clinical Trial Service Unit,  
University of Oxford

Professor Stuart Logan,  
Director of Health & Social  
Care Research, The Peninsula  
Medical School, Universities of  
Exeter and Plymouth

Dr Rafael Perera,  
Lecturer in Medical Statistics,  
Department of Primary Health  
Care, University of Oxford

Professor Ian Roberts,  
Professor of Epidemiology &  
Public Health, London School  
of Hygiene and Tropical  
Medicine

Professor Mark Sculpher,  
Professor of Health Economics,  
University of York

Professor Helen Smith,  
Professor of Primary Care,  
University of Brighton

Professor Kate Thomas,  
Professor of Complementary &  
Alternative Medicine Research,  
University of Leeds

Professor David John  
Torgerson,  
Director of York Trials Unit,  
University of York

Professor Hywel Williams,  
Professor of Dermato-  
Epidemiology, University of  
Nottingham

### Observers

Ms Kay Pattison,  
Section Head, NHS R&D  
Programme, Department of  
Health

Dr Morven Roberts,  
Clinical Trials Manager,  
Medical Research Council

## Diagnostic Technologies and Screening Panel

### Members

**Chair,**  
**Professor Paul Glasziou,**  
Professor of Evidence-Based  
Medicine, University of Oxford

**Deputy Chair,**  
**Dr David Elliman,**  
Consultant Paediatrician and  
Honorary Senior Lecturer,  
Great Ormond Street Hospital,  
London

Professor Judith E Adams,  
Consultant Radiologist,  
Manchester Royal Infirmary,  
Central Manchester &  
Manchester Children's  
University Hospitals NHS  
Trust, and Professor of  
Diagnostic Radiology, Imaging  
Science and Biomedical  
Engineering, Cancer &  
Imaging Sciences, University of  
Manchester

Mr A S Arunkalaivanan,  
Honorary Senior Lecturer,  
University of Birmingham and  
Consultant Urogynaecologist  
and Obstetrician, City Hospital

Dr Dianne Baralle,  
Consultant & Senior Lecturer  
in Clinical Genetics, Human  
Genetics Division & Wessex  
Clinical Genetics Service,  
Southampton, University of  
Southampton

Dr Stephanie Dancer,  
Consultant Microbiologist,  
Hairmyres Hospital, East  
Kilbride

Dr Ron Gray,  
Consultant, National Perinatal  
Epidemiology Unit, Institute of  
Health Sciences, University of  
Oxford

Professor Paul D Griffiths,  
Professor of Radiology,  
Academic Unit of Radiology,  
University of Sheffield

Mr Martin Hooper,  
Service User Representative

Professor Anthony Robert  
Kendrick,  
Professor of Primary  
Medical Care, University of  
Southampton

Dr Susanne M Ludgate,  
Director, Medical Devices  
Agency, London

Dr Anne Mackie,  
Director of Programmes, UK  
National Screening Committee

Dr David Mathew  
Service User Representative

Dr Michael Millar, Lead  
Consultant in Microbiology,  
Department of Pathology &  
Microbiology, Barts and The  
London NHS Trust, Royal  
London Hospital

Mr Stephen Pilling,  
Director, Centre for Outcomes,  
Research & Effectiveness,  
University College London

Mrs Una Rennard,  
Service User Representative

Ms Jane Smith,  
Consultant Ultrasound  
Practitioner, Ultrasound  
Department, Leeds Teaching  
Hospital NHS Trust, Leeds

Dr W Stuart A Smellie,  
Consultant, Bishop Auckland  
General Hospital

Professor Lindsay Wilson  
Turnbull,  
Scientific Director of the  
Centre for Magnetic Resonance  
Investigations and YCR  
Professor of Radiology, Hull  
Royal Infirmary

Dr Alan J Williams,  
Consultant in General  
Medicine, Department of  
Thoracic Medicine, The Royal  
Bournemouth Hospital

### Observers

Dr Tim Elliott,  
Team Leader, Cancer  
Screening, Department of  
Health

Dr Catherine Moody,  
Programme Manager,  
Neuroscience and Mental  
Health Board

Dr Ursula Wells,  
Principal Research Officer,  
Department of Health

## Disease Prevention Panel

### Members

**Chair,**  
**Dr Edmund Jessop,**  
Medical Adviser, National  
Specialist Commissioning  
Advisory Group (NSCAG),  
Department of Health

**Deputy Chair,**  
**Professor Margaret  
Thorogood,**  
Professor of Epidemiology,  
University of Warwick Medical  
School, Coventry

Dr Robert Cook  
Clinical Programmes Director,  
Bazian Ltd, London

Dr Elizabeth Fellow-Smith,  
Medical Director, West London  
Mental Health Trust, Middlesex

Dr Colin Greaves  
Senior Research Fellow,  
Peninsular Medical School  
(Primary Care)

Dr John Jackson,  
General Practitioner, Parkway  
Medical Centre, Newcastle  
upon Tyne

Dr Russell Jago,  
Senior Lecturer in Exercise,  
Nutrition and Health, Centre  
for Sport, Exercise and Health,  
University of Bristol

Dr Chris McCall,  
General Practitioner, The  
Hadleigh Practice, Corfe  
Mullen, Dorset

Miss Nicky Mullany,  
Service User Representative

Dr Julie Mytton,  
Locum Consultant in Public  
Health Medicine, Bristol  
Primary Care Trust

Professor Irwin Nazareth,  
Professor of Primary Care  
and Director, Department of  
Primary Care and Population  
Sciences, University College  
London

Professor Ian Roberts,  
Professor of Epidemiology and  
Public Health, London School  
of Hygiene & Tropical Medicine

Professor Carol Tannahill,  
Glasgow Centre for Population  
Health

Mrs Jean Thurston,  
Service User Representative

Professor David Weller,  
Head, School of Clinical  
Science and Community  
Health, University of  
Edinburgh

### Observers

Ms Christine McGuire,  
Research & Development,  
Department of Health

Ms Kay Pattison  
NHS R&D Programme/DH,  
Leeds

Dr Caroline Stone,  
Programme Manager, Medical  
Research Council

## External Devices and Physical Therapies Panel

### Members

<b>Chair,</b> <b>Dr John Pounsford,</b> Consultant Physician North Bristol NHS Trust, Bristol	Dr Dawn Carnes, Senior Research Fellow, Barts and the London School of Medicine and Dentistry, London	Dr Peter Martin, Consultant Neurologist, Addenbrooke's Hospital, Cambridge	Dr Pippa Tyrrell, Stroke Medicine, Senior Lecturer/Consultant Stroke Physician, Salford Royal Foundation Hospitals' Trust, Salford
<b>Deputy Chair,</b> <b>Professor E Andrea Nelson,</b> Reader in Wound Healing and Director of Research, University of Leeds, Leeds	Dr Emma Clark, Clinician Scientist Fellow & Cons. Rheumatologist, University of Bristol, Bristol	Dr Lorraine Pinnigton, Associate Professor in Rehabilitation, University of Nottingham, Nottingham	Dr Sarah Tyson, Senior Research Fellow & Associate Head of School, University of Salford, Salford
Professor Bipin Bhakta Charterhouse Professor in Rehabilitation Medicine, University of Leeds, Leeds	Mrs Anthea De Barton-Watson, Service User Representative	Dr Kate Radford, Division of Rehabilitation and Ageing, School of Community Health Sciences, University of Nottingham, Nottingham	Dr Nefyn Williams, Clinical Senior Lecturer, Cardiff University, Cardiff
Mrs Penny Calder Service User Representative	Professor Christopher Griffiths, Professor of Primary Care, Barts and the London School of Medicine and Dentistry, London	Mr Jim Reece, Service User Representative	
Professor Paul Carding, Professor of Voice Pathology, Newcastle Hospital NHS Trust, Newcastle	Dr Shaheen Hamdy, Clinical Senior Lecturer and Consultant Physician, University of Manchester, Manchester	Professor Maria Stokes, Professor of Neuromusculoskeletal Rehabilitation, University of Southampton, Southampton	

### Observers

Dr Phillip Leech, Principal Medical Officer for Primary Care, Department of Health, London	Ms Kay Pattison, Section Head R&D, DH, Leeds	Dr Morven Roberts, Clinical Trials Manager, MRC, London	Dr Ursula Wells PRP, DH, London
---	---	---	------------------------------------

## Interventional Procedures Panel

### Members

<b>Chair,</b> <b>Professor Jonathan Michaels,</b> Consultant Surgeon & Honorary Clinical Lecturer, University of Sheffield	Mr Seamus Eckford, Consultant in Obstetrics & Gynaecology, North Devon District Hospital	Dr Nadim Malik, Consultant Cardiologist/ Honorary Lecturer, University of Manchester	Dr Ashish Paul, Medical Director, Bedfordshire PCT
Mr David P Britt, Service User Representative, Cheshire	Professor David Taggart, Consultant Cardiothoracic Surgeon, John Radcliffe Hospital	Mr Hisham Mehanna, Consultant & Honorary Associate Professor, University Hospitals Coventry & Warwickshire NHS Trust	Dr Sarah Purdy, Consultant Senior Lecturer, University of Bristol
Mr Sankaran ChandraSekharan, Consultant Surgeon, Colchester Hospital University NHS Foundation Trust	Dr Matthew Hatton, Consultant in Clinical Oncology, Sheffield Teaching Hospital Foundation Trust	Dr Jane Montgomery, Consultant in Anaesthetics and Critical Care, South Devon Healthcare NHS Foundation Trust	Mr Michael Thomas, Consultant Colorectal Surgeon, Bristol Royal Infirmary
Professor Nicholas Clarke, Consultant Orthopaedic Surgeon, Southampton University Hospitals NHS Trust	Dr John Holden, General Practitioner, Garswood Surgery, Wigan	Dr Simon Padley, Consultant Radiologist, Chelsea & Westminster Hospital	Professor Yit Chiun Yang, Consultant Ophthalmologist, Royal Wolverhampton Hospitals NHS Trust
			Mrs Isabel Boyer, Service User Representative, London

## Pharmaceuticals Panel

### Members

**Chair,**  
**Professor Imti Choonara,**  
Professor in Child Health,  
University of Nottingham

**Deputy Chair,**  
**Dr Lesley Wise,**  
Unit Manager,  
Pharmacoepidemiology  
Research Unit, VRMM,  
Medicines & Healthcare  
Products Regulatory Agency

Mrs Nicola Carey,  
Senior Research Fellow,  
School of Health and Social  
Care, The University of  
Reading

Mr John Chapman,  
Service User Representative

Dr Peter Elton,  
Director of Public Health,  
Bury Primary Care Trust

Professor Robin Ferner,  
Consultant Physician and  
Director, West Midlands Centre  
for Adverse Drug Reactions,  
City Hospital NHS Trust,  
Birmingham

Dr Ben Goldacre,  
Research Fellow, Division of  
Psychological Medicine and  
Psychiatry, King's College  
London

Dr Bill Gutteridge,  
Medical Adviser, London  
Strategic Health Authority

Dr Dyfrig Hughes,  
Reader in Pharmacoeconomics  
and Deputy Director, Centre  
for Economics and Policy in  
Health, IMSCaR, Bangor  
University

Dr Yoon K Loke,  
Senior Lecturer in Clinical  
Pharmacology, University of  
East Anglia

Professor Femi Oyeboode,  
Consultant Psychiatrist  
and Head of Department,  
University of Birmingham

Dr Andrew Prentice,  
Senior Lecturer and Consultant  
Obstetrician and Gynaecologist,  
The Rosie Hospital, University  
of Cambridge

Dr Martin Shelly,  
General Practitioner, Leeds,  
and Associate Director, NHS  
Clinical Governance Support  
Team, Leicester

Dr Gillian Shepherd,  
Director, Health and Clinical  
Excellence, Merck Serono Ltd

Mrs Katrina Simister,  
Assistant Director New  
Medicines, National Prescribing  
Centre, Liverpool

Mr David Symes,  
Service User Representative

### Observers

Ms Kay Pattison,  
Section Head, NHS R&D  
Programme, Department of  
Health

Mr Simon Reeve,  
Head of Clinical and Cost-  
Effectiveness, Medicines,  
Pharmacy and Industry Group,  
Department of Health

Dr Heike Weber,  
Programme Manager,  
Medical Research Council

Dr Ursula Wells,  
Principal Research Officer,  
Department of Health

## Psychological and Community Therapies Panel

### Members

**Chair,**  
**Professor Scott Weich,**  
Professor of Psychiatry,  
University of Warwick

Professor Jane Barlow,  
Professor of Public Health in  
the Early Years, Health Sciences  
Research Institute, Warwick  
Medical School

Dr Sabyasachi Bhaumik,  
Consultant Psychiatrist,  
Leicestershire Partnership NHS  
Trust

Mrs Val Carlill,  
Service User Representative,  
Gloucestershire

Dr Steve Cunningham,  
Consultant Respiratory  
Paediatrician, Lothian Health  
Board

Dr Anne Hesketh,  
Senior Clinical Lecturer in  
Speech and Language Therapy,  
University of Manchester

Dr Yann Lefevre,  
GP Partner, Burrage Road  
Surgery, London

Dr Jeremy J Murphy,  
Consultant Physician &  
Cardiologist, County Durham &  
Darlington Foundation Trust

Mr John Needham,  
Service User, Buckinghamshire

Ms Mary Nettle,  
Mental Health User Consultant,  
Gloucestershire

Professor John Potter,  
Professor of Ageing and Stroke  
Medicine, University of East  
Anglia

Dr Greta Rait,  
Senior Clinical Lecturer and  
General Practitioner, University  
College London

Dr Paul Ramchandani,  
Senior Research Fellow/Cons.  
Child Psychiatrist, University of  
Oxford

Dr Howard Ring,  
Consultant & University  
Lecturer in Psychiatry,  
University of Cambridge

Dr Karen Roberts,  
Nurse/Consultant, Dunston Hill  
Hospital, Tyne and Wear

Dr Karim Saad,  
Consultant in Old Age  
Psychiatry, Coventry &  
Warwickshire Partnership Trust

Dr Alastair Sutcliffe,  
Senior Lecturer, University  
College London

Dr Simon Wright,  
GP Partner, Walkden Medical  
Centre, Manchester

### Observers

Ms Kay Pattison,  
Section Head, R&D, DH, Leeds

Dr Morven Roberts,  
Clinical Trials Manager, MRC,  
London

Professor Tom Walley,  
HTA Programme Director,  
Liverpool

Dr Ursula Wells,  
Policy Research Programme,  
DH, London

## Expert Advisory Network

### Members

Professor Douglas Altman,  
Professor of Statistics in  
Medicine, Centre for Statistics  
in Medicine, University of  
Oxford

Professor John Bond,  
Professor of Social Gerontology  
& Health Services Research,  
University of Newcastle upon  
Tyne

Professor Andrew Bradbury,  
Professor of Vascular Surgery,  
Solihull Hospital, Birmingham

Mr Shaun Brogan,  
Chief Executive, Ridgeway  
Primary Care Group, Aylesbury

Mrs Stella Burnside OBE,  
Chief Executive, Regulation  
and Improvement Authority,  
Belfast

Ms Tracy Bury,  
Project Manager, World  
Confederation for Physical  
Therapy, London

Professor Iain T Cameron,  
Professor of Obstetrics and  
Gynaecology and Head of the  
School of Medicine, University  
of Southampton

Dr Christine Clark,  
Medical Writer and Consultant  
Pharmacist, Rossendale

Professor Collette Clifford,  
Professor of Nursing and  
Head of Research, The  
Medical School, University of  
Birmingham

Professor Barry Cookson,  
Director, Laboratory of Hospital  
Infection, Public Health  
Laboratory Service, London

Dr Carl Counsel, Clinical Senior Lecturer in  
Neurology, University of  
Aberdeen

Professor Howard Cuckle,  
Professor of Reproductive  
Epidemiology, Department  
of Paediatrics, Obstetrics &  
Gynaecology, University of  
Leeds

Dr Katherine Darton,  
Information Unit, MIND – The  
Mental Health Charity, London

Professor Carol Dezateux,  
Professor of Paediatric  
Epidemiology, Institute of Child  
Health, London

Mr John Dunning,  
Consultant Cardiothoracic  
Surgeon, Papworth Hospital  
NHS Trust, Cambridge

Mr Jonothan Earnshaw,  
Consultant Vascular Surgeon,  
Gloucestershire Royal Hospital,  
Gloucester

Professor Martin Eccles,  
Professor of Clinical  
Effectiveness, Centre for Health  
Services Research, University of  
Newcastle upon Tyne

Professor Pam Enderby,  
Dean of Faculty of Medicine,  
Institute of General Practice  
and Primary Care, University of  
Sheffield

Professor Gene Feder,  
Professor of Primary Care  
Research & Development,  
Centre for Health Sciences,  
Barts and The London School  
of Medicine and Dentistry

Mr Leonard R Fenwick,  
Chief Executive, Freeman  
Hospital, Newcastle upon Tyne

Mrs Gillian Fletcher,  
Antenatal Teacher and Tutor  
and President, National  
Childbirth Trust, Henfield

Professor Jayne Franklyn,  
Professor of Medicine,  
University of Birmingham

Mr Tam Fry,  
Honorary Chairman, Child  
Growth Foundation, London

Professor Fiona Gilbert,  
Consultant Radiologist and  
NCRN Member, University of  
Aberdeen

Professor Paul Gregg,  
Professor of Orthopaedic  
Surgical Science, South Tees  
Hospital NHS Trust

Bec Hanley,  
Co-director, TwoCan Associates,  
West Sussex

Dr Maryann L Hardy,  
Senior Lecturer, University of  
Bradford

Mrs Sharon Hart,  
Healthcare Management  
Consultant, Reading

Professor Robert E Hawkins,  
CRC Professor and Director  
of Medical Oncology, Christie  
CRC Research Centre,  
Christie Hospital NHS Trust,  
Manchester

Professor Richard Hobbs,  
Head of Department of Primary  
Care & General Practice,  
University of Birmingham

Professor Alan Horwich,  
Dean and Section Chairman,  
The Institute of Cancer  
Research, London

Professor Allen Hutchinson,  
Director of Public Health and  
Deputy Dean of ScHARR,  
University of Sheffield

Professor Peter Jones,  
Professor of Psychiatry,  
University of Cambridge,  
Cambridge

Professor Stan Kaye,  
Cancer Research UK Professor  
of Medical Oncology, Royal  
Marsden Hospital and Institute  
of Cancer Research, Surrey

Dr Duncan Keeley,  
General Practitioner (Dr Burch  
& Ptnts), The Health Centre,  
Thame

Dr Donna Lamping,  
Research Degrees Programme  
Director and Reader in  
Psychology, Health Services  
Research Unit, London School  
of Hygiene and Tropical  
Medicine, London

Mr George Levy,  
Chief Executive, Motor  
Neurone Disease Association,  
Northampton

Professor James Lindesay,  
Professor of Psychiatry for the  
Elderly, University of Leicester

Professor Julian Little,  
Professor of Human Genome  
Epidemiology, University of  
Ottawa

Professor Alistaire McGuire,  
Professor of Health Economics,  
London School of Economics

Professor Rajan Madhok,  
Medical Director and Director  
of Public Health, Directorate  
of Clinical Strategy & Public  
Health, North & East Yorkshire  
& Northern Lincolnshire  
Health Authority, York

Professor Alexander Markham,  
Director, Molecular Medicine  
Unit, St James's University  
Hospital, Leeds

Dr Peter Moore,  
Freelance Science Writer,  
Ashted

Dr Andrew Mortimore,  
Public Health Director,  
Southampton City Primary  
Care Trust

Dr Sue Moss,  
Associate Director, Cancer  
Screening Evaluation Unit,  
Institute of Cancer Research,  
Sutton

Professor Miranda Mugford,  
Professor of Health Economics  
and Group Co-ordinator,  
University of East Anglia

Professor Jim Neilson,  
Head of School of Reproductive  
& Developmental Medicine  
and Professor of Obstetrics  
and Gynaecology, University of  
Liverpool

Mrs Julietta Patnick,  
National Co-ordinator, NHS  
Cancer Screening Programmes,  
Sheffield

Professor Robert Peveler,  
Professor of Liaison Psychiatry,  
Royal South Hants Hospital,  
Southampton

Professor Chris Price,  
Director of Clinical Research,  
Bayer Diagnostics Europe,  
Stoke Poges

Professor William Rosenberg,  
Professor of Hepatology  
and Consultant Physician,  
University of Southampton

Professor Peter Sandercock,  
Professor of Medical Neurology,  
Department of Clinical  
Neurosciences, University of  
Edinburgh

Dr Susan Schonfield,  
Consultant in Public Health,  
Hillingdon Primary Care Trust,  
Middlesex

Dr Eamonn Sheridan,  
Consultant in Clinical Genetics,  
St James's University Hospital,  
Leeds

Dr Margaret Somerville,  
Director of Public Health  
Learning, Peninsula Medical  
School, University of Plymouth

Professor Sarah Stewart-Brown,  
Professor of Public Health,  
Division of Health in the  
Community, University of  
Warwick, Coventry

Professor Ala Szczepura,  
Professor of Health Service  
Research, Centre for Health  
Services Studies, University of  
Warwick, Coventry

Mrs Joan Webster,  
Consumer Member, Southern  
Derbyshire Community Health  
Council

**Professor Martin Whittle,**  
Clinical Co-director, National  
Co-ordinating Centre for  
Women's and Children's  
Health, Lymington





### **Feedback**

The HTA programme and the authors would like to know your views about this report.

The Correspondence Page on the HTA website ([www.hta.ac.uk](http://www.hta.ac.uk)) is a convenient way to publish your comments. If you prefer, you can send your comments to the address below, telling us whether you would like us to transfer them to the website.

***We look forward to hearing from you.***