



BEST PRACTICE IN UNDERTAKING AND REPORTING HTA

***TO DEVELOP AND DISSEMINATE BEST PRACTICE IN
UNDERTAKING AND REPORTING ASSESSMENTS. IDENTIFYING
NEEDS FOR METHODOLOGICAL DEVELOPMENT***

Chair: Reinhard Busse

Co-chair: Jacques Orvain

Participants: Michael Drummond, Felix Gurtner,
Torben Jørgensen, Albert Jovell, Jim Malone, Matthias Perleth,
Claudia Wild

Prepared for and in close collaboration with the working group by
Reinhard Busse, Marcial Velasco, Matthias Perleth and Jacques Orvain

July 2001

Contents

- Chapter 0. Introduction
- Chapter 1. Objectives of this report
- Chapter 2. Methods of this report
- Chapter 3. Methodological framework for conducting HTA
 - 3.1 Characteristics of HTA
 - 3.2 Process of HTA
- Chapter 4. “Best Practice” in undertaking HTA reports
 - 4.1 Policy question
 - 4.2 HTA protocol
 - 4.3 Background information
 - 4.3.1 Condition and target group
 - 4.3.2 Technology
 - 4.4 Research question(s)
 - 4.5 Answering the questions/ General methodology
 - 4.5.1 Sources of information
 - 4.5.2 Inclusion and exclusion criteria/ Appraisal of the evidenc
 - 4.5.3 Non-quantitative and quantitative synthesis
 - 4.6 Specific methodological considerations
 - 4.6.1 Safety
 - 4.6.2 Efficacy and effectiveness
 - 4.6.2.1 Therapeutic interventions
 - 4.6.2.2 Diagnostic interventions
 - 4.6.2.3 Health organisation/system related interventions
 - 4.6.2.4 Preventive interventions
 - 4.6.3 Psychological, social, and ethical considerations
 - 4.6.4 Organisational and professional implications
 - 4.6.5 Economic issues
 - 4.7 Discussion of methods and results
 - 4.8 Conclusions and recommendations
 - 4.9 Other relevant issues
 - 4.9.1 Review process
 - 4.9.2 Updating of assessments
- Chapter 5. “Best Practice” in reporting HTA
 - 5.1 Abstract
 - 5.2 Scientific Summary Report
 - 5.3 Technical report
- Chapter 6. Conclusions
- Chapter 7. Recommendations
- Reference
- s
- Appendice A1. Toolkits and documents available
- s
 - A2. Sources of information
 - A3. Search filters
 - A4. Appraisal checklists
 - A5. Software for data synthesis
 - A6. Levels of evidence and grades of recommendations

List of Figures

- Figure 1. Assessment process
- Figure 2. Flow diagram of literature selection process
- Figure 3. Quality assessment presentation (example)

List of Boxes

- Box 1. Content of a HTA
- Box 2. Aspects included in the policy question
- Box 3. Key steps in finding background information
- Box 4. Questions to be addressed as background information on condition and target group
- Box 5. Questions to be addressed as background information on the technology
- Box 6. Examples of outcomes for different aspects of HTA
- Box 7. Characteristics of research questions
- Box 8. General methodological steps for addressing each aspect of assessment
- Box 9. Documentation of the sources
- Box 10. Issues addressed in inclusion and exclusion criteria
- Box 11. Quality items/criteria
- Box 12. Transparency in Quality Assessment
- Box 13. Elements to include in evidence tables
- Box 14. Factors to consider when using Quantitative Synthesis (meta-analysis)
- Box 15. Uses of modelling
- Box 16. Modelling
- Box 17. Definitions of “efficacy” and “effectiveness”
- Box 18. Key issues in assessing systematic review articles
- Box 19. Evaluation of efficacy and effectiveness for diagnostic interventions
- Box 20. Study designs used for assessing health care organisation and system related interventions
- Box 21. Criteria for assessment of studies using qualitative research methods
- Box 22. Types of costs in an economic analysis
- Box 23. Types of economic analysis
- Box 24. Economic evaluation
- Box 25. Discussion
- Box 26. Conclusions
- Box 27. Review process
- Box 28. Identification of the need for update
- Box 29. Update of HTA
- Box 30. Data to be included in English structured abstract
- Box 31. Differences between “Executive Summary” and “Scientific Summary Report”
- Box 32. Elements to be addressed in the Scientific Summary Report (not necessarily addressed in this order)
- Box 33. Statement on Conflict of Interest
- Box 34. Structure example for a HTA technical report

- Box 35. Criteria for the assessment of the quality of HTA reports
- Box A4-1 Checklist for an article about therapy
- Box A4-2 Checklist for an article about diagnostic tests
- Box A4-3 Checklist for an article about harm
- Box A4-4 Checklist for an article about prognosis
- Box A4-5 Checklist for a review article
- Box A4-6 Checklist for a clinical decision analysis
- Box A4-7 Checklist for clinical practice guidelines
- Box A4-8 Checklist based on the “Appraisal of Guidelines for Research & Evaluation (AGREE) Instrument”
- Box A4-9 Checklist for an article reporting variations in the outcomes of health services research
- Box A4-10 Checklist for a clinical utilisation review
- Box A4-11 Checklist for an article about health-related quality of life measurements
- Box A4-12 Checklist for qualitative research in health care
- Box A4-13 Checklist for an economic analysis article

Chapter 0. Introduction

Health technology assessment (HTA) is a multidisciplinary activity which systematically examines the technical performance, safety, clinical efficacy and effectiveness, cost, cost-effectiveness, organisational implications, social consequences, legal and ethical considerations of the application of a health technology (EUR-ASSESS 1997). HTA activity has been continuously increasing over the last few years. A number of HTA agencies and other institutions (termed in this report “HTA doers”) across Europe are producing an important and growing amount of HTA information. The objectives of HTA vary considerably between HTA agencies and other actors, from a strictly political decision making-oriented approach regarding advice on market licensure, coverage in benefits catalogue or investment planning to information directed to providers or to the public. However, there seems to be a broad agreement in the general elements which belong to the HTA process and HTA doers in Europe are using similar principles (Mears et al. 2000), although often a different language or terminology makes it difficult to see this.

In addition, the reporting of the findings from the assessments differs considerably. This reduces comparability and makes it difficult for those undertaking HTA assessments to integrate previous findings from other HTA doers in a subsequent evaluation of the same technology. Transparent and clear reporting is an important step towards disseminating the findings of a HTA, thus standards which ensure high quality reporting may contribute to a wider dissemination of results.

The EUR-ASSESS methodological subgroup already proposed a framework for conducting and reporting HTA (EUR-ASSESS 1997) which served as the basis for the current working group.

New developments have occurred in the last five years which necessitate revisiting that framework and providing a solid structure for future updates. Giving due attention to these methodological developments, this report presents a description of current “best practice” in both undertaking and reporting HTA, followed by the identification of needs for methodological development. It concludes with specific recommendations as well as tools for implementing them, e.g. by providing the structure for English-language scientific summary reports and a checklist to assess the methodological and reporting quality of HTA reports.

Specifically, this report is structured as follows: In chapter 1, the objectives of this report are stated. In chapter 2, the methods applied and the material used by the working group are briefly described. Chapter 3 characterises the HTA process which served as an outline for structuring the information of this report. In chapters 4 and 5, “Best practice” in undertaking and reporting HTA are identified and described. Chapter 6 gives the conclusions, which include a checklist for assessing quality and relevance of a HTA report, and place particular emphasis on identifying methodological gaps and needs for further development. Finally, in chapter 7, some recommendations are made.

Chapter 1. Objectives of the working group and this report

In the overall framework of the ECHTA project, the objectives of the working group and this report are as follows:

- To develop best practice in undertaking assessments
- To develop best practice in reporting assessments
- To disseminate best practice in undertaking assessments
- To disseminate best practice in reporting assessments
- To identify needs for methodological development

The report addresses the first two objectives in chapters 4 and 5; the two objectives of disseminating this best practice are addressed both by writing this report as well as through providing the structure for a scientific summary report and a checklist for assessing the quality of HTA reports. The final objective is addressed in chapter 6 of this report.

When reading the report, several caveats should be kept in mind:

- The report tries to outline current “best practice” covering all (possible) aspects, ordering them in a logical sequence and using an understandable terminology for the concepts. Actual practice regarding completeness, sequence and terminology of HTA doers will, however, vary, which does not per se constitute “bad practice”.
- While the report serves to identify “best practice,” the strength of the evidence to identify certain practices as “best” varies. In this respect, the degree to which they can be recommended also varies – this is clearly indicated in the text. The report makes recommendations, e.g. for methodological development, which are summarised at the end.

Chapter 2. Methodology applied by the working group

As mentioned, the EUR-ASSESS methodological subgroup proposed a framework for conducting and reporting HTA (EUR-ASSESS 1997), which served as the point of departure for the current working group. In its two formal meetings in June 2000 and January 2001, the working group decided to provide a methodological framework based on existing guidelines from HTA agencies and other institutions in order to enhance comparability among European HTA. In the discussion, particular importance was given to the need for a structured way of reporting, especially stressing the need for a structured/standard summary, to make HTA findings from European agencies and other institutions more available to the HTA community. In addition, specific issues which the group felt were underrepresented thus far (e.g. the HTA process, the use of qualitative methods, factors responsible for differences between efficacy and effectiveness) were identified as requiring special attention.

Considering the recommendations and consensus reported in discussion papers from the INAHTA Annual Meeting 2000 at Loosdrecht on a similar issue (Hailey 2001, personal communication), guidance documents and tool kits from different institutions involved in HTA were examined and summarised into an outline. Putting emphasis on freely available documents, the following tool kits and guidelines were identified via personal searches/ contacts of the working group members and a search of the websites of European and other HTA institutions and were taken into account for elaborating the methodological framework (in chronological order):

- EUR-ASSESS Project Subgroup Report on Methodology: Methodological guidance for the conduct of health technology. *Int J Technol Assess Health Care* 1997;13(2):186-219 (EUR-ASSESS 1997).
- Various reports from the NHS R&D HTA Programme, UK, 1998-2001 (for details see appendix A1).
- Guía para la elaboración de informes de evaluación de tecnologías sanitarias. Agencia de evaluación de tecnologías sanitarias, Madrid, Spain, 1999 (Imaz-Iglesia et al. 1999).
- Development and Evaluation Committee Guidelines. The Wessex Institute for Health Research and Development, Southampton, UK (DEC undated [2000]).

- West Midlands Development and Evaluation Service (DES) Handbook. Department of Public Health and Epidemiology, University of Birmingham, UK, DPHE Report No 8, 2000 (Burls et al. 2000).
- Guide d'Analyse de la littérature et gradation des recommandations. Agence Nationale d'Accréditation et d'Évaluation en Santé, Paris/ France, 2000 (Durocher et al. 2000).
- Undertaking Systematic Reviews of Research on Effectiveness. CRD's Guidance for Carrying Out or Commissioning Reviews. CRD Report 4, NHS-CRD, University of York, UK, 2000 (Khan et al. 2000).
- Guide to the Technology Appraisal Process. National Institute for Clinical Excellence, UK, 2000.
- German tool kit and checklist for the conducting and appraisal of HTA reports. German Scientific Working Group on Technology Assessment for Health Care, last updated 2000.
- Funding for new medical technologies and procedures: application and assessment guidelines. Medicare Services Advisory Committee, Canberra, 2000 (MSAC 2000).
- Health Technology Assessment Handbook. Danish Institute for Health Technology Assessment, Copenhagen, Denmark, 2001 (Kristensen et al. 2001).

In addition, based on working group members' experience as well as reference lists, specific guidance and key references for the identified specific issues – as well as for gaps which became obvious while drafting this report – were identified and selected for inclusion into the report. To achieve a consensus process, a core group drafted a first version of this report in April 2001 for discussion among the other working group members (Mike Drummond, Felix Gürtner, Torben Jørgenson, Albert Jovell, Alric Rüther, Claudia Wild) as well as other persons. This final version reflects the amendments, comments and discussion.

The authors are indebted to Wendy Wisbaum (European Observatory on Health Care Systems) for providing English-language editing.

Chapter 3. Methodological framework for conducting HTA

3.1 Characteristics of HTA

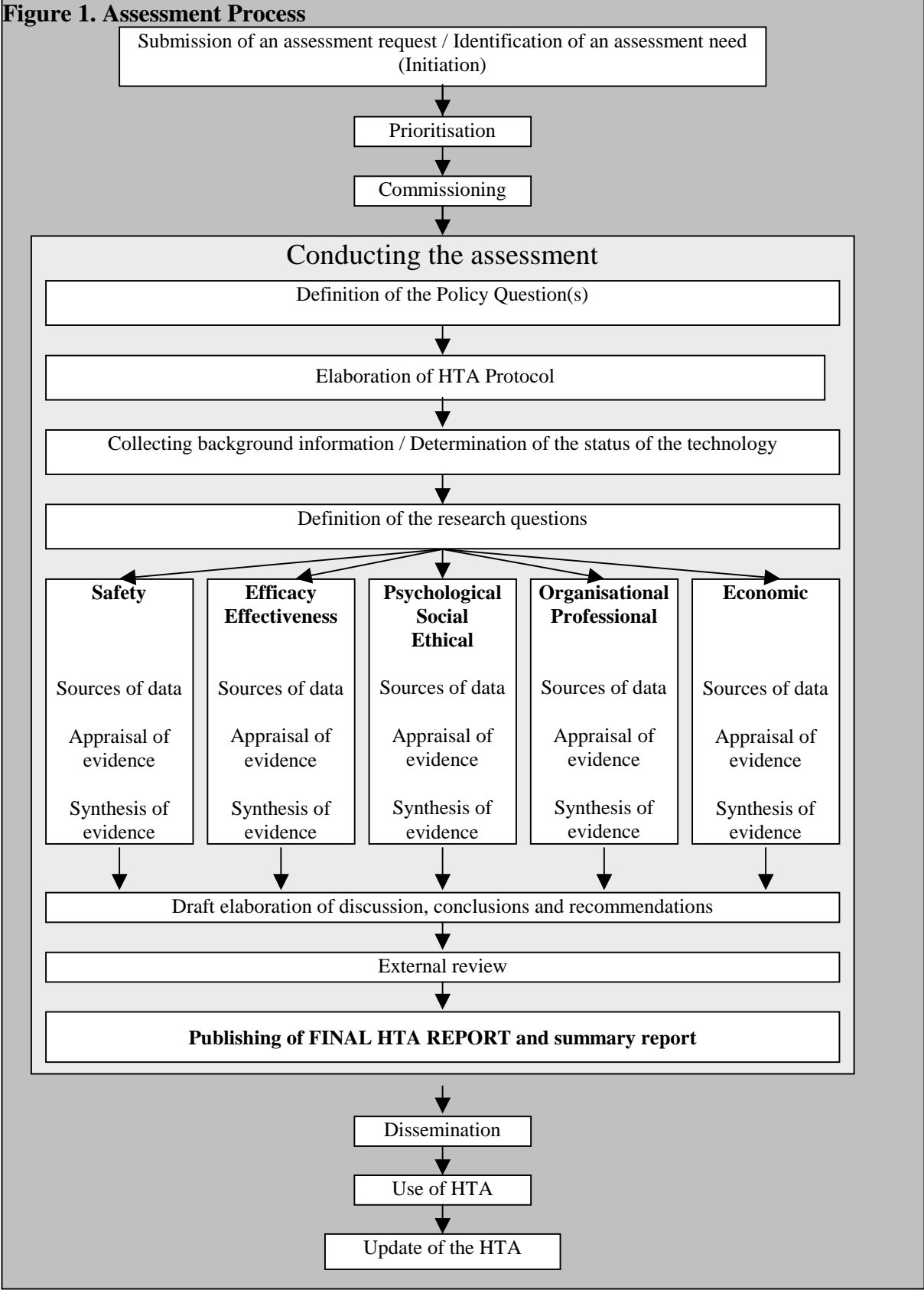
Health technology assessment, a multidisciplinary activity which systematically examines the technical performance, safety, clinical efficacy and effectiveness, cost, cost-effectiveness, organisational implications, social consequences, legal and ethical considerations of the application of a health technology (EUR-ASSESS 1997), has to take into consideration all aspects which might be influenced by the technology as well as those influencing the technology. In this context, health technology is a broad concept which includes drugs, devices, procedures and the organisational and support systems within which health care is delivered.

As with Evidence Based Medicine (EBM) and Clinical Practice Guidelines (CPG), HTA belongs to the group of best practice activities in the health care sector (European Commission 1999). These kinds of activities are characterised by a systematic and structured way of answering questions by evaluating and synthesising available evidence. Even though certain institutions (e.g. ANAES, NICE) use all three approaches, they differ in some aspects. The primary audience of HTA consists of decision makers at the policy level, while other activities aim at the clinical level (EBM, CPG). In addition, the sources of information and the methods used are broader in HTA than in the other approaches. It is now accepted that the characteristics of HTA are: a clear formulation of the problem, an explicit methodology and a wide scope on the technology, i.e. not only dealing with safety or efficacy/effectiveness (EUR-ASSESS 1997). Besides a systematic methodology, the strength of HTA relies on transparency of the process and in the reporting which also improves the usefulness and generalisability of the findings.

3.2 Process of HTA

When performing health technology assessments, all European doers seem to follow a similar process. Nevertheless, the way assessments are initiated, priorities are set, and reports are commissioned and later disseminated may differ substantially among agencies and other institutions (which is outside the scope of the current report). Although the aim of this report is not an analysis of the whole HTA process, it has to be pointed out that the way the different steps are undertaken influences the elaboration of the HTA report, which can be seen as a step in the overall

assessment process and represents the deliverable product of the assessment (Fig. 1). The “HTA Report Box” is the scope of this report.



After a report is commissioned, the first step to be taken is the definition of the policy question, if that has not been clearly formulated during the prioritisation and/or commissioning process. The next step consists of the gathering of background information (part of which may have already been collected during the prioritisation process). When collecting background information, possibly after (re-)contacting the commissioner, the researcher will be able to decide which aspects of the problem (e.g. efficacy, ethical considerations etc.) should be further assessed. Concise research questions will be posed and the methodology will be outlined.

In HTA, the five columns reflecting the main types of outcomes should all be considered relevant; thus, they are presented in a parallel way. However, it seems plausible to start with assessing safety first, then efficacy and so on, as subsequent aspects of the assessment might not be needed if previous ones already provided a negative answer. To illustrate, for instance, if the technology shows a safety deficit or proves to be not efficacious at all, evaluation of further aspects will not be necessary.

Chapter 4. “Best practice” in undertaking HTA reports

The EUR-ASSESS Subgroup proposed a framework with the following elements to be included in a HTA report (Box 1).

Box 1. Content of a HTA (modified from EUR-ASSESS 1997)	
	<ul style="list-style-type: none">• Policy Question (4.1)
	<ul style="list-style-type: none">• Background Information on: target group, target condition, technology (technical aspects, diffusion, and current practice) (4.3)
	<ul style="list-style-type: none">• Research Questions (4.4)
Findings & Methodology (4.5/ 4.6)	<ul style="list-style-type: none">• Safety• Efficacy / Effectiveness• Psychological, social and ethical considerations• Organisational and professional implications• Economic issues
	<ul style="list-style-type: none">• Policy conclusions and recommendations (4.8)

For each of the aspects of the HTA, it is important that the sources of data, the methods for searching and gathering data, and their synthesis are clearly stated. If some aspects are not being addressed, the reason for omission (e.g. sufficient data available from other HTA reports) should also be included.

The following sections will provide a general methodological framework, in terms of what could be considered best practice, following the structure shown in Fig. 1 and Box 1. Other important issues concerning the HTA process, like the review process, updates of the HTA and possible conflicts of interest cannot be clearly ordered in the structure proposed in Fig. 1 and will therefore be considered afterwards.

4.1. Policy question

HTA is policy-driven research, aimed to support decision-making. Thus, the commissioners’ scope of the problem has to be clearly documented in the report. Ideally, the policy question should be worded with close co-operation between the commissioners and the researchers.

The policy question reflects the context in which the assessment was carried out. This context is defined by the following aspects (Box 2).

Box 2. Aspects included in the policy question	
Question	Examples
<ul style="list-style-type: none"> • Who initiated the report? 	<ul style="list-style-type: none"> • Policy makers • Health care providers • Third party payers • Patients´ advocate
<ul style="list-style-type: none"> • Who commissioned it? 	
<ul style="list-style-type: none"> • Why is an assessment needed right now? 	<ul style="list-style-type: none"> • New technology • Changes in old technology • New indications for old technology • New findings • Structural/ Organisational changes • Safety concerns • Ethical concerns • Economic concerns
<ul style="list-style-type: none"> • Which decision is it going to support? 	<ul style="list-style-type: none"> • Investment decisions • Market licensure • Inclusion in/Exclusion from benefits catalogue • Planning of capacities • Guidance on best-practice • Investment in further research
<ul style="list-style-type: none"> • Who represents the primary target audience for the report? 	<ul style="list-style-type: none"> • Political decision makers • Third party payers • Hospital managers/administrators • Clinicians • Citizens / Patients

The context in which the research is carried out may lead to some financial or time constraints which determine the methods used and the extent/comprehensiveness of the assessment. The scope and level of detail of HTA vary considerably depending on who commissioned a study and why. Therefore, it is crucial to clearly explain that context, so that readers of HTA (other than those who initiated and commissioned the study) can better assess whether the report can be also relevant for their own problems. The scope of the assessment and its recommendations are determined by the policy question.

The policy question should be clearly stated in the HTA protocol (cf. section 4.2) as well as in both the technical report, i.e. the detailed document (cf. section 5.3), and the scientific summary report (cf. section 5.2). The questions listed in Box 2 should be answerable when reading any of this documents.

4.2 HTA protocol

As soon as the policy question is clear, a HTA Protocol should be developed to define how the whole assessment is going to be carried out. A HTA Protocol is not a Systematic Review Protocol, as this usually refers only to one of the possible aspects to be reviewed in the assessment. A HTA Protocol has to be understood as the elaboration of the plan for both undertaking the whole process of the assessment and for writing the HTA Report. The utilisation of such a protocol should be seen as an important component for achieving best practice in undertaking and reporting HTA. HTA Protocols are sometimes referred to as Project Plans (DEC 2000).

In a simplified way, the development of a HTA protocol can be divided into two phases, with the first one at the beginning of the assessment. Here, the problem will be stated and the way of gathering the background information will be defined. While synthesising the background information, the research questions will be posed. Then the protocol should be completed by stating:

- which aspects of the problem are going to be assessed,
- how each aspect will be addressed, i.e. which and how data sources will be searched and used,
- which methodology for the appraisal will be followed, and,
- what kind of synthesis of evidence is planned.

In this regard, a HTA protocol should include guidelines on when and how to undertake a systematic review of one or more of the aspects (if no standing operating procedures exist for such a decision within the commissioning agency or the institution undertaking the HTA). Additionally, it will most likely state timelines and division of competencies within the group of persons involved. The HTA protocol should document the way the whole process explained in Fig. 1 was carried out.

4.3 Background information

After defining the policy question, the HTA doers need to gather information about the target condition, the target group and the technology to be assessed.

The background information helps translate the policy question into a research question. The process of gathering background information is intimately related to the definition of the research questions, which can only be stated satisfactorily after the background information is reviewed.

Most of the agencies and other institutions recommend preliminary research to address the background issues. If a literature search is conducted, it is strongly recommended that it is carried out separate from the systematic literature search done later to address the research question(s). The scope of this first search is to learn the epidemiology, natural history and clinical presentation of the condition, possible target group(s) (see section 4.3.1), as well as background information on the technology, e.g. technological characteristics (see section 4.3.2). Review articles (not necessarily systematic) and text books can be helpful in giving an idea as to the condition and treatment alternatives.

Further information sources, such as routinely collected data, expert contacts, guidelines on diagnosis and management, patient opinions (e.g. websites of associations of persons suffering from the condition), or information from manufacturers of the technology are also valuable for an idea about the status of the technology. Previous HTA reports are another important source of background information.

Key steps and sources of data for the elaboration of background information are summarised in Box 3.

Box 3. Key steps in finding background information (DES; Burls et al. 2000)

- Perform this parallel with defining research question
- Search for and record information on the:
 - Nature of the health problem or disease
 - Epidemiology and burden of the disease
 - Treatments for the disease (alternatives)
 - Current practice
 - Technology status
- Sources:
 - Research literature (search strategies targeting “reviews,” “prevalence,” “incidence,” etc.)
 - Routinely collected data (on utilisation, costs, etc.)
 - Guidelines
 - Special sources (disease registers, organisations of affected people, experts, manufacturers) [some of those sources are accessible through the www]
 - Other HTA reports (searchable in INAHTA database, or in the websites of HTA agencies)

The elaboration of the background information does not necessarily imply systematic research, as other approaches may deliver sufficient information for elaborating the research questions.¹ However, for the transparency of the HTA, the approach(es) and sources used when elaborating the background information should be documented.

4.3.1 Condition and target group

The essential information needed to understand the nature of the health problem or disease and its consequences should be provided. The target group(s) to which the assessment refers should also be clearly stated. In this step of the assessment, the following questions concerning the condition and the target group should be addressed (Box 4).

Box 4. Questions to be addressed as background information on condition and target group (Adapted from DES; Burls et al. 2000)	
Questions	Example
<ul style="list-style-type: none"> • Condition(s) <p>What are the mechanisms of disease?</p> <p>What is the course and prognosis of the condition?</p> <p>What are the consequences? (Outcomes)</p> <p>Treatment alternatives and current practice</p>	<ul style="list-style-type: none"> • Health problem • Disease • Causes • Pathology • Clinical presentation • Stages • Time course • Physical disabling • Psychological consequences • Death • Drugs • Surgical • Current service provision
<ul style="list-style-type: none"> • Target group(s) (epidemiology, burden of disease) • How many people are affected? • Who is affected? 	<ul style="list-style-type: none"> • Patients • Healthy subjects (for prevention) • Incidence • Prevalence • Age • Gender • Social factors • Risk factors

¹ When drafting the full report, these sections of the background sections should be revisited to check whether they need any amendments due to the identified evidence. This could, for example, be the case if a technology is highly effective for an indication originally not included in the assessment.

These issues should be addressed briefly and clearly, keeping in mind that not all HTA readers are experts in the given field. The background information serves also to clarify and explain the concepts which are going to be used in the assessment on safety, efficacy, effectiveness, as well as the other relevant outcomes. The description of the appropriate outcomes and how they are measured is therefore an important issue too.

4.3.2 Technology

It is best practice to concisely describe the following aspects of the technology (Box 5), keeping in mind that the technology assessed may be a drug, a device (therapeutic/diagnostic), a community intervention, a medical aid, a procedure, an organisational process, a support system or a combination of these.

Box 5. Questions to be addressed as background information on the technology	
Question	Aspects / examples
<ul style="list-style-type: none"> • How does it work? What kind of intervention is it? 	<ul style="list-style-type: none"> • If a device, explain technical characteristics, functioning • If a community/system related intervention, explain its crucial features
<ul style="list-style-type: none"> • What are the requirements for its use? 	<ul style="list-style-type: none"> • Setting for use/implementation • Special measures needed for use/implementation • Qualification required • Maintenance
<ul style="list-style-type: none"> • What is the status of the technology? 	<ul style="list-style-type: none"> • Diffusion/distribution • Patterns of use • Current indications for use • Current utilisation • Costs • Regulatory status • Manufacturers and market shares

The description of the technology should be concise and understandable, with particular emphasis on those aspects of the technology that directly affect the safety, efficacy or effectiveness (e.g. doses of drugs, material in implants, image characteristics of diagnostic devices, etc.). Technical details of the technology, which have no influence on the outcomes, do not need to be described in detail.

A description of the *status quo* of the technology can be considered an important part of the assessment. Current practice, indications (if given) for use of the technology, frequency of utilisation, as well as associated costs should be described

here. Some of these issues are directly related to the point where the technology is on the learning curve of the technology². Sometimes these issues may not need serious consideration, depending on the status of technology (e.g. utilisation patterns if assessment is prior to approval for use).

4.4 Research Question(s)

Formulating the research question(s) means specifying the policy question in terms of safety, efficacy, effectiveness, psychological, social, ethical, organisational, professional and economic aspects. These aspects may be able to be addressed with available evidence and data, but they either have not yet been sufficiently answered or have answers that are not accessible and/ or appropriate for the use of decision-making.

The research questions can also be drawn from previous HTAs that were unable to answer them because of lack of evidence, and which stated that further research was required.

The research questions have to specify the target group, the (disease) condition and the aspects of the technology that are going to be assessed. Thus, formulation of the research questions is closely related to the gathering of background information. The examined guiding documents agree that both steps have to be taken in parallel.

The formulation of the research questions also implies defining the outcomes of interest for the assessment. The outcomes of interest for the evaluation are different for the different aspects of the assessment. Some of them may be easier to define than others. *Safety, efficacy* and *effectiveness* of an intervention should be always measured with health related outcomes. These should be patient-related (e.g. quality of life, mortality, morbidity). Outcomes for the assessment of *psychological, social* and ethical considerations are, for example, satisfaction or acceptance. *Organisational* and *professional implications* can be addressed with system-related outcomes, such as length of stay or required personnel. Finally, for the *economic issues*, costs and cost in relation to outcomes (cost-effectiveness, cost-utility, cost-benefit) are the main categories of interest. Box 6 provides examples of outcomes for the different aspects.

² Methods to statistically assess the learning curve have been gathered and evaluated by Ramsay et al. 2001.

Box 6. Examples of outcomes for different aspects of HTA	
Aspect of assessment	Outcomes
Safety	<ul style="list-style-type: none"> • Mortality directly related to the use of technology • Morbidity/disability directly related to the use of technology
Efficacy/Effectiveness	<ul style="list-style-type: none"> • Change in overall/ condition-specific mortality • Change in morbidity/ disability/ disease-free interval • Change in quality of life • Change in quality-/disability-adjusted life years (QALYs/DALYs)
Psychological/ Social/ Ethical	<ul style="list-style-type: none"> • Compliance • Acceptance • Satisfaction • Demand • Preferences • Information/patient advice requirements
Organisational/ Professional	<ul style="list-style-type: none"> • Utilisation of service • Change in the treatment location • Change in length of hospital stay • Change in required personnel, material inputs (e.g. hospital beds) and organisational structure • Training requirements
Economic	<ul style="list-style-type: none"> • Costs and changes in cost compared to current practice (if applicable) • Cost-effectiveness, cost-utility, cost-benefit

The research question(s) drive(s) how the rest of the assessment is going to be conducted, the aspects which will be evaluated and those which will not. The inclusion and exclusion criteria for literature or other sources of data to be reviewed in the assessment also depend on the formulation of the research questions. The documents and recommendations reviewed all agree that this is a crucial part of the assessment, as other aspects (e.g. methodological) of the evaluation flow from it. If possible and where relevant, there should be a feed-back loop to the commissioner(s) to ensure that the research questions a useful “translation” of the policy question(s).

The research questions need to be formulated in an understandable and answerable way, and should be limited in number (Box 7).

Box 7. Characteristics of research questions

- Clearly worded
- Answerable
- Limited in number
- Address meaningful outcomes
- Address other relevant treatment alternatives

4.5 Answering the questions/ General methodology

Once the research question(s) have been formulated, the next step is to answer them. As shown in Fig. 1, there are some general methodological steps which apply to all aspects of the HTA (i.e. safety, efficacy/effectiveness, psychological/social/ethical, organisational/professional, economic). Most of the methodology has been developed under the scope of systematic reviews on efficacy/effectiveness; however, some principles of this methodology are applicable to other aspects. These common principles are discussed in sections 4.5.1 to 4.5.3. Specific methodological considerations concerning each aspect of the assessment are then addressed in section 4.6.

The common methodology for addressing the different aspects can be summarised in three steps (Box 8).

Box 8. General methodological steps for addressing each aspect of assessment

- Searching for sources of information (4.5.1)
- Selecting and evaluating information (application of inclusion and exclusion criteria)/ appraising the evidence (4.5.2)
- Synthesising the obtained data (4.5.3)

4.5.1 Sources of information.

For different aspects of the assessment, different sources of data may be useful or appropriate. Sources of data do not always have to be published literature.

Databases, registries of routine data or even one's own primary research³ may be also appropriate, depending on the aspect being assessed.

One or more of the aspects of the current assessment may have been already addressed by other HTA reports. A first approach to answer the question(s) can thus be the search for previous HTA reports, even if one or more should have been already identified during the search for background information. Search for HTA reports has to be systematic and also clearly documented⁴. Identified HTA reports should also be critically appraised (see 4.5.2).

Systematic reviews may already cover some of the aspects and answer some of the questions posed. This may be the case for aspects like safety, efficacy, effectiveness or economic evaluation. Thus, a search for this kind of research has to be an integral component of all searches.

If primary scientific literature is going to be used, the principles of the systematic primary literature search, developed for example by the Cochrane Collaboration, can be applied to all aspects of the assessment, and not only to efficacy/effectiveness. In order to identify the evidence, a search strategy has to be developed, based on the research questions and to some extent, on inclusion and exclusion criteria (e.g. study design). Key words related to the condition, the technology, types of publication etc. will be combined, forming the search strategy in order to obtain the biggest number of hits. It is recommended that the language of publication *not* be used as a search criterion, as relevant literature in other languages will be missed (see also section 4.5.2).

A systematic approach can be also applicable for psychological/social/ethical⁵, organisational/professional or economic issues if literature is going to be used. Search strategies and databases searched will differ, depending on the aspect, and, as a result, they should be documented separately.

If other sources of information or evidence are used, a systematic approach should be followed. The strategies used to identify them, the way in which the information was obtained etc., should also be documented.

³ *Own primary research* refers here to primary research conducted within the assessment in order to address some aspects of it, e.g. a survey to assess the satisfaction after a treatment.

⁴ Appendix 2 provides further information on different databases for identifying HTA reports or systematic reviews.

⁵ This systematic approach can be applied when outcomes such as acceptance or satisfaction are being addressed. However, if more general philosophical issues are being assessed, the systematic

The documentation of the information sources is of utmost importance for the transparency of a HTA report. Both sources which provided useful information and those which did not should be included in the documentation (DIHTA; Kristensen et al. 2001).

Box 9. Documentation of the sources (DIHTA; Kristensen et al. 2001)

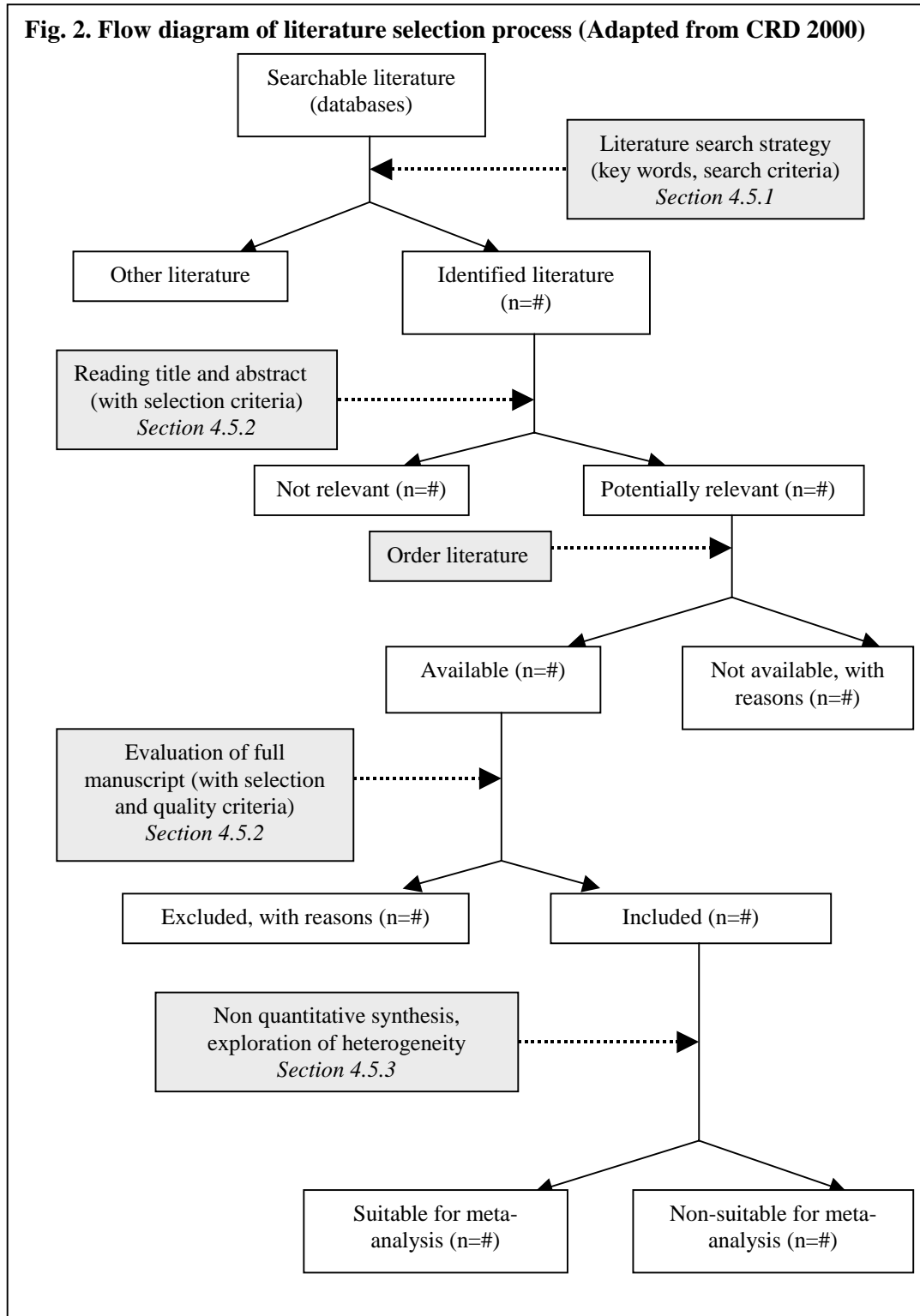
- Which sources have been consulted?
- Which period did the performed search cover?
- How was the search performed? (Strategies, key words, search criteria)
- When was the search conducted?

4.5.2 Inclusion and exclusion criteria/ Appraisal of the evidence.

The selection of the literature which will be definitely included to answer the research questions is a process with consecutive steps to be taken, as summarised in Fig. 2.

approach may not be possible at all, as disciplines may be involved which, for example, do not have databases such as those of the medical literature.

Fig. 2. Flow diagram of literature selection process (Adapted from CRD 2000)



With the systematic literature search, a big number of hits will be obtained. Applying selection criteria (inclusion and exclusion criteria) to the titles and abstracts of articles, these will be separated into relevant and not relevant. This first selection refers more to relevance than to quality of studies. Studies considered to be relevant will be ordered, but not all ordered studies will be actually retrieved (e.g. delayed delivery). The available studies will then be critically appraised for quality. Those which fulfil the defined quality standards will be definitively selected for inclusion in the synthesis. It is recommended that this process is reported in an understandable and transparent way, e.g. by using Fig. 2 as a guide.

It is also recommended that two reviewers select the literature to be included, however, this may not always be possible. When reporting on the methodology, it should be stated whether this step was performed by one or more reviewers, and how contradictions were handled.

Inclusion and exclusion criteria should be defined for all kinds of evidence, and not only for the literature on efficacy and effectiveness. Selection criteria should be developed in a prospective way in order to avoid bias when selecting the evidence. Inclusion and exclusion criteria flow from the background information, the research questions and the availability of evidence. They refer to patients being treated, outcomes being measured, aspects of the technology being studied, etc. Selection criteria also may refer to study design or other methodological issues. Those criteria (may) differ for each of the aspects being assessed. For instance, when assessment of efficacy issues is based on RCTs, study design will be an inclusion criterion. However, if, for example, routine register data are used to assess safety, the size and follow-up-time of the register might be the selection criterion (Box 10).

Box 10. Issues addressed in inclusion and exclusion criteria

- Patient characteristics (e.g. age, gender)
- Condition characteristics (e.g. stage of disease)
- Technology aspects
- Methodological issues (e.g. number of patients, length of follow-up, study design)
- Outcomes measured
- Publication type

Depending on the aspects being assessed, selection criteria may be narrower or wider. The selection of the literature or other sources has to be transparent, thus, the explicit stating of these criteria should be mandatory in a HTA report. Inclusion and exclusion criteria have to be documented in both the technical report and the scientific summary report. They have to be explained (especially if they might not seem to be justified) and they have to be compatible with the research questions.

Every effort should be made to include relevant evidence independent of the language available. This means that language should be used very cautiously as a selection criterion. Rather, potentially relevant studies published in languages not familiar to the HTA doers should be ordered. Possibly, tables or other pieces of information will indicate the relevance of the study and justify a translation. If the HTA doers are not able to handle potentially relevant publications in unfamiliar languages, these studies should be explicitly listed and their number later taken into account when discussing the results. This is important because the selection of literature/information sources based on language of publication may lead to bias in conclusions or results (Egger & Smith 1998).

Once the literature is ordered, the available references will be checked again for their relevance by carefully evaluating the full document. At this point, some studies will be excluded because they are not actually deemed relevant to the research questions, even though they were identified as relevant when the abstract was read.

The quality and relevance of all sources of data need to be critically assessed. Again, most of the work done here refers to the critical appraisal of the medical literature referring to efficacy and effectiveness (primary and secondary research), for which different checklists⁶ have been developed. Some doers have adapted these checklists and provide them in their guidance documents (ANAES [Durocher et al. 2000], German Toolkit 2000, MSAC 2000). However, every source of evidence should be appraised under the scope of validity, e.g. if a source of routine data, such as registry of side-effects, is going to be used, the quality and validity of the retrieved data should also be critically appraised and discussed. There are no standards or guidelines on how quality of sources of information other as the medical literature should be appraised. The tools and criteria developed for the medical literature are

⁶ In Appendix 4, validated appraisal tools for different study designs are collected.

not applicable to other sources of information, so there is a gap here which needs to be addressed in the future.

Hierarchies of study design have been developed, referred to as levels of evidence, where RCTs or meta-analysis from RCTs are usually classified as the highest level of evidence, as they are the study design less likely to provide biased results⁷. The inclusion threshold for studies can rely on those hierarchies; however, it may depend on the average quality of all the evidence (e.g. if no RCTs have been done, other kinds of studies may be included). For certain aspects, e.g. psychological/social/ethical considerations, the existing hierarchies may not be applicable at all.

Besides hierarchies of evidence, several quality checklists have been developed to assess the quality of studies (Moher et al. 1999a). Although standard quality assessment instruments/checklists/scores exist, such as the validated Jadad-Score (Jadad et al. 1996), some agencies recommend developing specific instruments for each assessment, as some quality issues are closely related to special aspects of the technology being assessed. The criteria should cover both generic and specific methodological aspects. Generic methodological aspects refer to study characteristics which if present, for example, indicate good quality of a study independent from the subject being studied (e.g. concealment of allocation). Specific methodological aspects refer to characteristics, which if present, for example, indicate good quality of the study for evaluating the specific question (e.g. length of follow up needed to assess relapses varies with the condition/intervention) (Box 11).

Box 11. Quality items/criteria
<ul style="list-style-type: none">• Generic methodological issues (e.g. study design, allocation of concealment, prospective, randomisation, drop-out-rate, etc.)• Specific methodological issues (e.g. length of follow up, methods for assessing outcomes, ways of applying technology, etc.)

This step should be reported in a transparent way. For each study, how or whether it fulfils the different quality items should be documented. An overall score that synthesises all the items might be also used, and if so, the way the score is

⁷ A comprehensive hierarchy of levels of evidence for different kinds of interventions has been developed by the Centre for Evidence Based Medicine at the Oxford University. This is provided in Appendix 6.

constructed should be explained. If a score is used, studies not reaching a defined threshold score will be excluded. However, since different overall scores may lead to different thresholds for excluding studies, possibly resulting in unexplained differences in the results of meta-analyses, a detailed checklist with ratings of the different quality items (component scale) should be used (Jüni et al. 1999).

Some criteria for appraising quality may be so-called “knock-out” criteria, which means that studies not fulfilling them will be automatically excluded, even if they fulfil all other quality criteria. If knock-out criteria are being used, which they are and why they were chosen should be clearly stated. Studies originally retrieved which do not fulfil the quality criteria will be excluded; documentation of excluded studies should be provided, along with the reasons for exclusions (Box 12).

<p>Box 12. Transparency in Quality Assessment</p> <ul style="list-style-type: none"> • Document and explain quality criteria and items included in assessment • If a score is used, describe how it is constructed • List retrieved studies which were not included with reasons for exclusion • Fully report results of quality assessment (tabulation)

A good approach for reporting the quality assessment is the use of tables, as recommended by the DES, where quality items assessed are listed and the degree to which studies meet the criteria is documented. These tables could be completed with a statement about whether a study was subsequently included or excluded. The use of such tables allows readers of HTA to assess and decide on the quality of the studies themselves (Fig. 3).

Fig 3. Quality assessment presentation (example) (DES 2000)				
	Prospective	Concealment	Follow-up sufficient	Included in assessment
Study 1	Yes	Yes	Yes	Yes
Study 2	Yes	No	Yes	Yes
Study 3	No	No	No	No
Study 4	No	Yes	Yes	No
Study 5	Yes	Yes	No	Yes
Study 6	Yes	No	No	Yes

4.5.3 Non-quantitative and quantitative synthesis

The next step to be taken is the extraction of the relevant data for the assessment from included studies and its synthesis in a way that allows comparison among studies. Data to be extracted is mainly determined by the research questions. It is strongly recommended that customised extraction sheets are used. As with the selection of studies, the process of data extraction should be done by more than one person; however, this is not always possible. The way the data were extracted should be reported.

The information will then be synthesised and presented in a clear and understandable way. This should be done for all aspects assessed. A clear methodology has been developed for the quantitative synthesis of data on efficacy and effectiveness of therapeutic interventions, and, to some extent, for therapeutic interventions. For the synthesis of data concerning other kinds of technologies or other aspects of the assessment, a methodology is being developed but no clear standards are yet available. If no quantitative synthesis can be made, the narrative way of summarising information can be used.

In HTA, synthesis should be transparent. A way to enhance transparency, even if synthesis is narrative, is the use of *evidence tables*. These tables are commonly used to summarise medical literature, but they can also be applied to other sources of information. The information contained in evidence tables may vary depending on what kinds of studies are being used and also on the scope of the assessment. The rationale for such tables is to present in a structured way the sources of information/data, the issues concerning their validity and quality, and their results (Box 13).

Box 13. Elements to include in evidence tables

- Reference, year
- Study type and design issues (if not a study, characteristics of the data source, e.g. registry of routine data)
- Setting
- Patient characteristics, subgroups
- Interventions, characteristics of the intervention
- Outcomes measured and methods
- Results
- Overall quality score, if used
- If appropriate, statement as to whether study was included in meta-analysis

If such kinds of tables are used, readers can easily compare sources and results and make their own judgements about their validity.

In order to include all the information needed in the tables, different tables may be constructed for study design issues, patient characteristics, results, etc. A standard way of constructing evidence tables has not been identified, mainly because this depends on the assessment problem. However, all results and characteristics of the included studies, which may have influenced the results or which are relevant for the generalisability of results, should be presented in a way that enables easy comparison between included studies.

When recommending the use of evidence tables to summarise study characteristics and study results as the best way to synthesise the evidence in a non-quantitative form (which always precedes a quantitative synthesis), agencies and other institutions coincide. In a non-quantitative synthesis, consistency of results throughout studies or heterogeneity among studies (e.g. differences among patients or relevant details of the intervention) can be explored. Furthermore, lack of valid or relevant evidence can also be identified. In the non-quantitative synthesis of information, explicit criteria for validity and quality of the studies have to be followed. Thus, the non-quantitative synthesis is closely related to the appraisal process (section 4.5.2).

An important issue here is also identifying possible duplicate publications of results. Studies may be reported several times and it is often difficult to detect which reports refer to the same trial (Cochrane Collaboration [Clarke & Oxman 2000]). These issues may only be clarified by contacting the principal investigators of the studies in question. In addition, results of studies may be reported in a fragmented way in several publications, referring to different outcomes, different patient groups or different lengths of follow-up (so called “salami-publication”). Sometimes it can be very difficult to assess how and to which extent publications of the same studies overlap. This is especially a problem in trials of rare diseases which may lead to repeat publications of sequential case series. Again, the principal investigators of the trials should be contacted directly to clarify overlap between study populations.

The decision as to whether a quantitative synthesis can be performed and if so, which results can be pooled into what comparisons, will be made from the results of the non-quantitative summary of the available evidence. If significant heterogeneity

among studies or lack of validity of results are identified, a quantitative synthesis may not be indicated.

There are different methods for performing a quantitative synthesis for HTA doers⁸. However, the most extended one is the use of meta-analysis. Box 14 gives an overview of the factors that should be taken into consideration when choosing a method of meta-analysis.

Box 14. Factors to consider when using Quantitative Synthesis (meta-analysis)* (adapted from QUOROM statement [Moher et al. 1999b] and Egger et al. 2001)

- Why does the meta-analysis approach seem possible and appropriate?
- Which studies are being included in meta-analysis and why?
- Which comparisons are going to be made and why?
- Which outcome measures are chosen and why?
- Which summary statistics (OR, RR, WMD, etc.) are chosen and why?
 - type of data (e.g. binary, continuous)
 - consistency of treatment effects across trials
 - ease / plausibility of interpretation of summary estimate
- Which weighting method is used?
 - reliability when sample sizes are small
 - reliability when events are rare
 - degree of imbalance in allocation ratios among groups
- Is heterogeneity explored? Possibilities to consider heterogeneity:
 - meaning of a meta-analysis depending on degree of disagreement between studies
 - use of random effects model
 - accounting for variations in treatment effects (e.g. meta-regression, stratified analysis)
- Is the presence and possible effect of publication bias taken into account?
- Is a sensitivity analysis carried out?

*Some of the issues listed should have been already specified in the review protocol; however, after the qualitative approach of the evidence, it may be necessary to modify some of these. Modifications should be clearly stated and justified.

⁸ A comprehensive review on quantitative synthesis methods is found in: Systematic reviews of trials and other studies (Sutton et al. 1998). An up-to-date review of the methods of meta-analysis of binary and continuous results is available in Egger et al. 2001.

In addition to assessing the problem of publication bias, robustness of results of a meta-analysis should be tested. This is done through a sensitivity analysis which enables an assessment of how sensitive results are to changes in included studies (e.g. studies of lower quality, or studies suspect of double publication) or in statistical methods of synthesis (random effects model, fixed effects model).

Certain types of modelling are other tools for quantitatively summarising information (AETS; Imaz-Iglesia et al. 1999). The use of models has usually been discussed as a part of the economic analysis; however, it also constitutes a way of comparing different options by quantifying their final results. By quantifying the results of different alternatives, the decision regarding which to choose can be simplified, as the more favourable way will be identified by the means of an overall score.

In addition, the use of modelling can be useful for other purposes, many of which aim at providing more information than “just” a quantitative synthesis of available evidence (Box 15).

Box 15. Uses of modelling (Adapted from EUR-ASSESS 1997)

- Include different sources of evidence in a structured way
- Generalise results to other settings and extrapolate data from studies to populations
- Include several aspects which influence the final outcomes

There are different methods for modelling, such as *decision-trees*, *Markov-models* or *threshold* analyses (Sloan 1995, Gold et al. 1996). The use of mathematical models implies some assumptions, which have to be explained. A model needs to be fed with probabilities (e.g. of having an illness, of suffering an event), which will be taken from different sources (e.g. meta-analyses, single studies, experts opinions), thus having different grades of validity. Therefore, the sources of data which feed the model have to be transparently stated. The results of models should be carefully interpreted, taking into account the validity of the data introduced in them and the assumptions made. A sensitivity analysis, conducted by varying the values from particular variables or by modifying the underlying assumptions, should always be made in order to explore how these influence the final results of the model. A comparison of results with other approaches or other models should also be made (Box 16).

Box 16. Modelling

- Why has the modelling approach been chosen?
- What kind of modelling method is used? and Why?
- Variables used (Which ones? Why? Sources?)
- Assumptions being made (e.g. pathways)
- Sensitivity analysis
- Comparison with other models' results

The different methods of quantitative synthesis provide complementary information and do not substitute each other.

4.6 Specific methodological considerations

In the following sections, methodological considerations concerning sources of information, outcomes or ways to synthesise will be addressed for specific aspects of an assessment.

4.6.1 Safety

Assessing safety implies a wide scope in order to identify all possible harm caused through the use of a technology and should be based on all available data for assessing adverse outcomes of an intervention (MSAC 2000). In its guidelines, the MSAC recommends reporting all possible harm related to the use of a technology in the form of a summary table. Outcomes relevant to safety may be adverse effects, morbidity or mortality caused by the use of the technology.

Data sources for outcomes related to safety are the medical literature and routinely collected data (e.g. from regulatory authorities such as the FDA, from clinical databases, from quality assurance projects).

Although severe adverse effects of a technology may lead to a reduction in efficacy or effectiveness (e.g. because of less survival) in an RCT designed to assess those aspects, this study design, first, is not always able to identify all possible harm caused by the use of the technology. In RCTs, only what was looked for will be seen. Second, the reporting of RCTs in regard to quality and quantity of safety (adverse effects and laboratory-determined toxicity) is currently largely inadequate (Ioannidis & Lau 2001); thus, it is extremely important to carefully examine the reasons why

subjects leave the study, as the presence of adverse effects might have been an exclusion criteria.

Other study designs, such as observational studies, have an important role in identifying infrequent but serious adverse effects. This is because these designs can provide reliable evidence about adverse effects when the outcome of interest is rare among those not exposed, the excess risk among the exposed is large or there are no obvious sources of bias likely to account for the observed association (MacMahon & Collins 2001). As a result, these study designs should also be considered when assessing safety. Also, as case-reports of adverse effects of a technology may be useful when describing its safety, the MSAC recommends a special literature search for such a publication type.

Routinely collected data can complement the ones obtained from the literature. The quality and validity of these data are variable. Often these databases are generic and may not contain enough information. However, they have advantages, such as bigger size or coverage over long periods of time.

The different sources of data on safety should be documented, taking into consideration their quality and validity. Presentation through tables is transparent and may be helpful in summarising the different data.

When discussing the safety of a technology, the way adverse effects are caused should be described. Harm may be device dependent or related to the application of the technology. The occurrence of adverse effects may be also operator or setting dependant (e.g. learning curve of surgeons), which also need to be also taken into consideration and discussed. Timing (short-term, long-term) and severity of adverse effects should be considered, too. Another important aspect of safety is the identification of differences in risk among different groups of patients.

When possible, quantification of harm into quality- or disability-adjusted life years (QALYs, DALYs) should be made (DEC 2000). Safety can be summarised as frequency of adverse effects, relative risk or as the number needed to treat to produce one episode of harm (NNH)⁹. Sometimes it may not be possible to calculate frequency, and, in this case, harmful effects should then be listed.

⁹ Currently known as Number Needed to Harm (NNH).

4.6.2 Efficacy and effectiveness

Efficacy of a health technology refers to its performance under ideal circumstances, such as study conditions. Effectiveness is the extent to which the technology works in day-to-day practice (see Box 17).

Box 17. Definitions of “efficacy” and “effectiveness”		
Efficacy	Effectiveness	Source
the ability of a particular medical action in altering the natural history of a particular disease for the better, under ideal conditions	the ability of a particular medical action in altering the natural history of a particular disease for the better, under actual conditions of practice and use	Cochrane 1971
the probability of benefit to individuals in a defined population from a medical technology applied for a given medical problem under ideal circumstances of use.	the benefit of a technology under average conditions of use	U.S. Congress 1978
maximum achievable benefit	achieved benefit	Williamson 1978
Can it work? Does the maneuver, procedure, or service do more good than harm to people who fully comply with the associated recommendations or treatment?	Does it work? Does the maneuver, procedure, or service do more good than harm to those people to whom it is offered?	Sackett 1980
what works under carefully controlled conditions, such as randomized clinical trials	what works in day-to-day clinical practice	Rettig 1997

The accepted methodology for assessing **efficacy** is to conduct a systematic review following the principles of the Cochrane Collaboration. It is also accepted that reviews are based on the findings of RCTs. Many areas of health care, however, have not been and often cannot be evaluated with RCTs, and, in these cases, assessment based on other study designs is justified. Besides this fact, another problem concerning RCTs is that the patients included in them do not necessarily represent the assessment’s target population. Even if the clinical characteristics were the same, however, they are different because patients included in RCTs gave consent to participate in the trial, and differences among those who choose to participate and those who choose not have been observed. Thus, effects observed

in a RCT represent an “ideal world” and do not necessarily have to be observed in the target population, or the “real world” (DIHTA; Kristensen et al. 2001).

Before conducting a systematic review, the need for it should be carefully assessed. At this point of the assessment, when the research questions have already been clearly formulated, a search for systematic reviews which could contain answers for those questions should be made. An important source of this kind of literature is the Cochrane Library (see Appendices). Search filters to identify systematic reviews have been developed and may be useful (CRD; Khan et al. 2000 and at <http://www.york.ac.uk/inst/crd/revs.htm>). If systematic reviews on efficacy are found which may be suitable for answering the questions of the current assessment, their quality and relevance have to be assessed, in order to decide if they can be included in the assessment. Checklists to critically appraise systematic reviews have been developed and are summarised in Box 18.

Box 18. Key issues in assessing systematic review articles (adapted from Oxman et al. 1994, Greenlagh 1997)

- What are the review questions? Are they relevant for the current research questions?
- Which sources were searched? How were they searched?
- Are selection criteria explicit and appropriate?
- What criteria were used to assess study quality?
- How were the data extracted?
- How were the data synthesised?
- Are the results of the review transferable to my context?
- Should the review be updated?

If an identified systematic review contains all information needed to assess efficacy, undertaking a new one might not be justified. An existing systematic review of good quality may only need to be updated.

If there is no relevant or usable secondary research, a systematic review is justified. When conducting a systematic review, a review protocol has to be formulated. The questions, the outcomes to be measured, the inclusion and exclusion criteria for studies, the search strategy and the planned analyses should be prospectively stated. Some of those points (e.g. the research questions) have already been defined in the HTA protocol, but others (e.g. inclusion/exclusion criteria) need to be

refined when undertaking the review. The review protocol can be seen as a part of the HTA protocol. Comprehensive methodological guidelines already exist on how to conduct systematic reviews of primary research¹⁰.

In contrast to these guidelines, little consensus exists in regard to how to measure **effectiveness**, especially “**community effectiveness**”. Tugwell et al. (1984) have proposed that the latter should be calculated as “efficacy x diagnostic accuracy x health professional compliance x patient compliance x coverage”. More systematically, one could differentiate between factors influencing the access to a procedure and factors influencing the actual process of the procedure. Regarding the former, important variables relate to the health care system (e.g. availability of health insurance, inclusion of service in benefits catalogue, geographical access), providers (e.g. appropriate/ inappropriate indication for service, which may be influenced by payment system) and patients (e.g. felt need for service, availability of information). Regarding the latter, important variables mainly relate to providers (especially technical quality of service) and patients (especially compliance) (Busse 1998). “Effectiveness” is thus the result of a complex interrelationship of efficacy with system-, provider- and patient-related variables. Many of these variables are the outcomes explored under different aspects of the assessment (especially psycho/social/ethical considerations and organisational/professional implications) and a solid estimation of “community effectiveness” is therefore possibly better placed in the conclusions section which brings together the evidence from the various strands.

4.6.2.1 Therapeutic interventions

In the slightly differing models which define levels of evidence, RCTs are always seen as the most valid approach for evaluating therapeutic interventions. However, evidence from RCTs will not always be available. Furthermore, RCTs may not always be suitable for the evaluation of some therapeutic interventions (e.g. if randomisation is not ethically justifiable). In such cases, the HTA doers will have to use evidence from other kinds of study designs. Optimised standard search

¹⁰ Systematic reviews of trials and other studies (Sutton et al. 1998); Undertaking systematic reviews of research on effectiveness (Khan et al. 2000).

procedures have been developed to find RCTs¹¹ and thus other search strategies may be needed if other study designs are to be included.

As mentioned above, when assessing efficacy and effectiveness of therapeutic interventions, health-related outcomes (e.g. mortality) should be used. Using physiological or biochemical outcomes (= “surrogate” outcomes) should be avoided as far as possible as they may not correlate with the health-related outcomes. Thus, if surrogate outcomes are used, the underlying assumptions have to be clearly stated and results should be regarded carefully. Reliance on surrogate outcomes may be harmful and even lethal (Gotzsche et al. 1996).

The methodology of meta-analysis has been mainly developed for combining the results of RCTs on therapeutic interventions and is comprehensively described elsewhere¹². However, the meta-analytical approach can also be applied to other study designs, such as observational ones.

As already mentioned, the main steps of a meta-analysis include pooling results, testing heterogeneity, carrying out a sensitivity analysis and testing for publication bias. A meta-analysis should only be conducted after the adequacy of statistically combining results has been assessed by means of a non-quantitative synthesis. Results of meta-analysis of therapeutic studies should be graphically presented using the forest plot, including confidence intervals.

The discussion of the results of a meta-analysis is an essential element, and should not be too superficially addressed. Here, the effects of a possible publication bias or of heterogeneity among studies should be addressed. In addition, the relevance and generalisability of results for the questions of the HTA should also be considered, taking into account the characteristics of patients and settings involved in the studies pooled in meta-analysis.

4.6.2.2 Diagnostic interventions

There are two kinds of technologies which aim at identifying conditions of patients: *diagnostic tests* and *screening tests*. *Screening* is the detection of disease in an asymptomatic population, whereas *diagnosis* is the confirmation of the presence or

¹¹ Optimal procedures are described in the manuals listed in Appendix 3 or are available at <http://www.york.ac.uk/inst/crd/revs.htm>.

¹² Cochrane Reviewers Handbook 4.1.1 (Clarke & Oxman 2000).

absence of disease in a symptomatic patient (CRD; Khan et al. 2000). The evaluation of both follows similar principles.

For the assessment of diagnostic and screening tests, a hierarchical model can be followed (Box 19).

Box 19. Evaluation of efficacy and effectiveness for diagnostic interventions (Adapted from Fryback & Thornbury 1991, Flynn & Adams 1996)	
“Level”	Typical measures
<ul style="list-style-type: none"> • Technical efficacy 	<ul style="list-style-type: none"> • Physical parameters describing technical performance of the test (e.g. image quality)
<ul style="list-style-type: none"> • Diagnostic accuracy efficacy 	<ul style="list-style-type: none"> • Sensitivity (% of positives among ill) • Specificity (% of negatives among healthy) • Accuracy (% of correct diagnoses) • Likelihood ratio (likelihood for a given test result in a patient with the target disorder compared to the likelihood of the same result in a patient without the target disorder; details at http://cebm.jr2.ox.ac.uk/docs/likerat.html)
<ul style="list-style-type: none"> • Diagnostic thinking efficacy/ effectiveness 	<ul style="list-style-type: none"> • Post-test odds/ probability compared to pre-test odds/ probability in target population • % of cases in which test is judged “helpful” to making diagnosis
<ul style="list-style-type: none"> • Therapeutic effectiveness 	<ul style="list-style-type: none"> • % of cases in which test is judged “helpful” in planning therapy • % of therapeutic procedures avoided due to test information
<ul style="list-style-type: none"> • Health-related effectiveness (Patient outcomes) 	<ul style="list-style-type: none"> • Mortality/morbidity avoided with test • Changes in quality of life through use of test

This hierarchy does not represent a hierarchy of levels of evidence (see Appendix 6), but a hierarchy of outcomes evaluated. Each level requires establishing evidence on the prior level. For the evaluation at each of the stages, studies belonging to different levels of evidence can be conducted.

In HTA, the evaluation of diagnostic technologies should be based on patient related outcomes, as they represent the actual effects of such tests in the health of patients. However, such evidence is not always available and efficacy of the technology is assessed based on test accuracy, sensitivity, specificity or likelihood ratios, which can be seen in this context as “surrogate parameters” for the real effect on the outcomes of the patients. When assessing any of these parameters, it is crucial that the diagnostic technology is evaluated against the “gold standard” (which is not in every case well established). The diagnostic technology should be ideally evaluated

in a patient sample that includes an appropriate spectrum of patients with the target condition plus an representative group of individuals without the disease (Flynn & Adams 1996). Both the positively and the negatively tested patients should be compared with the diagnostic gold standard, i.e. not only those who are tested positively (though, depending on the invasiveness of the gold standard, this might raise ethical issues). Ideally, the allocation of positively and negatively tested persons to the gold standard technology should be randomised and the examiners blinded regarding the result obtained with the diagnostic technology.

For the quantitative synthesis of studies on diagnostic tests, several methods have been proposed. The choice of the method depends mainly on homogeneity of results, type of outcome (binary, continuous) and variation in diagnostic thresholds. Nevertheless, all available meta-analytical methods summarise results of diagnostic accuracy.

Most frequently, studies on diagnostic accuracy use different study populations, different settings and different cut-points (diagnostic thresholds). For this situation, the method of Littenberg and Moses (SROC curves) has been proposed as standard approach (Irwig et al. 1994, 1995, Egger et al. 2001). In SROC curves, the area under the curve represents the accuracy of the test to diagnose the condition. This approach is attractive since it is easy to calculate and presents the results in a graphically appealing way. Another approach can be to pool the LR of the studies into a summary LR. This approach should be used only in cases of homogeneity of study results. There is still an ongoing debate as to which is the most suitable statistical method to pool test-accuracy studies. Thus, a good approach is to use several methods and test the sensitivity of the summary results to the method chosen (CRD; Khan et al. 2000).

When assessing a diagnostic test or strategy, outcomes deriving from misclassification / misdiagnosis of patients can also be considered as harm (MSAC 2000).

4.6.2.3 Health care organisation and system related interventions

Organisational, financial or regulatory interventions can also be considered as health technologies. As defined by the EPOC Group¹³, different types of interventions, such

¹³ Effective Practice and Organisation of Care Review Group, within the Cochrane Collaboration, which is elaborating some guidelines on how to review such kind of interventions. The guidelines from this group can be found at <http://www.abdn.ac.uk/hsru/epoc/down.hti>.

as professional (e.g. educational program on prescription), financial (e.g. co-payment), organisational (e.g. changes in medical record system) and regulatory (e.g. licensure) are included here. These interventions are not to be confused with organisational, professional and economic implications of introducing or applying a health technology (cf. sections 4.6.4 and 4.6.5).

For the evaluation of professional, financial, organisational or regulatory interventions, the HTA doers need often to be more flexible in their inclusion criteria for studies. Transparency in the selection process is of utmost importance as generalisability/ transferability to other settings will be highly context-dependent. Box 20 lists available study design by their methodological strength (with the weakest designs towards the lower left, marked in grey).

Box 20. Study designs used for assessing health care organisation and system related interventions (adapted from Busse 1998)			
	Cross-sectional	Longitudinal	
	1 point of measurement	2 points of measurement	Regular/ continuous measurements
<i>Experimental designs – often not feasible for evaluating health care organisation and system related interventions</i>			
Researcher has control over intervention and allocation of subjects/ institutions/ areas etc. into at least 2 groups; randomisation possible		classical experiment (randomised controlled trial = RCT)	
Researcher has control over intervention and allocation of subjects/ institutions/ areas etc. into at least 2 groups; randomisation not possible	post-test only with non-equivalent groups – <i>weak design</i>	control group design with pre- and post-test/ controlled before and after study	time series with non-equivalent control group/ cohort study
<i>Quasi-experimental designs – feasible for evaluating health care organisation and system related interventions</i>			
Natural experiment (i.e. intervention not determined by researcher) with randomised allocation of subjects/ institutions etc. into at least 2 groups through researcher		quasi-RCT – <i>theoretically possible and desirable but de-facto hardly ever used; requires 1. a dialogue between health politicians and researchers and 2. enough time before the intervention to prepare evaluation</i>	
Natural experiment with non-randomised allocation of subjects/ institutions etc. into at least 2 groups	post-test only with non-equivalent groups – <i>weak design</i>	control group design with pre- and post-test/ controlled before and after study	time series with non-equivalent control group/ cohort study
Natural experiment without prior allocation of subjects/ institutions etc.; control group existing		case-control study – <i>not ideal but a compromise if pre-intervention measurements were not possible</i>	
<i>Simple, methodologically weak designs</i>			
Intervention but no control group	one-group post-test only design	one-group pre-test-post-test design	simple interrupted time series – <i>acceptable if at least three data points before and three after the intervention</i>

Effectiveness of such interventions can be measured using patient health outcomes, but usually other, more process-related outcomes are measured (e.g. number of drugs prescribed, number of patient-physician contacts).

4.6.2.4 Preventive interventions

Preventive interventions intend to avoid having a target condition appear in a target group. They may be implemented at an individual level, making them comparable to therapeutic interventions (e.g. use of aspirin to prevent stroke), and thus evaluated using the same methodology (see section 4.6.2.1). Others, such as screening programmes, are more diagnostic and have to be implemented at a community-level; these have to incorporate the considerations listed both for diagnostic interventions (see section 4.6.2.2) as well as for organisational and system related interventions (see section 4.6.2.3). Other community-based interventions include health promotion programs or public health strategies aiming at the population or environmental factors (e.g. fluoridation of drinkable water). Common methodological problems when assessing these kinds of interventions are the need for a long-follow-up time (e.g. several years), the use of big observation units (e.g. regions, communities, etc.) instead of individuals, and the difficulty of establishing clear causal relationships between intervention and outcomes.

Regarding the process and methodology of evaluating preventive technologies, the “Current methods of the Third U.S. Preventive Services Task Force” (Harris et al. 2001) can be regarded as “best-practice”. Building upon previous work (especially Battista & Fletcher 1988), the Taskforce uses two “analytic frameworks” to map out the specific linkages in the evidence that must be present for a preventive technology to be considered effective. The frameworks make explicit the populations, technologies (e.g. counselling, diagnostic or therapeutic interventions), intermediate and health outcomes to be considered in a review. Most often evidence is only available for individual components of a whole chain of technologies of interventions necessary for a preventive technology to be effective.

In its paper, the Task Force also describes issues such as literature search and abstraction, assessing magnitude of benefits and harms as well as translating the evidence into recommendations including the codes and wording of statements (see Appendix 6).

4.6.3 Psychological, social and ethical considerations

The assessment of the impact of the use or no-use of a technology in terms of psychological, social and ethical benefits or harm is an important part of HTA. Effectiveness of an intervention is influenced by the way it is experienced by those to

whom it is directed, by the way they value it, etc. (e.g. if there is no acceptance, compliance will be reduced and thus effectiveness too). Such aspects should therefore also be included in a structured way in a HTA.

Psychological effects of a technology refer to a range of possible subjective effects, such as fear, anxiety, feeling labelled, satisfaction, etc. caused by the use of the technology by the individual. Under social effects of a technology, changes in equity or access to care produced by the implementation of a technology can be addressed. The introduction of a technology may, for example, improve the lot of the rich or middle-class while not touching the poor, so that the poor become relatively more disadvantaged. Addressing ethical implications of a technology refers more to the exploration of all possible effects of technology on values (e.g. the use of a technology may foster judgements: for example, discrimination of handicapped life through the use of pre-natal diagnostic tests).

The way to approach these issues in HTA depends on the degree of available knowledge. For some of these aspects, information may already be available in the form of studies. The scientific approach for addressing these topics has been included in the field of the so called "Qualitative Research," involving areas of knowledge such as psychology or the social sciences. Following a rigorous methodology, these approaches allow important variables and effects of the technology from the point of view of the patients and the society to be explored and described. Now, some work is being done in order to make possible the inclusion of qualitative research in a systematic way when assessing health care¹⁴.

Evidence on these topics can be available to some extent from the medical literature and optimal search strategies, similar to the ones used to identify RCTs, which are being developed now to allow systematic search of studies using the methods of qualitative research in Medline¹⁵. Comprehensive databases exist for social sciences, which also include literature on psychological and sociological aspects of health interventions (e.g. PsycINFO, Sociological Abstracts¹⁶). If such a literature search is done, the origin of the data and the strategies followed to find the evidence

¹⁴ For instance, in 1998 the Cochrane/Campbell Qualitative Methods Group (CQMN) was established, which focuses on including qualitative research in systematic reviews and developing methods to search for and critically appraise such studies. This group is also developing some methodological checklists for qualitative research (accessible at <http://www.salford.ac.uk/iphrp/cochrane/homepage.htm>).

¹⁵ Grant MJ. Searching for qualitative research studies on the Medline database. Presented at the Qualitative Evidence Based Practice Conference, Coventry, 14th-16th May 2001.

should be clearly stated. Literature found should then be assessed for their validity, quality and transferability. Some criteria for appraising qualitative research used in health care research have been proposed and are summarised in Box 21; however, debates on this are still ongoing.

Box 21. Sets of criteria for assessment of studies using qualitative research methods (updated from CRD; Khan et al. 2000)¹⁷

I. (Popay et al. 1998)

- A primary marker: is the research aiming to explore the subjective meanings that people give to particular experiences of interventions?
- Context sensitive: has the research been designed in such a way as to enable it to be sensitive/flexible to changes occurring during the study?
- Sampling strategy: has the study sample been selected in a purposeful way shaped by theory and/or attention to the diverse contexts and meanings that the study is aiming to explore?
- Data quality: are there comparisons of different sources of knowledge/understanding about the issues being explored?
- Theoretical adequacy: do the researchers make explicit the process through which they move from data to interpretation?
- Generalisability: if claims are made to generalisability, do these follow logically and/or theoretically from the data?

II. (Mays & Pope 1996)

- Adequate description: Is sufficient detail given about the theoretical framework of the study and the methods used? Is the description of the context for the study clear? Is there an adequate justification and description of the sampling strategy? Is the description of the fieldwork clear?
- Data analysis: Are procedures for analysis clearly described? Is the analysis repeated by more than one researcher? Are findings from quantitative research used to 'test' qualitative findings? Is there evidence that the researchers have looked for contradictory observations?
- Link to theory: Is the study design and sampling strategy theoretically grounded? Does the link to theory inform the analysis and any claims for generalisability? Is sufficient original evidence provided to support relationship between interpretation and evidence?

III. (BSA Medical Sociology Group 1996)

- Are research methods appropriate to the question being asked?
- Is there a clear connection to an existing body of knowledge/wider theoretical framework?
- Are the criteria for/approach to sample selection, data collection and analysis clear and systematically applied?
- Is the relationship between the researcher and the researched considered and have the latter been fully informed?
- Is sufficient consideration given to how findings are derived from the data and how the validity of the findings was tested?
- Has evidence for and against the researcher's interpretation been considered?
- Is the context for the research adequately described and accounted for?
- Are findings systematically reported and is original evidence reported to justify a relationship between evidence and conclusions sufficient?

¹⁶ For more see Appendix 2.

¹⁷ A further checklist, based on Giacomini & Cook 2000a/b, is provided in Box A4-12 (appendix 4).

- Are the researchers clear about their own position in relation to the research topic?

IV. (Mays & Pope 2000)

- Triangulation (comparison of results from two or more different methods)
- Respondent validation (comparison of investigator's account with those of research subjects to establish level of correspondence)
- Clear exposition of methods of data collection and analysis
- Reflexivity (discussion of the ways the researcher and research process have shaped collected data)
- Attention to negative cases
- Fair dealing (incorporation of a wide range of perspectives)

In the sense of levels of evidence, no hierarchy of study designs in qualitative research has yet been proposed. In fact, the use of more than one of the methods available in one study (triangulation of methods) is seen as a sign of high quality in a study (Mays & Pope 2000).

If no evidence from the literature is available, the HTA doers may need to conduct primary research by themselves, in order to include the patient perspective when assessing a technology. Some of the methods which can be applied for this purpose are participant observation, individual interviews, focus group discussions, Delphi method or future workshops¹⁸. If such primary research is going to be conducted within the HTA, expertise is needed in the use of this methodology, highlighting the multidisciplinary nature of HTA. The criteria exposed in Box 21 are also applicable to primary research.

Another source of data can be surveys or questionnaires about some aspects, e.g. satisfaction, acceptance. These sources may give more representative data, but they may only be useful to map phenomena which are already known (DIHTA; Kristensen et al. 2001). The knowledge gained through qualitative research can be complemented with quantitative approaches.

However, time and financial constraints may not allow such a comprehensive approach to address psychological or social aspects, and the HTA doers may use other sources of information like patient organisation websites to gain knowledge about the perspective of the patients or make some assumptions about the possible psychological/social implications and the ethical considerations of a technology. Such an approach can be considered as a "document analysis", which is part of the

¹⁸ A comprehensive review of qualitative methods is found in: Qualitative research methods in health technology assessment: a review of the literature (Murphy et al. 1998). Some of these methods are also described in the Handbook of DIHTA (Kristensen et al. 2001).

methodological tool kit available in qualitative research. Thus, it should also be systematic. It is important to clearly state the sources of data, methods used, and assumptions made when approaching these aspects, in order to maintain the principle of transparency and warrant that all positions are represented. Furthermore, HTA doers have to be careful not to rely on their own moral stance (EUR-ASSESS 1997).

In summary, assessment of psychological, social and ethical considerations refers to the inclusion of the public perspective in a structured way in HTA. These aspects determine public preferences about technologies and thus, their assessment could also be considered a tool of HTA¹⁹.

4.6.4 Organisational and professional implications²⁰

The scope of a HTA report should also include organisational and professional changes induced by the technology and predict their further consequences, especially if the background information indicates important implications (cf. Section 4.3). For instance, the use of a new surgical procedure may imply training of staff, but also reduce hospital length of stay, the need for hospital beds, and potentially the cost for treating patients with this condition. (This may or may not lead to conclusions and/or recommendations for reducing the number of hospitals beds, or alternatively, for using for patients with other indications.)

Organisational issues to be assessed may, for example, address changes in:

- utilisation of service (for example, if the introduction of a pharmaceutical therapy reduces or even replaces surgical interventions),
- change in the treatment location (for example, if a traditional in-patient treatment, by means of the new technology, can be performed as an out-patient procedure),
- training/ qualification requirements (for example, if the application of a health technology – in contrast to its alternatives – presuppose the skills of a special medical expert),

¹⁹A review on methods for assessing public preferences is included in: Eliciting public preferences in HTA: a systematic review (Ryan et al. 2001).

²⁰ The issues discussed here, i.e. impact and effects of the technology under consideration on organisational and regulatory issues should not be confused with the issues discussed in 4.6.2.3, i.e. efficacy and effectiveness (in terms of health outcomes) of organisational interventions.

- channels of co-operation/ communication (for example, if the effective use of a health technology presuppose extra communication between hospital and general practice), and
- job satisfaction (for example, if a new procedure presuppose such a high throughput that the physicians have insufficient time for following the patients' progress).

As an organisation is a social interaction, within given frames, between persons who have one or more common ends but also individual goals and aspirations, it is useful to start analysing organisational issues by identifying the stakeholders and their interests (for a review of stakeholder analysis see Brugha & Varvasovsky 2000).

An assessment of such issues gives the first picture of the technology's (potential) organisational impact. It may be relevant then to assess – often even to propose and then assess – a strategy for implementing the technology. Some stakeholders may be very interested in promoting diffusion of the technology, whereas others display resistance to change.

Evidence from available studies may have addressed organisational changes induced by a health technology. Often results from such studies are not directly transferable due to for example social or cultural differences, but issues identified, and methods applied to assess them may be relevant and useful. Therefore, in addition to a critical survey of literature, doers often have to collect data from the organisation in which the technology is considered implemented.

Observational studies and individual interviews may be applied, but more often methods used for this data collection are:

- questionnaires, mainly concerning existing technologies, for factual issues, when the doer knows what kind of information is needed,
- focus group interviews, mainly concerning existing technologies, when only some of the issues are known to the doer, and others are searched for (Morgan 1993),
- structured group processes such as future workshop or Delphi method (especially when trying to identify and evaluate future changes of organisational structure and processes, or when trying to predict reactions of people involved in the implementation.

Recommendations of manufacturers and current legislation may be consulted in order to establish which changes are needed as well.

4.6.5 Economic issues

Assessments of economic issues in HTA implies first collecting information on resource consumption from the use of the technology (costs). The next step will be to conduct an analysis comparing costs to other outcomes, such as efficacy or effectiveness.

Most of the existing guidelines focus on the second aspect; Baladi (1996) provides a useful guide on the identification of resources, the measurement of resources, cost valuation and dealing with possible bias in estimating costs. DIHTA also provides helpful hints for HTA doers (Kristensen et al. 2001).

Generally, there are different types of costs which need to be taken into account depending on purpose and perspective (Box 22). For all of them, the importance of measuring physical units first, before multiplying them with unit costs/ prices to get total costs cannot be over-emphasised in order to help interpreting results regarding their transferability to other settings – not only from one country to another (Drummond et al. 1992) but also within one country across different providers (Coyle & Drummond 2001). If the data have been collected alongside a clinical trial, protocol-driven costs should be identified and excluded to make the results useful for HTA (Rittenhouse 1997).

Box 22. Types of costs in an economic analysis (modified from DIHTA; Kristensen et al. 2001)

Perspectives		Types of costs	Examples
Societal Perspective	Health care payer	Hospital	Direct costs Health care staff, medicine, tests, capital costs (equipment and buildings), inpatient stay (hotel), outpatient visits, overhead costs (e.g. food, light, heat), possibly research and education
		Ambulatory care	Direct costs Visits with general practitioner, ambulatory specialist, physiotherapist etc., prescription drugs (the share paid by the health care payer), screening programmes
		Direct costs (possibly in other sectors)	Rehabilitation, home care and nursing care at home, social arrangements
		Direct costs (for the patient and family)	User payment (medicine, dentist), cost for travelling, time costs due to patients time used for the treatment, family or friends (unpaid) use of time of the patient
		Lost production in the society	The patient's temporary absence from work due to illness, reduced working capacity due to illness and disablement, or lost production due to an early death
		Future health care costs	Future unrelated health care costs caused by curing the patient with the present treatment

The types of costs and the perspectives used in the analysis should be clearly stated in the report. Data on costs may be obtained from different sources; thus, the evidence used to calculate the costs has to be stated and assessed for quality. After calculating costs, economic evaluation is necessary to put these into relation with the other outcomes. Depending on the purpose and availability of data, different types of economic evaluations are available (Box 23).

Box 23. Types of economic analysis (DIHTA; Kristensen et al. 2001)

Type of economic analysis	When should the specific type of analysis be chosen?
<ul style="list-style-type: none"> • Cost-minimisation analysis 	<ul style="list-style-type: none"> • If the compared technologies are equally effective, then it is only necessary to collect data about costs
<ul style="list-style-type: none"> • Cost-effectiveness analysis 	<ul style="list-style-type: none"> • If the effectiveness of the compared technologies are different (e.g. the difference in costs have to be weighted against the difference in effectiveness) • If activities with the same aim and measure of effectiveness are compared

<ul style="list-style-type: none"> • Cost-utility analysis 	<ul style="list-style-type: none"> • If health-related quality of life is an important health outcome • If activities across specialities or departments in the health care sector have to be compared
<ul style="list-style-type: none"> • Cost-benefit analysis 	<ul style="list-style-type: none"> • If non-health effects also are of importance (e.g. the treatment process itself, utility of information) • If only one technology is assessed (net-benefit) • If there is a wish that individual lives are valued in monetary units • If activities across society have to be compared

Guidelines on economic evaluation are numerous, though they are not tailored for use within the context of HTA (e.g. Canadian Coordinating Office for Health Technology Assessment 1997, Drummond et al. 1997a, 1997b, Gold et al. 1996, Guyatt et al. 1986, O'Brien et al. 1997). The EUROMET project, i.e. the "European Network on Methodology and Application of Economic Evaluation Techniques", reviewed the contents of guidelines for economic evaluation of medical technologies from Australia, Canada, France, Germany, Italy, Spain, Switzerland and the United Kingdom regarding stated purpose, comparator, study design, time horizon, perspective, data sources, cost measurement, outcome measurement, discounting and sensitivity analysis (von der Schulenburg & Hoffmann 2000). The recommendations in guidelines regarding discounting only were recently compared by Smith and Gravelle (2001).

The EUROMET group also developed an consensus on a framework for European guidelines which is useful in the context of HTA (von der Schulenburg & Hoffmann 2000). Box 24 summarises the main issues for economic evaluation in HTA.

Box 24. Economic evaluation (based on the EUROMET consensus; von der Schulenburg & Hoffmann 2000)

- **Study frame:** clearly stated research question, identification of target population, explanation of choices and assumptions made etc.
- **Analytical technique:** choice to be explained
- **Study perspective:** societal perspective if the study does not require a narrower perspective
- **Selection of alternatives:** description and justification of choice; recommendation to use currently most effective or efficient alternative

- **Data collection:** to be described in detail; must include systematic review of literature; various types of studies and data sources are suitable
- **Costing:** all relevant direct and indirect costs should be identified, collected and reported; physical units should be reported separately from costs of resources; use of average values only if marginal data are not available
- **Outcome measurement:** primary outcome measures to be reported clearly; if values for health states are used, individual utilities should be distinct from modelling society's valuation
- **Time frame:** long enough to capture all effects; modelling can be used to estimate long-term costs and outcomes if real data are unavailable; shortening of time horizon has to be justified and possible bias estimated
- **Discounting:** necessary if costs and consequences occur at different times; use of standard rate (5%) plus national recommendation
- **Sensitivity analysis:** should be conducted to test robustness of results to a variation of assumptions, cost and outcome parameters and discounting rate
- **Equity:** values and preferences important but more valid indicators are needed

4.7 Discussion of methods and results

The discussion is an important part of a HTA. When addressing the different aspects of the assessment, part of the discussion will be possibly already carried out, as a part of the appraisal process and the non-quantitative synthesis (see sections 4.5.2 and 4.5.3). However a structured summary discussion should be always included in an assessment as a separate section. This section should include following parts (Box 25).

Box 25. Discussion

- Methodology of the assessment
- Evidence used (quality, validity, generalisability)
- Assumptions made
- Discrepancies and uncertainties identified
- Expected changes (in technology, in evidence)

The methodology followed to address the different aspects and its appropriateness for assessing those aspects should be discussed (e.g. meta-analysis, modelling).

Possible limitations of the approaches used should be discussed with special attention to their influence on the results. The evidence available should also be discussed. Possible sources of bias from the type of evidence used (e.g. study design issues) and their possible influence on the findings should be discussed. Discrepant findings from different sources of information (e.g. if a meta-analysis and a large RCT with discrepant results were included) and the way that the discrepancies were handled should be also addressed. The areas where weak or no evidence is available should be presented, pointing out areas in which future research is needed. It is important to state the degree to which objectives and questions posed at the beginning of the assessment were fulfilled with the chosen approach.

When different outcomes were used, the possible interrelations among them should be addressed in the discussion.

For the issue of generalisability, in addition to the characteristics of the participants in the studies, the identified practice differences between studies and actual practice should also be discussed. Furthermore, identified upcoming changes in the use of the technology or in the evidence (e.g. identified ongoing studies) which could influence the findings of the assessment should also be addressed.

In the discussion, relationships among the findings on the different aspects assessed should be explored, trying to find the ways in which they may influence each other, and discussing how the different findings may be transferable to the real setting in which the assessed technology will be and/or is being implemented. It is also important to discuss which aspects may have an influence on the implementation of the technology and on its effectiveness in the real settings.

In summary, the discussion should point out the limitations (from the method used, from the evidence/lack of evidence) of the assessment and their possible effects on the findings. The discussion can be seen as a needed previous step to formulating conclusions and/or recommendations.

4.8 Conclusions and recommendations

The *conclusions* of the assessment aim primarily at providing *answers* to the research questions. They should be brief, clear and explicit, highlighting the most relevant aspects so they can be easily understood and used.

Derivation from the evidence found in the assessment should also be clear; in this respect, the NHS recommends to report conclusions always starting with: “Based on the evidence...”.

Conclusions are often the most read part of an assessment, so they should contain a summary of the most relevant findings taking into consideration the issues of the discussion (Box 26).

Box 26. Conclusions

- Related primarily to the research question(s)
- Summarise quality/origin of the evidence
- Summarise evidence on all aspects assessed
- Give size of effect (benefit/adverse)
- Highlight differences among groups of patients (if found)
- Highlight variations of effect with varying characteristics of technology (if found)
- Discuss applicability of evidence for national/local context and “community effectiveness”
- Point out fields where further research is needed

Note: There are good reasons, although there is no consensus yet, to view the estimation or calculation of the *community effectiveness* of the technology as an issue for this section as it not confined to the efficacy/effectiveness dimension but needs to take into account psychological/social/ethical, organisational/professional and economic considerations. For example, if a technology with a high efficacy has low or absent acceptance in the population, or if professional training requirements are extremely high, then the community effectiveness will be very low or even zero.

An important aspect of the conclusions is to clearly point out the fields in which further research is needed (e.g. because no or weak evidence was found). This has to be seen as a major relevant finding of a HTA.

The elaboration of *recommendations* depends on the original *policy* questions and objectives of the assessment, as well as on the policy of the HTA commissioners (e.g. the NHS-CRD HTA-Programme explicitly prohibits making recommendations about policy or about clinical care), so this is a facultative component of an assessment. If recommendations are given, the audience of focus should be clear (e.g. for decision makers, clinicians). Recommendations have to be consistent with

the findings of the assessment and take into account the kind of evidence they rely on. The gradation of recommendations using hierarchies, which consider the quality of the underlying evidence, represents the best practice when giving recommendations. There are different gradation scales, so the HTA doers have to state which one was used and the way it is constructed²¹.

Besides recommendations for the policy-makers, clinicians, etc., recommendations referring to the need for further research or further aspects to be assessed should be made, if such needs were identified.

4.9 Other relevant issues

The following issues should also be taken into account when undertaking a HTA. A transparent HTA should include statements on all of these, as they are important when assessing the quality of the work and, to some extent, might be helpful in interpreting its results.

4.9.1 Review process

Agreement exists that some kind of external review is needed before publication and dissemination of the assessment. Undergoing such a review is seen as a quality attribute of HTA reports, although no clear best practice could be identified among the different models of review²². The review processes of different institutions should be evaluated in order to make further recommendations on this issue. For the purpose of future evaluation, it would be very helpful to always clearly state whether an external review was done or not, and, if so, to document the comments from reviewers and the way in which they were incorporated (if so) to the final report (Box 27).

Box 27. Review process
<ul style="list-style-type: none">• Did the report undergo an expert review before publication?• Who reviewed the report (disciplines)? Were there possible conflict(s) of interest?• Were the comments from reviewers incorporated into the final report? How?• How many comments were usable? How many were not usable?

²¹ In Appendix 6, scales for gradation of recommendations related to levels of evidence and quality of data (internal validity) are given.

²² Review models range from individual reviewers giving comments on the report to a comprehensive review process, including institutional boards and consensus finding approaches.

Ideally, a preliminary version of the report should be reviewed by experts in the methodology and in the field which is being evaluated. The aim of the experts review is to assure the quality, accuracy and validity of the report. The external review process is also seen as a way to improve acceptance of the report among professionals (German Toolkit 2000). Within ANAES, for example, the review process takes place in two stages. The draft report may first be reviewed by a panel of experts who did not participate in the working group. Afterwards, the report is always reviewed by the Agency's Scientific Committee. This committee is nominated by the government from a list of representatives of the different health care providers.

4.9.2 Updating of assessment

The validity of the findings of a HTA is limited, and, as a result, it is generally accepted that updating is an important component in the process of HTA. However, it seems to be difficult to determine when a HTA report should be updated. Some institutions (NICE/DES) use a set of different criteria to decide how long a report is valid, and when it needs to be updated. Depending on how the assessment was conducted it might be very difficult to give an exact expiration date for the report. It seems much more important to provide information about the updating process itself, and not about when. In the report, it should be made clear whether an update is planned, and, if so, how the need of an update is going to be identified (e.g. periodical literature search, hearings, etc). Box 28 shows an example of the way DEC decided on an update.

Box 28. Identification of the need for update (DEC 2000)

- **New Evidence:** Screening searches can be regularly made (e.g. annually if rapid change is expected) to assess whether new evidence relevant to the problem has appeared.
- **Controversy:** If interested parties communicate disagreement with report after publication, revision may be indicated.
- **Interest:** If interest is communicated by the public, update may be undertaken.

The update timing depends on expected changes in the evidence for the technology (e.g. ongoing relevant trials which could not be included, but were already identified).

It could also be indicated when there are organisational or regulatory changes which may influence utilisation or even effectiveness.

An update is typically made through the original search strategy again, for the period of time subsequent to the original assessment. Original selection criteria should be applied to the literature found. If there have been many changes, the original search strategy, selection criteria and approach may no longer be acceptable, making a full new assessment necessary.

To provide an assessment with a expiration date does not seem to make much sense, as the need for an update may present itself earlier or later, and to determine this in a prospective way does not seem possible. It is of much more interest to provide information on the mechanisms used to identify the need for update. As with the review process (see section 4.9.1), documentation of the updating process can be helpful for the future evaluation of different approaches. Information about updating the HTA should include the following aspects (Box 29).

Box 29. Update of HTA
<ul style="list-style-type: none">• Is an update planned?• How will be the timing / the need for the update be assessed?• If an update need is identified, how should the update be conducted?

If a standard institutional policy on updating exists, which is always the same, this does not necessarily need to be always reported, as it may be enough to refer to the source in which the process is described.

Chapter 5. “Best practice” in reporting HTA

The reporting of an assessment should include at least three kinds of documents:

1. “Abstract”,
2. “Scientific Summary Report” and
3. “Technical Report”.

Besides the “Scientific Summary Report”, the doers (or commissioners) of the assessment may also publish other summaries targeted at specific audiences (e.g. an “Executive Summary” aimed at decision-makers or a “Patient Information”), with different lengths and content. In general, the common structure of reporting scientific work should be followed: “Objectives/Questions”, “Methods to answer those questions”, “Answers found/ Results” and “Discussion/ Conclusions.” The three types of documents mentioned will differ above all on length and target audience.

In terms of making these documents available for a wide audience, it is now best practice (as practised by most HTA institutions, even though the toolkits/ guidelines do not mention this) to place them freely available in the internet (usually, in pdf-Format). It is however still necessary to print executive summaries, patient information etc. in order to reach the desired target audience.

In the following sections, the main characteristics of these three documents will be described, with special attention to the concept of “Scientific Summary Report”.

5.1 Abstract

Recommendations already exist on how to write a structured abstract for the INAHTA databank (<http://agatha.york.ac.uk/hta/hp.htm>). The “Abstract” has to be written in English. In its present form, it is usually too short to contain all aspects of interest when assessing the relevance and quality of a HTA report. The aspects to be included in the “Abstract” are listed in Box 30.

Box 30. Data to be included in English structured abstract (AETS; Imaz-Iglesia et al. 1999)
--

- | |
|---|
| <ul style="list-style-type: none">• Title: first title in English, then original title in brackets• Author/s: according to Vancouver style• Organisation: organisation commissioning the report• Contact person: name and address• Date: month and year of publication• Language: language(s) of publication |
|---|

- **Abstract:** specify whether summaries other than structured abstract are included and their language (e.g. “patient information summary in Dutch”)
- **Publication type:** report, clinical practice guideline
- **Pages**
- **References:** number of references cited
- **ISBN:** International Standard Book Number.
- **Technology type:** e.g. screening, diagnostic, therapeutic, organisational
- **Subject index terms:** it is recommended to use terms from Index Medicus, indicating the Major Descriptors with *. State which terms are Non MeSH: e.g. *Aortic Aneurysm – epidemiology; *Stents; Blood Vessel Prosthesis; Kharkov Stent (Non MeSH)
- **Objectives:** general and specific objectives
- **Methods:** Data sources: Data used and sources. Criteria for study inclusion: Inclusion and exclusion criteria used. Primary data collection: Specify whether primary data were collected. Secondary data analysis: Specify whether secondary data (e.g. clinical registers) were used. Literature review and integration of evidence: Sources of literature and other sources of data used. Method of synthesis: (non-quantitative, meta-analysis, modelling, economic evaluation)
- **Results:** Main results
- **Recommendations:** if given
- **Peer review process:** Specify: Yes / No / Internal / External / Both

5.2 Scientific Summary Report (and other summaries)

Although HTA reports are primarily addressed to local agents (decision makers, clinicians etc.), their findings may also be of interest for the international scientific/HTA community (one of the underlying assumptions of the ECHTA project). Those readers need to be able to assess the relevance and quality of previous HTA reports when they are considering previous HTA knowledge in their assessment. Up to now, only the technical reports (“full” HTA report) contain (and not always) all the information needed to assess their quality and relevance.

Usually those technical reports are written in the official tongue(s) of the commissioning/writing agency. For Europe, (but also for other parts of the world) this means that a large amount of HTA knowledge is currently being produced in languages other than English, making them difficult to access for the European and international audience (which often restricts itself to English and the national language).

Aside from the abstract, the Executive Summary may be, if at all, the only part of a report written in a language (usually English) other than the official tongue(s) of an agency, representing the only information easily accessible for the scientific

community and the “rest of the world”. However, not all HTA doers and agencies provide English summaries of all their publications.

Besides language, another difficulty of validly assessing relevance and results arises from the fact that an (good) Executive Summary is (should be) actually addressed to local decision makers (“executives”), stressing a summary of conclusions and recommendations, as these are the kinds of information sought by local decision makers. Methodological aspects of the assessment are usually underrepresented in the Executive Summary, as they are not of much interest to the target audience.

Only a comprehensive and structured summary available in English could warrant that all information needed to assess the relevance of a report for can be found. This could be termed “*Scientific Summary Report*”, in order to distinguish this kind of summary from the well known “Executive Summary”, as they actually differ in their purpose and content (Box 31).

Box 31. Differences between “Executive Summary” and “Scientific Summary Report”	
Executive Summary	Scientific Summary Report
<ul style="list-style-type: none"> • Addressed to local decision makers (“executives”) • Focuses on recommendations and conclusions • Written in agencies’/institutions’ official tongue(s) • Quickly informs decisions 	<ul style="list-style-type: none"> • Addressed to the HTA and Scientific Community • Stresses the context of the HTA and methodological aspects, in addition to conclusions and recommendations • Available in English • Allows for critical appraisal of relevance, quality, and main findings

The Scientific Summary Report is a comprehensive summary of a HTA technical report, available in English and structured around five main questions (Who?, Why?, What?, How? and What are the findings?) in order to allow for a quick assessment of the report’s relevance, quality and main findings to determine its further consideration. The target audience of such a Scientific Summary Report is mainly other researchers undertaking HTA/other HTA doers.

All questions listed in Box 32 should be addressed in the Scientific Summary Report (though not necessarily in this order). The length should be enough to warrant that all items are covered sufficiently and adequately.

Box 32. Elements to be addressed in the Scientific Summary Report

Question	Aspects
<ul style="list-style-type: none"> Who? 	<ul style="list-style-type: none"> Who initiated the HTA? Who commissioned it? – statement on conflict of interest Who conducted it? – statement on conflict of interest Who paid for it? – statement on conflict of interest To whom is it addressed? Who will receive it?
<ul style="list-style-type: none"> Why? 	<ul style="list-style-type: none"> Why was the HTA commissioned/conducted? Why right now? What decision(s) is it going to inform?
<ul style="list-style-type: none"> What? 	<ul style="list-style-type: none"> What technology or which aspects of a technology are going to be assessed? Which aspects are relevant to the outcomes? For what target group? For what target condition? What outcomes were considered and why? What are the questions to be answered in the assessment?
<ul style="list-style-type: none"> How? 	<ul style="list-style-type: none"> Was a HTA protocol followed? How was the assessment approached? Which aspects were assessed? Sources and synthesis of background information? Was safety assessed? <ul style="list-style-type: none"> How was the evidence/data identified? Which were the sources? How were data sources/studies selected (inclusion/exclusion criteria)? How was quality of data/studies appraised? What data were extracted and why? How were the results synthesised? How was the efficacy/effectiveness assessed? <ul style="list-style-type: none"> How was the evidence/data identified? Which were the sources? How were data sources/studies selected (inclusion/exclusion criteria)? How was quality of data/studies appraised? What data were extracted and why? Was a qualitative review conducted? How was it conducted? Was a meta-analysis conducted? <ul style="list-style-type: none"> What comparisons were made? What effect measures were used? What pooling method was used? How was heterogeneity accounted for? Was publication bias assessed and taken into account in the analysis? Was a sensitivity analyses done? Were psychological/social/ethical considerations assessed? <ul style="list-style-type: none"> How was the evidence/data identified? Which were the sources? How were data sources/studies selected (inclusion/exclusion criteria)? How was quality of data/studies appraised? What data were extracted and why? How were the results synthesised? Were organisational/professional implications assessed? <ul style="list-style-type: none"> How was the evidence/data identified? Which were the sources? How were data sources/studies selected (inclusion/exclusion criteria)? How was quality of data/studies appraised? What data were extracted and why? How were the results synthesised? Was an economic evaluation conducted? <ul style="list-style-type: none"> What were the alternatives which were compared? What perspective was assumed? What were the underlying assumptions? What kind of analyses was made and why?
	<ul style="list-style-type: none"> Did the HTA undergo an external review process before publication?

Results	<ul style="list-style-type: none"> • What are the main findings of the research?
Conclusions/ Discussion	<ul style="list-style-type: none"> • Relate results to questions posed • For which aspects of the assessment are there information lacking/uncertain? • Discuss transferability issues of results
Recommendations	<ul style="list-style-type: none"> • If recommendations are given and graded, what gradation scale was it used?
Update	<ul style="list-style-type: none"> • Is an update of the report planned? • What criteria will be used to decide on it?
General aspects	<ul style="list-style-type: none"> • Keywords • Bibliographic info

The Scientific Summary Report could improve the dissemination and use of HTA findings among the HTA community, preventing duplication of work when assessing a technology.

As already mentioned, other summaries addressed to other groups (e.g. executives, patients) may be elaborated. For such summaries, no recommendation nor standards are given here. The way in which such summaries are elaborated should be left up to the commissioning institutions, as they better know their needs.

5.3 Technical Report

The technical report should include comprehensive information on all issues covered under Chapter 4. The questions listed in Box 30 also apply to the technical report; however, as there are no space limitations, information should be more comprehensive.

The technical report can be seen as the deliverable product of the assessment. The steps undertaken, tools used (e.g. protocols), and evidence included and excluded should be documented in this comprehensive report. There are different elements which can be included in the technical report in order to enhance transparency and comprehensiveness in an understandable way (Box 31).

The description of the methods followed cannot limit itself to the methodology of a systematic review of the literature on efficacy/effectiveness. Instead, it refers much more to the methodology used to conduct and write the whole HTA report, referring to methods used to approach the (HTA protocol) and methods used to assess each of the aspects. Generally, the methodology part should be as detailed as to allow other researchers/ doers to replicate exactly what has been done. If a HTA protocol was used, this, along with the extent to which it was followed, should be documented. The HTA protocol can also be included as a part of the appendices.

The same is true for the documentation of the sources. All sources (e.g. medical literature, databanks, experts opinions) used to obtain information on the different aspects should be documented in a structured way.

Background information can be accompanied by a glossary, which helps non-specialists understand the terms being used. Such a glossary is strongly recommended when the issues under study are highly specialised.

The results for each aspect should be presented in a structured way, using evidence tables. Sometimes, graphical presentation (e.g. forest-plot by meta-analysis) can be very helpful for understanding the results of a synthesis.

Another important issue which should be included in the technical report is a clear statement on possible conflicts of interest. Who performed the report, who commissioned it, and who financed it should be clearly stated. A description of relations and possible conflicts of interests of the HTA doers, commissioners and financiers of the assessment have to be transparently documented in the full HTA report (Box 33).

Box 33. Statement on Conflict of Interest
<ul style="list-style-type: none">• Who performed the report?• Who financed it?• Who commissioned it?• Are there any conflicts of interest for the performers, commissioners or payers?

The declaration of conflict(s) of interest makes the reader aware of the possibility of judgements which are influenced by the motives of the persons involved. Although some of these aspects (e.g. who commissioned the report) might also be addressed under the policy question, a separate statement on conflict of interest is strongly recommended. The importance for doing this should not be underestimated, as possible distrust and/or perceived bias is an important barrier for the credibility of studies (Hoffmann & von der Schulenburg 2000).

The way of organising the technical report depends on the assessment and, as a result, no standard is recommended. However, a general structure is given as an example which may be altered depending on the needs of the HTA doers – or the specifications of the commissioners – for each assessment (Box 34).

Box 34. Structure example for a HTA technical report

(in brackets the section of this report where further explanation is given)

- Title
- Authors
- Statement on Conflict of Interest
- Policy Question (*Section 4.1*)
 - Who commissioned the assessment?, Why?, What decision(s) is it supporting?
- Methodology of the HTA report
 - HTA-Protocol (*Section 4.2*)
 - Review process (*Section 4.9.1*)
 - Sources of data* (*Section 4.5.1*)
 - Appraisal of data/studies (inclusion/exclusion criteria)* (*Section 4.5.2*)
 - Method of synthesis* (*Section 4.5.3*)
- Background Information (*Section 4.3*)
 - Target Condition, Target Group, Outcomes of Interest, Technology aspects
- Research questions (*Section 4.4*)
- Results**
 - Safety (*Section 4.6.1*)
 - Efficacy/effectiveness (*Section 4.6.2*)
 - Psychological/social/ethical considerations (*Section 4.6.3*)
 - Organisational/professional implications (*Section 4.6.4*)
 - Economic issues (*Section 4.6.5*)
- Discussion (*Section 4.7*)
 - Methodology of the assessment
 - Quality of evidence / Types of evidence (studies/data)*
 - Uncertainties / lack of information*
 - Generalisability, applicability of findings*
- Conclusions (*Section 4.8*)
- Recommendations (*Section 4.8*)
- Appendices***
 - Documentation of sources (search protocols, key words used, etc.)
 - Selection process documentation
 - Tables of evidence for included studies (including study characteristics, quality, and results)
 - Excluded studies with reasons for exclusion
 - Reference lists (included, excluded, other references used)
 - Tables of evidence from other sources of data included (e. g routine registers)
 - Appraisal tools used
 - Levels of evidence / grading of recommendations used
 - Glossary
 - Update Plan

*For each of the aspects of the assessment.

**Results can be presented with the help of tables and graphics.

***Information contained in Appendices can also be included in the body of the report. This is up to HTA doers, who should choose the most comprehensible way to report their work.

Chapter 6. Conclusions

The members of the Working-Group 4 of ECHTA have come to the conclusion that an improvement in the methodology currently employed by European HTA agencies and other institutions is best served by providing this report on current “best practice” as well as an instrument for assessing the quality of reports, rather than prescribing a methodology in a rigid way. Particular emphasis should be given to the reporting of findings in order to enhance comparability and allow for a better cross-border dissemination of results.

During its work, the working group identified a number of methodological gaps and needs:

- There is a great deal of work on isolated methodological aspects relevant to HTA, but there is little done on how to apply the individual methodological tool kits when conducting HTA. Only a few of the identified documents provided methodological guidelines for carrying out HTA; most of the reports focused on specific issues.
- Transparency of the whole HTA process has to be achieved, which is warranted by clear reporting and explanations of all steps undertaken in the assessment. Up until now, transparency has been concentrated on the evaluation of efficacy/effectiveness or in economic evaluations, while the handling of other important aspects of HTA has not been in a very systematic way.
- Other aspects of HTA are not being treated in a structured way at present. These range from the elaboration of the background information and formulation of research questions, to the assessment of important aspects such as psychological, social or ethical implications. A systematic approach might not be possible (or needed at all) for all aspects, but a structured and transparent approach should be warranted.
- Further research needs to be conducted in order to shed light on how underrepresented aspects can be better approached and included in HTA. Some aspects of HTA can be assessed with the help of qualitative research. However, no clear standards exist on how to include this in HTA. Further work should be done in this field.
- The systematic review on efficacy of therapeutic interventions has been accepted as the core of HTA. Methodological guidance concentrates mostly on such

aspects, distracting from a balanced approach to all aspects. However, with expanding work of the Cochrane Collaboration and similar groups, it can be expected that the HTA doers will not need to carry out systematic reviews on efficacy by themselves all the time, as they will be able to use this work.

- There is currently no methodology to project or even calculate the community effectiveness of a technology even if the evidence on efficacy is of the highest level. This is an urgent need, as the main function of HTA is to provide sound evidence on effectiveness taking system-, provider- and patient-side issues into account. The identified gap might possibly dealt with through a methodological advancement of modelling techniques.
- Some work is being done to develop systematic reviews of diagnostic, preventive community-based and health system-related interventions; however, the methodological debate is still open.
- Important issues of an assessment, such as the review process or update process are being conducted in different ways, but there is a need for further evaluation of different alternatives in order to find out what could be “best practice”.
- No appraisal tool to assess the quality of HTA reports exists. The working group is therefore proposing such an instrument in Box 35.

Box 35. Proposal for a Checklist/ Criteria for the assessment of the quality of HTA reports

Criterion		Questions
A	<i>Basic information</i>	<ul style="list-style-type: none"> • Are the authors of the report stated? • Is/Are any possible conflict(s) of interest stated? • Is there any information about who financed the report? • Was the report externally reviewed?
B	<i>General methodological aspects of the assessment</i>	<ul style="list-style-type: none"> • Was there a stated HTA report protocol? Was it followed, if not why not? • Is the scope of the assessment specified? Is there an explanation given for aspects not being assessed? • Are there clear research questions posed? • Are sources of information used for each aspect stated? Is it described how was the information for the different aspects gathered? • Are selection criteria for the different kinds of information used stated? • Are validity/quality criteria for appraisal of information clearly stated for each aspect? • Were evidence tables used?
C	<i>Description of the context of the assessment</i>	<ul style="list-style-type: none"> • Is the reason why the HTA was conducted stated? • Is the timing of the HTA explained (e.g. inappropriate extension of indication)? • Is what decision(s) the HTA is intended to support stated? • Is there any information given of who has commissioned the HTA?
D	<i>Background information</i>	<ul style="list-style-type: none"> • Were conditions, target group, relevant interventions or comparisons between interventions and relevant outcomes appropriately defined?

E	<i>Data about the status quo of the technology</i>	<ul style="list-style-type: none"> • Are patterns of utilisation, diffusion, indications, time trends adequately described? • Is an analysis of the regulatory status of the technology provided (e.g. market admission, status in other countries)?
F	<i>Technical description of the technology</i>	<ul style="list-style-type: none"> • Is there any consideration of when and how technical characteristics affect the outcomes? • Description of additional influencing factors (e.g. qualification requirements of staff, quality assurance, risks)?
G	<i>Safety</i>	<ul style="list-style-type: none"> • Are sources of data stated? • Are selection criteria for material stated? • Is there a transparent assessment of validity/quality of data? • Are the results transparently presented?
H	<i>Efficacy / effectiveness</i>	<ul style="list-style-type: none"> • Is the literature search done in a systematic way and documented accordingly (including search strategies, data sources and years)? • Are inclusion / exclusion criteria for primary studies defined? • Are included studies checked for quality and validity? • Is there a description of data extraction of included studies? • Is there a listing of excluded studies with reasons for exclusion given? • Are the results properly documented (e.g. tables, graphs, meta-analysis plots)? • Do the conclusions match the results?
I	<i>Psychological, social, and ethical considerations</i>	<ul style="list-style-type: none"> • Are psychological/social/ethical implications of the technology under consideration adequately discussed? • Are sources of data stated? • Are selection criteria for material stated? • Is there a transparent assessment of validity/quality of data? • Are the results transparently presented? • Are assumptions made, clearly stated?
J	<i>Organisational and professional implications</i>	<ul style="list-style-type: none"> • Were organisational and regulatory issues discussed (e.g. responsibility, necessary investments, financing, regulation, personnel, need, demand)? • Are the methods used for assessing these aspects stated?
K	<i>Economic evaluation</i>	<ul style="list-style-type: none"> • Is there a proper documentation of the methods used (see above)? • Is the perspective of the economic evaluation clarified (e.g. social insurance, societal)? • Are assumptions (e.g. for discounting rates, sensitivity analysis) justified? • Are issues of transferability (e.g. prices, cost structures, remuneration) across countries or settings adequately discussed?
L	<i>Discussion of generalisability / applicability of the findings</i>	<ul style="list-style-type: none"> • Are aspects of the generalisability of the results discussed (e.g. for populations not included in clinical trials or in different settings)? • Are aspects of the transferability of the results to different settings discussed (with regard to epidemiology, diffusion, structure of health care delivery, reimbursement, access)?

Chapter 7. Recommendations

- While some of the methodological gaps identified in chapter 6 are relatively minor and could be solved through research efforts by individual HTA agencies or other institutions, others are of such magnitude or require consensus to be meaningfully filled (e.g. the issue of community effectiveness), that they should be addressed at a European level.
- To overcome two of the main barriers in European collaboration in HTA (i.e. the non-availability of structured reports and the language barrier), the use of a Scientific Summary Report, as described in this paper, should be seen as a sign of “Best Practice in Reporting HTA”; for all assessments conducted within Europe, such a Scientific Summary Report should be available in English.
- A European HTA Database could be built using the Scientific Summary Reports of European HTA reports to facilitate accessibility to the HTA findings to the European scientific community. In order to promote the use of such a summary, its use could be a requisite for reports to be included in the proposed database.

References

1. Baladi J-F. A guidance document for the costing process. Canadian Coordinating Office of Health Technology Assessment (CCOHTA), 1996.
2. Battista RN, Fletcher WS. Making recommendations on preventive services: methodological issues. In: Battista RN, Lawrance RS (eds.) Implementing preventive services. New York: Oxford University Press, 1988, p. 53-67.
3. Brugha R, Varvasovszky Z. Stakeholder analysis: a review. *Health Policy and Planning* 2000;15:239-246.
4. BSA Medical Sociology Group. Criteria for the evaluation of qualitative research papers. *Medical Sociology News* 1996;22
5. Burls A, Cummins C, Fry-Smith A, Gold L et al. West Midlands Development and Evaluation Service Handbook. West Midlands Development and Evaluation Service (DES), 2000.
6. Busse R. Anwendung und Weiterentwicklung von Konzepten und Methoden der Gesundheitssystemforschung im deutschen Kontext. Habilitationsschrift, Medizinische Hochschule Hannover, 1998.
7. Canadian Coordinating Office for Health Technology Assessment. Guidelines for economic evaluation of pharmaceuticals: Canada. Canadian Coordinating Office of Health Technology Assessment (CCOHTA), 1997.
8. Clarke M, Oxman AD (eds.). *Cochrane Reviewers Handbook* 4.1.1 [updated Dec 2000]. The Cochrane Library, Issue 1, 2001.
9. Cochrane AL. *Effectiveness and Efficiency*. London: The Nuffield Provincial Hospitals Trust, 1971.
10. Coyle D, Drummond MF. Analyzing differences in the costs of treatment across centers within economic evaluation. *Int J of Technol Assess Health Care* 2001;17:155-163.
11. DEC. DEC Guidelines. Wessex Institute for Health Research and Development, Development and Evaluation Committee, undated [2000].
12. Drummond MF, Bloom BS, Carrin G et al. Issues in the cross-national assessment of health technology. *Int J Technol Assess Health Care* 1992;8:671-682.
13. Drummond MF, O'Brien B, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes*, Oxford-New York: Oxford University Press, 1997a.
14. Drummond MF, Richardson WS, O'Brien BJ, Levine M, Heyland D, for the Evidence-Based Medicine Working Group. User's guides to the medical literature. XIII: How to use an article on economic analysis of clinical practice. A: Are the results of the study valid? *JAMA* 1997b;277:1552-1557.
15. Durocher A, Pazart L, Dosquet P, Moquet MJ et al. Guide d'analyse de la littérature et gradation des recommandations. Agence Nationale d'Accréditation et d'Évaluation en Santé (ANAES), 2000.
16. Egger M, Smith GD. Meta-analysis bias in location and selection of studies. *BMJ* 1998;316:61-66.
17. Egger M, Smith GD, Altman DG. *Systematic reviews in health care. Meta-analysis in context*, London: BMJ Publishing Group, 2001.
18. EUR-ASSESS (1997). Report from the EUR-ASSESS Project. *Int J Technol Assess Health Care* 13(2).

19. European Commission, Directorate-General for Employment, Industrial Relations and Social Affairs (ed.). "Best practice": State of the art and perspectives in the EU for improving the effectiveness and efficiency of European health systems. Written by E. Jakubowski, M. Perleth and R. Busse. Luxembourg: Office for Official Publications of the European Communities, 1999.
20. Flynn KL, Adams EJ. Assessing Diagnostic Technologies. Boston: U.S. Department of Veterans Affairs (VATAP) Technology Assessment Program Report No. 1, 1996.
21. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Medical decision making* 1991;11:88-94.
22. Giacomini MK, Cook DJ for the Evidence-Based Medicine Working Group. User's guides to the medical literature. XXIII. Qualitative research in health care. A: Are the results of the study valid? *JAMA* 2000a;284:357-362.
23. Giacomini MK, Cook DJ for the Evidence-Based Medicine Working Group. User's guides to the medical literature. XXIII. Qualitative research in health care. B: What are the results and how do they help me care for my patients? *JAMA* 2000b;284:478-482.
24. Gold MR, Siegel JE, Russell LB, Weinstein MC et al., editors. Cost-effectiveness in health and medicine. Oxford-New York: Oxford University Press, 1996.
25. Gotzsche PC, Liberati A, Torri V, Rossetti L. Beware of surrogate endpoints. *Int J Technol Assess Health Care*. 1996;12:238-246.
26. Greenlough T. How to read a paper: Papers that summarise other papers (systematic reviews and meta-analyses). *BMJ* 1997;315:672-675.
27. Guyatt G, Drummond M, Feeny D, Tugwell P, Stoddart G, Haynes RB, Bennett K, Labelle R. Guidelines for the clinical and economic evaluation of health care technologies. *Soc Sci Med* 1986;22:393-408.
28. Harris RP, Helfand M, Woolf SH et al. Current Methods of the U.S. Preventive Services Task Force: A Review of the Process. *Am J Prev Med* 2001;20(3S):21-35.
29. Hoffmann C, von der Schulenburg J-M Graf on behalf of the EUROMET group. The influence of economic evaluation studies on decision making. A European survey. *Health Policy* 2000;52(3):179-192.
30. Imaz-Iglesia I, Gonzalez-Enriquez J, Alcaide-Jimenez JF, Conde-Olasagasti JL. Guía para la elaboración de informes de evaluación de tecnologías sanitarias. Agencia de Evaluación de Tecnologías Sanitarias (AETS) Informe de Evaluación de Tecnologías Sanitarias No. 19, 1999.
31. Ioannidis JPA, Lau J. Completeness of safety reporting in randomized trials. An evaluation of 7 medical areas. *JAMA* 2001;285:437-443.
32. Irwig L, Tosteson AN, Gatsonis C, Lau J, Colditz G, Chalmers TC, Mosteller F. Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med* 1994;120:667-676.
33. Irwig L, Macaskill P, Glasziou P, Fahey M. Meta-analytic methods for diagnostic test accuracy. *J Clin Epidemiol* 1995;48:119-130.
34. Jadad AR, Moore RA, Carrol D et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996;17:1-12.
35. Jüni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999;282:1054-1060.
36. Khan KS, Ter-Riet G, Glanville J, Sowden AJ, Kleijnen J (eds.). Undertaking Systematic Reviews of Research on Effectiveness. CRD's Guidance for Carrying

- out or Commissioning Reviews. NHS Center for Reviews and Dissemination (CRD), Report No 4, 2000.
37. Kristensen FB, Hørder M, Poulsen PB (eds.). Health Technology Assessment Handbook. Danish Institute for Health Technology Assessment (DIHTA), 2001.
 38. MacMahon S, Collins R. Reliable assessment of the effects of treatment on mortality and major morbidity, II: observational studies. *Lancet* 2001;357:455-462.
 39. Mays N, Pope C. Qualitative research in health care. London: BMJ Publishing Group; 1996.
 40. Mays N, Pope C. Assessing quality in qualitative research. *BMJ* 2000;320:50-2.
 41. Mears R, Taylor R, Littlejohns P, Dillon A. Review of International Health Technology Assessment (IHTA): Project Report. National Institute for Clinical Excellence, 2000.
 42. Moher D, Cook DJ, Jadad AR, Tugwell P, Moher M, Jones A, Pham B, Klassen TP. Assessing the quality of reports of randomised trials: Implications for the conduct of meta-analyses. *Health Technology Assessment* 1999a;3:1-90.
 43. Moher D, Cook DJ, Eastwood S, Olkin I, Drummond R, Stroup DF, for the QUOROM Group. Improving the quality of meta-analyses of randomised controlled trials: the QUOROM statement. *Lancet* 1999b;354:1896-900.
 44. Morgan D (ed.). Successful focus groups. Advancing the state of the art. Newsbury Park: Sage, 1993.
 45. MSAC. Funding for new medical technologies and procedures: application and assessment guidelines. Medicare Services Advisory Committee (MSAC), 2000.
 46. Murphy E, Dingwall R, Greatbatch D, Parker S, Watson P. Qualitative research methods in health technology assessment: a review of the literature. *Health Technol Assess* 1998;2 (16).
 47. O'Brien BJ, Heyland D, Richardson WS, Levine M, Drummond MF, for the Evidence-Based Medicine Working Group. User's guides to the medical literature. XIII. How to use an article on economic analysis of clinical practice. B: What are the results and will they help me in caring for my patients? *JAMA* 1997;277:1802-1806.
 48. Oxman AD, Sackett DL, Guyatt GH, for the Evidence-Based Medicine Working Group. User's guides to the medical literature. VI: How to use an overview. *JAMA* 1994;272:1367-1371.
 49. Popay J, Rogers A, Williams G. Rationale and standards in the systematic review of qualitative literature in health services research. *Qualitative Health Research* 1998;8:341-351.
 50. Ramsay CR, Grant AM, Wallace SA, Garthwaite PH, Monk AF, Russell IT. Statistical assessment of the learning curves of health technologies. *Health Technol Assess* 2001;5(12).
 51. Rettig RA. Health care in transition: technology assessment in the private sector. Santa Monica: RAND, 1997.
 52. Rittenhouse BE. Exorcising protocol-induced spirits: Making the clinical trial relevant for economics. *Med Decis Making* 1997;17:331-339.
 53. Ryan M, Scott DA, Reeves C, Bate A et al. Eliciting public preferences for health care: a systematic review of techniques. *Health Technol Assess* 2001;5(5).
 54. Sackett DL. Evaluation of health services. In: Last J (ed.) *Maxcy-Rosenau Public Health and Preventive Medicine*. Norwalk, CT: Appleton-Century-Crofts, 1980, p. 1800-1823

55. Sloan FA (Ed.). Valuing Health Care, Costs, Benefits, and Effectiveness of Pharmaceuticals and Other Medical Technologies. Cambridge-New York-Melbourne: Cambridge University Press, 1995.
56. Smith DH, Gravelle H. The practice of discounting in economic evaluations of healthcare interventions. *Int J of Technol Assess Health Care* 2001;17:236-243.
57. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. Systematic reviews of trials and other studies. *Health Technol Assess* 1998;2(19).
58. Tugwell P, Bennett K, Feeny D, Guyatt G, Haynes RB. A framework for the evaluation of technology: the technology assessment iterative loop. In: Feeny D, Guyatt G, Tugwell P (eds.) *Health Care Technology: Effectiveness, Efficiency and Public Policy*. The Canadian Medical Association and The Institute for Research on Public Policy, 1984, p. 41-56.
59. U.S. Congress, Office of Technology Assessment. Assessing efficacy and safety of medical technologies. OTA-H-75. Washington, DC: U.S. Government Printing Office, 1978.
60. von der Schulenburg J-M Graf, Hoffmann C. Review of European guidelines for economic evaluation of medical technologies and pharmaceuticals. *Health Economics in Prevention and Care* 2000;1(1):2-8.
61. Williamson JW. The estimation of achievable health care benefit. In: Williamson JW (ed.) *Assessing and Improving Health Outcomes*. Cambridge, MA: Ballinger Publishers, 1978, p. 51-69

Appendices

- A1. Toolkits and methodological guidance documents
 - Table A1-1. Toolkits on HTA
 - Table A1-2. Methodological toolkits on specific topics
- A2. Sources of information
 - Table A2-1. Sources of HTA reports and systematic reviews
 - Table A2-2. Bibliographic sources
 - Table A2-3. Other sources of data/information
- A3. Search filters
- A4. Appraisal checklists
 - Box A4-1. Checklist for an article about therapy
 - Box A4-2. Checklist for an article about diagnostic tests
 - Box A4-3. Checklist for an article about harm
 - Box A4-4. Checklist for an article about prognosis
 - Box A4-5. Checklist for a review article
 - Box A4-6. Checklist for a clinical decision analysis
 - Box A4-7. Checklist for clinical practice guidelines
 - Box A4-8. Checklist based on the “Appraisal of Guidelines for Research & Evaluation (AGREE) Instrument”
 - Box A4-8. Checklist for a clinical utilisation review
 - Box A4-9. Checklist for an article reporting variations in the outcomes of health services research
 - Box A4-10. Checklist for an article about health-related quality of life measurements
 - Box A4-11. Checklist for qualitative research in health care
 - Box A4-12. Checklist for an economic analysis article
 - Box A4-13. Complete list of the Users’ Guide
- A5. Software for data synthesis
- A6. Levels of evidence and grades of recommendations

NB: All websites cited in appendices 1, 2, 3 and 5 were available as of late April 2001 while the ones in appendices 4 and 6 were available as of mid July 2001.

A1. Toolkits and methodological guidance documents

Table A1-1. Available toolkits for HTA which refer to the whole assessment process.

Reference	Source	Language	Comments
Burls A, Cummins C, Fry-Smith A, Gold L et al. West Midlands Development and Evaluation Service Handbook . West Midlands Development and Evaluation Service (DES), 2000.	http://www.bham.ac.uk/WMidsDES/	English	Description and methodological guidance of all steps undertaken when performing an assessment for the DES. Provides comprehensive guidance on how to elaborate background information and research questions, on how to report appraisal and selection of the data and on how to summarise the evidence found in a non-quantitatively way.
DEC. DEC Guidelines . Wessex Institute for Health Research and Development, Development and Evaluation Committee, undated [2000].		English	Description of the process of assessment for the DEC ("rapid HTA"), with special focus on the costs aspects.
Imaz-Iglesia I, Gonzalez-Enriquez J, Alcaide-Jimenez JF, Conde-Olasagasti JL. Guía para la elaboración de informes de evaluación de tecnologías sanitarias . Agencia de Evaluacion de Tecnologias Sanitarias (AETS) Informe de Evaluacion de Tecnologias Sanitarias No. 19, 1999.	http://www.isciii.es/unidad/aet/caet.html	Spanish	Description of the process of HTA and elaboration of HTA reports, including an overview of methods of synthesis of evidence and a comprehensive list of sources of data.
Kristensen FB, Hørder M, Poulsen PB (eds.). Health Technology Assessment Handbook . Danish Institute for Health Technology Assessment (DIHTA), 2001.	http://147.29.115.214/publikationer/docs/Metodehaandbog/MethodologyHandbook180601.pdf or via http://www.dihta.dk	English	Provides an overview of qualitative research methods, measurement of quality of life, methods to address the organisational aspects and economic evaluation methods which can be applied in HTA.

MSAC. Funding for new medical technologies and procedures: application and assessment guidelines. Medicare Services Advisory Committee (MSAC), 2000.	http://www.health.gov.au/haf/msac	English	Description of the assessment process and elaboration of HTA reports.
---	---	---------	---

Table A1-2. Methodological toolkits on specific topics. The documents listed here refer only to some aspects of HTA.

Reference	Source	Language	Comments
Baladi J-F. A guidance document for the costing process. Canadian Coordinating Office of Health Technology Assessment (CCOHTA), 1996.	http://www.ccohta.ca/newweb/pubapp/pdf/costing_e.pdf	English (French)	Deals with the identification of resources, the measurement of resources, cost valuation, possible bias in estimating costs, and proposes a reporting format for these issues.
Canadian Coordinating Office for Health Technology Assessment. Guidelines for economic evaluation of pharmaceuticals: Canada. Canadian Coordinating Office of Health Technology Assessment (CCOHTA), 1997.	http://www.ccohta.ca/newweb/pubapp/pdf/peg_e.pdf	English (French)	Focuses on the economic evaluation of drugs, giving also guidelines for reporting economic analyses.
Clarke M, Oxman AD (eds.). Cochrane Reviewers Handbook 4.1.1 [updated Dec 2000]. The Cochrane Library, Issue 1/ 2001.	http://www.cochrane.org	English	Comprehensive methodological guidance on how to conduct systematic reviews and meta-analyses of RCTs of therapeutic interventions.
Durocher A, Pazart L, Dosquet P, Moquet MJ et al. Guide d'analyse de la littérature et gradation des recommandations. Agence Nationale d'Accréditation et d'Évaluation en Santé (ANAES), 2000.	http://www.anaes.fr/ANAES/anaesparametrage.nsf/HomePage?ReadForm	French	Focuses on literature search and appraisal, including a set of checklists and literature appraisal criteria for different types of medical literature.
Egger M, Smith GD, Altman DG. Systematic reviews in health care. Meta-analysis in context. London: BMJ, 2001.		English	Comprehensive and updated review of methods for meta-analyses of binary and continuous results.

Flynn KL, Adams EJ. Assessing Diagnostic Technologies. U.S. Department of Veterans Affairs (VATAP) Technology Assessment Program Report No. 1, 1996.	http://www.va.gov/resdev/ps/pshsrd/mdrc.htm#HealthCareTechnologyAssessment	English	Provides methodological guidance on how to conduct systematic reviews on accuracy of diagnostic tests.
Harris RP, Helfand M, Woolf SH et al. Current Methods of the U.S. Preventive Services Task Force: A Review of the Process. Am J Prev Med 2001;20(3S):21-35.	Via http://www.ahcpr.gov/clinic/ajpm.htm	English	Detailed description of process and methods applied by the Third U.S. Preventive Services Task Force for assessing preventive technologies including useful analytic frameworks, its principles for making recommendations etc.
Khan KS, Ter-Riet G, Glanville J, Sowden AJ, Kleijnen J (eds.). Undertaking Systematic Reviews of Research on Effectiveness. CRD's Guidance for Carrying out or Commissioning Reviews. NHS Center for Reviews and Dissemination (CRD). Report No 4, 2000.*	http://www.york.ac.uk/inst/crd	English	Comprehensive methodological guidance on conducting systematic reviews of literature referring to effectiveness of therapeutic interventions and to some extent of diagnostic interventions.
Billingham LJ, Abrams KR, Jones DR. Methods for the analysis of quality-of-life and survival data in health technology assessment. Health Technol Assess 1999;3 (10).**	http://www.hta.nhsweb.nhs.uk/fullmono/mon310.pdf	English	
Lewsey JD, Leyland AH, Murray GD, Boddy FA. Using routine data to complement and enhance the results of randomised controlled trials. Health Technol Assess 2000;4 (22).**	http://www.hta.nhsweb.nhs.uk/fullmono/mon422.pdf	English	
Murphy E, Dingwall R, Greatbatch D, Parker S, Watson P. Qualitative research methods in health technology assessment: a review	http://www.hta.nhsweb.nhs.uk/fullmono/mon216.pdf	English	Comprehensive review of qualitative research methods applicable in HTA.

of the literature. Health Technol Assess 1998;2 (16).**			
Ramsay CR, Grant AM, Wallace SA, Garthwaite PH, Monk AF, Russell IT. Statistical assessment of the learning curves of health technologies. Health Technol Assess 2001;5 (12).**	http://www.hta.nhsweb.nhs.uk/fullmono/mon512.pdf	English	
Ryan M, Scott DA, Reeves C, Bate A et al. Eliciting public preferences for health care: a systematic review of techniques. Health Technol Assess 2001;5 (5).**	http://www.hta.nhsweb.nhs.uk/fullmono/mon505.pdf	English	Review of methods to include the public preferences perspective on HTA.
Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. Systematic reviews of trials and other studies. Health Technol Assess 1998;2 (19).**	http://www.hta.nhsweb.nhs.uk/fullmono/mon219.pdf	English	Comprehensive methodological guidance on conducting systematic reviews of literature. Presents a comprehensive overview of different meta-analytic approaches.

*Other methodological documents on how to conduct systematic reviews are collected in the CRMD Cochrane Reviews Methodology Database available at <http://www.update-software.com/ccweb/cochrane/cdsr.htm>.

**Besides the documents listed here, the Health Technology Assessment Series of the NHS includes further methodological reviews on more specific topics concerning HTA. A complete list of them is available at <http://www.hta.nhsweb.nhs.uk/htapubs.htm>.

A2. Sources of information

In the following tables a selection of sources of information and literature is presented. The tables were elaborated with information obtained from the Handbooks of AETS, DES, DIHTA and own research. The sites listed below are only a selection of providers (free or for fee) of access to the mentioned databases. Many of the databases may be also available in CD-ROM or online, through databases providers (e.g. <http://www.silverplatter.com>, <http://www.ovid.com>, <http://www.dialog.com>, <http://www.fiz-karlsruhe.de/stn.html>) It is recommended to consult documentation specialist for further details on access and use of the different databases.

Table A2-1. Sources of HTA reports and systematic reviews.

Name of the Source	Available at	Comments
INAHTA Members	http://www.inahta.org	Provides access to HTA agencies members of INAHTA. Many HTA Agencies allow online-retrieving of their HTA reports.
HSTAT Health Services/Technology Assessment Text	http://text.nlm.nih.gov	Includes the technology assessments and evidence reports of the Agency for Health Care Policy and Research/ Agency for Healthcare Research and Quality.
HTA Database	http://agatha.york.ac.uk/hta hp.htm	Abstracts of publications and projects from INAHTA members and other organisations.
ISTAHC Database	http://www.istahc.org/en/database.html	Includes abstracts, journal citations, meeting programs, post conference courses and articles related to health technology assessment.
Cochrane Database of Systematic Reviews	http://www.update-software.com/ccweb/cochrane/cdsr.htm	Systematic reviews elaborated by members of the Cochrane Collaboration.
DARE Database of abstracts of reviews of effectiveness	http://agatha.york.ac.uk/darehp.htm	A collection of structured abstracts and bibliographic references of systematic reviews assembled by the NHS Centre for Reviews and Dissemination (NHS CRD).
TRIP Database	http://www.tripdatabase.com	Allows searching in evidence based medicine related databases, including guidelines.

HSRProj Health Services Research Projects in Progress	http://igm.nlm.nih.gov	Database of ongoing research and projects referring to health services research including health technology assessment and the development and use of clinical practice guidelines (will be replaced by NLM Gateway later in 2001).
---	---	---

Table A2-2. Bibliographic sources

Name of the Source	Available at	Comments
<i>General</i>		
MEDLINE	Usually available at university libraries or through Internet: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi	Covers the whole field of medical information, including dentistry and medical psychology. If using optimised search filters, systematic reviews can also be found.
NLM Gateway	http://gateway.nlm.nih.gov	Contains MEDLINE plus citations of monographs (LOCATORplus) and meeting abstracts, e.g. those of the ISTAHC meetings (previously available via HealthStar). The Gateway will, from late 2001, also include all unique journal citations which are currently available at AIDSLINE, BIOETHICSLINE and other databases not relevant to HTA.
HealthSTAR	All citations are available through NLM Gateway: http://gateway.nlm.nih.gov	Focused on the clinical (e.g evaluation of patient outcomes, effectiveness of procedures, programs, products, services, and processes) and the non-clinical (health care administration, economics, planning, and policy) aspects of health care delivery (specific database was dismantled early in 2001 as information is now available through the NLM Gateway).
EMBASE	http://www.embase.com	Covers the whole field of medical literature, including health policy, management and pharmacoeconomics.
UNCOVER Database	http://uncweb.carl.org	Provides access to multidisciplinary journals (English speaking).
Science Citation Index	http://www.isinet.com/isi/products/inde	Provides access to bibliographic information, author abstracts,

	x.html#sdb	and cited references found in technical and science journals.
<i>Specific</i>		
AIDSLINE	Currently accessible through GratefulMed: http://igm.nlm.nih.gov	Acquired immunodeficiency syndrome (AIDS) and related topics (to be replaced by NLM Gateway).
AIDSDRUGS/ AIDSTRIALS	http://www.actis.org/	Clinical trials of substances being tested for use against AIDS, HIV infection, and AIDS-related opportunistic diseases.
BIOETHICSLINE	Currently accessible through GratefulMed: http://igm.nlm.nih.gov	Ethics and related public policy issues in health care and biomedical research (to be replaced by NLM Gateway).
CANCERLIT	http://cancernet.nci.nih.gov	Literature related to cancer.
DIRLINE	http://dirline.nlm.nih.gov	Focuses primarily on health and biomedical information resources including organizations, government agencies, information centers, professional societies, voluntary associations, support groups, academic and research institutions, and research facilities and resources.
CINAHL Cumulative Index to Nursing and Allied Health Literature	http://www.CINAHL.com	Database of information concerning nursing, physiotherapy and related topics.
AMED Allied and Complementary Medicine Database	http://www.bl.uk/services/stb/amed.html	Covers topics related to complementary medicine physiotherapy occupational therapy, rehabilitation and palliative care.
PsycINFO Psychological Abstracts	http://www.apa.org/psycinfo	Literature on psychology, medicine, education and social science.
ASSIA (Applied Social Sciences Index and Abstracts)	http://www.bowker-saur.co.uk/products/catalog/a_and_i/assia_plus_c.htm	Includes abstracts and references from literature on social science applied to medicine and health care system.
Social Science Citation Index	http://www.isinet.com/isi/products/index.html#sdb	Provides access to bibliographic information, author abstracts, and cited references found in social science journals.
Sociological Abstracts	http://www.silverplatter.com/catalog/soci.htm	Covers sociological aspects of medicine and health among many others including interdisciplinary research in social

		sciences issues.
NHSEED NHS Economic Evaluation Database	http://agatha.york.ac.uk/nhsdhp.htm	Database of economic evaluations studies of health care interventions.
ECONLit	http://econlit.org	Database of general economic literature, including health economics and technological change.
ECONbase	http://www.elsevier.nl/homepage/sae/econbase/menu.sht	Database of general economic literature, including health economics topics.
HEED Health Economics Evaluation Database	http://www.ohe-heed.com	Contains information on studies of cost-effectiveness and other forms of economic evaluation of medicines and other treatments and medical interventions.
<i>Grey literature/Ongoing Research</i>		
SIGLE System for Information on Grey Literature	http://www.fiz-karlsruhe.de/stn/Databases/sigle.html	Covers many research fields including health, social science and economics. Limited to Europe.
Conference Papers Index	http://www.csa1.co.uk	Abstracts of conference papers. Multidisciplinary.
<i>Registries of trials and other ongoing research</i>		
CCTR Cochrane Register of Controlled trials	http://www.update-software.com/ccweb/cochrane/cdsr.htm	Includes RCTs and other controlled studies identified by contributors to the Cochrane Collaboration. It includes many sources not included in MEDLINE or other bibliographic databases.
Controlled Trials (USA)	http://clinicaltrials.gov	
Glaxo Wellcome register	http://ctr.glaxowellcome.co.uk	
Meta-register of controlled trials	http://www.controlled-trials.com	
UKCCCR registry of cancer trials	http://www.ctu.mrc.ac.uk/ukcccr/	
NTIS National Technical Information Service	http://www.ntis.gov	Contains information about ongoing research on different fields
NRR National Research Register	http://www.doh.gov.uk/research/nrr.htm	Set of databases containing information on ongoing research of interest for the NHS.

Table A2-3. Other sources of data/information*

Name of the Source	Available at	Comments
WHO World Health Organisation	http://www.who.org	Access to multiple health statistics.
FDA Food and Drug Administration	http://www.fda.gov	US Approval Agency for medical devices and drugs, contains information on safety for different medical technologies.
OECD	http://www.oecd.org	Access to the OECD Health Data Database, which can be useful for the elaboration of the background information.
CORDIS Community Research and Development Information Service	http://www.cordis.lu	Information about research and development activities within the EU.
EUROSTAT European Union Statistics Office	http://europa.eu.int/en/comm/eurostat/eurostat.ht	Statistical service of the EU.
WHO, Regional Office for Europe	http://www.who.dk/country/country.htm	Contains epidemiological information on European countries.

*The sources cited here aim at providing a general idea of sources other than the literature. Statistical agencies, ministries, epidemiological registers, manufacturers and professional, consumers and patient associations at the national, regional or local level are not listed here but are also useful sources of information, which the HTA doers can consider when under taking an assessment.

A3. Search filters

In this section a selection of websites is presented where validated search strategies are available.

Source	Available at	Search filters provided for		
		Database	Software	Topics
University of Rochester, USA	http://www.urmc.rochester.edu/Miner/Ed uc/Expertsearch.html	MEDLINE CINAHL	Ovid	diagnostic devices, aetiology, harm, prognosis/natural history, therapy, meta-analysis/systematic reviews and qualitative research
NHS CRD, UK	http://www.york.ac.uk/inst/crd/search.htm	MEDLINE CINAHL	Ovid Silverplatter	meta-analyses and systematic reviews
Oxford University, UK	http://wwwlib.jr2.ox.ac.uk/caspfew/filters	MEDLINE CINAHL EMBASE PsycInfo	Ovid Silverplatter	aetiology, diagnostic, prognosis and therapy
BMJ Publishing Group, UK	http://www.evidence.org/what-is-ce/search-strategy-appraisal.htm	MEDLINE	Ovid	systematic reviews, RCTs, cohort studies

A4. Appraisal checklists

In this section a selection of checklists for appraisal of the medical literature are presented. More checklists and appraisal tools have been developed by other authors and also by HTA institutions. This thus not a comprehensive collection but an example. Except for the one in box A4-8, all the checklists presented here have been originally published in the JAMA-series "Users' guide to the medical literature" (complete list in Box A4-14).

Internet source of http://www.cche.net/principles/content_all.asp
checklists:

Box A4-1. Checklist for an article about therapy (Guyatt et al. 1993, 1994)

I. Are the results of the study valid?

Primary Guides:

- Was the assignment of patients to treatments randomised?
- Were all patients who entered the trial properly accounted for and attributed at its conclusion?
- Was follow up complete?
- Were patients analysed in the groups to which they were randomised?

Secondary Guides:

- Were patients, health workers, and study personnel "blind" to treatment?
- Were the groups similar at the start of the trial?
- Aside from the experimental intervention, were the groups treated equally?

II. What were the results?

- How large was the treatment effect?
- How precise was the estimate of the treatment effect?

III. Will the results help in the clinical practice?

- Can the results be applied to my patient group?
- Were all clinically important outcomes considered?
- Are the likely treatment benefits worth the potential harms and costs?

Box A4-2. Checklist for an article about diagnostic tests (Jaeschke et al. 1994a, 1994b)

I. Are the results of the study valid?

Primary Guides:

- Was there an independent, blind comparison with a reference standard?
- Did the patient sample include an appropriate spectrum of patients to whom the diagnostic test will be applied in clinical practice?

Secondary Guides:

- Did the results of the test being evaluated influence the decision to perform the reference standard?
- Were the methods for performing the test described in sufficient detail to permit replication?

II. What were the results?

- Are likelihood ratios presented or data necessary for their calculation provided?

III. Will the results help in the clinical practice?

- Will the reproducibility of the test result and its interpretations be satisfactory in my setting?
- Are the results applicable to my patient group?
- Will the results change management of the patient all?
- Will patients be better off as a result of the test?

Box A4-3. Checklist for an article about harm (Levine et al. 1994)

I. Are the results of the study valid?

Primary Guides:

- Were there clearly identified comparison groups that were similar with respect to important determinants of outcome, other than the one of interest?
- Were the outcomes and exposures measured in the same way in the groups being compared?
- Was follow up sufficiently long and complete?

Secondary Guides:

- Is the temporal relationship correct?
- Is there a dose response gradient?

II. What are the results?

- How strong is the association between exposure and outcome?
- How precise is the estimate of the risk?

III. Will the results help in the clinical practice?

- Are the results applicable to my patient group?
- What is the magnitude of the risk?
- Should it be attempted to stop the exposure?

Box A4-4. Checklist for an article about prognosis (Laupacis et al. 1994)

I. Are the results of the study valid?

Primary Guides:

- Was there a representative and well-defined sample of patients at a similar point in the course of the disease?
- Was follow up sufficiently long and complete?
- Did the patient sample include an appropriate spectrum of patients to whom the diagnostic test will be applied in clinical practice?

Secondary Guides:

- Were objective and unbiased outcome criteria used?
- Was there adjustment for important prognostic factors?

II. What were the results?

- How large is the likelihood of the outcome event(s) in a specified period of time?
- How precise are the estimates of likelihood?

III. Will the results help in the clinical practice?

- Were the study patients similar to my patient group?
- Will the results lead directly to selecting or avoiding therapy?
- Are the results useful for reassuring or counselling patients?

Box A4-5. Checklist for a review article (Oxman et al. 1994)

I. Are the results of the study valid?

Primary Guides:

- Did the overview address a focused clinical question?
- Were the criteria used to select articles for inclusion appropriate?

Secondary Guides:

- Is it unlikely that important, relevant studies were missed?
- Was the validity of the included studies appraised?
- Were assessments of studies reproducible?
- Were the results similar from study to study?

II. What are the results?

- What are the overall results of the review?
- How precise are the results?

III. Will the results help in the clinical practice?

- Are the results applicable to my patient group?
- Were all clinically important outcomes considered?
- Are the benefits worth the harms and costs?

Box A4-6. Checklist for a clinical decision analysis (Richardson & Detsky 1995a, 1995b)

I. Are the results of the study valid?

Primary Guides:

- Were all important strategies and outcomes included?
- Were all of the realistic clinical strategies compared?
- Were all clinically relevant outcomes considered?
- Was an explicit and sensible process used to identify, select and combine the evidence into probabilities?
- Were the utilities obtained in an explicit and sensible way from credible sources?
- Was the potential impact of any uncertainty in the evidence determined?

Secondary Guides:

- Were objective and unbiased outcome criteria used?
- Was there adjustment for important prognostic factors?

II. What were the results?

- In the baseline analysis, does one strategy result in a clinically important gain for patients? If not, is the result a toss-up?
- How strong is the evidence used in the analysis?
- Could the uncertainty in the evidence change the result?

III. Will the results help in the clinical practice?

- Do the probability estimates fit my patients' clinical features?
- Do the utilities reflect how my patients would value the outcomes of the decision?

Box A4-7. Checklist for clinical practice guidelines (Hayward et al. 1995, Wilson et al. 1995)

I. Are the recommendations valid?

Primary Guides:

- Were all important options and outcomes included?
- Was an explicit and sensible process used to identify, select, and combine evidence?

Secondary Guides:

- Was an explicit and sensible process used to consider the relative value of different outcomes?
- Is the guideline likely to account for important recent developments?
- Has the guideline been subjected to peer review and testing?
- Were the results similar from study to study?

II. What are the recommendations?

- Are practical, clinically important, recommendations made?
- How strong are the recommendations?
- What is the impact of uncertainty associated with the evidence and values used in the guidelines?

III. Will the recommendations help in the clinical practice?

- Is the primary objective of the guideline consistent with your objectives?
- Are the recommendations applicable to your patients?

Box A4-8. Checklist based on the “Appraisal of Guidelines for Research & Evaluation (AGREE) Instrument” (June 2001; available at www.agreecollaboration.org)

1. Are the overall objectives of the guidelines specifically described?
2. Are the clinical questions covered by the guideline specifically described?
3. Are the patients to whom the guideline is meant to apply specifically described?
4. Does the guideline development group include individuals from all the relevant professional groups?
5. Have the patients' views and preferences been sought?
6. Are the target users of the guideline clearly defined?
7. Has the guideline been piloted among end users?
8. Were systematic methods used to search for the evidence?
9. Are the criteria for selecting the evidence clearly described?
10. Are the methods for formulating the recommendations clearly described?
11. Have the health benefits, side effects and risks been considered in formulating the recommendations?
12. Is there an explicit link between recommendations and the supporting evidence?
13. Has the guideline been externally reviewed by experts prior to its publication?
14. Is a procedure for updating the guideline provided?
15. Are the recommendations specific and unambiguous?
16. Are the different options for the management of the condition clearly presented?
17. Are key recommendations easily identifiable?
18. Is the guideline supported with tools for application (e.g. a summary document, a quick reference guide, educational tools, patients' leaflets, computer support)?
19. Have the potential organisational barriers in applying the recommendations been discussed?
20. Have the potential cost implications of applying the recommendations been considered?
21. Does the guideline present key review criteria for monitoring and/ or audit purposes?
22. Is the guideline editorially independent from the funding body?
23. Have conflicts of interest of guideline development members been recorded?

Box A4-9. Checklist for an article reporting variations in the outcomes of health services research (Naylor & Guyatt 1996a)

I. Are the recommendations valid?

- Are the outcome measures accurate and comprehensive?
- Were the comparison groups similar with respect to important determinants of outcome, other than the one of interest, and were residual differences adjusted for in the analysis?

II. What are the recommendations?

III. Will the recommendations help you in caring for your patients?

- How will the recommendations help you?

Box A4-10. Checklist for a clinical utilisation review (Naylor & Guyatt 1996b)

I. Are the criteria valid?

- Was an explicit and sensible process used to identify, select, and combine evidence for the criteria?
- What is the quality of the evidence used in framing the criteria?
- Was an explicit and sensible process used to consider the relative values of different outcomes?
- Are the judgements of the clinical experts who established the criteria reproducible?
- If the quality of the evidence used in originally framing the criteria was weak, have the criteria been prospectively evaluated in an implementation study and shown to improve patient outcome?

II. Were the criteria applied appropriately?

- Did the process of applying the criteria meet scientific standards?
- What is the impact of uncertainty associated with evidence and values on the criteria-based ratings of process of care?
- Could the uncertainty in the evidence change the result?

III. Can you use the criteria on your own setting?

- Have the criteria been field-tested for feasibility of use in diverse settings?
- Are the criteria up-to-date?

Box A4-11. Checklist for an article about health-related quality of life measurements (Guyatt et al. 1997)

I. Are the recommendations valid?

Primary Guides:

- Have the investigators measured aspects of patients' lives that patients consider important?
- Did the HRQL instruments work in the way they are supposed to?

Secondary Guides:

- Are there important aspects of HRQL that have been omitted?
- If there were tradeoffs between quality and quantity of life, or an economic evaluation, have they used the right measures?

II. What were the results?

- What was the magnitude of effect on HRQL?

III. Will the recommendations help in the clinical practice?

- Will the information from the study help me inform my patients?
- Did the study design simulate clinical practice?

Box A4-12. Checklist for qualitative research in health care (Giacomini & Cook 2000a, 2000b)

I. Are the results valid?

- Were participants relevant to the research question and was their selection well reasoned?
- Were the data collection methods appropriate for the research objectives and setting?
- Was the data collection comprehensive enough to support rich and robust descriptions of the observed events?
- Were the data appropriately analyzed and the findings adequately corroborated?

II. What were the results?

- How evocative and thorough is the description?
- How comprehensive and relevant are the theoretical conclusions?
- What major and minor concepts does the theory entail, and how well-defined are they?
- What are the relationships between the conceptual categories, are these dynamics clearly described, and do they make sense?
- Are the concepts adequately developed and illustrated?
- Where does the empirically-generated theory fit in relation to existing theory and beliefs in the field?

III. How do the results help in the clinical practice?

- Does this study help to understand the context of the clinical practice?
- Does this study help to understand the relationships with the patients and their families?

Box A4-13. Checklist for an economic analysis article (Drummond et al. 1997, O'Brien et al. 1997)

I. Are the results of the study valid?

- Did the analysis provide a full economic comparison of health care strategies?
- Were the costs and outcomes properly measured and valued?
- Was appropriate allowance made for uncertainties in the analysis?
- Are estimates of costs and outcomes related to the baseline risk in the treatment population?

II. What were the results?

- What were the incremental costs and outcomes of each strategy?
- Do incremental costs and outcomes differ between subgroups?
- How much does allowance for uncertainty change the results?

III. Will the results help in the clinical practice?

- Are the treatment benefits worth the harms and costs?
- Could my patients expect similar health outcomes?
- Could I expect similar costs?

Box A4-14. Complete list of the User's Guides

Oxman AD, Sackett DL, Guyatt GH. **Users' guides to the medical literature. I. How to get started.** Evidence-Based Medicine Working Group. JAMA 1993;270:2093-5.

Guyatt GH, Sackett DL, Cook DJ. **Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid?** Evidence-Based Medicine Working Group. JAMA 1993;270:2598-601.

Guyatt GH, Sackett DL, Cook DJ. **Users' guides to the medical literature. II. How to use an article about therapy or prevention. B. What were the results and will they help me in caring for my patients?** Evidence-Based Medicine Working Group. JAMA 1994;271:59-63.

Jaeschke R, Guyatt G, Sackett DL. **Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid?** Evidence-Based Medicine Working Group. JAMA 1994a;271:389-91.

Jaeschke R, Guyatt GH, Sackett DL. **Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients?** Evidence-Based Medicine Working Group. JAMA 1994b;271:703-7.

Levine M, Walter S, Lee H, Haines T, Holbrook A, Moyer V. **Users' guides to the medical literature. IV. How to use an article about harm.** Evidence-Based Medicine Working Group. JAMA 1994;271:1615-9.

Laupacis A, Wells G, Richardson WS, Tugwell P. **Users' guides to the medical literature. V. How to use an article about prognosis.** Evidence-Based Medicine Working Group. JAMA 1994;272:234-7.

Oxman AD, Cook DJ, Guyatt GH. **Users' guides to the medical literature. VI. How to use an overview.** Evidence-Based Medicine Working Group. JAMA 1994;272:1367-71.

Richardson WS, Detsky AS. **Users' guides to the medical literature. VII. How to use a clinical decision analysis. A. Are the results of the study valid?** Evidence-Based Medicine Working Group. JAMA 1995a;273:1292-5.

Richardson WS, Detsky AS. **Users' guides to the medical literature. VII. How to use a clinical decision analysis. B. What are the results and will they help me in caring for my patients?** Evidence Based Medicine Working Group. JAMA 1995b;273:1610-3.

Hayward RS, Wilson MC, Tunis SR, Bass EB, Guyatt G. **Users' guides to the medical literature. VIII. How to use clinical practice guidelines. A. Are the recommendations valid?** Evidence-Based Medicine Working Group. JAMA 1995;274:570-4.

Wilson MC, Hayward RS, Tunis SR, Bass EB, Guyatt G. **Users' guides to the Medical Literature. VIII. How to use clinical practice guidelines. B. what are the recommendations and will they help you in caring for your patients?** Evidence-Based Medicine Working Group. JAMA 1995;274:1630-2.

Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ. **Users' guides to the medical literature. IX. A method for grading health care recommendations.** Evidence-Based Medicine Working Group. JAMA 1995;274:1800-4.

Naylor CD, Guyatt GH. **Users' guides to the medical literature. X. How to use an article reporting variations in the outcomes of health services.** Evidence-Based Medicine Working Group. JAMA 1996a;275:554-8.

Naylor CD, Guyatt GH. **Users' guides to the medical literature. XI. How to use an article about a clinical utilization review.** Evidence-Based Medicine Working Group. JAMA 1996b;275:1435-9.

Guyatt GH, Naylor CD, Juniper E, Heyland DK, Jaeschke R, Cook DJ. **Users' guides to the medical literature. XII. How to use articles about health-related quality of life.** Evidence-Based Medicine Working Group. JAMA 1997;277:1232-7.

Drummond MF, Richardson WS, O'Brien BJ, Levine M, Heyland D. **Users' guides to the medical literature. XIII. How to use an article on economic analysis of clinical practice. A. Are the results of the study valid?** Evidence-Based Medicine Working Group. JAMA 1997;277:1552-7.

O'Brien BJ, Heyland D, Richardson WS, Levine M, Drummond MF. **Users' guides to the medical literature. XIII. How to use an article on economic analysis of clinical practice. B. What are the results and will they help me in caring for my patients?** Evidence-Based Medicine Working Group. JAMA 1997;277:1802-6.

Dans AL, Dans LF, Guyatt GH, Richardson S. **Users' guides to the medical literature: XIV. How to decide on the applicability of clinical trial results to your patient.** Evidence-Based Medicine Working Group. JAMA 1998;279:545-9.

Richardson WS, Wilson MC, Guyatt GH, Cook DJ, Nishikawa J. **Users' guides to the medical literature: XV. How to use an article about disease probability for differential diagnosis.** Evidence-Based Medicine Working Group. JAMA 1999;281:1214-9.

Guyatt GH, Sinclair J, Cook DJ, Glasziou P. **Users' guides to the medical literature: XVI. How to use a treatment recommendation.** Evidence-Based Medicine Working Group and the Cochrane Applicability Methods Working Group. JAMA 1999;281:1836-43.

Barratt A, Irwig L, Glasziou P, Cumming RG, Raffle A, Hicks N, Gray JA, Guyatt GH. **Users' guides to the medical literature: XVII. How to use guidelines and recommendations about screening.** Evidence-Based Medicine Working Group. JAMA 1999;281:2029-34.

Randolph AG, Haynes RB, Wyatt JC, Cook DJ, Guyatt GH. **Users' guides to the medical literature: XVIII. How to use an article evaluating the clinical impact of a computer-based clinical decision support system.** Evidence-Based Medicine Working Group. JAMA 1999;282:67-74.

Bucher HC, Guyatt GH, Cook DJ, Holbrook A, McAlister FA. **Users' guides to the medical literature: XIX. Applying clinical trial results. A. How to use an article measuring the effect of an intervention on surrogate end points.** Evidence-Based Medicine Working Group. JAMA 1999;282:771-8.

McAlister FA, Laupacis A, Wells GA, Sackett DL. **Users' guides to the medical literature: XIX. Applying clinical trial results B. Guidelines for determining whether a drug is exerting (more than) a class effect.** Evidence-Based Medicine Working Group. JAMA 1999;282:1371-7.

Hunt DL, Jaeschke R, McKibbin KA. **Users' guides to the medical literature: XXI. Using electronic health information resources in evidence-based practice.** Evidence-Based Medicine Working Group. JAMA 2000;283:1875-9.

McAlister FA, Straus SE, Guyatt GH, Haynes RB. **Users' guides to the medical literature: XX. Integrating research evidence with the care of the individual patient.** Evidence-Based Medicine Working Group. JAMA 2000;283:2829-36.

McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, Richardson WS. **Users' guides to the medical literature: XXII. How to use articles about clinical decision rules.** Evidence-Based Medicine Working Group. JAMA 2000;284:79-84.

Giacomini MK, Cook DJ. **Users' guides to the medical literature: XXIII. Qualitative research in health care A. Are the results of the study valid?** Evidence-Based Medicine Working Group. JAMA 2000a;284:357-62.

Giacomini MK, Cook DJ. **Users' guides to the medical literature: XXIII. Qualitative research in health care B. What are the results and how do they help me care for my patients?** Evidence-Based Medicine Working Group. JAMA 2000b;284:478-82.

Richardson WS, Wilson MC, Williams JW, Moyer VA, Naylor CD. **Users' guides to the medical literature: XXIV. How to use an article on the clinical manifestations of disease.** Evidence-Based Medicine Working Group. JAMA 2000;284:869-75.

Guyatt GH, Haynes RB, Jaeschke RZ, Cook DJ, Green L, Naylor CD, Wilson MC, Richardson WS. **Users' guides to the medical literature: XXV. Evidence-based medicine: principles for applying the users' guides to patient care.** Evidence-Based Medicine Working Group. JAMA 2000;284:1290-6.

A5. Software for data synthesis

A selection of useful software for the synthesis of data is here provided. The list was elaborated with information obtained from the CRD Report No. 4, Egger et al. 2001 and from "Netting the Evidence" (<http://www.shef.ac.uk/~scharr/ir/netting>):

Software	Available at	Comments
Epi Meta	http://www.cdc.gov/epo/dpram/epimeta/epimeta.htm	Meta-Analysis
Meta	http://www.fu-berlin.de/gesund/gesu_engl/meta_e.htm	Basic meta-analysis procedures, based on DOS
Meta-Analyst	Available on request from: Dr J Lau, New England Medical Center, Box 63, 750 Washington St, Boston, MA 02111, USA. e-mail: joseph.lau@es.nemc.org	Basic meta-analysis procedures, based on DOS
EasyMA	http://www.spc.univ-lyon1.fr/~mcu/easyrna/	DOS based, performs basic procedures, standard and cumulative MA
Meta-Test	http://www.cochrane.org/cochrane/sadt.htm	Meta-analysis of diagnostic test data, based on DOS
Metaxis	http://www.update-software.com/metaxis/metaxis-frame.html	Commercial package
Review Manager	http://www.cochrane.org/cochrane/revman.htm	Manages the whole systematic review process
Clinical decision making	http://www.ccc.nottingham.ac.uk/~mczwww/tltp/decis.htm	Decision making trees
StatsDirect	http://www.statsdirect.co.uk	Statistical package for epidemiology and health research
EpiInfo	http://www.cdc.gov/epiinfo	Statistical package for epidemiology

Meta-analyses may also be performed with comprehensive statistical packages such as SAS or STATA, for which meta-analytic procedures are available.

A6. Levels of Evidence and Grades of Recommendations

Table A6-1. Levels of Evidence (Centre for Evidence Based Medicine, Oxford - version May 2001)

Level of Evidence	Therapy/Prevention, Aetiology/Harm	Prognosis	Diagnosis	Differential diagnosis/symptom prevalence study	Economic and decision analyses
1a	SR (with homogeneity*) of RCTs	SR (with homogeneity*) of inception cohort studies; CDR† validated in different populations	SR (with homogeneity*) of Level 1 diagnostic studies; CDR† with 1b studies from different clinical centres	SR (with homogeneity*) of prospective cohort studies	SR (with homogeneity*) of Level 1 economic studies
1b	Individual RCT (with narrow Confidence Interval‡)	Individual inception cohort study with > 80% follow-up; CDR† validated in a single population	Validating** cohort study with good††† reference standards; or CDR† tested within one clinical centre	Prospective cohort study with good follow-up****	Analysis based on clinically sensible costs or alternatives; systematic review(s) of the evidence; and including multi-way sensitivity analyses
1c	All or none [§]	All or none case-series	Absolute SpPins and SnNouts††	All or none case-series	Absolute better-value or worse-value analyses ††††
2a	SR (with homogeneity*) of cohort studies	SR (with homogeneity*) of either retrospective cohort studies or untreated control groups in RCTs.	SR (with homogeneity*) of Level >2 diagnostic studies	SR (with homogeneity*) of 2b and better studies	SR (with homogeneity*) of Level >2 economic studies
2b	Individual cohort study (including low quality RCT; e.g., <80% follow-up)	Retrospective cohort study or follow-up of untreated control patients in an RCT; Derivation of CDR† or validated on split-sample§§§ only	Exploratory** cohort study with good††† reference standards; CDR† after derivation, or validated only on split-sample§§§ or databases	Retrospective cohort study, or poor follow-up	Analysis based on clinically sensible costs or alternatives; limited review(s) of the evidence, or single studies; and including multi-way sensitivity analyses
2c	"Outcomes" Research; Ecological studies	"Outcomes" Research		Ecological studies	Audit or outcomes research
3a	SR (with homogeneity*) of case-control studies		SR (with homogeneity*) of 3b and better studies	SR (with homogeneity*) of 3b and better studies	SR (with homogeneity*) of 3b and better studies
3b	Individual Case-Control Study		Non-consecutive study; or without consistently applied reference standards	Non-consecutive cohort study, or very limited population	Analysis based on limited alternatives or costs, poor quality estimates of data, but including sensitivity analyses incorporating clinically sensible variations.
4	Case-series (and poor quality cohort and case-control studies ^{§§})	Case-series (and poor quality prognostic cohort studies ^{***})	Case-control study, or non-independent reference standard	Case-series or superseded reference standards	Analysis with no sensitivity analysis
5	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on economic theory or "first principles"

Source: Centre for Evidence Based Medicine, Oxford, UK. <http://cebm.jr2.ox.ac.uk/docs/levels.html>

SR Systematic Review

RCT Randomised controlled trial

*By homogeneity we mean a systematic review that is free of worrisome variations (heterogeneity) in the directions and degrees of results between individual studies. Not all systematic reviews with statistically significant heterogeneity need be worrisome, and not all worrisome heterogeneity need be statistically significant. As noted above, studies displaying worrisome heterogeneity should be tagged with a "-" at the end of their designated level.

†Clinical Decision Rule. (These are algorithms or scoring systems which lead to a prognostic estimation or a diagnostic category.)

**An appropriate spectrum is a cohort of patients who would normally be >tested for the target disorder. An inappropriate spectrum compares patients already known to have the target disorder with patients diagnosed with another condition.

‡See note #2 above for advice on how to understand, rate and use trials or other studies with wide confidence intervals.

§Met when all patients died before the Rx became available, but some now survive on it; or when some patients died before the Rx became available, but none now die on it.

††An "Absolute SpPin" is a diagnostic finding whose Specificity is so high that a Positive result rules-in the diagnosis. An "Absolute SnNout" is a diagnostic finding whose Sensitivity is so high that a Negative result rules-out the diagnosis.

‡‡Good, better, bad and worse refer to the comparisons between treatments in terms of their clinical risks and benefits.

§§By poor quality cohort study we mean one that failed to clearly define comparison groups and/or failed to measure exposures and outcomes in the same (preferably blinded), objective way in both exposed and non-exposed individuals and/or failed to identify or appropriately control known confounders and/or failed to carry out a sufficiently long and complete follow-up of patients. By poor quality case-control study we mean one that failed to clearly define comparison groups and/or failed to measure exposures and outcomes in the same (preferably blinded), objective way in both cases and controls and/or failed to identify or appropriately control known confounders.

§§§Split-sample validation is achieved by collecting all the information in a single tranche, then artificially dividing this into "derivation" and "validation" samples.

***By poor quality prognostic cohort study we mean one in which sampling was biased in favour of patients who already had the target outcome, or the measurement of outcomes was accomplished in <80% of study patients, or outcomes were determined in an unblinded, non-objective way, or there was no correction for confounding factors.

****Good follow-up in a differential diagnosis study is >80%, with adequate time for alternative diagnoses to emerge (eg 1-6 months acute, 1 - 5 years chronic)

†††Good reference standards are independent of the test, and applied blindly or objectively to applied to all patients. Poor reference standards are haphazardly applied, but still independent of the test. Use of a non-independent reference standard (where the 'test' is included in the 'reference', or where the 'testing' affects the 'reference') implies a level 4 study.

†††Better-value treatments are clearly as good but cheaper, or better at the same or reduced cost. Worse-value treatments are as good and more †expensive, or worse and the equally or more expensive.

Table A6-2. "Traditional" EBM hierarchy of reseach design/ quality of evidence

I:	Evidence obtained from at least one properly randomized controlled trial.
II-1:	Evidence obtained from well-designed controlled trials without randomization.
II-2:	Evidence obtained from well-designed cohort or case-control analytic studies, preferably from more than one center or

	research group.
II-3:	Evidence obtained from multiple time series with or without the intervention. Dramatic results in uncontrolled experiments (such as the results of the introduction of penicillin treatment in the 1940s) could also be regarded as this type of evidence.
III:	Opinions of respected authorities, based on clinical experience; descriptive studies and case reports; or reports of expert committees.

Table A6-3. Grades of Recommendations (Centre for Evidence Based Medicine, Oxford - version May 2001)

A	consistent level 1 studies
B	consistent level 2 or 3 studies or extrapolations from level 1 studies
C	level 4 studies or extrapolations from level 2 or 3 studies
D	level 5 evidence or troublingly inconsistent or inconclusive studies of any level

"Extrapolations" are where data is used in a situation which has potentially clinically important differences than the original study situation.
Source: Centre for Evidence Based Medicine, Oxford, UK. <http://cebm.jr2.ox.ac.uk/docs/levels.html>

Table A6-4. Recommendation grid and standard recommendation language (based on Third U.S. Preventive Services Task Force)

Quality of evidence	Net benefit			
	Substantial	Moderate	Small	Zero/ Negative
Good	A	B	C	D
Fair	B	B	C	D
Poor	I			
A	... strongly recommends that clinicians routinely provide [X] to eligible patients. (... found good evidence that [X] improves important health outcomes and concludes that benefits substantially outweigh harms.)			
B	... recommends that clinicians routinely provide [X] to eligible patients. (... found at least fair evidence that [X] improves important health outcomes and concludes that benefits outweigh harms.)			
C	... makes no recommendation for or against routine provision of [X]. (... found at least fair evidence that [X] can improve health outcomes but concludes the balance of the benefits and harms is too close to justify a general recommendation.)			
D	... recommends against routinely providing [X] to asymptomatic patients. (... found at least fair evidence that [X] is ineffective or that harms outweigh benefits.)			

I	... concludes that the evidence is insufficient to recommend for or against routinely providing [X]. (Evidence that [X] is effective is lacking, of poor quality, or conflicting and the balance of benefits and harms cannot be determined.)
---	--

Source: Harris RP, Helfand M, Woolf SH et al. Current Methods of the U.S. Preventive Services Task Force: A Review of the Process. Am J Prev Med 2001;20(3S):21-35