International Journal of Technology Assessment in Health Care, **18:2** (2002), 361–422.

BEST PRACTICE IN UNDERTAKING AND REPORTING HEALTH TECHNOLOGY ASSESSMENTS

Working Group 4 Report

Reinhard Busse, Chair
European Observatory on Health Care Systems, Spain & Technische Universïtät Berlin, Germany

Jacques Orvain, Co-Chair
National Agency for Accreditation and Evaluation in Health (ANAES), France

Marcial Velasco
Technische Universïtät Berlin, Germany

Matthias Perleth
AOK-Bundesverband, Germany

Michael Drummond
University of York, Center for Health Economics, United Kingdom

Felix Gürtner
Medical Technology Unit, Federal Social Insurance Office, Switzerland (MTU-FSIOS)

Torben Jørgensen
Danish Centre for Evaluation and Health Technology Assessment (DACEHTA), Denmark

Albert Jovell
Fundació – Biblioteca Josep Laporte Casa de Convalèscencia, Spain

Jim Malone
St. James's Hospital, Ireland

Alric Rüther
German Agency for Health Technology Assessment (DAHTA@DIMDI), Germany

Claudia Wild
HTA-Unit of the Institute of Technology Assessment (ITA), Austrian Academy of Science, Austria

Prepared for and in close collaboration with the working group by Reinhard Busse, Marcial Velasco, Matthias Perleth, and Jacques Orvain. The authors are indebted to Wendy Wisbaum (European Observatory on Health Care Systems) for providing English-language editing.

EXECUTIVE SUMMARY

The aim of Working Group 4 has been to develop and disseminate best practice in undertaking and reporting assessments, and to identify needs for methodologic development.

Health technology assessment (HTA) is a multidisciplinary activity that systematically examines the technical performance, safety, clinical efficacy, and effectiveness, cost, cost-effectiveness, organizational implications, social consequences, legal, and ethical considerations of the application of a health technology (18). HTA activity has been continuously increasing over the last few years. Numerous 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. Although there seems to be broad agreement on the general elements that belong to the HTA process, and although HTA doers in Europe use similar principles (41), this is often difficult to see because of differences in language and terminology.

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 toward disseminating the findings of a HTA; thus, standards that ensure high quality reporting may contribute to a wider dissemination of results.

The EUR-ASSESS methodologic subgroup already proposed a framework for conducting and reporting HTA (18), which served as the basis for the current working group. New developments in the last 5 years necessitate revisiting that framework and providing a solid structure for future updates. Giving due attention to these methodologic developments, this report describes the current "best practice" in both undertaking and reporting HTA and identifies the needs for methodologic development. It concludes with specific recommendations and tools for implementing them, e.g., by providing the structure for English-language scientific summary reports and a checklist to assess the methodologic and reporting quality of HTA reports.

OBJECTIVES OF WORKING GROUP 4

Within the overall framework of the ECHTA project, the objectives of Working Group 4 and this report are:

- 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; and
- To identify needs for methodologic development.

This report addresses the first two objectives, and the two objectives of disseminating this best practice are addressed both by writing this report and 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 Conclusions 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 methodologic development, which are summarized at the end.

METHODOLOGY APPLIED BY THE WORKING GROUP

As mentioned, the EUR-ASSESS methodologic subgroup proposed a framework for conducting and reporting HTA (18), 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 methodologic framework based on existing guidelines from HTA agencies and other institutions 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 that 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, personal communication, 2001), guidance documents and tool kits from different institutions involved in HTA were examined and summarized 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 methodologic 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;13:186-219. (reference 18).
- Various reports from the NHS R&D HTA Programme, 1998–2001 (for details see Appendix 1).
- Guía para la elaboración de informes de evaluación de tecnologías sanitarias. Agencia de evaluación de tecnologías sanitarias, Madrid, Spain, 1999 (reference 30).
- Development and Evaluation Committee Guidelines. The Wessex Institute for Health Research and Development, Southampton, UK (reference 11).
- West Midlands Development and Evaluation Service (DES) Handbook. Department of Public Health and Epidemiology, University of Birmingham, DPHE Report No. 8, 2000 (reference 5).
- Guide d'Analyse de la littérature et gradation des recommandations. Agence Nationale d'Accréditation et d'Évaluation en Santé, Paris, 2000 (reference 15).
- Undertaking Systematic Reviews of Research on Effectiveness. CRD's Guidance for Carrying Out or Commissioning Reviews. CRD Report 4, NHS-CRD, University of York, 2000 (reference 36).
- Guide to the Technology Appraisal Process. National Institute for Clinical Excellence, 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 (reference 45).
- Health Technology Assessment Handbook. Danish Institute for Health Technology Assessment, Copenhagen, Denmark, 2001 (reference 37).

In addition, based on working group members' experience and reference lists, specific guidance, and key references for the identified specific issues—and for gaps that 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) and others. This final version reflects the amendments, comments, and discussion.

METHODOLOGIC FRAMEWORK FOR CONDUCTING HTA

Characteristics of HTA

Health technology assessment, a multidisciplinary activity that systematically examines the technical performance, safety, clinical efficacy and effectiveness, cost, cost-effectiveness, organizational implications, social consequences, and legal and ethical considerations of the application of a health technology (18), has to take into consideration all aspects that might be influenced by the technology and those influencing the technology. In this context, health technology is a broad concept that includes drugs, devices, procedures, and the organizational 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 healthcare sector (19). These kinds of activities are characterized by a systematic and structured way of answering questions by evaluating and synthesizing 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 (18). Besides a systematic methodology, the strength of HTA relies on transparency of the process and in the reporting, which also improves the usefulness and generalizability of the findings.

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 should be pointed out that the way the different steps are undertaken influences the elaboration of the HTA report, which can be seen as one layer in the overall assessment process and represents the deliverable product of the assessment (Figure 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 prioritization and/or commissioning process. The next step consists of the gathering of background information (part of which may have already been collected during the prioritization 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)

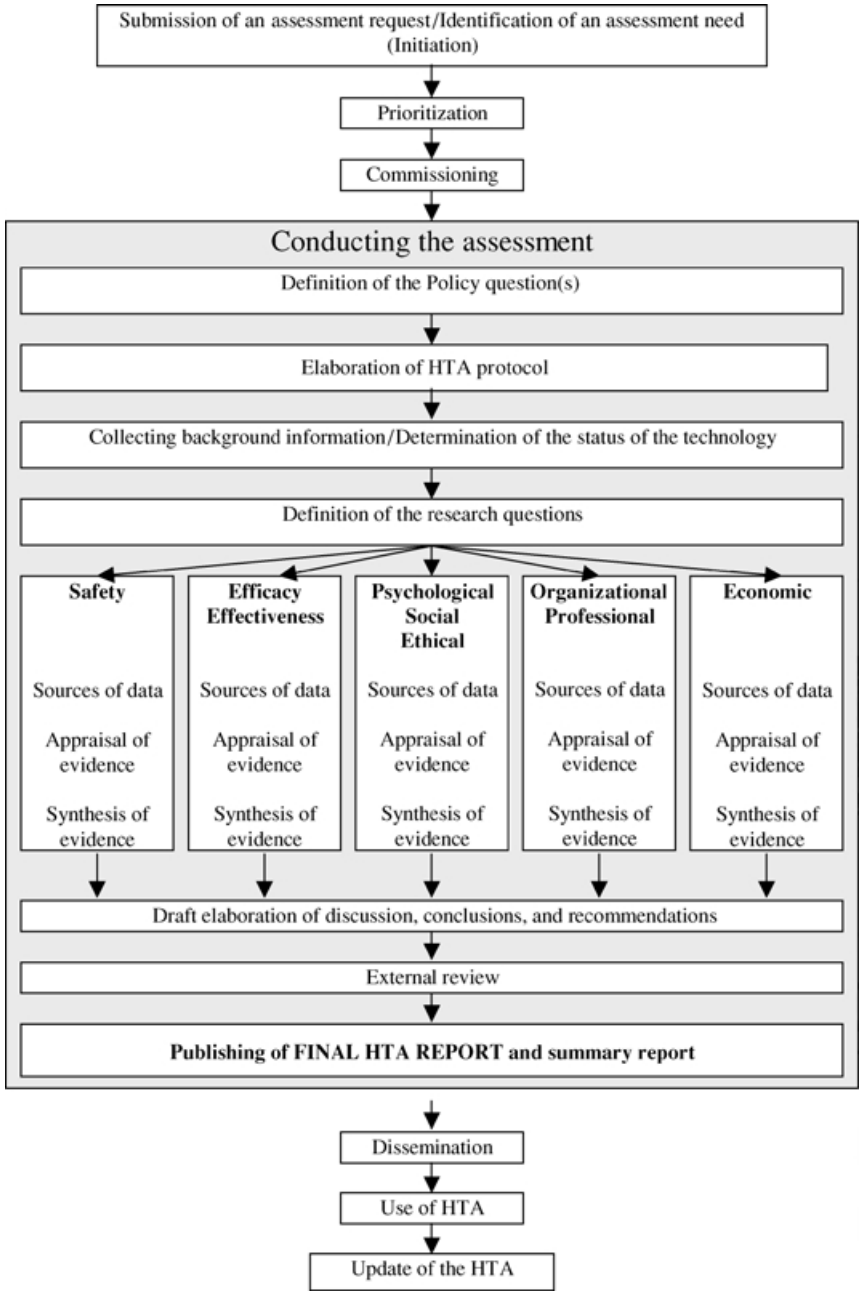

Figure 1. Assessment process.

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, since 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.

Table 1. Content of an HTA (Modified from EUR-ASSESS 1997)

Policy question	
Background information on target group, target condition, technology (technical aspects, diffusion, and current practice)	
Research questions	
Findings and Methodology	• Safety • Efficacy/effectiveness • Psychological, social, and ethical considerations • Organizational and professional implications • Economic issues
Policy conclusions and recommendations	

BEST PRACTICE IN UNDERTAKING HTA REPORTS

The EUR-ASSESS Subgroup proposed a framework with the elements that should be included in an HTA report (Table 1). 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 methodologic framework, in terms of what could be considered best practice, following the structure shown in Figure 1 and Table 1. Other important issues concerning the HTA process, such as the review process, updates of the HTA, and possible conflicts of interest, cannot be clearly ordered in the structure proposed in Figure 1 and will therefore be considered later.

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 cooperation 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 (Table 2).

The context in which the research is carried out may lead to some financial or time constraints that 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 as well as in the technical report (i.e., the detailed document), and the scientific summary report. The questions listed in Table 2 should be answerable when reading any of these documents.

HTA Protocol

As soon as the policy question is clear, an HTA protocol should be developed to define how the whole assessment is going to be carried out. An HTA protocol is not a systematic

Table 2. Aspects Included in the Policy Question

Question	Examples
Who initiated the report?	Policy makers
	Healthcare providers
	Third-party payers
	Patients' advocate
Who commissioned it?	
Why is an assessment needed right now?	New technology
	Changes in old technology
	New indications for old technology
	New findings
	Structural/organizational changes
	Safety concerns
	Ethical concerns
	Economic concerns
Which decision is it going to support?	Investment decisions
	Market licensure
	Inclusion in/exclusion from benefits catalogue
	Planning of capacities
	Guidance on best practice
	Investment in further research
Who represents the primary target audience for the report?	Political decision makers
	Third-party payers
	Hospital managers/administrators
	Clinicians
	Citizens/patients

review protocol, as this usually refers only to one of the possible aspects to be reviewed in the assessment. An 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 utilization 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 (11).

In a simplified way, the development of an 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 synthesizing 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, an 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 Figure 1 was carried out.

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 be carried out separately 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) and background information on the technology (e.g., technological characteristics). Review articles (not necessarily systematic) and textbooks 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 include (5;11):

1. Perform this parallel with defining research question;
2. 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; and
3. Sources:
   - Research literature (search strategies targeting "reviews," "prevalence," "incidence," etc.);
   - Routinely collected data (on utilization, costs, etc.);
   - Guidelines;
   - Special sources (disease registers, organizations of affected people, experts, manufacturers; some of those sources are accessible through the internet);
   - 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, since 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.

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 (Table 3).

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 that are going to be used in the assessment on safety, efficacy,

Table 3. Questions To Be Addressed as Background Information on Condition and Target Group

Questions	Example
Condition(*s*)	Health problem Disease
What are the mechanisms of disease?	Causes Pathology
What is the course and prognosis of the condition?	Clinical presentation Stages Time course
What are the consequences? (outcomes)	Physically disabling Psychological consequences Death
Treatment alternatives and current practice	Drugs Surgery Current service provision
Target group(*s*) (*epidemiology*, *burden of disease*)	Patients Healthy subjects (for prevention)
How many people are affected?	Incidence Prevalence
Who is affected?	Age Gender Social factors Risk factors

Source: Adapted from Burls et al. (5).

Table 4. Questions To Be Addressed as Background Information on the Technology

Question	Aspects/examples
How does it work? What kind of intervention is it?	If a device, explain technical characteristics, functioning If a community/system-related intervention, explain its crucial features
What are the requirements for its use?	Setting for use/implementation Special measures needed for use/implementation Qualification required Maintenance
What is the status of the technology?	Diffusion/distribution Patterns of use Current indications for use Current utilization Costs Regulatory status Manufacturers and market shares

effectiveness, and the other relevant outcomes. The description of the appropriate outcomes and how they are measured is therefore an important issue too.

Technology. It is best practice to concisely describe the following aspects of the technology (Table 4), keeping in mind that the technology assessed may be a drug, device (therapeutic/diagnostic), community intervention, medical aid, procedure, organizational process, support system, or a combination of these.

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). 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 utilization, and 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 technology.[2] Sometimes these issues may not need serious consideration, depending on the status of the technology (e.g., utilization patterns if assessment is prior to approval for use).

Research Question(s)

Formulating the research question(s) means specifying the policy question in terms of safety, efficacy, effectiveness, psychological, social, ethical, organizational, 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. *Organizational* 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. Table 5 provides examples of outcomes for the different aspects.

The research question drives how the rest of the assessment is going to be conducted, the aspects that will be evaluated, and those that 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 indicate that this is a crucial part of the assessment, since other aspects (e.g., methodologic) of the evaluation flow from it. If possible and where relevant, there should be a feedback 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. Characteristics of research questions include:

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

Table 5. Examples of Outcomes for Different Aspects of HTA

Aspect of assessment	Outcomes
Safety	Mortality directly related to the use of technology
	Morbidity/disability directly related to the use of technology
Efficacy/effectiveness	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	Compliance
	Acceptance
	Satisfaction
	Demand
	Preferences
	Information/patient advice requirements
Organizational/professional	Utilization of service
	Change in the treatment location
	Change in length of hospital stay
	Change in required personnel, material inputs (e.g., hospital beds) and organizational structure
	Training requirements
Economic	Costs and changes in cost compared to current practice (if applicable)
	Cost-effectiveness, cost-utility, cost-benefit

Answering the Questions/General Methodology

Once the research question(s) have been formulated, the next step is to answer them. As shown in Figure 1, there are some general methodologic steps that apply to all aspects of the HTA (i.e., safety, efficacy/effectiveness, psychological/social/ethical, organizational/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 below. Specific methodologic considerations concerning each aspect of the assessment are addressed in the next section.

The common methodology for addressing the different aspects can be summarized in three steps:

1. Searching for sources of information;
2. Selecting and evaluating information (application of inclusion and exclusion criteria)/appraising the evidence; and
3. Synthesizing the obtained data.

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 research[3] 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 identified during the search for background information. Search for HTA reports has to be systematic and clearly documented.[4] Identified HTA reports should also be critically appraised. Systematic reviews may already cover some of the aspects and answer some of the questions posed. This may be the case for aspects such as 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, such as those developed by the Cochrane Collaboration, can be applied to all aspects of the assessment, and not only to efficacy/effectiveness. 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). Keywords related to the condition, the technology, and types of publication will be combined, forming the search strategy to obtain the biggest number of hits. It is recommended that the language of publication *not* be used as a search criterion, because relevant literature in other languages will be missed.

A systematic approach can also be applicable for psychological/social/ethical,[5] organizational/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 and the way in which the information was obtained should also be documented.

The documentation of the information sources is of utmost importance for the transparency of an HTA report. Both sources that provided useful information and those that did not should be included in the documentation (37). Documentation of the sources include:

- 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?

Inclusion and Exclusion Criteria/Appraisal of the Evidence. The selection of the literature that will be definitely included to answer the research questions is a process with consecutive steps to be taken, as summarized in Figure 2. With a 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 that fulfill the defined quality standards will be definitively selected for inclusion in the synthesis. It is recommended that this process be reported in an understandable and transparent way, e.g., by using Figure 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, rather than only for the literature on efficacy and effectiveness. Selection criteria should be developed in a prospective way to avoid bias when selecting the evidence. Inclusion and exclusion criteria flow from the background information, the research questions, and the availability of evidence. The criteria refer to, for example, patients being treated, outcomes being measured, and aspects of the technology being studied. Selection criteria also may refer to study design or other methodologic issues. Those criteria (may) differ for each of the aspects being assessed. For instance, when assessment of efficacy issues is based on randomized controlled trials (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

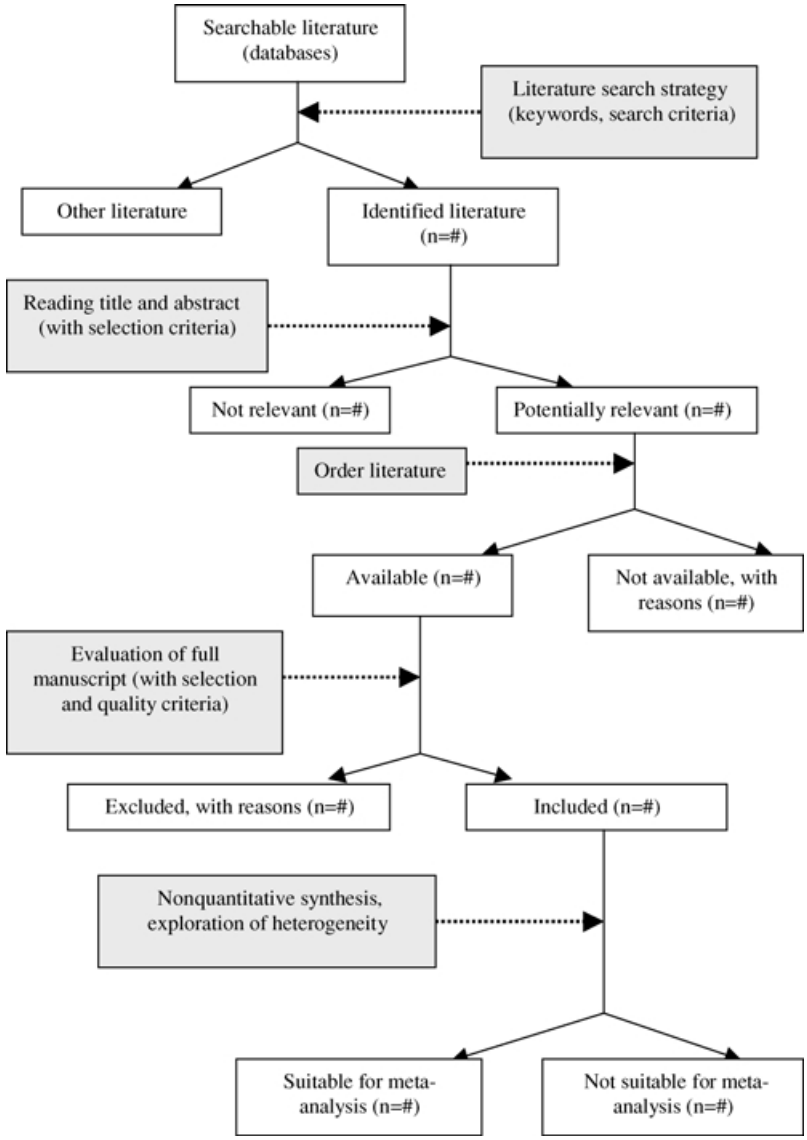

Figure 2. Flow diagram of literature selection process (adapted from Khan et al. (3b).)

register might be the selection criterion. Issues addressed in inclusion and exclusion criteria include:

- Patient characteristics (e.g., age, gender);
- Condition characteristics (e.g., stage of disease);
- Technology aspects;
- Methodologic issues (e.g., number of patients, length of follow-up, study design);
- Outcomes measured; and
- 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, explicit description

of these criteria should be mandatory in a HTA report. Inclusion and exclusion criteria must 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 must 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 (16).

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 checklists[6] have been developed. Some doers have adapted these checklists and provide them in their guidance documents (15;45). 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 than the medical literature, should be appraised. The tools and criteria developed for the medical literature are not applicable to other sources of information, so there is a gap here that 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 because they are the study design less likely to provide biased results.[7] 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 such as 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 (43). Although standard quality assessment instruments/checklists/scores exist, such as the validated Jadad score (34), some agencies recommend developing specific instruments for each assessment, since some quality issues are closely related to special aspects of the technology being assessed. The criteria should cover both generic and specific methodologic aspects. Generic methodologic aspects refer to study characteristics that, if present, for example, indicate good quality of a study independent from the subject being studied (e.g., concealment of allocation). Specific methodologic aspects refer to characteristics that 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). Quality items/criteria include:

- Generic methodologic issues (e.g., study design, allocation of concealment, prospective, randomization, dropout rate); and
- Specific methodologic issues (e.g., length of follow-up, methods for assessing outcomes, ways of applying technology).

Table 6. Quality Assessment Presentation (Example)

	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

Source: Burls et al. (5).

This step should be reported in a transparent way. For each study, how or whether it fulfills the different quality items should be documented. An overall score that synthesizes all the items also might be used, and if so, the way the score is 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 (35). 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 fulfill all other quality criteria. If knock-out criteria are being used, which criteria and why they were chosen should be clearly stated. Studies originally retrieved that do not fulfill the quality criteria will be excluded; documentation of excluded studies should be provided, along with the reasons for exclusions. Transparency in quality assessment should include the following steps:

- Document and explain quality criteria and items included in assessment;
- If a score is used, describe how it is constructed;
- List retrieved studies that were not included, with reasons for exclusion; and
- Fully report results of quality assessment (tabulation).

A good approach for reporting the quality assessment is the use of tables, as recommended by the West Midlands Development and Evaluation Service (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 (Table 6).

Nonquantitative 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 are mainly determined by the research questions. It is strongly recommended that customized extraction sheets be 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 synthesized 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 summary of information can be used.

In HTA, synthesis should be transparent. A way to enhance transparency, even if the synthesis is narrative, is the use of *evidence tables*. These tables are commonly used to summarize 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. Elements to include in evidence tables are:

- 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; and
- 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 judgments about their validity.

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 generalizability of results, should be presented in a way that enables easy comparison between included studies.

When recommending the use of evidence tables to summarize study characteristics and study results as the best way to synthesize the evidence in a nonquantitative form (which always precedes a quantitative synthesis), agencies and other institutions coincide. In a nonquantitative 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 nonquantitative synthesis of information, explicit criteria for validity and quality of the studies have to be followed. Thus, the nonquantitative synthesis is closely related to the appraisal process.

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 (8). 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 what 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 nonquantitative 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.

Table 7. Factors To Consider When Using Quantitative Synthesis (Meta-analysis)[a]

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?

Source: Adapted from Egger et al. (17) and Moher et al. (42).

[a]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.

There are different methods for performing a quantitative synthesis for HTA doers.[8] However, the most extended one is the use of meta-analysis. Table 7 gives an overview of the factors that should be taken into consideration when choosing a method of meta-analysis.

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 modeling are other tools for quantitatively summarizing information (30). 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 favorable way will be identified by means of an overall score.

In addition, the use of modeling can be useful for other purposes, many of which aim at providing more information than "just" a quantitative synthesis of available evidence. Uses of modeling (adapted from EUR-ASSESS) (18):

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

There are different methods for modeling, such as decision trees, Markov models, or threshold analyses (24;55). The use of mathematical models implies some assumptions that should be explained. A model needs to be fed probabilities (e.g., having an illness, suffering an event), which will be taken from different sources (e.g., meta-analyses, single studies, expert opinions), thus having different grades of validity. Therefore, the sources of data that feed the model should 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 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, using the following guidelines:

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

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

Specific Methodologic Considerations

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

Safety. Assessing safety implies a wide scope 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 (45). In its guidelines, the Medical Services Advisory Committee (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, or 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 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. Also, the reporting of RCTs in regard to quality and quantity of safety (adverse effects and laboratory-determined toxicity) is currently largely inadequate (31); thus, it is extremely important to carefully examine the reasons why subjects leave the study, since the presence of adverse effects might have been an exclusion criterion.

Other study designs, such as observational studies, play 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 (38). 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 summarizing the different data.

Table 8. 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 (9)
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 (59)
Maximum achievable benefit	Achieved benefit	Williamson (61)
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 (54)
What works under carefully controlled conditions, such as RCTs	What works in day-to-day clinical practice	Rettig (51)

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 also may be operator- or setting-dependent (e.g., learning curve of surgeons), which also need to be taken into consideration and discussed. Timing (short-term, long-term) and severity of adverse effects should also be considered. Another important aspect of safety is the identification of differences in risk among different groups of patients.

When possible, quantification of harm into QALYs or DALYs should be made (11). Safety can be summarized 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.

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 (Table 8).

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 (37).

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 that could contain answers for those questions should be made. An important source of this kind of literature is the Cochrane Library (see

Appendixes). Search filters to identify systematic reviews have been developed and may be useful (http://www.york.ac.uk/inst/crd/revs.htm) (36). If systematic reviews on efficacy are found that may be suitable for answering the questions of the current assessment, their quality and relevance should be assessed to decide whether they can be included in the assessment. Checklists to critically appraise systematic reviews have been developed and are summarized as follows (adapted from Greenlagh [26] and Oxman et al. [48]):

- 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 synthesized?
- 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 methodologic guidelines already exist on how to conduct systematic reviews of primary research.[10] In contrast to these guidelines, little consensus exists in regard to how to measure effectiveness, especially "community effectiveness." Tugwell et al. (58) proposed that the latter should be calculated as "efficacy × diagnostic accuracy × health professional compliance × patient compliance × 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 healthcare 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) (6). *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 organizational/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.

THERAPEUTIC INTERVENTIONS

In the slightly differing models that 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 randomization is not ethically justifiable). In such cases, the HTA doers will have to use evidence from other kinds of

study designs. Optimized standard search procedures have been developed to find RCTs,[11] 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 (25). The methodology of meta-analysis has been mainly developed for combining the results of RCTs on therapeutic interventions and is comprehensively described elsewhere (8). 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 nonquantitative 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 generalizability 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.

DIAGNOSTIC INTERVENTIONS

There are two kinds of technologies that 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 absence of disease in a symptomatic patient (36). The evaluation of both follows similar principles. For the assessment of diagnostic and screening tests, a hierarchical model can be followed (Table 9).

Table 9. Evaluation of Efficacy and Effectiveness for Diagnostic Interventions

Level	Typical measures
Technical efficacy	• Physical parameters describing technical performance of the test (e.g., image quality)
Diagnostic accuracy efficacy	• 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/likerats.html)
Diagnostic thinking efficacy/effectiveness	• Post-test odds/probability compared to pre-test odds/probability in target population • % of cases in which test is judged "helpful" to making diagnosis
Therapeutic effectiveness	• % of cases in which test is judged "helpful" in planning therapy • % of therapeutic procedures avoided due to test information
Health-related effectiveness (patient outcomes)	• Mortality/morbidity avoided with test • Changes in quality of life through use of test

Source: Adapted from Flynn and Adams (20) and Fryback and Thornbury (21).

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, because 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 well established in every case). The diagnostic technology should be ideally evaluated in a patient sample that includes an appropriate spectrum of patients with the target condition plus a representative group of individuals without the disease (20). Both patients who tested positive as well as those who tested negative should be compared with the diagnostic gold standard, i.e., not only those who tested positive (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 randomized 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 summarize 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 the standard approach (17;32;33). 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 logistic regression of the studies into a summary logistic regression. 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 (36). When assessing a diagnostic test or strategy, outcomes deriving from misclassification/misdiagnosis of patients can also be considered as harmful.

HEALTHCARE ORGANIZATION AND SYSTEM-RELATED INTERVENTIONS

Organizational, financial or regulatory interventions can also be considered as health technologies. As defined by the EPOC Group,[12] different types of interventions, such as professional (e.g., educational program on prescription), financial (e.g., copayment), organizational (e.g., changes in medical record system), and regulatory (e.g., licensure), are included here. These interventions are not to be confused with organizational, professional, and economic implications of introducing or applying a health technology.

For the evaluation of professional, financial, organizational or regulatory interventions, the HTA doers often need to be more flexible in their inclusion criteria for studies. Transparency in the selection process is of utmost importance as generalizability/transferability to other settings will be highly context-dependent. Table 10 lists available study designs by their methodologic strength (with the weakest designs toward the lower left). 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).

Table 10. Study Designs Used for Assessing Healthcare Organization and System-related Interventions

	Cross-sectional 1 point of measurement	Longitudinal 2 points of measurement	Regular/ continuous measurements
Experimental designs—often not feasible for evaluating healthcare organization and system-related interventions			
Researcher has controt over intervention and allocation of subjects/ institutions/areas, etc. into at least 2 groups; randomization possible		Classic experiment	
Researcher has control over intervention and allocation of subjects/institutions/ areas, etc. into at least 2 groups; randomization not possible	Post-test only with nonequivalent groups—weak design	Control group design with pre- and post-test/ controlled before and after study	Time series with nonequivalent control group/ cohort study
Quasi-experimental designs—feasible for evaluating healthcare organization and system-related interventions			
Natural experiment (i.e., intervention, not determined by researcher) with randomized allocation of subjects/institutions, etc. into at least 2 groups through researcher		Quasi-RCT—theoretically possible and desirable but de-facto hardly ever used; requires a dialogue between health politicians and researchers and enough time before the intervention to prepare evaluation	
Natural experiment with nonrandomized allocation of subjects/institutions, etc. into at least 2 groups	Post-test only with nonequivalent groups—weak design	Control group design with pre- and post-test/ controlled before and after study	Time series with nonequivalent control group/ cohort study
Natural experiment without prior allocation of subjects/institutions etc.; control group existing		Case-control study—not ideal but a compromise if pre-intervention measurements were not possible	
Simple, methodologically weak designs			
Intervention but no control group	One group post-test only design	One group pre-test post-test design	Simple interrupted time series—acceptable if at least three data points before and three after the intervention

Source: Adapted from Busse (6).

PREVENTIVE INTERVENTIONS

Preventive interventions are used 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. Other interventions such as screening programs are more diagnostic and must be implemented at a community level; these incorporate the considerations listed both for diagnostic interventions and for organizational and system-related interventions. Other community-based interventions include health promotion programs or public health strategies aimed at the population or environmental factors (e.g., fluoridation of drinkable water). Common methodologic 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 (28) can be regarded as best practice. Building upon previous work (2), the task force 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., counseling, 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 (Appendix 6).

Psychological, Social, and Ethical Considerations. The assessment of the impact of the use or non-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 and by the way they value it (e.g., if there is no acceptance, compliance will be reduced, and thus effectiveness too). Such aspects should therefore be included in a structured way in an HTA.

Psychological effects of a technology refer to a range of possible subjective effects, such as fear, anxiety, feeling labeled, and satisfaction, 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 judgments; for example, discrimination of handicapped life through the use of prenatal 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 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. Some work is being done to enable the inclusion of qualitative research in a systematic way when assessing health care.[13]

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 MEDLINE.[14] Comprehensive databases exist for social sciences, which also include literature on psychological and sociological aspects of health interventions (e.g., PsycINFO, Sociological Abstracts[15]). If such a literature search is done, the origin of the data and the strategies followed to find the evidence should be clearly stated. Literature found should then be assessed for their validity, quality, and transferability. Some criteria for appraising qualitative research used in healthcare research have been proposed and are summarized in Table 11; however, debates on this are still ongoing.

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 (40). If no evidence from the literature is available, the HTA doers may need to conduct primary research themselves in order to include the patient perspective when assessing a technology. Some of the methods that can be applied for this purpose are participant observation, individual interviews, focus group discussions, Delphi method, or future workshops.[16] 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 Table 11 are also applicable to primary research.

Another source of data can be surveys or questionnaires about some aspects such as satisfaction and acceptance. These sources may give more representative data, but they may only be useful to map phenomena that are already known (37). 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 such as patient organization 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 methodologic 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 should be careful not to rely on their own moral stance (18). 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 HTA.[17]

Organizational and Professional Implications.[18] The scope of an HTA report should also include organizational and professional changes induced by the technology and predict their further consequences, especially if the background information indicates important implications. For instance, the use of a new surgical procedure may imply training of staff, but may 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.) Organizational issues to be assessed may, for example, address changes in:

- Utilization 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 inpatient treatment, by means of the new technology, can be performed as an outpatient procedure);

Table 11. Sets of Criteria for Assessment of Studies Using Qualitative Research Methods

Popay et al. (49)
- 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?
- Generalizability: If claims are made to generalizability, do these follow logically and/or theoretically from the data?

Mays and Pope (39)
- 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 generalizability? Is sufficient original evidence provided to support the relationship between interpretation and evidence?

BSA Medical Sociology Group (4)
- 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?
- Are the researchers clear about their own position in relation to the research topic?

Mays and Pope (40)
- 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)

Source: Updated from Khan et al. (36).
A further checklist, based on Giacomini and Cook (22;23), is provided in Appendix 4.

- Training/qualification requirements (for example, if the application of a health technology—in contrast to its alternatives—presupposes the skills of a special medical expert);
- Channels of cooperation/communication (for example, if the effective use of a health technology presupposes extra communication between hospital and general practice); and
- Job satisfaction (for example, if a new procedure presupposes such a high throughput that the physicians have insufficient time for following the patients' progress).

Because an organization is a social interaction, within given frames, between persons who have one or more common ends as well as individual goals and aspirations, it is useful to start analyzing organizational issues by identifying the stakeholders and their interests (for a review of stakeholder analysis, see reference 3).

An assessment of such issues gives the first picture of the technology's (potential) organizational 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 organizational 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 organization through 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 (44); and
- Structured group processes such as future workshops or the Delphi method, especially when trying to identify and evaluate future changes of organizational structure and processes or when trying to predict reactions of people involved in the implementation.

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

Economic Issues. Assessments of economic issues in HTA imply first collecting information on resource consumption from the use of the technology (costs). The next step is 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 (1) 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 (37).

Generally, there are different types of costs that need to be taken into account, depending on purpose and perspective (Table 12). For all of them, the importance of measuring physical units first, before multiplying them with unit costs/prices to obtain total costs, cannot be overemphasized in order to help interpret results regarding their transferability to other settings—not only from one country to another (12) but also within one country across different providers (10). 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 (52).

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 must 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 (Table 13).

Guidelines on economic evaluation are numerous, although they are not tailored for use within the context of HTA (e.g., 7;13;14;24;27;47). The EUROMET project reviewed the contents of guidelines for economic evaluation of medical technologies from Australia,

Table 12. Types of Costs in an Economic Analysis

Perspectives	Types of costs	Examples
Healthcare payer		
Hospital	Direct costs	Healthcare 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 healthcare payer), screening programs
Societal perspective		
	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 traveling, time costs due to patient's time used for the treatment, family or friends' (unpaid) use of time of the patient
	Lost production in 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 healthcare costs	Future unrelated healthcare costs caused by curing the patient with the present treatment

Source: Modified from Kristenson et al. (37).

Table 13. Types of Economic Analysis

Type of economic analysis	When should the specific type of analysis be chosen?
Cost-minimization analysis	If the compared technologies are equally effective, then it is only necessary to collect data about costs
Cost-effectiveness analysis	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
Cost-utility analysis	If health-related quality of life is an important health outcome
	If activities across specialties or departments in the healthcare sector are compared
Cost-benefit analysis	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 individual lives are valued in monetary units
	If activities across society are compared

Source: Kristensen et al. (37).

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 (60). The recommendations in guidelines regarding discounting were recently compared in a study by Smith and Gravelle (56).

The EUROMET group also developed a consensus on a framework for European guidelines that is useful in the context of HTA (60). Table 14 summarizes the main issues for economic evaluation in HTA.

Table 14. Economic Evaluation

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 modeling society's valuation.
Time frame: long enough to capture all effects; modeling can be used to estimated 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 are important but more valid indicators are needed.

Source: Based on the EUROMET consensus (60).

Discussion of Methods and Results

The discussion is an important part of an 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 nonquantitative synthesis. However, a structured summary discussion should be always included in an assessment as a separate section, which should include the following:

- Methodology of the assessment;
- Evidence used (quality, validity, generalizability);
- Assumptions made;
- Discrepancies and uncertainties identified; and
- 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, modeling). 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 generalizability, 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) that 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 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 step before formulating conclusions and/or recommendations.

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 reporting 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. Conclusions should include the following points:

- Related primarily to the research question(s);
- Summarize quality/origin of the evidence;
- Summarize 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"; and
- Point out fields where further research is needed.

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 because it not confined to the efficacy/effectiveness dimension but also needs to take into account psychological/social/ethical, organizational/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). These are considered a major relevant finding of an 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 must 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 constructed.[19]

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

Other Relevant Issues

The following issues should also be taken into account when undertaking an 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.

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 review.[20] 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, and if so, to document the comments from reviewers and the way in which they were incorporated (if so) in to the final report. The review process should assess the following:

- 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?

Ideally, a preliminary version of the report should be reviewed by experts in the methodology and in the field that 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. 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. Afterward, the report is always reviewed by the Agency's Scientific Committee, which is nominated by the government from a list of representatives of the different healthcare providers.

Updating of Assessment. The validity of the findings of an 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 an 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 for an update is going to be identified (e.g., periodical literature search, hearings). The following shows an example of how the DEC identifies the need for an update (11):

- *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; and
- *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 that could not be included but were already identified). It could also be indicated when there are organizational or regulatory changes that may influence utilization 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 new full assessment necessary. To provide an assessment with an 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 an update. As with the review process, 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:

- Is an update planned?
- How will 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 always be reported, since it may be enough to refer to the source in which the process is described.

"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 practiced by most HTA institutions, even though the toolkits/guidelines do not mention this) to place them freely available on the Internet (usually, in pdf format). It is, however, still necessary to print executive summaries and patient information 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.

Abstract

Recommendations already exist on how to write a structured abstract for the INAHTA Database (http://agatha.york.ac.uk/htahp.htm). The abstract must be written in English. In

Table 15. Data To Be Included in an English Structured Abstract

Title: first title in English, then original title in brackets
Author/s: according to Vancouver style
Organization: organization 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, organizational
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 Prothesis; 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: nonquantitative, meta-analysis, modeling, economic evaluation
Results: main results
Recommendations: if given
Peer review process: specify: Yes/No/Internal/External/Both

Source: Imaz-Iglesia et al. (30).

its present form, it is usually too short to contain all aspects of interest when assessing the relevance and quality of an HTA report. The aspects to be included in the abstract are listed in Table 15.

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 (but 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,

Table 16. Differences Between an Executive summary and a Scientific Summary Report

Executive summary	Scientific summary report
Addressed to local decision makers ("executives")	Addressed to the HTA and the scientific community
Focuses on recommendations and conclusions	Stresses the context of the HTA and methodologic aspects, in addition to conclusions and recommendations
Written in agencies'/institutions' official tongue(s)	Available in English
Quickly informs decisions	Allows for critical appraisal of relevance, quality, and main findings

because these are the kinds of information sought by local decision makers. Methodologic aspects of the assessment are usually underrepresented in the executive summary, since 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 can be found. This could be termed a *scientific summary report*, to distinguish this kind of summary from the well-known executive summary, since they actually differ in their purpose and content Table 16.

The scientific summary report is a comprehensive summary of an HTA technical report, available in English and structured around five main questions (Who?, Why?, What?, How?, and What are the findings?) to allow for a quick assessment of the report's relevance, quality, and main findings to determine its further consideration. Additionally, both methodologic and contents-oriented keywords should be included to help to identify the report in database searches. The target audience of such a scientific summary report is mainly other researchers undertaking HTA or other HTA doers.

All questions listed in Table 17 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.

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 or standards are given here.

Table 17. Elements To Be Addressed in the Scientific Summary Report

Question	Aspects
Who?	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?
Why?	Why was the HTA commissioned/conducted?
	Why right now?
	What decision(s) is it going to inform?
What?	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?

(*continued*)

Table 17. (*Continued*)

Question	Aspects
How?	Was a HTA protocol followed? How was the assessment approached? Which aspects were assessed?
	Sources and synthesis of background information?
	Was safety assessed?
	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 synthesized?
	How was the efficacy/effectiveness assessed?
	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?
	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 analysis done?
	Were psychological/social/ethical considerations assessed?
	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 synthesized?
	Were organizational/professional implications assessed?
	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 synthesized?
	Was an economic evaluation conducted?
	What were the alternatives that were compared?
	What perspective was assumed?
	What were the underlying assumptions?
	What kind of analyses was made and why?
	Did the HTA undergo an external review process before publication?
Results	What are the main findings of the research?
Conclusions/ Discussion	Relate results to questions posed
	For which aspects of the assessment are there information lacking/ uncertain?
	Discuss transferability issues of results
Recommendations	If recommendations are given and graded, what gradation scale was it used?
Update	Is an update of the report planned?
	What criteria will be used to decide on it?
General aspects	Key words
	Bibliographic information

The way in which such summaries are elaborated should be left up to the commissioning institutions, as they better know their needs.

Technical Report

The technical report should include comprehensive information on all issues covered in the section on "Best Practice" in Undertaking HTA Reports. The questions listed in Table 15 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 that can be included in the technical report to enhance transparency and comprehensiveness in an understandable way (Table 16).

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 should be as detailed as to allow other researchers/doers to replicate exactly was has been done. If an 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 appendixes.

The same is true for the documentation of the sources. All sources (e.g., medical literature, data banks, expert 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 nonspecialists understand the terms being used. Such a glossary is strongly recommended when the issues under study are highly specialized. 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 that 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 should be transparently documented in the full HTA report. A statement on conflict of interest should answer the following questions:

- 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 that 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, because possible distrust and/or perceived bias is an important barrier for the credibility of studies (29).

The way of organizing 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 (Table 18).

Table 18. Structure Example for an HTA Technical Report

- Title
- Authors
- Statement on conflict of interest
- Policy question
  Who commissioned the assessment? Why? What decision(s) is it supporting?
- Methodology of the HTA report
  HTA Protocol
  Review process
  Sources of data[a]
  Appraisal of data/studies (inclusion/exclusion criteria)[a]
  Method of synthesis[a]
- Background information
  Target condition, target group, outcomes of interest, technology aspects
- Research questions
- Results[b]
  Safety
  Efficacy/effectiveness
  Psychological/social/ethical considerations
  Organizational/professional implications
  Economic issues
- Discussion
  Methodology of the assessment
  Quality of evidence/types of evidence (studies/data)[a]
  Uncertainties/lack of information[a]
  Generalizability, applicability of findings[a]
- Conclusions
- Recommendations
- Appendixes[c]
  Documentation of sources (search protocols, keywords 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

[a]For each of the aspects of the assessment.
[b]Results can be presented with the help of tables and graphics.
[c]Information contained in appendixes 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.

CONCLUSIONS

The members of Working Group 4 have reached 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 and an instrument for assessing the quality of reports, rather than prescribing a rigid methodology. Particular emphasis should be given to the reporting of findings to enhance comparability and allow for a better cross-border dissemination of results.

During its work, the working group identified several methodologic gaps and needs:

- Considerable work has been done on isolated methodologic aspects relevant to HTA, but little is done on how to apply the individual methodologic tool kits when conducting HTA. Only a few of the identified documents provided methodologic guidelines for carrying out HTA; most of the reports focused on specific issues.

- Transparency of the entire HTA process should be achieved, which is warranted by clear reporting and explanations of all steps undertaken in the assessment. To date, transparency has been concentrated on the evaluation of efficacy/effectiveness or in economic evaluations, while other important aspects of HTA have not been handled 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 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. Methodologic guidance concentrates mostly on such aspects, distracting from a balanced approach to all aspects. However, with the 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, since they will be able to use this work.
- Currently, no methodology is available 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, since the main function of HTA is to provide sound evidence on effectiveness, taking system-, provider-, and patient-orientated issues into account. The identified gap might possibly be dealt with through methodologic advancement of modeling techniques.
- Some work is being done to develop systematic reviews of diagnostic, preventive, community-based, and health system-related interventions; however, the methodologic debate is still open.
- Important issues of an assessment, such as the review process or update process, are being conducted in different ways, but further evaluation of different alternatives is needed to identify what could be best practice.

Table 19. Proposal for a Checklist/Criteria To Assess the Quality of HTA Reports

Criterion	Questions
Basic information	Are the authors of the report stated? Is any possible conflict(s) of interest stated? Is there any information about who financed the report? Was the report externally reviewed?
General methodologic aspects of the assessment	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 the information for the different aspects was 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?
Description of the context of the assessment	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 on who has commissioned the HTA?

(*continued*)

Table 19. (*Continued*)

Criterion	Questions
Background information	Were conditions, target group, relevant interventions or comparisons between interventions and relevant outcomes appropriately defined?
Data about the status quo of the technology	Are patterns of utilization, 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)?
Technical description of the technology	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)?
Safety	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?
Efficacy/effectiveness	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?
Psychological, social, and ethical considerations	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?
Organizational and professional implications	Were organizational and regulatory issues discussed (e.g., responsibility, necessary investments, financing, regulation, personnel, need, demand)?
	Are the methods used for assessing these aspects stated?
Economic evaluation	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?
Discussion of generalizability/ applicability of the findings	Are aspects of the generalizability 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 healthcare delivery, reimbursement, access)?

- No appraisal tool exists to assess the quality of HTA reports. The working group therefore proposes such an instrument (Table 19).

RECOMMENDATIONS

- While some of the methodologic gaps identified in the Conclusions 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 unavailability of structured reports and the language barrier), the use of a scientific summary report, as described in this paper, should be viewed as a sign of best practice in reporting HTA. For all assessments conducted within Europe, such a scientific summary report should be available in the working languages of the EU, and at the least 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. To promote the use of such a summary, its use could be a requisite for reports of assessments that receive EU funding.

NOTES

[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.

[2] Methods to statistically assess the learning curve have been gathered and evaluated by Ramsay et al. (50).

[3] *Own primary research* refers here to primary research conducted within the assessment 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 approach may not be possible at all, since disciplines may be involved that for example, do not have databases such as those of the medical literature.

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

[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.

[8] A comprehensive review on quantitative synthesis methods is found in the study by Sutton et al. (57). An up-to-date review of the methods of meta-analysis of binary and continuous results is available in the study Egger et al. (17).

[9] Currently known as the number needed to harm (NNH).

[10] Systematic reviews of trials and other studies (Sutton et al. 57); undertaking systematic reviews of research on effectiveness (36).

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

[12] 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.

[13] 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 methodologic checklists for qualitative research (accessible at http://www.salford.ac.uk/iphrp/cochrane/homepage.htm).

[14] Grant MJ. Searching for qualitative research studies on the MEDLINE database. Presented at the Qualitative Evidence Based Practice Conference, Coventry, May 14–16, 2001.

[15] See Appendix 2 for more examples.

[16] A comprehensive review of qualitative methods is found in Murphy et al. (46). Some of these methods are also described in Kristensen et al. (37).

[17] A review on methods for assessing public preferences is included in Ryan et al. (53).

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

[19] Scales for gradation of recommendations related to levels of evidence and quality of data (internal validity) are given in Appendix 6.

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

REFERENCES

1. Baladi J-F. *A guidance document for the costing process*. Ottawa: 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; 1988:53-67.
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, 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*. Ottawa: CCOHTA; 1997.
8. Clarke M, Oxman AD, eds. *Cochrane reviewers handbook 4.1.1*. 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 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*. 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, et al. *Guide d'analyse de la littérature et gradation des recommendations*. Paris: 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. Report from the EUR-ASSESS Project. *Int J Technol Assess Health Care.* 1997;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*. Luxembourg: Office for Official Publications of the European Communities, 1999.
20. Flynn KL, Adams EJ. *Assessing diagnostic technologies*. Boston: US Dept of Veterans Affairs; 1996. Technology Assessment Program Report No. 1.
21. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis 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, et al., eds. *Cost-effectiveness in health and medicine*. 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. Greenlagh 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, et al. 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;52:179-192.
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 Evaluacion de Tecnologias Sanitarias (AETS) Informe de Evaluacion de Tecnologias 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, Macaskill P, Glasziou P, Fahey M. Meta-analytic methods for diagnostic test accuracy. *J Clin Epidemiol*. 1995;48:119-130.
33. Irwig L, Tosteson AN, Gatsonis C, et al. Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med*. 1994;120:667-676.
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*. London: NHS Center for Reviews and Dissemination (CRD); 2000. Report no. 4.
37. Kristensen FB, Hørder M, Poulsen PB, eds. *Health technology assessment handbook*. 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;320:50-52.
41. Mears R, Taylor R, Littlejohns P, Dillon A. *Review of international health technology assessment (IHTA): Project report*. London: National Institute for Clinical Excellence; 2000.
42. Moher D, Cook DJ, Eastwood S, et al., for the QUOROM Group. Improving the quality of meta-analyses of randomised controlled trials: The QUOROM statement. *Lancet*. 1999b;354:1896-1900.
43. Moher D, Cook DJ, Jadad AR, et al. Assessing the quality of reports of randomised trials: Implications for the conduct of meta-analyses. *Health Technol Assess*. 1999a;3:1-90.
44. Morgan D, ed. *Successful focus groups*: *Advancing the state of the art*. Newbury Park, Calif: Sage Publications Inc; 1993.
45. MSAC. *Funding for new medical technologies and procedures*: *Application and assessment guidelines*. Medicare Services Advisory Committee, 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, et al. 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. *Public health and preventive medicine*. Norwalk, CT: Appleton-Century-Crofts; 1980:1800-1823.
55. Sloan FA, ed. *Valuing health care*: *Costs*, *benefits*, *and effectiveness of pharmaceuticals and other medical technologies*. New York: 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; 1984:41-56.
59. U.S. Congress, Office of Technology Assessment. *Assessing efficacy and safety of medical technologies*. Washington, DC: US Government Printing Office, 1978. Publication no. OTA-H-75.
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;1:2-8.
61. Williamson JW. The estimation of achievable health care benefit. In: Williamson JW, ed. *Assessing and improving health outcomes*. Cambridge, Mass: Ballinger Publishers; 1978:51-69.

APPENDIX 1

Toolkits and Methodologic Guidance Documents

Table A1-1. Available Toolkits for HTA that Refer to the Whole Assessment Process

Reference	Source	Language	Comments
Burls et al. (5)	http://www.bham.ac.uk/WMidsDES/	English	Description and methodologic guidance of all steps undertaken when performing an assessment for the DES. Provides comprehensive guidance on how to elaborate on background information and research questions, on how to report appraisal and selection of the data, and on how to summarize the evidence found in a nonquantitative way.

(*continued*)

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

Table A1-1. (*Continued*)

Reference	Source	Language	Comments
DEC (11)		English	Description of the process of assessment for the DEC (rapid HTA), with special focus on the cost aspects.
Imaz-Iglesia et al. (30)	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 et al. (37)	http://147.29.115.214/ publikationer/docs/ Metodehaandbog/ Methodology Handbook 180601.pdf or via http://www.dihta.dk	English	Provides an overview of qualitative research methods, measurement of quality of life, methods to address the organizational aspects, and economic evaluation methods that can be applied in HTA.
MSAC (45)	http://www.health.gov. au/haf/msac	English	Description of the assessment process and elaboration of HTA reports.

Table A1-2. Methodologic Toolkits on Specific Topics

Reference	Source	Language	Comments
Baladi (1)	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 (7)	http://www.ccohta.ca/ newweb/pubapp/ pdf/peg_e.pdf	English (French)	Focuses on the economic evaluation of drugs, also giving guidelines for reporting economic analyses.
Clarke et al. (8)	http://www.cochrane.org	English	Comprehensive methodologic guidance on how to conduct systematic reviews and meta-analyses of RCTs of therapeutic interventions.
Durocher et al. (15)	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 et al. (17)		English	Comprehensive and updated review of methods for meta-analyses of binary and continuous results.

(*continued*)

Table A1-2. (*Continued*)

Reference	Source	Language	Comments
Flynn and Adams (20)	http://www.va.gov/resdev/ps/pshsrd/mdrc.htm# HealthCareTechnology Assessment	English	Provides methodologic guidance on how to conduct systematic reviews on accuracy of diagnostic tests.
Harris et al. (28)	Via http://www.ahcpr.gov/clinic/ajpm.htm	English	Detailed description 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 et al. (36)[a]	http://www.york.ac.uk/inst/crd	English	Comprehensive methodologic 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).[b]	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).[b]	http://www.hta.nhsweb.nhs.uk/fullmono/mon422.pdf	English	
Murphy et al. (46)[b]	http://www.hta.nhsweb.nhs.uk/fullmono/mon216.pdf	English	Comprehensive review of qualitative research methods applicable in HTA .
Ramsay et al. (50)[b]	http://www.hta.nhsweb.nhs.uk/fullmono/mon512.pdf	English	

(*continued*)

Table A1-2. (*Continued*)

Reference	Source	Language	Comments
Ryan et al. (53)[b]	http://www.hta.nhsweb.nhs.uk/fullmono/mon505.pdf	English	Review of methods to include the public preferences perspective on HTA.
Sutton et al. (57)[b]	http://www.hta.nhsweb.nhs.uk/fullmono/mon219.pdf	English	Comprehensive methodologic guidance on conducting systematic reviews of literature. Presents a comprehensive overview of different meta-analytic approaches.

[a]Other methodologic 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.
[b]Besides the documents listed here, the Health Technology Assessment Series of the NHS includes further methodologic reviews on more specific topics concerning HTA. A complete list of them is available at http://www.hta.nhsweb.nhs.uk/htapubs.htm.

APPENDIX 2

Sources of Information

The following tables present a selection of sources of information and literature. The tables were elaborated with information obtained from the Handbooks of AETS, DES, DIHTA, and our 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 also may be available in CD-ROM or online through database 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 that you consult a 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 that are 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/htahp.htm	Abstracts of publications and projects from INAHTA members and other organizations.
ISTAHC Database	http://www.istahc.org/en/database.html	Includes abstracts, journal citations, meeting programs, postconference courses, and articles related to health technology assessment.

(*continued*)

Table A2-1. (*Continued*)

Name of the source	Available at:	Comments
Cochrane Database of Systematic Reviews	http://www.update-software.com/ccweb/cochrane/cdsr.htm	Systematic reviews were 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 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 HTA and the development and use of clinical practice guidelines (will be replaced by NLM Gateway later in 2001).

Table A2-2. Bibliographic Source

Name of the source	Available at:	Comments
General		
MEDLINE	Usually available at university libraries or through the Internet: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi	Covers the whole field of medical information, including dentistry and medical psychology. If using optimized 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 that are currently available at AIDSLINE, BIOETHICSLINE, and other databases not relevant to HTA.

(*continued*)

Table A2-2. (*Continued*)

Name of the source	Available at:	Comments
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 nonclinical (healthcare administration, economics, planning, and policy) aspects of healthcare 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/index. html#sdb	Provides access to bibliographic information, author abstracts, and cited references found in technical and science journals.
Specific		
AIDSLINE	Currently accessible through GratefulMed:http://igm.nlm.nih.gov	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 healthcare 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.

(*continued*)

Table A2-2. (*Continued*)

Name of the source	Available at:	Comments
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 healthcare 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 evaluation studies of healthcare 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.
Gray literature/ongoing research		
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.
Registries of trials and other ongoing research		
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
NNR 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[a]

Name of the source	Available at:	Comments
World Health Organization (WHO)	http://www.who.org	Access to multiple health statistics.
Food and Drug Administration (FDA)	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.
Community Research and Development Information Service (CORDIS)	http://www.cordis.lu	Information about research and development activities within the EU.
European Union Statistics Office (EUROSTAT)	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 epidemiologic information on European countries.

[a]The sources cited here aim at providing a general idea of sources other than the literature. Statistical agencies, ministries, epidemiologic registers, manufacturers, and professional, consumer, 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 undertaking an assessment.

APPENDIX 3

Search Filters

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

		Search filters provided for:		
Source	Available at:	Database	Software	Topics
University of Rochester, USA	http://www.urmc.rochester.edu/Miner/Educ/Expertsearch.html	MEDLINE CINAHL	Ovid	Diagnostic devices, etiology, 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	Etiology, 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

APPENDIX 4

Appraisal Checklists

This section presents a selection of checklists for appraisal of the medical literature. More checklists and appraisal tools have been developed by other authors and also by HTA institutions. Thus, this is not a comprehensive collection but an example. Except for 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 checklists: http://www.cche.net/principles/content_all.asp.

Box A4-1. Checklist for an Article About Therapy (27)

I. Are the results of the study valid?

Primary guides:

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

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?
- 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 counseling patients?

Box A4-5. Checklist for a review article (48)

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 and 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 and 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 organizational 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 and 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 Utilization Review (Naylor and 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 judgments 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 trade-offs 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 (22;23)

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 (13;47)

I. Are the results of the study valid?

- Did the analysis provide a full economic comparison of healthcare 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 (Listed by year of Publication)

Oxman AD, Sackett DL, Guyatt GH. Users' guides to the medical literature. I: How to get started. Evidence-Based Medicine Working Group. *JAMA*. 1993;270:2093-2095.

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;270:2598-2601.

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;271:389-391.

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;271:703-707.

Levine M, Walter S, Lee H, et al. Users' guides to the medical literature, IV: How to use an article about harm. Evidence-Based Medicine Working Group. *JAMA*. 1994;271:1615-1619.

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;272:234-237.

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;272:1367-1371.

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;273:1292-1295.

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;273:1610-1613.

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;274:570-574.

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;274:1630-1632.

Guyatt GH, Sackett DL, Sinclair JC, et al. Users' guides to the medical literature, IX: A method for grading health care recommendations. Evidence-Based Medicine Working Group. *JAMA*. 1995;274:1800-1804.

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;275:554-558.

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;275:1435-1439.

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;277:1232-1237.

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;277:1552-1557.

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;277:1802-1806.

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;279:545-549.

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;281:1214-1219.

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;281:1836-1843.

Barratt A, Irwig L, Glasziou P, et al. Users' guides to the medical literature, XVII: How to use guidelines and recommendations about screening. Evidence-Based Medicine Working Group. *JAMA*. 1999;281:2029-2034.

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;282:771-778.

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;282:1371-1377.

Hunt DL, Jaeschke R, McKibbon KA. Users' guides to the medical literature, XXI: Using electronic health information resources in evidence-based practice. Evidence-Based Medicine Working Group. *JAMA*. 2000;283:1875-1879.

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;283:2829-2836.

McGinn TG, Guyatt GH, Wyer PC, et al. 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;284:357-362.

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;284:478-482.

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;284:869-875.

Guyatt GH, Haynes RB, Jaeschke RZ, et al. 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;284:1290-1296.

APPENDIX 5

Software for Data Synthesis

A selection of useful software for the synthesis of data is provided here. 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.fuberlin.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.univlyon1.fr/~mcu/easyma/	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.updatesoftware.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.

APPENDIX 6

Levels of Evidence and Grades of Recommendations

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

Level of evidence	Therapy/prevention, etiology/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 centers	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 center	Prospective cohort study with good follow-up****	Analysis based on clinically sensible costs or alternatives; systematic review(s) of the evidence; and including multiway 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 multiway 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		Nonconsecutive study, or without consistently applied reference standards	Nonconsecutive 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

(*continued*)

Table A6-1. (*Continued*)

Level of evidence	Therapy/prevention, etiology/harm	Prognosis	Diagnosis	Differential diagnosis/ symptom prevalence study	Economic and decision analyses
4	Case series (and poor quality cohort and case control studies[§§])	Case series (and poor quality prognostic cohort studies[***])	Case control study or nonindependent 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 = randomized 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 that 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 *S*pecificity is so high that a *P*ositive result rules *in* the diagnosis. An "Absolute SnNout" is a diagnostic finding whose *Se*nsitivity is so high that a *N*egative 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 shape 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 nonexposed individuals and/or failed to identify or appropriately control known confounders and/or failed to carry out a sufficiently long and complete followup 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 control subjects 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 favor 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, nonobjective 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 (e.g., 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 nonindependent 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 Research 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, 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

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

Table A6-4. Recommendation Grid and Standard Recommendation Language

Quality of evidence	Net benefit substantial	Moderate	Small	Zero/negative
Good	A	B	C	D
Fair	B	B	C	D
Poor	I			

Source: Harris et al. (28).
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).