



# eunethta

EUROPEAN NETWORK FOR HEALTH TECHNOLOGY ASSESSMENT

## **SECOND DRAFT GUIDELINE**

**Therapeutic medical devices**

**October 2015**

1  
2

The primary objective of the EUnetHTA methodology guidelines is to focus on methodological challenges that are encountered by HTA assessors while performing a relative effectiveness assessment.

The guideline represents a consolidated view of non-binding recommendations of EUnetHTA network members and is in no case an official opinion of the participating institutions or individuals.

The research leading to these results has received funding from the European Community's Seventh Framework Programme under grant agreement HEALTH-F3-2012-305694 (Project MEDTECHTA "Methods for Health Technology Assessment of Medical Devices: a European Perspective").

Further the ADVANCE-HTA project ("Advancing and strengthening the methodological tools and policies relating to the application and implementation of Health Technology Assessment") shared the results on medical device specific methodological guidelines of HTA agencies within the EU and also their draft manuscript regarding a taxonomy of medical devices submitted for publication.

Although we built on the work of MedtechHTA, the recommendations given in this EUnetHTA guideline are independent results. They have been derived from an iterative process of internal discussions within the guideline authors' team, and also reflect reviewers' feedback from scheduled internal (EUnetHTA) and external consultations (Stakeholder Advisory Group, Public).

3  
4  
5

**Copyright © EUnetHTA 2015. All Rights Reserved. No part of this document may be reproduced without an explicit acknowledgement of the source. Reproductions for commercial use need EUnetHTA's expressed consent. Translations have to be indicated to EUnetHTA.**

10  
11  
12

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22

This guideline “Therapeutic medical devices” has been developed by  
UMIT (University for Health Sciences, Medical Informatics and Technology), Austria

With assistance from draft group members from  
IQWiG (Institute for Quality and Efficiency in Health Care), Germany  
G-BA (Federal Joint Committee), Germany  
Osteba (Basque Office for Health Technology Assessment, Ministry for Health), Basque  
Country, Spain

The guideline was also reviewed and validated by a group of dedicated reviewers from  
IER – Slovenia  
VASPVT – Lithuania  
ZIN – Netherlands

EUnetHTA’s contact person for this guideline document is Dr. Petra Schnell-Inderst  
( [petra.schnell-inderst@umit.at](mailto:petra.schnell-inderst@umit.at) )

# 1 Table of contents

2		
3	Table of contents .....	3
4	Acronyms - Abbreviations .....	4
5	Summary and table with main recommendations .....	5
6	1. Introduction.....	8
7	1.1. Definitions of central terms and concepts.....	8
8	1.2. Problem statement .....	11
9	1.3. Objective(s) and scope of the guideline .....	12
10	1.4. Related EUnetHTA documents .....	12
11	1.5. Other related documents.....	13
12	1.6. Methods.....	14
13	2. Analysis and discussion of the methodological issue .....	15
14	2.1. Results from literature review .....	15
15	2.2. The role of logic models in the HTA context.....	16
16	2.3. Systematic Reviews of MD: Framing the research question .....	17
17	2.3.1. Defining the intervention .....	18
18	2.3.2. Identifying context and user dependency and other potential effect	
19	modifying factors .....	20
20	2.4. Where to find information? .....	21
21	2.4.1. Searching bibliographic databases .....	22
22	2.4.2. Searching clinical trial registers .....	23
23	2.4.3. Specific issues .....	23
24	2.4.4. Selection of relevant studies .....	24
25	2.5. What kind of information is required? .....	25
26	2.5.1. Randomised study designs and analysis.....	25
27	2.5.2. Non-randomised study designs and analysis.....	29
28	2.6. Tools for critical appraisals .....	29
29	2.7. Analysing and synthesising evidence.....	30
30	2.8. Reporting and interpreting .....	31
31	3. Conclusion and main recommendations.....	33
32	Annexe 1. Bibliography .....	35
33	Annexe 2. Documentation of literature search .....	40
34		

# 1 **Acronyms - Abbreviations**

- 2 ACROBAT-NRSI - A Cochrane Risk of Bias Assessment Tool for Non-Randomized
- 3 Studies of Interventions
- 4 AMSTAR – Assessing the Methodological Quality of Systematic Reviews
- 5 CONSORT – Consolidated Standards of Reporting Trials
- 6 EuroScan - International Information Network on New and Emerging Health Technologies
- 7 EPAR - European public assessment reports
- 8 EUnetHTA – European network for Health Technology Assessment
- 9 FDA – Food and Drug Administration
- 10 G-BA – Gemeinsamer Bundesausschuss (Federal Joint Committee)
- 11 GRADE – Grading of Recommendations, Assessment, Development and Evaluation
- 12 HTA – Health Technology Assessment
- 13 IQWiG – Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (Institute for
- 14 Quality and Efficiency in Health Care)
- 15 ISPOR – International Society for Pharmacoeconomics and Outcomes Research
- 16 JA – Joint Action
- 17 MAUDE – Manufacturer and User Facility Device Experience
- 18 MD – Medical Device
- 19 MTA – Multi-Technology Assessment
- 20 Osteba – Osasun teknologien Ebaluazioaren Zerbitzua (Basque Office for Health
- 21 Technology Assessment, Ministry for Health)
- 22 PBAC – Pharmaceutical Benefits Advisory Committee
- 23 PICO – Population, Intervention, Comparison, Outcome
- 24 PRISMA – Preferred Reporting Items for Systematic Reviews and Meta-Analyses
- 25 RCT – Randomised Controlled Trial
- 26 REA – Relative Effectiveness Assessment
- 27 RoB – Risk of Bias
- 28 STA – Single Technology Assessment
- 29 SuRE Info – Summarized Research in Information Retrieval for HTA
- 30 UMIT – University for Health Sciences, Medical Informatics and Technology

# 1 Summary and table with main recommendations

2 INTRODUCTION: Health technology assessment (HTA) of medical devices (MD) may  
3 present specific challenges as compared to an assessment of pharmaceuticals.  
4 OBJECTIVES: The aim of this guideline was to identify those areas where specific HTA  
5 methods may be required and propose best-practice solutions for these problems.  
6 METHODS: On the basis of a literature review, we identified issues and methods that are  
7 specific or particularly relevant for MD assessment, drawing on the results from an EU FP7  
8 project.

9 RESULTS: The vast majority of standard HTA methodology (e.g. in selecting evidence and  
10 evaluating its validity) also is applicable when assessing medical devices. For some  
11 reasons, however, specific attention is required when defining, describing and evaluating  
12 MD interventions. First, the application of therapeutic MDs implies often further procedures  
13 and steps that may vary and the MD itself can be composed of several components  
14 undergoing frequent incremental modification. This makes it difficult to judge whether two  
15 MD interventions are sufficiently similar to be considered as representing the same  
16 medical intervention and whether and which analyses of subgroups of intervention  
17 characteristics may be relevant. Furthermore the rapid development poses also a  
18 challenge to trial design. Secondly, treatment effects may strongly depend on the skills  
19 and experience of the MD user, may it be physician, patient, nurse or other healthcare  
20 professional, as well as on the infrastructure of the providing institutions. From the  
21 perspective of the HTA assessor both problems can be addressed by a more detailed  
22 analysis of the available evidence considering these factors. To deal with these issues  
23 appropriately clinical prior information about former versions of the intervention and  
24 professional judgment will often be needed. However, appropriate primary studies are the  
25 basis for more conclusive evaluation of clinical effectiveness of MD. Knowledge about  
26 study designs addressing MD specific challenges is also necessary to assess MD  
27 interventions and to give advice for future research.

28

<b>Recommendations</b>	<b>The recommendation is based on arguments presented in the following publications and / or parts of the guideline text</b>
<b>1<sup>st</sup> recommendation: Specifics of HTA of MD</b>  HTA of medical device interventions should be done without unnecessary modification of currently established methods for finding, selecting, analysing, synthesizing and interpreting evidence on clinical effectiveness. A need for specific methods	2 Introduction

<p>mainly derives from the incremental development of MDs and their user and context dependency, and some implications of the physical mode of action.</p>	
<p><b>2<sup>nd</sup> recommendation: Framing the research question</b></p> <p>The more complex nature of MD interventions need a more elaborated development of the research question.</p> <p>A logic model (e. g. analytical framework) may help in describing the components of the intervention and comparators, outcomes and effect-modifying factors such as individual and institutional learning.</p> <p>Try to use clinical prior information about properties of the intervention that might influence treatment effects.</p> <p>Provide the sources / evidence for this information.</p>	2.3
<p><b>3<sup>rd</sup> recommendation: Defining the intervention</b></p> <p>Explicitly state whether the focus of the HTA report is the evaluation of one particular MD product (single technology assessment, STA) or of all MDs that can be used for a certain treatment method (multi technology assessment, MTA).</p> <p>If the aim is to perform a MTA, the review should take a broad scope for the definition of the intervention.</p> <p>Try to identify</p> <ul style="list-style-type: none"> <li>• all MD interventions,</li> <li>• which technologies are used in combination or alternatively,</li> <li>• potentially important differences.</li> </ul> <p>Redefinition of the intervention may become necessary during the course of the assessment.</p>	2.3.1
<p><b>4<sup>th</sup> recommendation: Information retrieval</b></p> <p>For information retrieval search strategies may include both general search terms such as the generic name of the device type as well as specific devices (proprietary or brand names).</p> <p>If randomized controlled trial (RCT) data are not available or for developing the research question, literature search can be broadened to include all types of study design, including case series and even case reports.</p> <p>In addition to the search in bibliographical databases, information about the MD may also be retrieved from device registries, incident reporting databases and administrative</p>	2.4

databases.	
<p><b>5<sup>th</sup> recommendation: Information requirements for clinical effectiveness</b></p> <p>Although RCT are to be preferred in the assessment of effectiveness, HTA assessors should anticipate that such evidence is frequently lacking for MD interventions. Thus, no definite conclusions should be expected, especially when assessing the effectiveness of very new MD interventions.</p> <p>HTA assessors should also be familiar with special RCT designs that take into account the specifics of MD (e.g. expertise-based trials, tracker designs).</p>	2.5
<p><b>6<sup>th</sup> recommendation: Information requirements for long-term effects</b></p> <p>In case of an assessment of long-term safety, it is useful to include disease-specific or MD-specific registries of high quality and incident reporting databases.</p> <p>Registry analyses should be considered to assess long-term outcomes but should not be routinely used for the assessment of treatment effects due to their susceptibility to bias.</p>	2.5
<p><b>7<sup>th</sup> recommendation: User dependency and context factors</b></p> <p>If it is likely that there is an influence of institutional expertise, learning and infrastructure (e. g. level of care, volume of interventions, case mix) and individual proficiency or learning (e. g. physician, patient, caregiver) on treatment effects. take this into account in the assessment.</p> <p>User proficiency and healthcare setting may affect both, intervention and comparator.</p>	2.3.2, 2.7
<p><b>8<sup>th</sup> recommendation: Applicability of findings</b></p> <p>When interpreting the review’s findings consider the influence of health care settings, user proficiency, and incremental treatment modification.</p> <p>In addition, systematically check the applicability by an applicability checklist (see EUnetHTA’s guideline “Applicability of evidence in the context of a relative effectiveness assessment”).</p>	2.8

1

2

# 1. Introduction

## 1.1. Definitions of central terms and concepts

- **Medical device:** any instrument, apparatus, appliance, software, material or other article,  
i) which is intended by the manufacturer to be used for human beings for the purpose of diagnosis, prevention, monitoring, treatment, disease alleviation, handicap or injury compensation, investigation, replacement or modification of the anatomy or of a physiological process, or control of conception and  
ii) which does not achieve its principal intended action in or on the human body by pharmacological, immunological or metabolic means, but which may be assisted in its function by such means.

MD are classified by the European Union Medical Devices Directive into four classes (I, IIa, IIb, III) according to the risk associated with their use. Eighteen rules guide classification based on the degree of invasiveness, duration of use, anatomical location, and other criteria. Class IIb (medium–high risk) MD are, e.g., infant incubators and external defibrillators, class III (high risk) MD are, e.g., heparin-coated catheters and biological heart valves.

- **Therapeutic medical device:** medical device, whether used alone or in combination with other medical devices, to support, modify, replace or restore biological functions or structures with a view to treatment or alleviation of an illness, injury or handicap.
- **Taxonomy of medical devices:** A novel taxonomic model that follows the logic of HTA combines classification according to risk aspects with the distinction between diagnostic and therapeutic devices, the user group (patients versus professionals), and the fields of application. It aims at providing decision-makers with a tool for considering device characteristics across multiple dimensions.  
Source: Henschke 2015(1)
- **Performance:** any technical characteristics, any effects and any benefit of the device when used for the intended purpose and in accordance with the instructions of use.  
Source: amendments adopted by the European Parliament on the proposal for a regulation of the European Parliament and of the Council on medical devices and amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 (COM(2012)0542 – C7-0318/2012 – 2012/0266(COD))
- **Treatment effect:** The effect on the subjects' health status or well-being attributable only to a treatment or intervention. Note: Investigators seek to estimate the true effect of a treatment or intervention by calculating the difference between the outcome obtained in the experimental group and the control group.  
Source: HTAGlossary.net
- **Clinical effectiveness:** The benefit of using a technology, programme or intervention to address a specific problem under general or routine conditions, rather than under controlled conditions, for example, by a physician in a hospital or by a patient at home.  
Source: HTAGlossary.net

- 1 • **Relative effectiveness:** Relative effectiveness can be defined as the extent to  
2 which an intervention does more good than harm compared with one or more  
3 alternative interventions for achieving the desired results when provided under the  
4 usual circumstances of health-care practice.
- 5 • **Safety:** Substantive evidence of an absence of harm. The term is often misused  
6 when there is simply absence of evidence of harm.  
7 Source: Ioannidis 2004 (2)
- 8 • **Internal validity:** the extent to which the (treatment) difference observed in a trial is  
9 likely to reflect the 'true' effect within the trial (or in the trial population) by  
10 considering methodological criteria.
- 11 • **Applicability** (also known as external validity, generalisability, or transposability):  
12 the extent to which the effects observed in clinical studies are likely to reflect the  
13 expected results when a specific intervention is applied to the population of interest.
- 14 • **Effect(-measure) modification:** Variation in the selected effect measure for the  
15 factor under study across levels of another factor. In statistical terminology this is  
16 called interaction. Effect-measure modification results in heterogeneity (see below).  
17 An effect-modifier may modify different effect-measures (e. g. relative risks, risk  
18 difference) for the same factor (e. g. age) in different directions and may modify one  
19 measure, but not another (3, 4).  
20
- 21 • **Heterogeneity (5, 6):**
- 22 ○ **Clinical:** Variation in study population characteristics, coexisting conditions,  
23 cointerventions, and outcomes evaluated across studies included in  
24 systematic reviews or comparative effectiveness research that may influence  
25 or modify the magnitude of the intervention measure of effect.
- 26 ○ **Methodologic:** In the context of systematic reviews on effectiveness,  
27 among-study differences in estimated effect sizes for the intervention that  
28 can be attributed to variability and quality of study designs and analyses.
- 29 ○ **Statistical:** Variability in the observed treatment effects beyond what would  
30 be expected by random error. Statistical heterogeneity may signal the  
31 presence of clinical heterogeneity, methodological heterogeneity, or chance.
- 32
- 33 • **Prognostic factor:** A prognostic factor is a measurement that is associated with  
34 clinical outcome in the absence of therapy or with the application of a standard  
35 therapy that patients are likely to receive. It can be thought of as a measure of the  
36 natural history of the disease (7).
- 37 • **Confounding by indication:** Patients are selected for different therapies based on  
38 clinical indications. If these indications are also prognostic factors, the estimates of  
39 treatment efficacy become confounded by these factors. This phenomenon is often  
40 referred to as confounding by indication (3).

- 1 • **Contextual factors:** The phenomena of health care and health, are complex  
2 systems that are fundamentally context-dependent. Contextual factors with potential  
3 influence on health outcomes could be for example: national, state, local, and  
4 organizational policies, community norms and resources, health care system  
5 organization, payment and incentive systems, practice culture, history, and staffing,  
6 historical factors and recent events, the culture and motivations surrounding the use  
7 of MD, changes in these factors over time (8).
- 8 • **User dependency:** the extent to which the treatment result in a clinical study or in  
9 clinical practice is influenced by the skills or experience of the people involved in the  
10 treatment.
- 11 • **Learning effect/ curve:** improvements in the technical performance of a new  
12 technique over time (9).  
13
- 14 • **Incremental development:** Many device classes are developed in a step-wise  
15 process with frequent technology changes each of which present only minor  
16 modifications, resulting in short product life cycles.

## 1.2. Problem statement

Health technology assessment (HTA) is aiming to inform decisions on adoption to benefit catalogues, reimbursement, best practice, and on disinvestment of health technologies. Guidelines on HTA methods were predominantly developed in the context of the evaluation of pharmaceuticals. Additionally the focus of regulation for market access of MD in Europe is on safety and performance and not on the assessment of clinical effectiveness. Typically, the small and medium sized technology enterprises have limited resources for clinical investigations. This all leads to a situation with a scarce evidence base on clinical effectiveness at market access for many MD (10). Assessors will often be in the situation that they have to assess a technology on the basis of the existing data. However, there are differences in the mandates and competencies of agencies, and additional data generation may be requested for coverage and reimbursement. Nevertheless, the regulatory situation cannot be an argument to lower the level of clinical evidence for HTA and decision making. Rather, HTA agencies may have the mandate to request additional data (generation) for their purposes (11).

We have identified three major issues relevant for relative effectiveness assessment (REA) in which MD differ from drugs:

### 1. The short life-cycle, and rapid, and predominantly incremental development

MD are developed in a highly dynamic market environment. Product life cycles are usually shorter than 3 years (12, 13). Each of the frequent technology updates of a product may represent only minor modifications and competitor “me too” products enter the European market soon after the first comer (14).

The short time frame and regulatory landscape limit the performance of randomized controlled trials with sufficient sample size and follow-up. Results may already be outdated when finally available and a new model of a product may be introduced during the course of a trial.(15) In addition, the reference technology is also subject to modification (16). The need for new clinical studies for small modifications is unclear (17). Similarity of products and how to define it is not only an issue for successive modifications of a specific product but also for products of different manufacturers. The question of which devices can be grouped into one “class” (e.g., in terms of technical comparability) is important in health technology assessment for the choice of comparator in the evaluation of new technologies (18).

### 2. Stronger user dependency of the treatment effect and learning curves

High risk MD are often combined with surgical procedures or other interventions (19). In such more complex procedures, a MD is an essential part based on a specific theoretical and scientific concept. These interventions usually require specific skills and training. The context of the intervention such as user characteristics, institutional knowledge, facilities and ancillary care may substantially influence the effect of treatment and it might be difficult to separate the contribution of the individual factors. Learning curves of individual users, but also of institutions have to be taken into account (20).

### 3. Evaluation of long-term effects for high risk devices

1 A third issue, which is not completely unique to MD, but is relevant to many high risk  
2 devices (e. g. implants), is their long-term use. Long-term effectiveness measures as well  
3 as adverse events have to be followed up (21).  
4

### 5 **1.3. Objective(s) and scope of the guideline** 6

7 MD encompass a huge variety of heterogeneous products which hardly could be  
8 addressed in one guideline. This guideline will primarily focus on therapeutic devices  
9 which are associated with high safety risks (class IIb and III, according to European  
10 regulatory framework) and therefore are in special need for thorough evaluation. But many  
11 recommendations will also apply to other MD.

12 In principle HTA methods for finding, selecting, analyzing, synthesizing and interpreting  
13 evidence on clinical effectiveness are also applicable to MD. MD differ only in some  
14 aspects from drugs. This methodology guideline has the goal to address MD-specific  
15 issues in relative effectiveness assessment (REA) of MD. The guideline will support HTA  
16 assessors, systematic reviewers and decision makers in HTA agencies by providing  
17 systematic review methodology advice for evaluating the clinical effectiveness of  
18 therapeutic medical devices. The focus is on:

- 19 1. Aspects deriving from the incremental development of MD
- 20 2. The greater importance of context and user dependence in the evaluation of MD  
21 compared to drugs

22 We will address these points through all steps of a systematic review of relative  
23 effectiveness. Cost-effectiveness will not be discussed, nor will other non-clinical benefits  
24 and harms (e.g., system/ organisation benefits/harms) due to limited time resources.

25

### 26 **1.4. Related EUnetHTA documents** 27

- 28 • EUnetHTA guidelines:
  - 29 - Methodological guideline for REA: Endpoints used for relative effectiveness  
30 assessment of pharmaceuticals: Clinical endpoints
  - 31 - Methodological guideline for REA: Endpoints used for relative effectiveness  
32 assessment of pharmaceuticals: Composite endpoints
  - 33 - Methodological guideline for REA: Endpoints used in relative effectiveness  
34 assessment of pharmaceuticals: Surrogate endpoints
  - 35 - Methodological guideline for REA: Endpoints used in relative effectiveness  
36 assessment of pharmaceuticals: Safety
  - 37 - Methodological guideline for REA: Endpoints used for relative effectiveness  
38 assessment of pharmaceuticals: Health-related quality of life and utility  
39 measures
  - 40 - Methodological guideline for REA: Comparators & comparisons: Criteria for the  
41 choice of the most appropriate comparator(s). Summary of current policies and  
42 best practice recommendations

- 1 - Methodological guideline for REA: Comparators & comparisons: Direct and  
2 indirect comparison  
3 - Methodological guideline for REA: Levels of Evidence: Internal validity of  
4 randomized controlled trials  
5 - Methodological guideline for REA: Levels of Evidence: Applicability of evidence  
6 in the context of a relative effectiveness assessment of pharmaceuticals  
7 - Internal validity of non-randomised studies (NRS) on interventions  
8 - Process of information retrieval for systematic reviews and health technology  
9 assessments on clinical effectiveness  
10  
11 • EunetHTA Core Model® (<http://www.eunethta.eu/hta-core-model>)  
12

### 13 **1.5. Other related documents**

14

- 15 • IDEAL framework (22-25)  
16 • JCE series on complex interventions (26-34)  
17 • ADVANCE-HTA (1)  
18 • HTA guidelines with reference to MD (35-44)

19

## 1 1.6. Methods

2

3 Two European Framework Programme 7 (FP7) projects recently searched for official  
4 guidance documents on MD assessment on European ( “Advancing and strengthening the  
5 methodological tools and policies relating to the application and implementation of Health  
6 Technology Assessment”, ADVANCE-HTA, <http://www.advance-hta.eu/>) and international  
7 levels ("Methods for Health Technology Assessment of Medical Devices: a European  
8 Perspective", MedtechHTA, <http://www.medtechta.eu/>), covering the period until mid-2014.  
9 Relevant methodological guidance of the seven identified documents (35, 36, 41-44) is  
10 considered in this guideline.

11

12 This guideline also includes results of work package 3 “Comparative effectiveness of  
13 medical devices” of MedtechHTA. In this project a targeted literature search for any  
14 methodological guidance on comparative effectiveness research issues relevant to MD  
15 evaluation was performed. It consisted of an initial systematic search in selected journals  
16 and was amended by screening the reference lists of included publications and  
17 consultation of experts. Methodological publications addressing MD as well as  
18 methodological publications addressing topics identified as being relevant for MD, such as  
19 learning curve, operator characteristics, non-inferiority studies, post-market surveillance  
20 were included. We use the information base identified by MedtechHTA and also results on  
21 topics relevant for the scope of this guideline. Most important, two recent reviews on study  
22 designs for MD assessment by HAS and KNAW were added (38, 40).

23 Regarding methods for information retrieval we used the experience of the institutions in  
24 the guideline authors group: We used a preliminary version (status May 2015) of  
25 EUnetHTA’s guideline on information retrieval as a starting point.

26 The recommendations of this guideline have been derived from an iterative process of  
27 internal discussions within the guideline authors’ team, and also reflect reviewers’  
28 feedback from scheduled internal (EUnetHTA) and external consultations (Stakeholder  
29 Advisory Group, Public).

30 Chapter 2 of this guideline is structured according to the methodology section of the  
31 clinical effectiveness domain of EUnetHTA’s HTA Core Model®. Specific aspects for  
32 primary studies of therapeutic MD are described under the heading “What kind of  
33 information is required. Primary studies for therapeutic MD” in section 2.3.

## 1 **2. Analysis and discussion of the methodological issue**

### 2 **2.1. Results from literature review**

3 Methods for finding, selecting, analysing, synthesizing and interpreting evidence on clinical  
4 effectiveness recommended for conducting systematic reviews such as the Cochrane  
5 handbook (45) or the CRD handbook (46) are in principle applicable to all health  
6 technologies, including MDs for therapeutic purposes. Both guidance documents are also  
7 endorsed by the EUnetHTA HTA Core Model® (see methodology section of clinical  
8 effectiveness domain). The results of the targeted literature review of WP3 of MedtechHTA  
9 on methods for the evaluation of clinical effectiveness of therapeutic MD showed that  
10 existing tools are fully applicable. Nevertheless, additional guidance and recommendations  
11 specific or more prominent for MD could be derived for

- 12 - the framing of the research question,
- 13 - requirements for primary studies for MD,
- 14 - selection and analysis of primary studies on MD,
- 15 - evidence synthesis and interpretation of the review's results.

16 No specific tools or methods were identified regarding appraisal tools for validity of studies,  
17 meta-analysis or decision-analytical modelling. With respect to information retrieval the  
18 review did not identify specific issues (47). However we used the experience of the  
19 institutions in the guideline authors group to give some MD-specific advice for information  
20 retrieval (see 1.6. methods).

21 A need for specific methods or specific guidance for applying well-known methods to  
22 therapeutic MD mainly derives from the incremental development of MD (16, 48), user and  
23 context dependency of the intervention, and some implications of the physical mode of  
24 action (e.g., blinding) of MD (47, 49).

25 High risk MD interventions, in particular implants, comprise multiple components often  
26 changing with fast pace over time (i.e. different parts of the implant itself and techniques  
27 and procedures to apply them) and typical contextual factors such as expertise and  
28 learning of operators, institutions or patients interact with treatment effect on the causal  
29 pathway between intervention and outcomes. These are properties of complex  
30 interventions at least compared to drugs (22, 47).

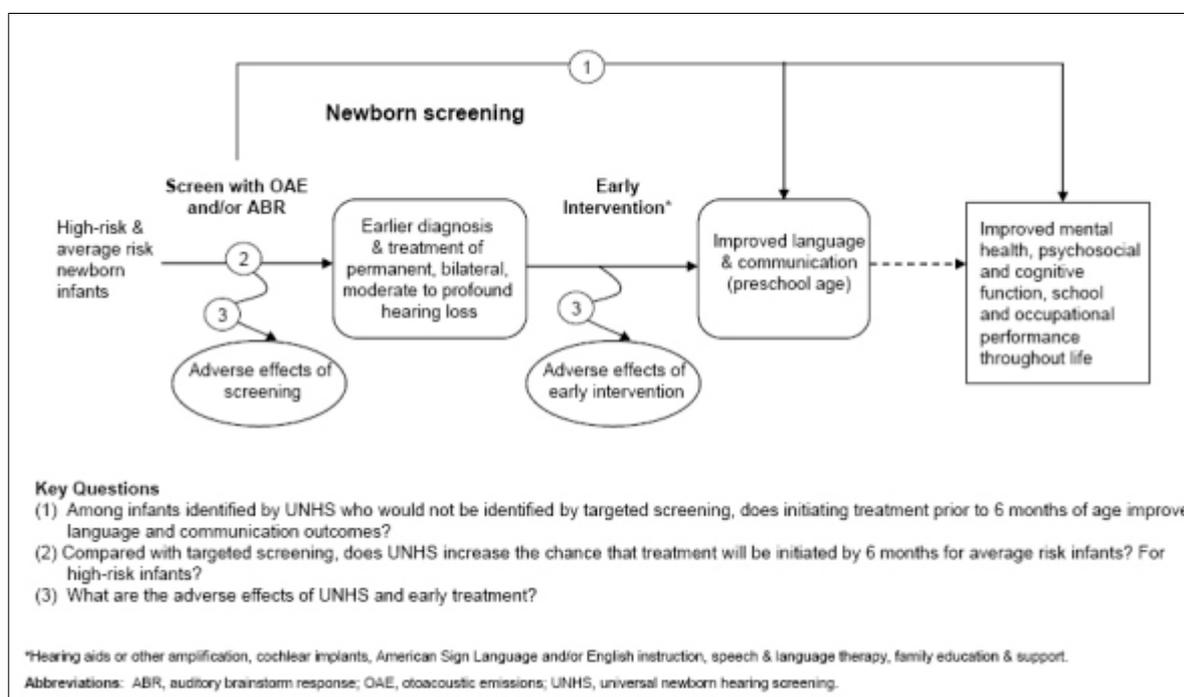
31 A framework for systematic reviews of complex interventions was published in 2013 by  
32 authors from Cochrane methods groups. It gives guidance for clarifying the review  
33 question, identifying what study types should be sought as evidence and which evidence  
34 synthesis methods can be chosen, and how applicability of findings can be assessed (26-  
35 34).

36 Although the authors mainly give examples from the field of health promotion showing a  
37 higher degree of complexity because further factors such as feedback loops between  
38 intervention and outcomes are involved that are usually not present in MD. However  
39 applying this framework as far as it is relevant to therapeutic MD may help to better take

1 into account the specifics of the evaluation of MD interventions in a systematic and  
2 transparent way.

### 3 2.2. The role of logic models in the HTA context

4 To get a better understanding of the intervention's components and the relation between  
5 intervention, modifying factors and outcome the use of logic models is recommended. "A  
6 logic model is a graphic description of a system and is designed to identify important  
7 elements and relationships within a system" (26). In the context of health care  
8 interventions logic models describe a theory of change that is how the intervention  
9 accomplishes beneficial and harmful changes in outcomes (27). A logic model can add to  
10 a common initial understanding in review production regarding evidence requirements  
11 before the evidence synthesis stage. It can be a tool to support a priori decisions regarding  
12 the proposed approach to evidence synthesis (27). The use of logic models is already a  
13 regular step in supporting the development of the analytical framework of systematic  
14 reviews on clinical effectiveness on public health interventions such as screening  
15 programs by the US Preventive Services Task Force (50). In the agency's reviews they  
16 are used to identify all relevant steps mediating beneficial and harmful treatment effects by  
17 screening, diagnostics and early intervention and related subcategories of research  
18 questions. With respect to MD a logic model could be used to clarify which elements  
19 belong to the intervention and to comparators and which contextual factors could  
20 potentially modify treatment effects. This can support defining search terms and inclusion  
21 and exclusion criteria for literature search and selection as well as better specifying which  
22 outcomes should be sought, which modifying factors should be extracted, and which  
23 subgroups of interventions or populations should be analyzed in evidence synthesis.



24

25 Figure 1: Example of a logic model: Analytical Framework for Universal Newborn Hearing Screening. From  
26 the US Preventive Services Task Force (51).

### 1 2.3. Systematic Reviews of MD: Framing the research question

2 The focus of a systematic review of clinical effectiveness should be determined by a well-  
3 formulated research question, because the research question will help to guide the further  
4 steps of the review process. General guidance for deriving a well-defined research  
5 question is provided in the Cochrane handbook (45). “Where possible the style should be  
6 of the form ‘To assess the effects of [*intervention or comparison*] for [*health problem*] in  
7 [*types of people, disease or problem and setting if specified*]’. This might be followed by  
8 one or more secondary objectives, for example relating to different participant groups,  
9 different comparisons of interventions or different outcome measures....The ‘clinical  
10 question’ should specify the types of population (participants), types of interventions (and  
11 comparisons), and the types of outcomes that are of interest. The acronym PICO  
12 (**P**articipants, **I**nterventions, **C**omparisons and **O**utcomes) helps to serve as a reminder of  
13 these” (45).

14 In addition to applying PICO to therapeutic MD, a logic model can address contextual  
15 factors that modify outcomes. Besides these - as for all other types of reviews of  
16 interventions - prognostic factors, co-therapies, etc. may also influence the outcomes of  
17 interest. Here the logic model can be used as an analytical tool to sum up and help to  
18 clarify relations between intervention, outcome and other factors.

19 This will also help to better explore the larger heterogeneity of treatment effects which has  
20 to be expected from the above mentioned greater number of effect-modifying factors.  
21 Because the number of studies is usually limited and only the influence of a small number  
22 of variables can be formally analyzed in subgroup analysis or meta-regression, thinking  
23 about pre-specification of these variables should start with framing the research question,  
24 to avoid spurious results by multiple testing. The information for the development of  
25 different parts of the logic model can stem from various sources, ranging from clinical  
26 studies with different designs to expert advice and patient opinion, for example information  
27 whether learning curves are relevant may be found in case series.

28 In Figure 2 we provide a template that can be used for framing the research question  
29 during subsequent steps.

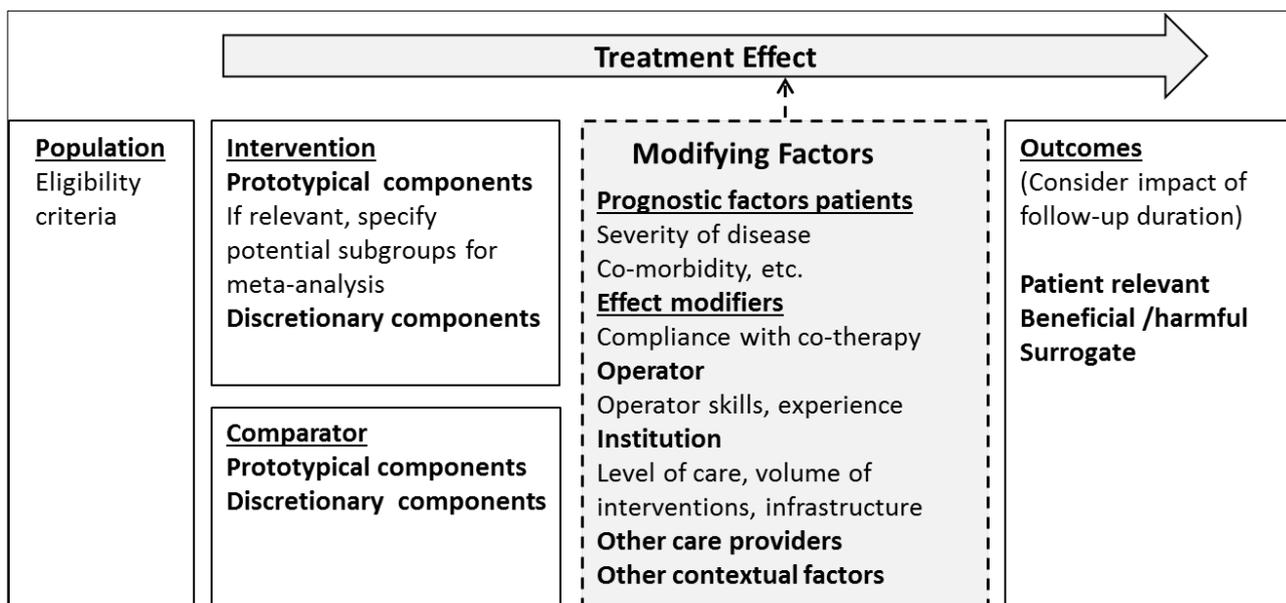


Figure 2. Template for a logic model for interventions with therapeutic medical devices (47). Prototypical components need to be present for the intervention to meet the working definition; discretionary components may be present but are not compulsory to meet the working definition (33).

### 2.3.1. Defining the intervention

Evaluation of MD can focus on only one particular product (STA), but in many cases HTA wants to investigate the clinical effectiveness of a whole group of MD that can be used for a certain treatment method or indication (MTA). This may encompass similar products from different manufacturers and different versions of one manufacturer's product. The way how the MD is applied, that is for example surgery, is also a part of the intervention (47). HTA authors should clearly define the aim of their evaluation and explicitly state whether their focus is the evaluation of one particular MD product or of all MDs that can be used for a certain treatment method. Example 1 in the box describes a case of STA and MTA.

#### Example 1: How to choose between device-specific and product class evaluation

Between 2007 and 2011, the French Agency HAS (Haute Autorité de Santé) assessed several total hip implants with metal-on-metal bearing surfaces, including the DePuy ASR XL Head system (52) and the Zimmer Metasul/Durom LDH (large diameter head) system (53). In each of these assessments, a product system of a single manufacturer was evaluated. For each individual product, quite different types of clinical evidence were available and used. If the manufacturer submitted unpublished studies, these were also accepted. Over time, some products were not accepted for reimbursement, because no or only very weak short-term clinical data was available. In some cases, data indicated potential wear of the component materials. Other products were accepted for coverage in France. Then in 2013, HAS reassessed all metal-on-metal hip implants. This step was triggered by worrying results observed in three national registries. The available data now indicated that wear and early implant failure was not a device-specific problem. The problem appeared to be present in all metal-on-metal hip implants, primarily in those

1 devices with large femoral head diameter ( $\geq 36$  mm). Therefore, it was appropriate to  
2 assess the whole class of products and to define new subgroups according to femoral  
3 head diameter (54). The 2013 report was based on a systematic review of randomized and  
4 non-randomized studies, and several orthopaedic surgeons participated in data  
5 interpretation (55). Finally, reimbursement of all large head metal-on-metal hip implants  
6 was discontinued.

7 This example shows that HTA of medical devices can be done on a device-specific or on a  
8 product class level. The criteria to define a class of devices may be adjusted or changed if  
9 data indicate this is necessary.

10

11 If an intervention that might be delivered with different MD or procedures has to be  
12 evaluated, the HTA report should use a broad approach for defining the intervention. For  
13 complex interventions it is recommended by Squires et al. to “use as broad an approach (i.  
14 e., lumping with subsequent explicit a priori subgroup analysis) as makes practical sense”  
15 and that “potentially important differences between the composition or intensity of the  
16 interventions in question should be specified in the review question”. For interventions  
17 involving the application of MD it often is not so obvious what can be classified as the  
18 same or similar intervention as for drugs.

19 Squires et al. recommend to “identify any prototypical and discretionary components of the  
20 intervention. Prototypical components are those that need to be present for the  
21 intervention to meet the working definition of the intervention. Discretionary components  
22 are those that may be present but do not need to be present to meet the working  
23 definition” (33). Assessors should try to identify which technologies are used together or  
24 alternatively. Redefinition of the intervention may become necessary during the course of  
25 the assessment.

26 When including variants of a MD or MD with similar functioning in the assessment, it  
27 should be made explicit whether the MD is assumed to be generic (‘genericisation’(18)),  
28 that is, that they belong to the same class of devices and are therefore in principal of  
29 comparable effectiveness. Assumptions about comparability of interventions are critical  
30 issues when decisions on the appropriateness of the combination of data are made. They  
31 should be clearly stated. Often it is also part of the research question to compare  
32 subgroups of the intervention. Furthermore, it often is useful also to search for, include and  
33 assess clinical studies that compared different variants of a MD intervention. Prior  
34 information from clinical studies on former versions of the technology or similar technology  
35 should be used to decide whether and which modifications of the MD could impact the  
36 clinical effect. This information can guide the definition of subgroups of intervention  
37 characteristics that may be relevant for the investigation of differences in treatment  
38 effects. The sources of this information used to decide about subgroups should be  
39 provided (47).

40 See an example how to discern between MD that are substantially equivalent or not for  
41 catheter ablation in patients with atrial fibrillation in “Example 2”.

### **Example 2: How to discern between medical devices that are substantially equivalent or not**

In 2012, the Belgian HTA body KCE (Belgian Health Care Knowledge Centre) published a report on catheter ablation in patients with atrial fibrillation (56). This procedure involves ablation of myocardial tissue in the left atrium by radiofrequency waves, freezing (cryoablation), high intensity focused ultrasound (HIFU) or laser beams. Several MD are (or were) available for this procedure on the European market. Although all devices have the same aim of destroying tissue through a catheter, it was concluded from the different types of energy sources that this large spectrum of devices is best assessed in subgroups. Accordingly, radiofrequency ablation was assessed separately from cryoablation, HIFU and laser ablation. This subgroup formation was based on medical knowledge, but turned out to be appropriate when safety and effectiveness data were examined. Randomised controlled trials had been completed only for those devices that used radiofrequency waves as energy source. Early-stopped randomised and non-randomised studies indicated that the other energy sources either were less effective or led to serious complications. However, there was also one radiofrequency device that was found to have a safety problem. The report concluded that catheter ablation should only be performed in rigorously selected patient subgroups. This example illustrates that expert knowledge, safety data and effectiveness results all can and should be used when deciding about subgroup analyses of MD.

1

### **2.3.2. Identifying context and user dependency and other potential effect modifying factors**

Exploring effect modifiers and critical factors for implementation may enhance the value of a review of clinical effectiveness for users (33). If heterogeneity within and between studies can be explained by effect modification, these factors should be considered in clinical practice.

For interventions with therapeutic MD implying surgery or other procedures individual and institutional expertise (including infrastructure) and learning effects/curve has to be taken into account as potential effect modifying factors (see definitions).

Cook et al. provide a framework that integrates the various factors that influence learning and highlight their hierarchical structure ((20), Figure 3). The first level is the clinical community that informs guidelines and protocols for the use of the MD. Next, the institution can adapt its organizational pathways and facilities to the new technology and thereby influence the surgeon's learning curve. Also the experience and type of people in the surgical team influence the MD performance. The case mix level reflects that more experienced surgeons may be likely to see more complicated cases that have poor outcomes which makes the performance appear becoming worse. Finally, the surgeon's abilities, attitudes and capacities determine his or her learning curve.

20

21

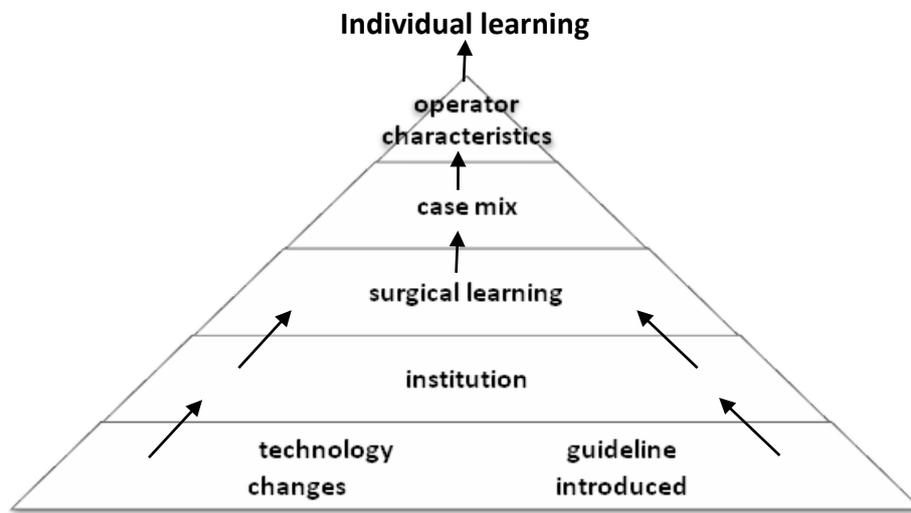


Figure 3: Hierarchical factors that influence learning (source: Cook 2004(20))

In practice, a three-tiered approach may be helpful to analyse user-dependency:

1. Screen whether a data analysis of an association between user proficiency (e.g. more or less skilled surgeons, patients with or without training) or healthcare setting (e.g. hospitals or study centers) has been done and reported **within included studies** or how studies tried to handle a possible effect modification by these factors (e. g. run-in periods). Particularly, pragmatic RCT may be considered.
2. If learning curves are not reported in RCT and no information can be retrieved by contacting the authors search and include additional non-randomized or even non-comparative evidence (e.g. administrative database analyses) in order to analyse the association between user proficiency or healthcare setting and treatment results in more detail.
3. Consider the influence of user proficiency and healthcare setting as source of heterogeneity of the treatment effect **between studies**. Try also to explore influence of learning curves across studies in meta-analysis, e. g. by subgroup analysis, meta-regression or more sophisticated methods. These analyses could compare among studies conducted in centers with a high or low level of user proficiency (defined for example by strict or less strict eligibility criteria for study physicians or study centers)..

#### 2.4. Where to find information?

In contrast to the European Clinical Trials Database (EudraCT) for medicines, there is no trial registry imposed by the regulators for MD, there is no list of MD on the market, and there is no document like an European Public Assessment Report (EPAR), and there is no personnel actively involved in MD vigilance in most EU countries. This poses challenges to information retrieval for MD assessments.

For principles and details of the process of information retrieval and the different sources listed above please see the EUnetHTA Guideline “Process of information retrieval for systematic reviews and health technology assessments on clinical effectiveness”(57).. In

1 case of new and emerging MD one can also refer to the EuroScan toolkit  
2 (<http://euroscan.org.uk/methods/>).

3 The search for MD-related literature should not be too restrictive, e.g. searching brand  
4 names of devices only would be inadequate, since relevant studies with similar devices  
5 could be missed. Hence, available evidence also on comparable interventions which have  
6 the same technical core features and a similar patient population should be included in the  
7 search terms.

8

### 9 **2.4.1. Searching bibliographic databases**

10 The two approaches of searching generic electronic databases, using controlled terms  
11 (such as MeSH in MEDLINE and Cochrane Library databases or Emtree in EMBASE) or  
12 using text words (including synonyms, abbreviations, and acronyms) should be combined  
13 to account for the variability of the many different ways a certain technology can be  
14 indexed in a database.

15 Search strategies may include both general search terms (such as the generic name of the  
16 device type, for example 'transcatheter aortic valve replacement') as well as specific  
17 devices (proprietary or brand names). The development of search strategies can be  
18 challenging and may involve several iterations to reach a strategy that captures the  
19 complex way, records may present concepts of a device-related procedure and the target  
20 condition (<http://vortal.htai.org/?q=node/339>). The search syntax needs to be adapted to  
21 specific databases. Additional features of a device may be included in the search strategy.  
22 These may include product codes (Global Medical Device Nomenclature), authorization  
23 holder, risk class, mechanism of action, invasiveness of approach (e.g. percutaneous,  
24 vascular, endoscopic), technical platform / additional equipment if required. In general,  
25 careful selection of search terms is crucial, since there is no public availability of an EPAR  
26 or similar information on specific devices.

27 Regarding the structuring of the search strategy along the PICO scheme, careful selection  
28 of search terms for the relevant patient population is of particular importance since (often  
29 different from drugs) the same devices may be used for several patient populations and  
30 indications. The intervention is not necessarily identical to the device, especially when a  
31 procedure to apply the MD is involved not just the MD. Comparator may be another  
32 procedure, a drug or quite often a sham device or procedure. Database filters may be  
33 used to restrict the search to certain study types.

34 In order to avoid language bias, no language restrictions should be applied to the search  
35 strategy. Date restrictions should be applied only if it is known that relevant studies could  
36 have been reported during a specific time period, for example if the device or procedure  
37 was only available after a given time point (Cochrane Handbook Chapter 6 (45),  
38 EUnetHTA (57)).

1 The literature search should be documented in a transparent way. As a minimum  
2 requirement, the databases included in the search, the respective search strategy, date of  
3 search, number of hits and applied limits / filters should be documented.

4

#### 5 **2.4.2. Searching clinical trial registers**

6 A search of publicly available studies in clinical trials registries (i.e.  
7 <http://apps.who.int/trialsearch/> including [clinicaltrials.gov](http://clinicaltrials.gov) and EU clinical trials register)  
8 complements database searches in order to ensure that ongoing, terminated or completed  
9 but not yet published trials are identified and eventually to identify study results. Study  
10 registries may provide valuable insights into the status of the development of innovative  
11 MD and procedures including relevant target diseases that are addressed. For more  
12 registries also check "Health Technology Assessment on the Net: 2014"  
13 ([http://www.ihe.ca/index.php?/publications/health-technology-assessment-on-the-net-](http://www.ihe.ca/index.php?/publications/health-technology-assessment-on-the-net-2014)  
14 [2014](http://www.nlm.nih.gov/nichsr/hta101/ta10109.html)) or "HTA 101" (<http://www.nlm.nih.gov/nichsr/hta101/ta10109.html>).

15

#### 16 **2.4.3. Specific issues**

##### 17 ***Further information sources***

18 In order to obtain full information on unpublished studies or unpublished data from  
19 published studies it is necessary to search further sources. It is known that many trials on  
20 high risk MD remain unpublished or publications give incomplete information (58). Thus,  
21 reporting bias should be considered. Besides searching in bibliographic databases and  
22 clinical trial registries additional information sources include documents from regulatory  
23 bodies, unpublished company documents, and other options like queries to authors or  
24 conference abstracts. Also experts and patients may be a valuable source of information.

##### 25 *Studies on efficacy and effectiveness*

26 Since the regulatory approval process for MD in Europe does not necessarily require  
27 conducting RCTs, literature search can be broadened to include the best available  
28 evidence. In absence of RCT, this may be all types of study designs, including case series  
29 and even case reports. In many cases, this applies also for devices licensed under the  
30 510(k) pathway in the USA (see Safety Guideline, chapter 2.3.1.2). It might be helpful to  
31 search the FDA website for clinical trial information of devices previously licensed for the  
32 US market, including data on conditional marketing approval and its status. Medical device  
33 databases at FDA are freely accessible  
34 (<http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/Databases/default.htm>)  
35 and searches can be performed separately for premarket approval, humanitarian device  
36 exemption, 510(k) and other legal procedures. Most of these databases are updated  
37 weekly. In addition it might be advisable to additionally search [Clinicaltrials.gov](http://Clinicaltrials.gov) for ongoing  
38 studies on medical devices using a combination of device and company-specific search  
39 terms, as trial registration is mandatory for FDA approval of medical devices.

1 Other regulatory agencies can be checked for information on regulatory status and  
2 marketing approval such as: PBAC (<http://www.pbs.gov.au>) or Health Canada  
3 (<http://www.hc-sc.gc.ca>). Horizon scanning or Early Awareness and alert programs and  
4 databases could be checked also for additional information (see Health Technology on the  
5 Net ([http://www.ihe.ca/index.php?/publications/health-technology-assessment-on-the-net-](http://www.ihe.ca/index.php?/publications/health-technology-assessment-on-the-net-2015)  
6 2015) and contact EuroScan (<http://euroscan.org.uk/>) members and active programs).

#### 7 *Data on safety*

8 In addition to the search in bibliographical databases, orienting/initial information on safety  
9 may also be retrieved from device registries, incident reporting databases (e.g. US  
10 Manufacturer and User Facility Device Experience Database [MAUDE]) and administrative  
11 databases. For details refer to the EUnetHTA Safety Guideline (chapter 2.3.5) and to the  
12 SuRE Info on the HTAi webpage (<http://vortal.htai.org/?q=node/50>).

#### 13 *User or setting dependency*

14 Retrieving data on user or setting dependency (e.g. learning curves) may require several  
15 approaches. One approach refers to the analysis of eligibility criteria of institutions and  
16 operators in clinical trials regarding their level of expertise and experience. If properly  
17 described, these criteria may provide the level of qualification needed to successfully use a  
18 MD. There may also be dedicated studies on user dependency and contextual factors  
19 available (infrastructure required, architectural requirements, etc.). These should be  
20 searched for and included using a combination of device-specific search terms, controlled  
21 vocabulary if available (e.g. MeSH “learning curve” and “clinical competence” in MEDLINE)  
22 and text words (e.g. “learning”, “learning curve”, “training”, “minimal experience”,  
23 “experience curve”, “experience effects”, “qualification”). Restrictions for study design  
24 should not be applied since many studies are low level evidence (mostly case series).

25

#### 26 **2.4.4. Selection of relevant studies**

27 The selection of relevant studies from the resulting pool of studies follows the items that  
28 have been set out in the Preferred Reporting Items for Systematic Reviews and Meta-  
29 Analyses PRISMA statement and flow diagram (59). Inclusion and exclusion criteria of  
30 individual studies should be stated in advance.

31 After removing duplicate records of the same report (e.g. by merging search results in a  
32 reference management software), the first step is to remove obviously irrelevant reports by  
33 examining title and abstract of the records. This should be done by two reviewers on the  
34 basis of pre-specified exclusion criteria. At this stage, reviewers can be over-inclusive  
35 (sensitive) in order not to miss potentially relevant entries. In other words, preferably  
36 exclusion criteria should be applied.

37

## 1 **2.5. What kind of information is required?**

2 In the assessment of MD, different information for different purposes is required. Framing  
3 the research question and filling all the parts of the logic model, for example, need prior  
4 information that may be different from the information needed for the effectiveness  
5 assessment. For the former, a broad range of sources should be consulted and used,  
6 including grey literature, expert advice, and patient opinion and data from various study  
7 designs, while the latter task requires more robust evidence.

8 According to the developmental stage of a technology, starting from the idea to the  
9 development and exploration, further to assessment and finally to the long-term study (22,  
10 25), one should consider the different aims and appropriate study designs for each stage.  
11 The recommended study design for the evaluation of comparative effectiveness of a MD is  
12 a RCT. Challenges may arise for performing the classical large double-blind RCT with  
13 respect to blinding, placebo arms, preferences of patients and investigators, rapid  
14 development of the device, and contextual factors (47).

15 Since the regulatory approval process for MD in Europe does not necessarily require  
16 conducting RCTs such evidence is frequently not available, therefore observational studies  
17 including case series and even case reports may be the only evidence available. Due to  
18 risk of bias such evidence usually does not allow drawing definite conclusions on  
19 treatment effects.

### 20 **2.5.1. Randomised study designs and analysis**

21 A comprehensive overview of available randomized and non-randomized study designs  
22 and their advantages and disadvantages with respect to MD assessment is given by  
23 Bernard 2014 (60), HAS 2013 (38), and KNAW 2014 (40).KNAW also gives examples of  
24 real MD studies for most designs. Table 1 lists study design description, advantages and  
25 disadvantages of different study designs that may help to address challenges more  
26 prominent in MD interventions.

27 To deal with incremental development in addition to studies comparing new vs standard  
28 intervention, also those studies that compare different variants of the new intervention  
29 should be included and assessed.

30 For the study design of RCT for therapeutic MD additional to the rapid incremental  
31 development of MD, the influence of contextual factors and user proficiency on treatment  
32 effects some further challenges have to be considered:

- 33 • The physical mode of action and the surgical procedures associated with MD  
34 implantation may be an obstacle for (double-) blinding.
- 35 • Preferences of patients or surgeons for one or the other compared treatment may  
36 hamper trial recruitment or lead to high rates of drop-out or cross-over.

37 While incremental development and preferences are addressed by different RCT study  
38 types, issues with blinding as well as contextual factors and learning can be handled within  
39 each study type.

## 1 **Rapid incremental development**

2 The fact that the MD under investigation may be subject to modification even during the  
3 course of the trial needs to be reflected (48). It has to be investigated whether these  
4 changes have the potential to make a difference to the device effectiveness or safety (16).

5 Tracker trials provide an opportunity to track device variations over time and allow  
6 comparisons at each stage. They are guided by flexible protocols for the collection of  
7 randomized data and require sophisticated interim analyses (61).

8 “Adaptive study designs allow for changes in sample size or randomization ratios  
9 throughout the trial as additional information about the performance of the device is  
10 gathered. They are in principle possible by frequentist and Bayesian approaches”(47).

11 For the analysis of MD trials, Bayesian methods are especially useful, because  
12 development in small steps with minor modifications makes it more plausible that former  
13 versions of the device might also be a source of information which should be taken into  
14 account as prior information (47, 62, 63).

## 15 **Preferences**

16 Patients as well as investigators may have strong preferences for one or the other  
17 compared treatment. This may be because medical management is conceived more  
18 convenient, or because the surgeon is particularly familiar with one method (22). These  
19 preferences are sources for patients’ crossing over between treatment arms. In case of  
20 superiority trials with intention-to-treat analysis this will dilute treatment effects (16, 47).  
21 Expertise-based randomization (64) and randomization of patients before obtaining  
22 consent (Zelen’s design (65, 66)) or groups available for randomisation also includes a  
23 preference group (Wennberg’s design (67)) are approaches for dealing with this problem.  
24 In the first case all physicians can perform their preferred treatment and therefore have an  
25 incentive to participate. In the latter case patients giving informed consent are certain to  
26 get the new device; patients who do not consent receive the usual treatment. In  
27 Wennberg’s design patients are randomised to a preference group (treatment of choice) or  
28 a randomisation group (new or control treatment). This reduces the number of patients that  
29 refuse to participate in trials either because they do not want to be randomised to placebo/  
30 standard treatment or have strong preferences.

## 31 **Blinding**

32 Generally, the comparator in a study can be no treatment, placebo, sham device treatment  
33 (blinding possible) or an active treatment and in principle, all these approaches are also  
34 possible for MD trials. A placebo can be any inactive fake treatment (no matter the route  
35 of administration) while a sham procedure in MD trials should mostly resemble the  
36 (invasive) experimental intervention, e.g., surgery, without achieving a treatment effect. A  
37 sham arm with blinded investigators and patients may not always be feasible or ethical. If  
38 blinding is impossible, at least blinded endpoint evaluation is recommended (16, 48, 60).

1 “In those situations, where all available evidence stems from open studies, objective  
2 outcomes (e.g. mortality, some morbidity endpoints) should have more influence on the  
3 HTA report’s conclusion than subjective outcomes (e.g. symptoms, quality of life). Large  
4 effect sizes (e.g. >25% relative difference), however, usually cannot be attributed only to  
5 information bias in an unblinded study” (68).

### 6 *Contextual factors and user proficiency*

7 The most general recommendation for clinical researchers is to collect and report the  
8 contextual factors in detail to allow for later interpretation and analysis of their influence on  
9 the MD intervention effectiveness (19, 20, 47, 48).

10 To reflect improvement in performance of individual users and institutions, learning curves  
11 should be taken into account (20). First the existence of learning effects needs to be  
12 assessed and then it should be quantified. For the identification of learning effects, all  
13 factors likely to have a learning curve effect need to be systematically collected. When  
14 there is evidence of a learning curve, in a trial one can either standardize baseline  
15 conditions or capture and evaluate variations to quantify learning effects. Approaches to  
16 deal with learning in an RCT are, for example, to standardize trainings given to the  
17 investigators or to define a certain level of expertise for investigators to assure that  
18 learning occurs outside the trial (15, 16). For the comparative effectiveness assessment,  
19 training and experience of operators would have to correspond to that of the personnel  
20 that will finally use the device (47, 48).

21 Existing approaches to describe and quantify learning curves can be applied for different  
22 data structure and at different levels of complexity (20, 69).

Table 1: Description of different experimental designs and adaptive methods that can address challenges prominent for the evaluation of medical devices (modified and adapted from Bernard 2014(60))

Technological changes	Provider preferences	Patient preferences	Design	Principle	Advantages	Disadvantages
		x	Zelen's design	Randomizing before requesting consent	<ul style="list-style-type: none"> <li>• Facilitates inclusion</li> </ul>	<ul style="list-style-type: none"> <li>• Selection bias possible</li> <li>• Loss of statistical power if many patients refuse treatment</li> <li>• Ethical problems</li> </ul>
		x	Wennberg's Design (67)	Randomizing to preference group (people can choose their treatment) or randomization group	<ul style="list-style-type: none"> <li>• Facilitates participation</li> </ul>	<ul style="list-style-type: none"> <li>• Blinding not possible</li> <li>• Statistical power low, when a high proportion of participants chooses the same treatment</li> <li>• Difficulty of knowing whether the observed difference is related to the expertise of the therapist</li> </ul>
	x		Expertise-based randomized trial	Randomizing patients to a specialized physician	<ul style="list-style-type: none"> <li>• Better acceptability</li> <li>• Reduces execution bias and protocol deviations</li> </ul>	<ul style="list-style-type: none"> <li>• Practical organization is complex</li> <li>• Higher budget</li> </ul>
x			Tracker trial design	Allowing changes in the study protocol during the trial	<ul style="list-style-type: none"> <li>• Early assessment of technological developments</li> </ul>	
	x		Cluster randomized trials	Randomizing clusters of individuals (hospital, department)	<ul style="list-style-type: none"> <li>• Easy to implement</li> </ul>	<ul style="list-style-type: none"> <li>• Lack of power</li> <li>• Selection bias possible</li> </ul>
	x		Sequential trials	Interim analysis (the results from patients already included are analysed before randomization of new patients)	<ul style="list-style-type: none"> <li>• Reduces the number of patients needed</li> </ul>	<ul style="list-style-type: none"> <li>• Lack of power for secondary endpoints or adverse effects</li> <li>• The time between the inclusion of patients and endpoint must be short</li> <li>• Independent data monitoring committee is necessary</li> </ul>
x			Adaptive randomization trials	<ul style="list-style-type: none"> <li>• Interim analysis</li> <li>• Adjustments are possible, related to the ratio of randomization or the re-evaluation of the number of patients required or interim analysis</li> </ul>	<ul style="list-style-type: none"> <li>• Reduces the number of patients needed</li> <li>• Greater flexibility</li> </ul>	<ul style="list-style-type: none"> <li>• Logistical constraints</li> <li>• Independent data monitoring committee</li> <li>• Internal validity has also been called into question</li> </ul>
	x		Bayesian methods	<ul style="list-style-type: none"> <li>• Combining prior information with information from the ongoing trial</li> <li>• A priori information is supplied by the literature or expert opinions</li> </ul>	<ul style="list-style-type: none"> <li>• Greater flexibility</li> <li>• Reduces the number of patients needed</li> </ul>	<ul style="list-style-type: none"> <li>• Risk of taking into account arbitrary and erroneous prior information</li> </ul>

## 1 **2.5.2. Non-randomised study designs and analysis**

2 Improved applicability in addition to RCT evidence motivates the need for large, rigorous  
3 observational studies for the long-term evaluation of MD for outcomes such as revision  
4 rates and also for safety outcomes and quality assurance. Prospective disease-based  
5 registries including all relevant real world treatment options can be particularly useful for  
6 this objective and are preferable over device-based registries (22). Observational studies  
7 and randomized trials can be nested within these registries (70). Registry data are  
8 sometimes the only source of evidence for MD for example for many brands used in  
9 practice getting market access without evidence from RCT. However, data analysis of  
10 observational studies regarding treatment effects is challenging because unadjusted  
11 results are prone to bias, especially to confounding by indication (71). Data necessary for  
12 adjustment may not have been collected. The Agency for Healthcare Research and  
13 Quality (AHRQ) issued a comprehensive guide for the planning, design, maintenance and  
14 quality assessment of patient registries (47, 72).

15 For identifying safety issues post-market surveillance data can also be taken into  
16 consideration,

17

## 18 **2.6. Tools for critical appraisals**

19 We did not identify specific appraisal tools for internal validity for primary studies or  
20 systematic reviews of MD. The same is true for reporting guidelines or checklists for  
21 investigating the applicability of findings. Existing tools can be applied such as

- 22 • for RCT: Cochrane risk of bias tool for randomized studies, EUnetHTA guideline for  
23 internal validity of RCT
- 24 • for non-randomized studies: Cochrane risk of bias tool for non-randomized studies  
25 of intervention ACROBAT-NRSI (A Cochrane Risk Of Bias Assessment Tool for  
26 Non-Randomized Studies of Interventions), EUnetHTA guideline for internal validity  
27 of non-randomized studies (71), a quality appraisal checklist for case series  
28 published recently by authors from the Institute of Health Economics, Canada (73).
- 29 • for systematic reviews: AMSTAR (74), Oxman and Guyatt index (75)
- 30 • for modelling studies: ISPOR questionnaire for modelling studies (76)
- 31 • network meta-analysis studies: ISPOR checklists for network meta-analysis studies  
32 (77), Grading of Recommendations, Assessment, Development and Evaluation  
33 (GRADE) approach for rating the quality of network meta-analysis studies (78, 79),  
34 PRISMA extension statement for reporting of systematic reviews incorporating  
35 network meta-analyses (80)
- 36 • EUnetHTA guideline on applicability, applicability checklist (28, 73)

## 1 **2.7. Analysing and synthesising evidence**

### 2 ***Data extraction***

3 To be able to take better into account incremental development, the complexity of MD  
4 interventions and its user and context dependency, in addition to the information that is  
5 usually extracted from the included studies (e. g. study design, patient characteristics,  
6 results on outcomes), characteristics of the intervention, their users and providing  
7 institutions should be thoroughly extracted. As guidance what to extract the items of the  
8 Template for Intervention Description and Replication (TIDieR) checklist (81) developed as  
9 reporting guideline for primary studies can be used or customised for a specific device.  
10 They also include eligibility criteria for intervention providers, their expertise, background  
11 and training. But also eligibility criteria for institutions where the intervention took place  
12 should be extracted. Further all potentially effect-modifying factors identified during framing  
13 the research question should also be extracted such as co-therapies and adherence. It  
14 should be kept in mind that user proficiency, healthcare setting, and incremental  
15 development may affect both, intervention and comparator.

### 16 ***Evidence synthesis***

17 One important goal of a systematic review is to combine and summarize the findings of  
18 individual studies quantitatively using meta-analysis or through a narrative approach. A  
19 pooled overall effect estimate is only meaningful to a user of the systematic review, when  
20 clinical and methodological heterogeneity contributing to statistical heterogeneity is not too  
21 large. Therefore investigation of possible sources of heterogeneity is an important part of  
22 the analysis. Variables contributing to clinical heterogeneity (effect modifiers) should be  
23 identified to find subgroups and settings where the intervention works most effective or has  
24 fewest side effects. Methodological heterogeneity (i. e. by different study design, conduct  
25 or analysis) can distort the true effect estimate. Sensitivity analyses can be applied by in-  
26 and excluding studies with different methodological features.

27 For interventions involving MD, incremental development, learning effects and contextual  
28 factors contribute to heterogeneity of treatment effects. Hence heterogeneity will usually  
29 be larger and its possible sources more numerous. All variables identified in the logic  
30 model - PICO and modifying factors – may contribute to heterogeneity. Some variables  
31 changing with time may not be identifiable individually such as small variation in MD due to  
32 incremental development, change in co-therapies and setting factors. Here study or  
33 publication year might serve as proxy variable. When surgical procedures are implied,  
34 different follow-up times may further contribute to heterogeneity of results, because the  
35 overall effect may be compounded of higher short-term risks but better long-term  
36 outcomes after surviving surgery (47). For reviews of complex interventions Pigott and  
37 Shepperd give advice on the identification, documentation and examination of  
38 heterogeneity (32, 47). In general there are no specific methods for evidence synthesis  
39 neither quantitative nor narrative for MD. However, with regard to meta-analytical  
40 approaches the compilation of methods by Petticrew, Rehfuss et al. (31) for systematic  
41 reviews of complex interventions adapted by the MedtechHTA project (47) can be applied to

1 investigate subgroups of the intervention / comparator or variables that operationalize user  
2 dependency (e. g. eligibility criteria of operators and providing institutions, volume of  
3 interventions in the providing institution), contextual factors (e. g. level of care) and study  
4 designs issues ( e. g. blinding, study type, research procedures, definition of outcomes)  
5 relevant to MD (47). Subgroup analyses and meta-regression as well as more  
6 sophisticated statistical approaches to analyse multiple covariates can be applied. But  
7 these methods are soon limited by an insufficient number of studies (a rule of thumb is ten  
8 studies per variable analysed (82)).

9

10 Example 3 shows how user dependency and other factors can be addressed in a  
11 quantitative analysis.

12

### **Example 3: How to address the user-dependency of medical device safety and effectiveness**

Early in the evolution of radiofrequency ablation (RFA) of liver tumors, Mulier and colleagues performed a systematic review of case series (83). Based on nearly 100 studies with over 5000 patients included, they were able to examine statistically whether local recurrence rates were associated with the experience of the surgeon performing RFA. The surgeon's previous experience with RFA was used to classify studies into four groups (<20, 21 to 50, 51 to 100, and more than 100 operations). However, it was necessary to contact primary study authors to collect this information for all primary studies. Furthermore, multicenter studies had to be excluded. In the meta-analysis, recurrence rates showed a stepwise decrease of recurrence with increasing surgeon experience (18%, 16%, 14%, and 10%, respectively in the four groups). Still, this association does not necessarily mean that good technical performance of the procedure itself influences the outcomes. Mulier and colleagues also looked at patient factors, which were also found to be significantly associated with outcomes. As only tumor size and surgical (versus percutaneous) approach remained significant in the multivariate meta-analysis, it appears as if the more experienced surgeons achieved better results mainly through a more careful selection of patients. This example shows that user-dependency can be incorporated into the assessment of MD. For obtaining useable statistical results, however, a large amount of primary study data is required.

13

## **2.8. Reporting and interpreting**

15 For reporting results the usual methods described in the Cochrane handbook, CRD's  
16 guidance for undertaking reviews in health care (45) or other textbooks on systematic  
17 reviews on clinical effectiveness in healthcare can be applied and the guidance of the  
18 PRISMA statement should be adhered to (59). Interpreting results of a systematic review  
19 may comprise weighing the body of evidence for example with GRADE, considering  
20 limitations including publication and related biases, the strength of evidence, applicability  
21 of results and implications for further research (84). One issue more demanding for MD  
22 interventions compared to drugs is to judge the applicability of review findings to target

1 populations and settings. Greater diversity of interactions within and between the intended  
2 population, intervention components, comparators, contextual factors and outcomes can  
3 challenge the assessment of applicability (47). EUnetHTA's guideline 'Applicability of  
4 evidence in the context of a relative effectiveness assessment' can be used for a  
5 systematic assessment. Another checklist of applicability criteria is provided in Burford et  
6 al 2013 (28). It is derived from a more recent systematic review on external validity,  
7 transferability and applicability criteria. Some threats to applicability very prominent in  
8 studies on MD or procedures can be summed up as Hartling et al. write: "It should not be  
9 assumed that the efficacy and safety seen in clinical trials conducted in highly select  
10 subsets of patients cared for by highly select providers from highly select institutions will  
11 translate into similar safety and effectiveness rates when applied in usual practice,  
12 particularly over time as devices and surgical techniques evolve" (85). Relevant issues are  
13 applicability of

- 14 • eligibility criteria for patients
- 15 • modifications of the study intervention (MD and surgical procedure and or the  
16 comparator since conduction of the study)
- 17 • eligibility criteria of providers and providing institutions

18 It is noteworthy that also MD studies may include run-in periods, which can affect eligibility  
19 and inclusion of patients and thus can also affect applicability of study results (86).

### 3. Conclusion and main recommendations

- This methodological guideline primarily focuses on the evaluation of the clinical effectiveness of therapeutic devices that are **associated with high risks** (class IIb and III, according to European regulatory framework) and therefore are in special need for thorough evaluation. But methodological recommendations also apply to other therapeutic MD.
- Standard methods for finding, selecting, analysing, synthesizing and interpreting evidence on clinical effectiveness are in principle also applicable to therapeutic MD and therefore **evaluation should be done without unnecessary modification of standard methods**. A need for specific methods mainly derives from the incremental development of MD, user and context dependence, and some implications of the physical mode of action (e. g. blinding may be difficult).
- Compared to drugs therapeutic **MD interventions are often more complex**: The intervention usually consists of several components and procedures and the effects of the intervention are more context and user dependent. In planning and conducting a systematic review this means that in framing the research question more effort is necessary 1) to systematically and clearly define the intervention and its potential subgroups and 2) to identify and characterize effect-modifying factors especially proficiency and learning of MD users and providers. If these aspects are transparently characterized, this will help to take them appropriately into account in information retrieval, data extraction and synthesis as well as in assessing the applicability of the review's results (see recommendations 1 to 3, 7, 8, ).
- Whether systematic reviews and HTA of clinical effectiveness can contribute to conclusive results for **decision making strongly depends on the quality of primary research**. Also primary studies have to address the specific challenges that result from the physical mode of MD action and the often invasive nature of the MD intervention. These challenges include the inability to blind the trial, the strong treatment preferences of patients and care providers, and the effect-modifying influence of MD user proficiency. Several modifications of the common two-armed double blinded RCT design exist, that can be used for therapeutic MD. The most challenging aspect of MD evaluation for an adaption of study design and analysis methods from drugs to MD is the fast pace of the development of modifications of MD interventions. Study design and analysis addressing rapid development often use Bayesian approaches but have not been used much so far despite being highly recommended by the FDA. HTA assessors should make themselves familiar with all study designs and analysis methods relevant for MD (see (40, 60), recommendation 5).

## 1 **Limitations**

2 This guideline has several limitations.

- 3 • Firstly this guideline is limited to the evaluation of clinical effectiveness. This does  
4 not mean that there are no specific issues in the evaluation of other domains. For  
5 instance regarding costs and health economic evaluation the different market  
6 situation compared to drugs leads to a more dynamic pricing and organizational  
7 implications of the use of MD may result in upfront costs, which are not present for  
8 drugs. Also an impact on the organizational domain is obvious. In fact, benefits in  
9 non-health outcomes also to others than the patient (e. g., institutions, environment)  
10 may be equally important and so are possible health benefits for the care providers.  
11 Therefore the MD-specific issues for the evaluation of other domains have to be  
12 addressed in an update of this guideline.
- 13 • Secondly the targeted literature search on which much of this work is based on did  
14 not allow investigating specific issues in depth, especially when there was no MD  
15 specific literature. The search did not include more general questions such as  
16 'where and how to find information', the patient's perspective on usability, or the  
17 handling of missing data. Therefore, our analysis lacks information e.g. on sources  
18 for description of technology and current use (specific databases, trial registries  
19 etc.), as well as MD user's preferences for device properties and handling. With  
20 regard to information retrieval we tried to compensate that by using the experience  
21 of the institutions of the guideline authors group.

## 1 Annexe 1. Bibliography

- 2 1. Henschke C, Panteli D, Perleth M, Busse R. A taxonomy of medical devices in  
3 the logic of HTA. *Int J Technol Assess Health Care*. 2015.
- 4 2. Ioannidis JP, Evans SJ, Gotzsche PC, O'Neill RT, Altman DG, Schulz K, et al.  
5 Better reporting of harms in randomized trials: an extension of the CONSORT  
6 statement. *Ann Intern Med*. 2004;141(10):781-8.
- 7 3. Rothman K, Greenland S, Lash T. *Modern Epidemiology*. 3rd Revised edition  
8 ed: Lippincott Williams and Wilkins; 2008.
- 9 4. Porta M, Last J. *Dictionary of Epidemiology*. 5th Revised edition ed: Oxford  
10 University Press Inc; 2008.
- 11 5. Fletcher J. What is heterogeneity and is it important? *BMJ*.  
12 2007;334(7584):94-6.
- 13 6. West S, Gartlehner G, Mansfield A, Poole C, Tant E, Lenfestey N, et al.  
14 Comparative Effectiveness Review Methods: Clinical Heterogeneity: Agency  
15 for Healthcare Research and Quality; 2010. Available from:  
16 <http://effectivehealthcare.ahrq.gov/>.
- 17 7. Clark GM. Prognostic factors versus predictive factors: Examples from a  
18 clinical trial of erlotinib. *Mol Oncol*. 2008;1(4):406-12.
- 19 8. Stange K, Glasgow R. *Considering and Reporting Important Contextual  
20 Factors in Research on the Patient-Centered Medical Home*. . Rockville, MD:  
21 Agency for Healthcare Research and Quality, 2013.
- 22 9. Cook JA, Ramsay CR, Fayers P. Using the literature to quantify the learning  
23 curve: a case study. *Int J Technol Assess Health Care*. 2007;23(2):255-60.
- 24 10. Eikermann M, Gluud C, Perleth M, Wild C, Sauerland S, Gutierrez-Ibarluzea I,  
25 et al. Commentary: Europe needs a central, transparent, and evidence based  
26 regulation process for devices. *BMJ*. 2013;346:f2771.
- 27 11. Baeyens H, Pouppez C, Slegers P, Vinck I, Hulstaert F, Neyt M. *Towards a  
28 guided and phased introduction of high-risk medical devices in Belgium:  
29 Health Services Research (HSR) Brussels: Belgian Health Care Knowledge  
30 Centre (KCE); 2015.*
- 31 12. Siebert M, Clauss LC, Carlisle M, Casteels B, de Jong P, Kreuzer M, et al.  
32 Health technology assessment for medical devices in Europe. What must be  
33 considered. *Int J Technol Assess Health Care*. 2002;18(3):733-40.
- 34 13. Schulenburg Graf v.d. JM, Mittendorf T, Kulp W, Greiner W. Health  
35 Technology Assessment (HTA) im Bereich der Medizinprodukte - gleiches  
36 Spiel mit gleichen Regeln? *Gesundh ökon Qual manag*. 2009;14(3):144-55.
- 37 14. Parquin F, Audry A. Clinical evaluation of medical devices: main constraints  
38 and specificities. *Therapie*. 2012;67(4):311-8.
- 39 15. Campbell G. Statistics in the world of medical devices: the contrast with  
40 pharmaceuticals. *J Biopharm Stat*. 2008;18(1):4-19.
- 41 16. Sedrakyan A, Marinac-Dabic D, Normand SL, Mushlin A, Gross T. A  
42 framework for evidence evaluation and methodological issues in implantable  
43 device studies. *Medical care*. 2010;48(6 Suppl):S121-8.
- 44 17. Konstam MA, Pina I, Lindenfeld J, Packer M. A device is not a drug. *J Card  
45 Fail*. 2003;9(3):155-7.
- 46 18. Sorenson C, Tarricone R, Siebert M, Drummond M. Applying health  
47 economics for policy decision making: do devices differ from drugs? *Europace*.  
48 2011;13 Suppl 2:ii54-8.

- 1 19. Ergina PL, Cook JA, Blazeby JM, Boutron I, Clavien PA, Reeves BC, et al.  
2 Challenges in evaluating surgical innovation. *Lancet*. 2009;374(9695):1097-  
3 104.
- 4 20. Cook JA, Ramsay CR, Fayers P. Statistical evaluation of learning curve  
5 effects in surgical trials. *Clin Trials*. 2004;1(5):421-7.
- 6 21. Lao CS, Bushar HF. Longitudinal data analysis in medical device clinical  
7 studies. *J Biopharm Stat*. 2008;18(1):44-53.
- 8 22. Cook JA, McCulloch P, Blazeby JM, Beard DJ, Marinac-Dabic D, Sedrakyan  
9 A. IDEAL framework for surgical innovation 3: randomised controlled trials in  
10 the assessment stage and evaluations in the long term study stage. *BMJ*.  
11 2013;346:f2820.
- 12 23. Ergina PL, Barkun JS, McCulloch P, Cook JA, Altman DG. IDEAL framework  
13 for surgical innovation 2: observational studies in the exploration and  
14 assessment stages. *BMJ*. 2013;346:f3011.
- 15 24. McCulloch P, Altman DG, Campbell WB, Flum DR, Glasziou P, Marshall JC,  
16 et al. No surgical innovation without evaluation: the IDEAL recommendations.  
17 *Lancet*. 2009;374(9695):1105-12.
- 18 25. McCulloch P, Cook JA, Altman DG, Heneghan C, Diener MK. IDEAL  
19 framework for surgical innovation 1: the idea and development stages. *BMJ*.  
20 2013;346:f3012.
- 21 26. Anderson L, Oliver S, Michie S, Rehfuss E, Noyes J, Shemilt I. Investigating  
22 complexity in systematic reviews of interventions by using a spectrum of  
23 methods. *J Clin Epidemiol*. 2013;66(11).
- 24 27. Anderson LM, Petticrew M, Chandler J, Grimshaw J, Tugwell P, O'Neill J, et  
25 al. Introducing a series of methodological articles on considering complexity in  
26 systematic reviews of interventions. *J Clin Epidemiol*. 2013;66(11):1205-8.
- 27 28. Burford B, Lewin S, Welch V, Rehfuss E, Waters E. Assessing the  
28 applicability of findings in systematic reviews of complex interventions can  
29 enhance the utility of reviews for decision making. *J Clin Epidemiol*.  
30 2013;66(11):1251-61.
- 31 29. Noyes J, Gough D, Lewin S, Mayhew A, Michie S, Pantoja T, et al. A research  
32 and development agenda for systematic reviews that ask complex questions  
33 about complex interventions. *J Clin Epidemiol*. 2013;66(11):1262-70.
- 34 30. Petticrew M, Anderson L, Elder R, Grimshaw J, Hopkins D, Hahn R, et al.  
35 Complex interventions and their implications for systematic reviews: a  
36 pragmatic approach. *J Clin Epidemiol*. 2013;66(11):1209-14.
- 37 31. Petticrew M, Rehfuss E, Noyes J, Higgins JP, Mayhew A, Pantoja T, et al.  
38 Synthesizing evidence on complex interventions: how meta-analytical,  
39 qualitative, and mixed-method approaches can contribute. *J Clin Epidemiol*.  
40 2013;66(11):1230-43.
- 41 32. Pigott T, Shepperd S. Identifying, documenting, and examining heterogeneity  
42 in systematic reviews of complex interventions. *J Clin Epidemiol*.  
43 2013;66(11):1244-50.
- 44 33. Squires JE, Valentine JC, Grimshaw JM. Systematic reviews of complex  
45 interventions: framing the review question. *J Clin Epidemiol*.  
46 2013;66(11):1215-22.
- 47 34. Tugwell P, Knottnerus JA, Idzerda L. Complex interventions-how should  
48 systematic reviews of their impact differ from reviews of simple or complicated  
49 interventions? *J Clin Epidemiol*. 2013;66(11):1195-6.

- 1 35. DACEHTA. Introduction to Mini-HTA – a management and decision support  
2 tool for the hospital service. The Danish Centre for Evaluation and Health  
3 Technology Assessment, 2005.
- 4 36. DECIT. Methodological Guidelines: Elaborating Studies for the Assessment of  
5 Medical Care Equipment. Brazilia: 2013.
- 6 37. HAS. Medical Device Assessment in France. Paris: Haute Autorité de Santé,  
7 2009.
- 8 38. HAS. Methodological Choices for the Clinical Development of Medical  
9 Devices. Haute Autorité de Santé, 2013.
- 10 39. KCE. The pre-market clinical evaluation of innovative high-risk medical  
11 devices. Federaal Kenniscentrum voor de Gezondheidszorg, 2011.
- 12 40. KNAW. Evaluation of new technology in health care. In need of guidance for  
13 relevant evidence. 2014.
- 14 41. NICE. Medical Technologies Evaluation Programme. 2011.
- 15 42. NICE. The Diagnostics Assessment Programme manual. National Institute for  
16 Health and Clinical Excellence, 2011.
- 17 43. CVZ. Medical tests (assessment of established medical science and medical  
18 practice). College voor zorgverzekeringen, 2011.
- 19 44. HAS. Rapid Assessment Method for Assessing Medical and Surgical  
20 Procedures. Haute Autorité de santé, 2007.
- 21 45. Higgins J, Green S. Cochrane handbook for systematic reviews of  
22 interventions version 5.1.0 2011. Available from: [http://www.cochrane-](http://www.cochrane-handbook.org)  
23 [handbook.org](http://www.cochrane-handbook.org).
- 24 46. Centre for Reviews and Dissemination. Systematic Reviews. CRD's guidance  
25 for undertaking reviews in health care. 2009.
- 26 47. Schnell-Inderst P, Hunger T, Arvandi M, Conrads-Frank A, Siebert U.  
27 Recommendations for the Assessment of Comparative Effectiveness of  
28 Medical Devices. A Framework for Evaluation. 2015.
- 29 48. FDA. Design Considerations for Pivotal Clinical Investigations for Medical  
30 Devices. Draft Guidance for Industry, Clinical Investigators, and Food and  
31 Drug Administration Staff. 2011.
- 32 49. Boutron I, Tubach F, Giraudeau B, Ravaud P. Blinding was judged more  
33 difficult to achieve and maintain in nonpharmacologic than pharmacologic  
34 trials. *J Clin Epidemiol*. 2004;57(6):543-50.
- 35 50. USPSTF. U.S. Preventive Services Task Force Procedure Manual. AHRQ,  
36 July 2008.
- 37 51. US Preventive Services Task Force. Universal Screening for Hearing Loss in  
38 Newborns: US Preventive Services Task Force Recommendation Statement.  
39 *Pediatrics*. 2008;122(1):143-8.
- 40 52. Commission d'Evaluation des Produits et Prestations. Avis de la Commission:  
41 ASR XL Head 2008. Available from: [http://www.has-](http://www.has-sante.fr/portail/upload/docs/application/pdf/2008-07/cepp_1737_asr_xl_head.pdf)  
42 [sante.fr/portail/upload/docs/application/pdf/2008-](http://www.has-sante.fr/portail/upload/docs/application/pdf/2008-07/cepp_1737_asr_xl_head.pdf)  
43 [07/cepp\\_1737\\_asr\\_xl\\_head.pdf](http://www.has-sante.fr/portail/upload/docs/application/pdf/2008-07/cepp_1737_asr_xl_head.pdf).
- 44 53. Commission d'Evaluation des Produits et Prestations. Avis de la Commission:  
45 Metasul 2009. Available from: [http://www.has-](http://www.has-sante.fr/portail/upload/docs/application/pdf/2009-02/cepp-1795_metasul.pdf)  
46 [sante.fr/portail/upload/docs/application/pdf/2009-02/cepp-1795\\_metasul.pdf](http://www.has-sante.fr/portail/upload/docs/application/pdf/2009-02/cepp-1795_metasul.pdf)
- 47 54. Smith AJ, Dieppe P, Vernon K, Porter M, Blom AW, National Joint Registry of  
48 E, et al. Failure rates of stemmed metal-on-metal hip replacements: analysis  
49 of data from the National Joint Registry of England and Wales. *Lancet*.  
50 2012;379(9822):1199-204.

- 1 55. Commission Nationale d'Évaluation des Dispositifs Médicaux et des  
2 Technologies de Santé. Rapport sur les prothèses totales de hanche à couple  
3 de frottement métal-métal 2013. Available from: [http://www.has-](http://www.has-sante.fr/portail/upload/docs/application/pdf/2013-07/rapport_protheses_totales_de_hanche_metal-metal.pdf)  
4 [sante.fr/portail/upload/docs/application/pdf/2013-](http://www.has-sante.fr/portail/upload/docs/application/pdf/2013-07/rapport_protheses_totales_de_hanche_metal-metal.pdf)  
5 [07/rapport\\_protheses\\_totales\\_de\\_hanche\\_metal-metal.pdf](http://www.has-sante.fr/portail/upload/docs/application/pdf/2013-07/rapport_protheses_totales_de_hanche_metal-metal.pdf) (English summary  
6 at: [http://www.has-sante.fr/portail/upload/docs/application/pdf/2013-](http://www.has-sante.fr/portail/upload/docs/application/pdf/2013-07/synthesis_metal_on_metal_hip_implants_dm-eval_54.pdf)  
7 [07/synthesis\\_metal\\_on\\_metal\\_hip\\_implants\\_dm-eval\\_54.pdf](http://www.has-sante.fr/portail/upload/docs/application/pdf/2013-07/synthesis_metal_on_metal_hip_implants_dm-eval_54.pdf).)
- 8 56. van Barabandt H, Neyt M, Devos C. Catheter ablation of atrial fibrillation. KCE  
9 report 184C [Internet]. 2012. Available from:  
10 <https://kce.fgov.be/publication/report/catheter-ablation-of-atrial-fibrillation>.
- 11 57. EUnetHTA. Process of information retrieval for systematic reviews and health  
12 technology assessments on clinical effectiveness: European network for  
13 Health Technology Assessment (EUnetHTA); 2015. Available from:  
14 [http://www.eunetha.eu/news/closed-public-consultation-draft-methodological-](http://www.eunetha.eu/news/closed-public-consultation-draft-methodological-guideline-process-information-retrieval-systema)  
15 [guideline-process-information-retrieval-systema](http://www.eunetha.eu/news/closed-public-consultation-draft-methodological-guideline-process-information-retrieval-systema).
- 16 58. Chang L, Dhruva SS, Chu J, Bero LA, Redberg RF. Selective reporting in trials  
17 of high risk cardiovascular devices: cross sectional comparison between  
18 premarket approval summaries and published reports. *BMJ*. 2015;350:h2613.
- 19 59. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items  
20 for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*.  
21 2009;339:b2535.
- 22 60. Bernard A, Vaneau M, Fournel I, Galmiche H, Nony P, Dubernard JM.  
23 Methodological choices for the clinical development of medical devices.  
24 *Medical devices*. 2014;7:325-34.
- 25 61. Lilford RJ, Braunholtz DA, Greenhalgh R, Edwards SJ. Trials and fast  
26 changing technologies: the case for tracker studies. *BMJ*. 2000;320(7226):43-  
27 6.
- 28 62. FDA. Guidance for the Use of Bayesian Statistics in Medical Device Clinical  
29 Trials Guidance for Industry and FDAStaff2010.
- 30 63. Bonangelino P, Irony T, Liang S, Li X, Mukhi V, Ruan S, et al. Bayesian  
31 approaches in medical device clinical trials: a discussion with examples in the  
32 regulatory setting. *J Biopharm Stat*. 2011;21(5):938-53.
- 33 64. Devereaux PJ, Bhandari M, Clarke M, Montori VM, Cook DJ, Yusuf S, et al.  
34 Need for expertise based randomised controlled trials. *BMJ*.  
35 2005;330(7482):88.
- 36 65. Horwitz RI, Feinstein AR. Advantages and drawbacks of the Zelen design for  
37 randomized clinical trials. *J Clin Pharmacol*. 1980;20(7):425-7.
- 38 66. Zelen M. A new design for randomized clinical trials. *N Engl J Med*.  
39 1979;300(22):1242-5.
- 40 67. Wennberg JE, Barry MJ, Fowler FJ, Mulley A. Outcomes research, PORTs,  
41 and health care reform. *Ann N Y Acad Sci*. 1993;703:52-62.
- 42 68. Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG, et al. Empirical  
43 evidence of bias in treatment effect estimates in controlled trials with different  
44 interventions and outcomes: meta-epidemiological study. *BMJ*.  
45 2008;336(7644):601-5.
- 46 69. Ramsay CR, Wallace SA, Garthwaite PH, Monk AF, Russell IT, Grant AM.  
47 Assessing the learning curve effect in health technologies. Lessons from the  
48 nonclinical literature. *Int J Technol Assess Health Care*. 2002;18(1):1-10.
- 49 70. Lauer MS, D'Agostino RB, Sr. The randomized registry trial--the next  
50 disruptive technology in clinical research? *N Engl J Med*. 2013;369(17):1579-  
51 81.

- 1 71. EUnetHTA. Internal validity of non-randomised studies (NRS) on interventions:  
2 European Network for Health Technology Assessment (EUnetHTA); 2015.  
3 Available from: [http://www.eunethta.eu/news/closed-public-consultation-draft-  
5 methodological-guideline-internal-validity-non-randomised-stud](http://www.eunethta.eu/news/closed-public-consultation-draft-<br/>4 methodological-guideline-internal-validity-non-randomised-stud).
- 6 72. Gliklich R, Dreyer N, Leavy M, eds. Registries for Evaluating Patient  
7 Outcomes: A User's Guide. Rockville: Agency for Healthcare Research and  
8 Quality (AHRQ); 2014. Available from:  
9 <http://www.effectivehealthcare.ahrq.gov/registries-guide-3.cfm>.
- 10 73. Guo B, Moga C, Harstall C, Schopflocher D. A quality appraisal checklist  
11 developed specifically for evaluating case series studies. *J Clin Epidemiol*.  
12 2015.
- 13 74. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al.  
14 Development of AMSTAR: a measurement tool to assess the methodological  
15 quality of systematic reviews. *BMC Med Res Methodol*. 2007;7:10.
- 16 75. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles.  
17 *J Clin Epidemiol*. 1991;44(11):1271-8.
- 18 76. Caro JJ, Briggs AH, Siebert U, Kuntz KM. Modeling good research practices--  
19 overview: a report of the ISPOR-SMDM Modeling Good Research Practices  
20 Task Force-1. *Medical decision making : an international journal of the Society  
21 for Medical Decision Making*. 2012;32(5):667-77.
- 22 77. Jansen JP, Trikalinos T, Cappelleri JC, Daw J, Andes S, Eldessouki R, et al.  
23 Indirect treatment comparison/network meta-analysis study questionnaire to  
24 assess relevance and credibility to inform health care decision making: an  
25 ISPOR-AMCP-NPC Good Practice Task Force report. *Value Health*.  
26 2014;17(2):157-73.
- 27 78. Puhan MA, Schunemann HJ, Murad MH, Li T, Brignardello-Petersen R, Singh  
28 JA, et al. A GRADE Working Group approach for rating the quality of treatment  
29 effect estimates from network meta-analysis. *BMJ*. 2014;349:g5630.
- 30 79. Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JP. Evaluating  
31 the quality of evidence from a network meta-analysis. *PLoS One*.  
32 2014;9(7):e99682.
- 33 80. Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, et al.  
34 The PRISMA extension statement for reporting of systematic reviews  
35 incorporating network meta-analyses of health care interventions: checklist  
36 and explanations. *Ann Intern Med*. 2015;162(11):777-84.
- 37 81. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al.  
38 Better reporting of interventions: template for intervention description and  
39 replication (TIDieR) checklist and guide. *BMJ*. 2014;348:g1687.
- 40 82. Borenstein M, Hedges L, Higgins J, Rothstein H. Introduction to meta analysis:  
41 statistics in practice. Chichester, UK: Wiley; 2009.
- 42 83. Mulier S, Ni Y, Jamart J, Ruers T, Marchal G, Michel L. Local recurrence after  
43 hepatic radiofrequency coagulation: multivariate meta-analysis and review of  
44 contributing factors. *Ann Surg*. 2005;242(2):158-71.
- 45 84. Egger M, Davey Smith G, Altman DG. *Systematic Reviews in Health Care:  
46 Meta-Analysis in Context, Second Edition*: Wiley; 2008.
- 47 85. Hartling L, McAlister FA, Rowe BH, Ezekowitz J, Friesen C, Klassen TP.  
48 Challenges in systematic reviews of therapeutic devices and procedures. *Ann  
49 Intern Med*. 2005;142(12 Pt 2):1100-11.
- 50 86. Pablos-Mendez A, Barr RG, Shea S. Run-in periods in randomized trials:  
51 implications for the application of results in clinical practice. *JAMA*.  
1998;279(3):222-5.

## 1 **Annexe 2. Documentation of literature search**

2 No systematic literature search has been conducted for the elaboration of this guideline.  
3 However, reference was made to the results of an existing literature search conducted  
4 within the MedtechHTA project WP 3. Details of this literature search are published in  
5 Schnell-Inderst et al 2015. (47)