eunethta

EUROPEAN NETWORK FOR HEALTH TECHNOLOGY ASSESSMENT

GUIDELINE

Therapeutic medical devices

November 2015

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 MedtecHTA, 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).

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).

Table of contents

Acronyms – Abbreviations

ACROBAT-NRSI - A Cochrane Risk of Bias Assessment Tool for Non-Randomized Studies of Interventions

AMSTAR – Assessing the Methodological Quality of Systematic Reviews

CONSORT – Consolidated Standards of Reporting Trials

EuroScan - International Information Network on New and Emerging Health Technologies

EPAR - European public assessment reports

EUnetHTA – European network for Health Technology Assessment

FDA – Food and Drug Administration

G-BA – Gemeinsamer Bundesausschuss (Federal Joint Committee)

GRADE – Grading of Recommendations, Assessment, Development and Evaluation

HTA – Health Technology Assessment

IQWiG – Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (Institute for Quality and Efficiency in Health Care)

ISPOR – International Society for Pharmacoeconomics and Outcomes Research

JA – Joint Action

MAUDE – Manufacturer and User Facility Device Experience

MD – Medical Device

MTA – Multi-Technology Assessment

Osteba – Osasun teknologien Ebaluazioaren Zerbitzua (Basque Office for Health Technology Assessment, Ministry for Health)

PBAC – Pharmaceutical Benefits Advisory Committee

PICO – Population, Intervention, Comparison, Outcome

PRISMA – Preferred Reporting Items for Systematic Reviews and Meta-Analyses

RCT – Randomised Controlled Trial

REA – Relative Effectiveness Assessment

RoB – Risk of Bias

STA – Single Technology Assessment

SuRE Info – Summarized Research in Information Retrieval for HTA

UMIT – University for Health Sciences, Medical Informatics and Technology

Summary and table with main recommendations

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

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

Recommendations	**The recommendation is based on arguments presented in the following publications and / or parts of the guideline text**
1st recommendation: Specifics of HTA of MD HTA of medical device interventions should generally be done with currently established methods for finding, selecting, analysing, synthesizing and interpreting evidence on clinical effectiveness. A need for specific methods mainly derives from the incremental development of MDs and their user and context dependency, and some implications of the physical mode of action.	2 Introduction

2nd recommendation: Framing the research question The more complex nature of MD interventions requires a more elaborated development of the research question. 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. Try to use clinical prior information about properties of the intervention that might influence treatment effects. Provide the sources / evidence for this information.	2.3
3rd recommendation: Defining the intervention 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). If the aim is to perform a MTA, the review should take a broad scope for the definition of the intervention. Try to identify • all MD interventions, • which technologies are used in combination or alternatively, • potentially important differences. Redefinition of the intervention may become necessary during the course of the assessment.	2.3.1
4th recommendation: Information retrieval 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). 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. In addition to the search in bibliographical databases, information about the MD may also be retrieved from device registries, incident reporting databases and administrative databases.	2.4

5th recommendation: Information requirements for clinical effectiveness 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. 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).	2.5
6th recommendation: Information requirements for long-term effects 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. Registry analyses should be considered to assess long-term outcomes but should only be used for the assessment of treatment effects when appropriate confounder control is possible. Also residual confounding has to be addressed.	2.5
7th recommendation: User dependency and context factors 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. User proficiency and healthcare setting may affect both, intervention and comparator.	2.3.2, 2.7
8th recommendation: Applicability of findings When interpreting the review’s findings consider the influence of health care settings, user proficiency, and incremental treatment modification. 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”).	2.8

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

- **Relative effectiveness:** Relative effectiveness can be defined as the extent to which an intervention does more good than harm compared with one or more alternative interventions for achieving the desired results when provided under the usual circumstances of health-care practice.

- **Safety:** Substantive evidence of an absence of harm. The term is often misused when there is simply absence of evidence of harm.
  Source: Ioannidis 2004 (2)

- **Internal validity:** the extent to which the (treatment) difference observed in a trial is likely to reflect the ‘true’ effect within the trial (or in the trial population) by considering methodological criteria.

- **Applicability** (also known as external validity, generalisability, or transposability): the extent to which the effects observed in clinical studies are likely to reflect the expected results when a specific intervention is applied to the population of interest.

- **Effect(-measure) modification**: Variation in the selected effect measure for the factor under study across levels of another factor. In statistical terminology this is called interaction. Effect-measure modification results in heterogeneity (see below). An effect-modifier may modify different effect-measures (e. g. relative risks, risk difference) for the same factor (e. g. age) in different directions and may modify one measure, but not another (3, 4).

- **Heterogeneity (5, 6):**

    - **Clinical:** Variation in study population characteristics, coexisting conditions, cointerventions, and outcomes evaluated across studies included in systematic reviews or comparative effectiveness research that may influence or modify the magnitude of the intervention measure of effect.

    - **Methodologic:** In the context of systematic reviews on effectiveness, among-study differences in estimated effect sizes for the intervention that can be attributed to variability and quality of study designs and analyses.

    - **Statistical:** Variability in the observed treatment effects beyond what would be expected by random error. Statistical heterogeneity may signal the presence of clinical heterogeneity, methodological heterogeneity, or chance.

- **Prognostic factor**: A prognostic factor is a measurement that is associated with clinical outcome in the absence of therapy or with the application of a standard therapy that patients are likely to receive. It can be thought of as a measure of the natural history of the disease (7).

- **Confounding by indication:** Patients are selected for different therapies based on clinical indications. If these indications are also prognostic factors, the estimates of treatment efficacy become confounded by these factors. This phenomenon is often referred to as confounding by indication (3).

- **Contextual factors:** The phenomena of health care and health, are complex systems that are fundamentally context-dependent. Contextual factors with potential influence on health outcomes could be for example: national, state, local, and

organizational policies, community norms and resources, health care system organization, payment and incentive systems, practice culture, history, and staffing, historical factors and recent events, the culture and motivations surrounding the use of MD, changes in these factors over time (8).

- **User dependency:** the extent to which the treatment result in a clinical study or in clinical practice is influenced by the skills or experience of the people involved in the treatment.

- **Learning effect/ curve**: improvements in the technical performance of a new technique over time (9).

- **Incremental development:** Many device classes are developed in a step-wise process with frequent technology changes each of which present only minor 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 medicinal products. 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

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

1.3. Objective(s) and scope of the guideline

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

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

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

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

1.4. Related EUnetHTA documents

- EUnetHTA guidelines (http://www.eunethta.eu/eunethta-guidelines):
  - Methodological guideline for REA: Endpoints used for relative effectiveness assessment of pharmaceuticals: Clinical endpoints
  - Methodological guideline for REA: Endpoints used for relative effectiveness assessment of pharmaceuticals: Composite endpoints
  - Methodological guideline for REA: Endpoints used in relative effectiveness assessment of pharmaceuticals: Surrogate endpoints
  - Methodological guideline for REA: Endpoints used in relative effectiveness assessment of pharmaceuticals: Safety
  - Methodological guideline for REA: Endpoints used for relative effectiveness assessment of pharmaceuticals: Health-related quality of life and utility measures
  - Methodological guideline for REA: Comparators & comparisons: Criteria for the choice of the most appropriate comparator(s). Summary of current policies and best practice recommendations
  - Methodological guideline for REA: Comparators & comparisons: Direct and indirect comparison
  - Methodological guideline for REA: Levels of Evidence: Internal validity of randomized controlled trials
  - Methodological guideline for REA: Levels of Evidence: Applicability of evidence in the context of a relative effectiveness assessment of pharmaceuticals
  - Internal validity of non-randomised studies (NRS) on interventions
  - Process of information retrieval for systematic reviews and health technology assessments on clinical effectiveness

- EunetHTA Core Model® (http://www.eunethta.eu/hta-core-model)

1.5. Other related documents

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

1.6. Methods

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

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

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

The recommendations of this guideline 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).

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

2. Analysis and discussion of the methodological issue

2.1. Results from literature review

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

- the framing of the research question,
- requirements for primary studies for MD,
- selection and analysis of primary studies on MD,
- evidence synthesis and interpretation of the review’s results.

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

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

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

A framework for systematic reviews of complex interventions was published in 2013 by authors from Cochrane methods groups (26-34). It gives guidance for clarifying the review question, identifying what study types should be sought as evidence and which evidence synthesis methods can be chosen, and how applicability of findings can be assessed. Although the authors mainly give examples from the field of health promotion showing a higher degree of complexity, applying this framework as far as it is relevant to therapeutic MD may help to better take into account the specifics of the evaluation of MD interventions in a systematic and transparent way.

2.2. The role of logic models in the HTA context

To get a better understanding of the intervention's components and the relation between intervention, modifying factors and outcome the use of logic models is recommended. "A logic model is a graphic description of a system and is designed to identify important elements and relationships within a system" (26). In the context of health care interventions logic models describe a theory of change, that is, how the intervention achieves beneficial and harmful changes in outcomes (27). A logic model can add to a common initial understanding in review production regarding evidence requirements before the evidence synthesis stage. It can be a tool to support a priori decisions regarding the proposed approach to evidence synthesis (27). The use of logic models is already a regular step in supporting the development of the analytical framework of systematic reviews on clinical effectiveness on public health interventions such as screening programs by the US Preventive Services Task Force (50). In the agency's reviews they are used to identify all relevant steps mediating beneficial and harmful treatment effects by screening, diagnostics and early intervention and related subcategories of research questions (see example in Figure 1). With respect to MD a logic model could be used to clarify which elements belong to the intervention and to comparators and which contextual factors could potentially modify treatment effects. This can support defining search terms and inclusion and exclusion criteria for literature search and selection as well as better specifying which outcomes should be sought, which modifying factors should be extracted, and which subgroups of interventions or populations should be analyzed in evidence synthesis.

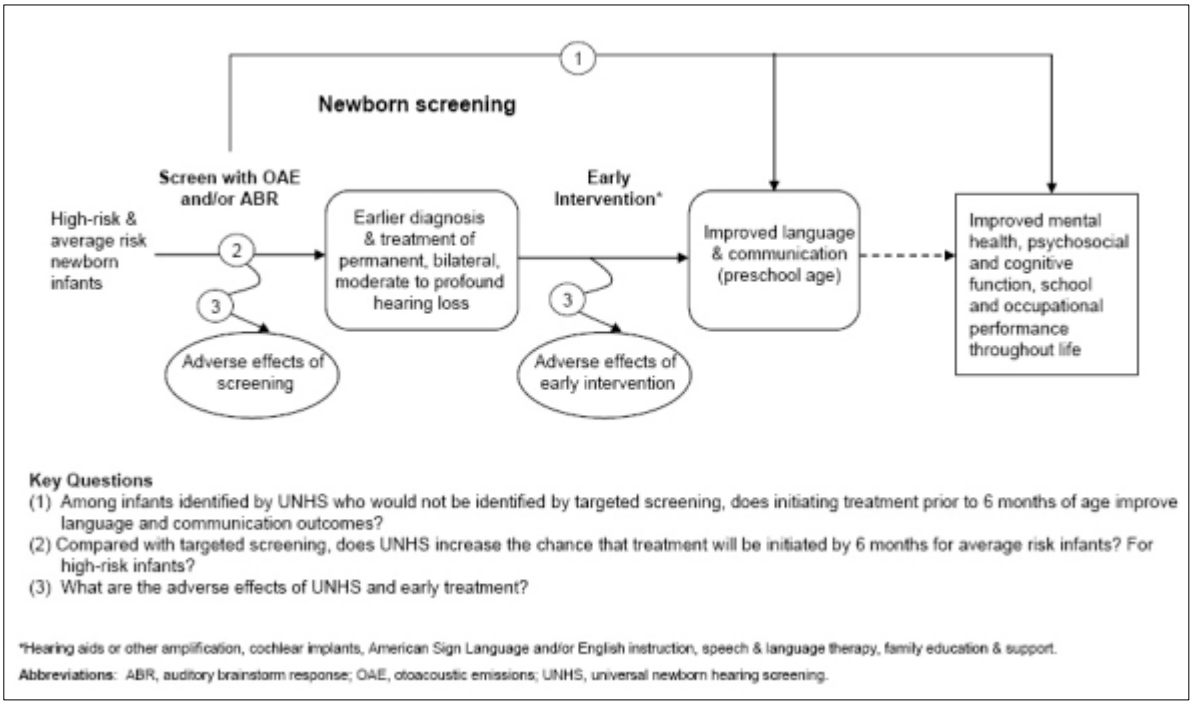

Figure 1: Example of a logic model: Analytical Framework for Universal Newborn Hearing Screening. From the US Preventive Services Task Force (51).

2.3. Systematic Reviews of MD: Framing the research question

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

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

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

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

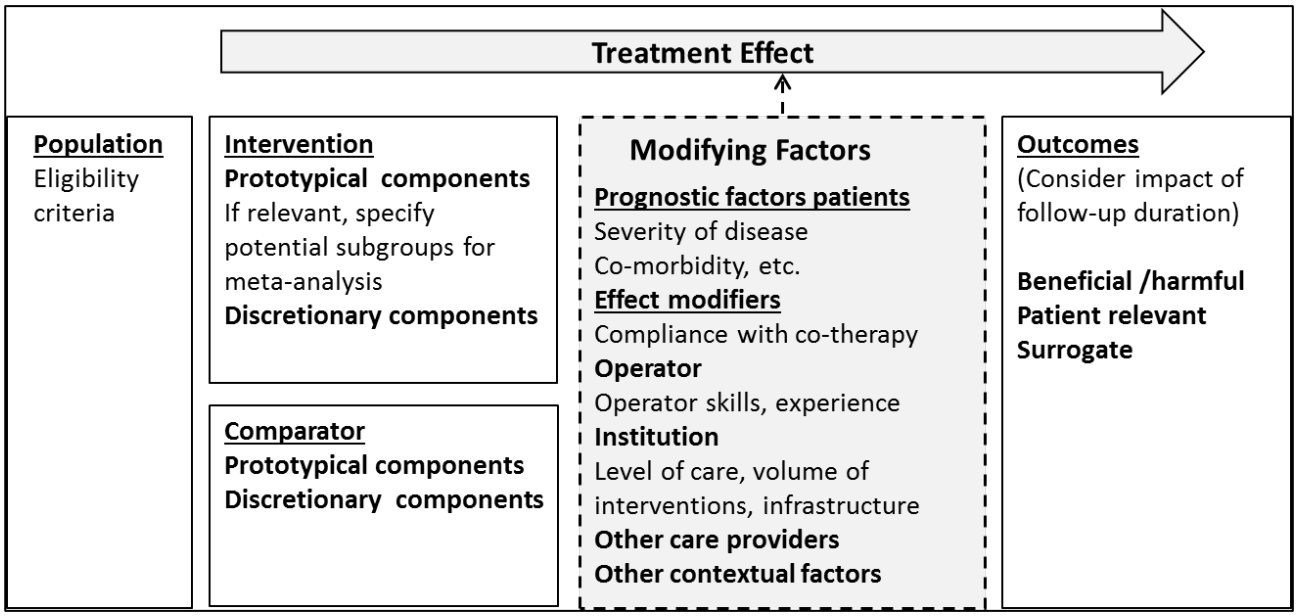

Figure 2. Template for a logic model for interventions with therapeutic medical devices (47). Prototypical components need to be present for the intervention to meet the working definition; discretionary components may be present but are not compulsory to meet the working definition (33).

2.3.1. Defining the intervention

Evaluation of MD can focus on only one particular product (STA), but in many cases HTA aims 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 devices with large femoral head diameter (≥ 36 mm). Therefore, it was appropriate to assess the whole class of products and to define new subgroups according to femoral head diameter (54). The 2013 report was based on a systematic review of randomized and non-randomized studies, and several orthopaedic surgeons participated in data interpretation (55). Finally, reimbursement of all large head metal-on-metal hip implants was discontinued.

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

If an intervention that might be delivered with different MD or procedures is evaluated, the HTA report should use a broad approach for defining the intervention. For complex interventions it is recommended by Squires et al. to “use as broad an approach (i. e., lumping with subsequent explicit a priori subgroup analysis) as makes practical sense” and that “potentially important differences between the composition or intensity of the interventions in question should be specified in the review question”(33). For interventions involving the application of MD it is often not so obvious what can be classified as the same or similar intervention as for drugs.

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

When including variants of a MD or MD with similar functioning in the assessment, it should be made explicit whether the MD is assumed to be generic (‘genericization’(18)), that is, that they belong to the same class of devices and are therefore in principal of

comparable effectiveness. Assumptions about comparability of interventions are critical issues when decisions on the appropriateness of the combination of data are made. They should be clearly stated. Often it is also part of the research question to compare subgroups of the intervention. Furthermore, it is often useful also to search for, include and assess clinical studies that compare different variants of a MD intervention. Prior information from clinical studies on former versions of the technology or similar technology should be used to decide whether and which modifications of the MD could impact the clinical effect. This information can guide the definition of subgroups of intervention characteristics that may be relevant for the investigation of differences in treatment effects. The sources of this information used to decide on subgroups should be provided (47).

See an example how to discern between MD that are substantially equivalent or not for 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.

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 have 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.

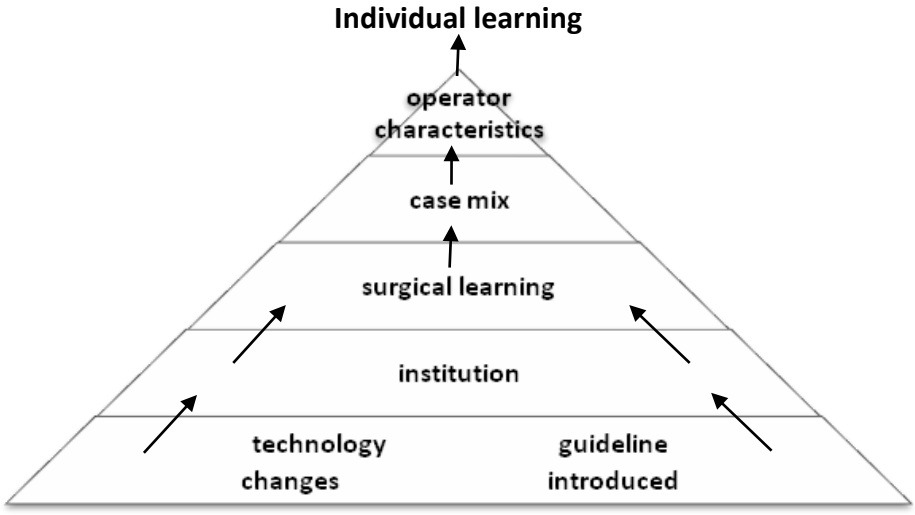

Figure 3: Hierarchical factors that influence learning (source: Cook 2004(20))

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

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

2.4.1. Searching bibliographic databases

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

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

Regarding the structuring of the search strategy along the PICO scheme, careful selection of search terms for the relevant patient population is of particular importance since (often different from drugs) the same devices may be used for several patient populations and indications. The intervention is not necessarily identical to the device, especially when a procedure to apply the MD is involved. Comparator may be another procedure, a drug or quite often a sham device or procedure. In order to avoid language bias, no language restrictions should be applied to the search strategy. Date restrictions should be applied

only if it is known that relevant studies could have been reported during a specific time period, for example if the device or procedure was only available after a given time point (Cochrane Handbook Chapter 6 (45), EUnetHTA (57)).

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

2.4.2. Searching clinical trial registers

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

2.4.3. Specific issues

Further information sources

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

Studies on efficacy and effectiveness

Since the regulatory approval process for MD in Europe does not necessarily require conducting RCTs, literature search can be broadened to include the best available evidence. In absence of RCT, this may be all types of study designs, including case series and even case reports. In many cases, this applies also for devices licensed under the 510(k) pathway in the USA (see Safety Guideline, chapter 2.3.1.2). It might be helpful to search the FDA website for clinical trial information of devices previously licensed for the US market, including data on conditional marketing approval and its status. Medical device databases at FDA are freely accessible (http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/Databases/default.htm) and searches can be performed separately for premarket approval, humanitarian device exemption, 510(k) and other legal procedures. Most of these databases are

updated weekly. In addition it might by advisable to additionally search Clinicaltrials.gov for ongoing studies on medical devices using a combination of device and company-specific search terms, as trial registration is mandatory for FDA approval of medical devices.

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

Data on safety

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

User or setting dependency

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

2.4.4. Selection of relevant studies

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

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

2.5. What kind of information is required?

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

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

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

2.5.1. Randomised study designs and analysis

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

For the study design of RCT for therapeutic MD, the rapid incremental development of MD, the influence of contextual factors and user proficiency on treatment effects, the MD’s physical mode of action including surgical procedures, as well as preferences of providers and patients have to be considered:

Rapid incremental development

The fact that the MD under investigation may be subject to modification even during the course of the trial needs to be reflected (48). , In addition to studies comparing new vs standard intervention also those studies that compare different variants of the new intervention should be included and assessed.

It has to be investigated whether these changes have the potential to make a difference to the device effectiveness or safety (16).

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

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

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

Preferences

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

Blinding

The physical mode of action of MD often challenges blinding patients and providers towards treatment and comparator. Generally, the comparator in a study can be no treatment, placebo, sham device treatment (blinding possible) or an active treatment and in principle, all these approaches are also possible for MD trials. A placebo can be any inactive fake treatment (no matter the route of administration) while a sham procedure in MD trials should mostly resemble the (invasive) experimental intervention, e.g., surgery, without achieving a treatment effect. A sham arm with blinded investigators and patients may not always be feasible or ethical. If blinding is impossible, at least blinded endpoint evaluation is recommended (16, 48, 60). In those situations, where all available evidence stems from open studies, objective outcomes (e.g. mortality, some morbidity endpoints) should have more influence on the HTA report’s conclusion than subjective outcomes (e.g. symptoms, quality of life). In an meta-epidemiological study, effect sizes were found to be exaggerated by 25 % in trials with subjective outcomes and lack of blinding (68) However,

the larger the effect sizes, the less they can be attributed only to information bias in an unblinded study.

Contextual factors and user proficiency

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

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

Existing approaches to describe and quantify learning curves can be applied for different 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	• Facilitates inclusion	• Selection bias possible • Loss of statistical power if many patients refuse treatment • Ethical problems
		x	Wennberg's Design (67)	Randomizing to preference group (people can choose their treatment) or randomization group	• Facilitates participation	• Blinding not possible • Statistical power low, when a high proportion of participants chooses the same treatment
	x		Expertise-based randomized trial	Randomizing patients to a specialized physician	• Better acceptability • Reduces execution bias and protocol deviations	• Difficulty of knowing whether the observed difference is related to the expertise of the therapist
x			Tracker trial design	Allowing changes in the study protocol during the trial	• Early assessment of technological developments	• Practical organization is complex • Higher budget
	x		Cluster randomized trials	Randomizing clusters of individuals (hospital, department)	• Easy to implement	• Lack of power • Selection bias possible
x			Sequential trials	Interim analysis (the results from patients already included are analysed before randomization of new patients)	• Reduces the number of patients needed	• Lack of power for secondary endpoints or adverse effects • The time between the inclusion of patients and endpoint must be short • Independent data monitoring committee is necessary
x			Adaptive randomization trials	• Interim analysis • Adjustments are possible, related to the ratio of randomization or the re-evaluation of the number of patients required or interim analysis	• Reduces the number of patients needed • Greater flexibility	• Logistical constraints • Independent data monitoring committee • Internal validity has also been called into question
x			Bayesian methods	• Combining prior information with information from the ongoing trial • A priori information is supplied by the literature or expert opinions	• Greater flexibility • Reduces the number of patients needed	• Risk of taking into account arbitrary and erroneous prior information

2.5.2. Non-randomised study designs and analysis

Improved applicability in addition to RCT evidence motivates the need for large, rigorous observational studies for the long-term evaluation of MD for outcomes such as revision rates and also for safety outcomes and quality assurance. Prospective disease-based registries including all relevant real world treatment options can be particularly useful for this objective and are preferable over device-based registries (22). Observational studies and randomized trials can be nested within these registries (70). Registry data are sometimes the only source of evidence for MD as many products used in practice get market approval without evidence from RCT. However, data analysis of observational studies regarding treatment effects is challenging because unadjusted results are prone to bias, especially to confounding by indication (71). Data necessary for adjustment may not have been collected. Appropriate bias-adjustment is required and the potential for residual confounding has also to be addressed (72). The Agency for Healthcare Research and Quality (AHRQ) issued a comprehensive guide for the planning, design, maintenance and quality assessment of patient registries (47, 73). The cross-border PAtients REgistries INitTative (PARENT; www.patientregistries.eu) provides support and methodological advice for interoperable patient registries within the EU. They found that about 10 % of EU patient registries are product-based, 80 % (n=83) thereof are for MD (74).

For identifying safety issues post-market surveillance data can also be taken into consideration.

2.6. Tools for critical appraisals

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

- for RCT: Cochrane risk of bias tool for randomized studies, EUnetHTA guideline for internal validity of RCT
- for non-randomized studies: Cochrane risk of bias tool for non-randomized studies of intervention ACROBAT-NRSI (A Cochrane Risk Of Bias Assessment Tool for Non-Randomized Studies of Interventions), EUnetHTA guideline for internal validity of non-randomized studies (71), a quality appraisal checklist for case series published recently by authors from the Institute of Health Economics, Canada (75).
- for systematic reviews: AMSTAR (76), Oxman and Guyatt index (77)
- for modelling studies: ISPOR questionnaire for modelling studies (78)
- for network meta-analysis studies: ISPOR checklists for network meta-analysis studies (79), Grading of Recommendations, Assessment, Development and Evaluation (GRADE) approach for rating the quality of network meta-analysis studies (80, 81), PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses (82)
- EUnetHTA guideline on applicability, applicability checklist (28, 75)

2.7. Analysing and synthesising evidence

Data extraction

To be able to take better into account incremental development, the complexity of MD interventions and its user and context dependency, in addition to the information that is usually extracted from the included studies (e. g. study design, patient characteristics, results on outcomes), characteristics of the intervention, their users and providing institutions should be thouroughly extracted. The Template for Intervention Description and Replication (TIDieR) checklist (83) was developed as reporting guideline for primary studies and can be used or customised for extracting items on a specific device. It also includes eligibility criteria for intervention providers, their expertise, background and training. But also eligibility criteria for institutions where the intervention was performed should be extracted. Further, all potentially effect-modifying factors identified during framing the research question should also be extracted such as co-therapies and adherence. It should be kept in mind that user proficiency, healthcare setting, and incremental development may affect both, intervention and comparator.

Evidence synthesis

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

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

dependency (e. g. eligibility criteria of operators and providing institutions, volume of interventions in the providing institution), contextual factors (e. g. level of care) and study designs issues (e. g. blinding, study type, research procedures, definition of outcomes) relevant to MD (47). Subgroup analyses and meta-regression as well as more sophisticated statistical approaches to analyse multiple covariates can be applied. But these methods are limited by an insufficient number of studies (a rule of thumb is ten studies per variable analysed (84)).

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

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 (85). 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.

2.8. Reporting and interpreting

For reporting results the usual methods described in the Cochrane handbook (45), CRD’s guidance for undertaking reviews in health care (45) or other textbooks on systematic reviews on clinical effectiveness in healthcare can be applied and the guidance of the PRISMA statement should be adhered to (59). Interpreting results of a systematic review may comprise weighing the body of evidence for example with GRADE, considering limitations including publication and related biases, the strength of evidence, applicability of results and implications for further research (86). One issue more demanding for MD interventions compared to drugs is to judge the applicability of review findings to target populations and settings. Greater diversity of interactions within and between the intended population, intervention components, comparators, contextual factors and outcomes can challenge the assessment of applicability (47). EUnetHTA’s guideline ‘Applicability of

evidence in the context of a relative effectiveness assessment’ can be used for a systematic assessment. Another checklist of applicability criteria is provided by Burford et al 2013 (28). It is derived from a systematic review on external validity, transferability and applicability criteria. Some threats to applicability very prominent in studies on MD or procedures can be summed up as Hartling et al. write: “It should not be assumed that the efficacy and safety seen in clinical trials conducted in highly select subsets of patients cared for by highly select providers from highly select institutions will translate into similar safety and effectiveness rates when applied in usual practice, particularly over time as devices and surgical techniques evolve” (87). Relevant issues are applicability of

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

It is noteworthy that MD studies may include run-in periods, which can affect eligibility and inclusion of patients and thus can also affect applicability of study results (88).

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 generally be done with currently established 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).

Limitations

This guideline has several limitations.

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

Annexe 1. Bibliography

1. Henschke C, Panteli D, Perleth M, Busse R. A taxonomy of medical devices in the logic of HTA. Int J Technol Assess Health Care. 2015.
2. Ioannidis JP, Evans SJ, Gotzsche PC, O'Neill RT, Altman DG, Schulz K, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Ann Intern Med. 2004;141(10):781-8.
3. Rothman K, Greenland S, Lash T. Modern Epidemiology. 3rd Revised edition ed: Lippincott Williams and Wilkins; 2008.
4. Porta M, Last J. Dictionary of Epidemiology. 5th Revised edition ed: Oxford University Press Inc; 2008.
5. Fletcher J. What is heterogeneity and is it important? BMJ. 2007;334(7584):94-6.
6. West S, Gartlehner G, Mansfield A, Poole C, Tant E, Lenfestey N, et al. Comparative Effectiveness Review Methods: Clinical Heterogeneity: Agency for Healthcare Research and Quality; 2010. Available from: http://effectivehealthcare.ahrq.gov/.
7. Clark GM. Prognostic factors versus predictive factors: Examples from a clinical trial of erlotinib. Mol Oncol. 2008;1(4):406-12.
8. Stange K, Glasgow R. Considering and Reporting Important Contextual Factors in Research on the Patient-Centered Medical Home. . Rockville, MD: Agency for Healthcare Research and Quality, 2013.
9. Cook JA, Ramsay CR, Fayers P. Using the literature to quantify the learning curve: a case study. Int J Technol Assess Health Care. 2007;23(2):255-60.
10. Eikermann M, Gluud C, Perleth M, Wild C, Sauerland S, Gutierrez-Ibarluzea I, et al. Commentary: Europe needs a central, transparent, and evidence based regulation process for devices. BMJ. 2013;346:f2771.
11. Baeyens H, Pouppez C, Slegers P, Vinck I, Hulstaert F, Neyt M. Towards a guided and phased introduction of high-risk medical devices in Belgium: Health Services Research (HSR) Brussels: Belgian Health Care Knowledge Centre (KCE); 2015.
12. Siebert M, Clauss LC, Carlisle M, Casteels B, de Jong P, Kreuzer M, et al. Health technology assessment for medical devices in Europe. What must be considered. Int J Technol Assess Health Care. 2002;18(3):733-40.
13. Schulenburg Graf v.d. JM, Mittendorf T, Kulp W, Greiner W. Health Technology Assessment (HTA) im Bereich der Medizinprodukte - gleiches Spiel mit gleichen Regeln? Gesundh ökon Qual manag. 2009;14(3):144-55.
14. Parquin F, Audry A. Clinical evaluation of medical devices: main constraints and specificities. Therapie. 2012;67(4):311-8.
15. Campbell G. Statistics in the world of medical devices: the contrast with pharmaceuticals. J Biopharm Stat. 2008;18(1):4-19.
16. Sedrakyan A, Marinac-Dabic D, Normand SL, Mushlin A, Gross T. A framework for evidence evaluation and methodological issues in implantable device studies. Medical care. 2010;48(6 Suppl):S121-8.
17. Konstam MA, Pina I, Lindenfeld J, Packer M. A device is not a drug. J Card Fail. 2003;9(3):155-7.
18. Sorenson C, Tarricone R, Siebert M, Drummond M. Applying health economics for policy decision making: do devices differ from drugs? Europace. 2011;13 Suppl 2:ii54-8.
19. Ergina PL, Cook JA, Blazeby JM, Boutron I, Clavien PA, Reeves BC, et al. Challenges in evaluating surgical innovation. Lancet. 2009;374(9695):1097-104.
20. Cook JA, Ramsay CR, Fayers P. Statistical evaluation of learning curve effects in surgical trials. Clin Trials. 2004;1(5):421-7.
21. Lao CS, Bushar HF. Longitudinal data analysis in medical device clinical studies. J Biopharm Stat. 2008;18(1):44-53.

22. Cook JA, McCulloch P, Blazeby JM, Beard DJ, Marinac-Dabic D, Sedrakyan A. IDEAL framework for surgical innovation 3: randomised controlled trials in the assessment stage and evaluations in the long term study stage. BMJ. 2013;346:f2820.
23. Ergina PL, Barkun JS, McCulloch P, Cook JA, Altman DG. IDEAL framework for surgical innovation 2: observational studies in the exploration and assessment stages. BMJ. 2013;346:f3011.
24. McCulloch P, Altman DG, Campbell WB, Flum DR, Glasziou P, Marshall JC, et al. No surgical innovation without evaluation: the IDEAL recommendations. Lancet. 2009;374(9695):1105-12.
25. McCulloch P, Cook JA, Altman DG, Heneghan C, Diener MK. IDEAL framework for surgical innovation 1: the idea and development stages. BMJ. 2013;346:f3012.
26. Anderson L, Oliver S, Michie S, Rehfuess E, Noyes J, Shemilt I. Investigating complexity in systematic reviews of interventions by using a spectrum of methods. J Clin Epidemiol. 2013;66(11).
27. Anderson LM, Petticrew M, Chandler J, Grimshaw J, Tugwell P, O'Neill J, et al. Introducing a series of methodological articles on considering complexity in systematic reviews of interventions. J Clin Epidemiol. 2013;66(11):1205-8.
28. Burford B, Lewin S, Welch V, Rehfuess E, Waters E. Assessing the applicability of findings in systematic reviews of complex interventions can enhance the utility of reviews for decision making. J Clin Epidemiol. 2013;66(11):1251-61.
29. Noyes J, Gough D, Lewin S, Mayhew A, Michie S, Pantoja T, et al. A research and development agenda for systematic reviews that ask complex questions about complex interventions. J Clin Epidemiol. 2013;66(11):1262-70.
30. Petticrew M, Anderson L, Elder R, Grimshaw J, Hopkins D, Hahn R, et al. Complex interventions and their implications for systematic reviews: a pragmatic approach. J Clin Epidemiol. 2013;66(11):1209-14.
31. Petticrew M, Rehfuess E, Noyes J, Higgins JP, Mayhew A, Pantoja T, et al. Synthesizing evidence on complex interventions: how meta-analytical, qualitative, and mixed-method approaches can contribute. J Clin Epidemiol. 2013;66(11):1230-43.
32. Pigott T, Shepperd S. Identifying, documenting, and examining heterogeneity in systematic reviews of complex interventions. J Clin Epidemiol. 2013;66(11):1244-50.
33. Squires JE, Valentine JC, Grimshaw JM. Systematic reviews of complex interventions: framing the review question. J Clin Epidemiol. 2013;66(11):1215-22.
34. Tugwell P, Knottnerus JA, Idzerda L. Complex interventions-how should systematic reviews of their impact differ from reviews of simple or complicated interventions? J Clin Epidemiol. 2013;66(11):1195-6.
35. The Danish Centre for Evaluation and Health Technology Assessment (DACEHTA). Introdcution to Mini-HTA – a management and decision support tool for the hospital service 2005. Available from: http://sundhedsstyrelsen.dk/publ/publ2005/cemtv/mini_mtv/introduction_mini_hta.pdf.
36. Department of Science and Technology (DECIT). Methodological Guidelines. Elaborating Studies for the Assessment of Medical Care Equipment [Diretrizes Metodológicas. Elaboração de Estudos para Avaliação de Equipamentos médico-assistenciais]. Brazilia: Ministry of Health; 2013. Available from: http://bvsms.saude.gov.br/bvs/publicacoes/diretrizes_metodologicas_elaboracao_estudos.pdf.
37. Haute Autorité de Santé (HAS). Medical Device Assessment in France. Paris 2009. Available from: http://www.has-sante.fr/portail/plugins/ModuleXitiKLEE/types/FileDocument/doXiti.jsp?id=c_930908.

38. Haute Autorité de Santé (HAS). Methodological Choices for the Clinical Development of Medical Devices 2013. Available from: http://www.has-sante.fr/portail/plugins/ModuleXitiKLEE/types/FileDocument/doXiti.jsp?id=c_1727556 .
39. Hulstaert F, Neyt M, Vinck I, Stordeur S, Huić M, Sauerland S, et al. The pre-market clinical evaluation of innovative high-risk medical devices. Brussels: Belgian Health Care Knowlegde Centre (KCE); 2011. Available from: https://kce.fgov.be/sites/default/files/page_documents/kce_158c_innovative_high-risk_medical_devices_0.pdf.
40. Royal Netherlands Academy of Arts and Sciences (KNAW). Evaluation of new technology in health care. In need of guidance for relevant evidence. Amsterdam2014. Available from: https://www.knaw.nl/en/news/publications/evaluation-of-new-technology-in-health-care-1/@@download/pdf_file/verkenning-new-technology-health-care.pdf.
41. National Institute for Health and Clinical Excellence (NICE). Medical Technologies Evaluation Programme. Methods Guide 2011. Available from: http://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-medical-technologies/Medical-technologies-evaluation-programme-methods-guide.pdf.
42. National Institute for Health and Clinical Excellence (NICE). The Diagnostics Assessment Programme manual 2011. Available from: http://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-diagnostics-guidance/Diagnostics-assessment-programme-manual.pdf.
43. College voor zorgverzekeringen (CVZ). Medical tests (assessment of established medical science and medical practice) 2011. Available from: http://www.zorginstituutnederland.nl/binaries/content/documents/zinl-www/documenten/publicaties/publications-in-english/2011/1101-medical-tests-assessment-of-established-medical-science-and-medical-practice/1101-medical-tests-assessment-of-established-medical-science-and-medical-practice/Medical+tests+%28assessment+of+established+medical+science+and+medical+practice%29.pdf.
44. Haute Autorité de Santé (HAS). Rapid Assessment Method for Assessing Medical and Surgical Procedures 2007. Available from: http://www.has-sante.fr/portail/plugins/ModuleXitiKLEE/types/FileDocument/doXiti.jsp?id=c_541199.
45. Higgins J, Green S. Cochrane handbook for systematic reviews of interventions version 5.1.0 2011. Available from: http://www.cochrane-handbook.org.
46. Centre for Reviews and Dissemination (CRD). Systematic Reviews. CRD’s guidance for undertaking reviews in health care 2009. Available from: https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf.
47. Schnell-Inderst P, Hunger T, Arvandi M, Conrads-Frank A, Siebert U. Recommendations for the Assessment of Comparative Effectiveness of Medical Devices. A Framework for Evaluation. 2015.
48. Food and Drug Administration (FDA). Design Considerations for Pivotal Clinical Investigations for Medical Devices. Guidance for Industry, Clinical Investigators, and Food and Drug Administration Staff 2013. Available from: http://www.fda.gov/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm373750.htm.
49. Boutron I, Tubach F, Giraudeau B, Ravaud P. Blinding was judged more difficult to achieve and maintain in nonpharmacologic than pharmacologic trials. J Clin Epidemiol. 2004;57(6):543-50.
50. U.S. Preventive Services Task Force Procedure Manual (USPSTF). U.S. Preventive Services Task Force Procedure Manual. AHRQ Publication No. 08-05118-EF:

Agency for Healthcare Reserach and Quality (AHRQ); 2008. Available from: http://www.uspreventiveservicestaskforce.org/Page/Name/procedure-manual.
51. U.S. Preventive Services Task Force. Universal Screening for Hearing Loss in Newborns: US Preventive Services Task Force Recommendation Statement. Pediatrics. 2008;122(1):143-8.
52. Commission d'Evaluation des Produits et Prestations. Avis de la Commission: ASR XL Head: Haute Autorité de Santé; 2008. Available from: http://www.has-sante.fr/portail/upload/docs/application/pdf/2008-07/cepp_1737_asr_xl_head.pdf.
53. Commission d'Evaluation des Produits et Prestations. Avis de la Commission: Metasul: Haute Autorité de Santé; 2009. Available from: http://www.has-sante.fr/portail/upload/docs/application/pdf/2009-02/cepp-1795_metasul.pdf
54. Smith AJ, Dieppe P, Vernon K, Porter M, Blom AW, National Joint Registry of E, et al. Failure rates of stemmed metal-on-metal hip replacements: analysis of data from the National Joint Registry of England and Wales. Lancet. 2012;379(9822):1199-204.
55. Commission Nationale d'Évaluation des Dispositifs Médicaux et des Technologies de Santé. Rapport sur les prothèses totales de hanche à couple de frottement métal-métal: Haute Autorité de Santé; 2013. Available from: http://www.has-sante.fr/portail/upload/docs/application/pdf/2013-07/rapport_protheses_totales_de_hanche_metal-metal.pdf (English summary at: http://www.has-sante.fr/portail/upload/docs/application/pdf/2013-07/synthesis_metal_on_metal_hip_implants_dm-eval_54.pdf.
56. van Barabandt H, Neyt M, Devos C. Catheter ablation of atrial fibrillation. KCE report 184C [Internet]. 2012. Available from: https://kce.fgov.be/publication/report/catheter-ablation-of-atrial-fibrillation.
57. European network for Health Technology Assessment (EUnetHTA). Process of information retrieval for systematic reviews and health technology assessments on clinical effectiveness 2015. Available from: http://www.eunethta.eu/news/closed-public-consultation-draft-methodological-guideline-process-information-retrieval-systema.
58. Chang L, Dhruva SS, Chu J, Bero LA, Redberg RF. Selective reporting in trials of high risk cardiovascular devices: cross sectional comparison between premarket approval summaries and published reports. BMJ. 2015;350:h2613.
59. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009;339:b2535.
60. Bernard A, Vaneau M, Fournel I, Galmiche H, Nony P, Dubernard JM. Methodological choices for the clinical development of medical devices. Medical devices. 2014;7:325-34.
61. Lilford RJ, Braunholtz DA, Greenhalgh R, Edwards SJ. Trials and fast changing technologies: the case for tracker studies. BMJ. 2000;320(7226):43-6.
62. Food and Drug Administration (FDA). Guidance for Industry and FDA Staff: Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials 2010. Available from: http://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm071121.pdf.
63. Bonangelino P, Irony T, Liang S, Li X, Mukhi V, Ruan S, et al. Bayesian approaches in medical device clinical trials: a discussion with examples in the regulatory setting. J Biopharm Stat. 2011;21(5):938-53.
64. Devereaux PJ, Bhandari M, Clarke M, Montori VM, Cook DJ, Yusuf S, et al. Need for expertise based randomised controlled trials. BMJ. 2005;330(7482):88.
65. Horwitz RI, Feinstein AR. Advantages and drawbacks of the Zelen design for randomized clinical trials. J Clin Pharmacol. 1980;20(7):425-7.

66. Zelen M. A new design for randomized clinical trials. N Engl J Med. 1979;300(22):1242-5.
67. Wennberg JE, Barry MJ, Fowler FJ, Mulley A. Outcomes research, PORTs, and health care reform. Ann N Y Acad Sci. 1993;703:52-62.
68. Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ. 2008;336(7644):601-5.
69. Ramsay CR, Wallace SA, Garthwaite PH, Monk AF, Russell IT, Grant AM. Assessing the learning curve effect in health technologies. Lessons from the nonclinical literature. Int J Technol Assess Health Care. 2002;18(1):1-10.
70. Lauer MS, D'Agostino RB, Sr. The randomized registry trial--the next disruptive technology in clinical research? N Engl J Med. 2013;369(17):1579-81.
71. European Network for Health Technology Assessment (EUnetHTA). Internal validity of non-randomised studies (NRS) on interventions 2015. Available from: http://www.eunethta.eu/news/closed-public-consultation-draft-methodological-guideline-internal-validity-non-randomised-stud.
72. Berger ML, Martin BC, Husereau D, Worley K, Allen JD, Yang W, et al. A questionnaire to assess the relevance and credibility of observational studies to inform health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force report. Value Health. 2014;17(2):143-56.
73. Gliklich R, Dreyer N, Leavy M, eds. Registries for Evaluating Patient Outcomes: A User’s Guide. Rockville: Agency for Healthcare Research and Quality (AHRQ); 2014. Available from: http://www.effectivehealthcare.ahrq.gov/registries-guide-3.cfm.
74. Zaletel M, Kralj M, [editors]. Methodological guidelines and recommendations for efficient and rational governance of patient registries. Ljubljana: National Institute of Public Health; 2015. Available from: http://patientregistries.eu/deliverables?p_p_id=110_INSTANCE_E2SKlJ9mcegy&p_p_lifecycle=0&p_p_state=normal&p_p_mode=view&p_p_col_id=column-2&p_p_col_count=1&_110_INSTANCE_E2SKlJ9mcegy_struts_action=%2Fdocument_library_display%2Fview_file_entry&_110_INSTANCE_E2SKlJ9mcegy_redirect=http%3A%2F%2Fpatientregistries.eu%2Fdeliverables%3Fp_p_id%3D110_INSTANCE_E2SKlJ9mcegy%26p_p_lifecycle%3D0%26p_p_state%3Dnormal%26p_p_mode%3Dview%26p_p_col_id%3Dcolumn-2%26p_p_col_count%3D1&_110_INSTANCE_E2SKlJ9mcegy_fileEntryId=35801.
75. Guo B, Moga C, Harstall C, Schopflocher D. A quality appraisal checklist developed specifically for evaluating case series studies. J Clin Epidemiol. 2015.
76. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol. 2007;7:10.
77. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. J Clin Epidemiol. 1991;44(11):1271-8.
78. Caro J, Eddy DM, Kan H, Kaltz C, Patel B, Eldessouki R, et al. Questionnaire to assess relevance and credibility of modeling studies for informing health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force report. Value Health. 2014;17(2):174-82.
79. Jansen JP, Trikalinos T, Cappelleri JC, Daw J, Andes S, Eldessouki R, et al. Indirect treatment comparison/network meta-analysis study questionnaire to assess relevance and credibility to inform health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force report. Value Health. 2014;17(2):157-73.
80. Puhan MA, Schunemann HJ, Murad MH, Li T, Brignardello-Petersen R, Singh JA, et al. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. BMJ. 2014;349:g5630.

81. Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JP. Evaluating the quality of evidence from a network meta-analysis. PLoS One. 2014;9(7):e99682.
82. Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. Ann Intern Med. 2015;162(11):777-84.
83. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ. 2014;348:g1687.
84. Borenstein M, Hedges L, Higgins J, Rothstein H. Introduction to meta analysis: statistics in practice. Chichester, UK: Wiley; 2009.
85. Mulier S, Ni Y, Jamart J, Ruers T, Marchal G, Michel L. Local recurrence after hepatic radiofrequency coagulation: multivariate meta-analysis and review of contributing factors. Ann Surg. 2005;242(2):158-71.
86. Egger M, Davey Smith G, Altman DG. Systematic Reviews in Health Care: Meta-Analysis in Context, Second Edition: Wiley; 2008.
87. Hartling L, McAlister FA, Rowe BH, Ezekowitz J, Friesen C, Klassen TP. Challenges in systematic reviews of therapeutic devices and procedures. Ann Intern Med. 2005;142(12 Pt 2):1100-11.
88. Pablos-Mendez A, Barr RG, Shea S. Run-in periods in randomized trials: implications for the application of results in clinical practice. JAMA. 1998;279(3):222-5.

Annexe 2. Documentation of literature search

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