CORE-MD

Coordinating Research and Evidence for Medical Devices

Recommendations for a hierarchy of clinical evidence for high-risk medical devices

Deliverable 4.3

Deliverable factsheet

Source Activity:	Work package n. 4, Task 4.2
Title:	Recommendations for a hierarchy of clinical evidence for high-risk medical devices
Lead Beneficiary:	ESC
Nature:	Report
Dissemination level:	Public
Editor:	Alan Fraser (ESC)
Authors:	CORE-MD Consortium
Status:	Final
Date:	31/03/2024
Contractual Delivery Date:	M36

Version Log

Issue Date	Version	Involved	Comments
15/11/2023	0.1	Alan Fraser (ESC)	First version prepared by the Coordinator as the basis for discussion in the stakeholder workshop held on November 21st, 2023
14/03/2024	1.1	All consortium members, in person	Final Project Board, revision with new tables, on agenda for consultation.
29/03/2024	1.2	Circulated to whole consortium (email)	Draft final version; many comments received; then refined in a series of conference calls.
31/05/2024	2.0	Alan Fraser (ESC) Valentina Tageo (ESC)	Final consolidated version
29/08/2024	*2.1*	*Alan Fraser (ESC)* *Valentina Tageo (ESC)*	*Final consolidated version, revised with addition of paragraph 2.2.1 as requested by reviewers (concerning recommendations for performing sham interventions).*

Table of Contents

Index of figures

Index of tables

Executive Summary

The EU Horizon 2020 call that was answered by CORE–MD requested that the successful consortium should carry out a review of the currently used clinical investigation designs for the evaluation of high-risk medical devices and provide a "hierarchy" of these approaches. Detailed systematic reviews were therefore performed (in Task 1.1) of the clinical evidence and study designs used for high-risk medical devices in cardiology and cardiac surgery, orthopaedic surgery, diabetic medicine, and paediatric practice. They all demonstrated that little information from clinical trials is publicly available at the time that new high-risk devices are granted market access in the EU, and secondly that a minority of all clinical studies are randomised trials.

Secondly, Task 1.4 of CORE–MD performed a very detailed **review of recommendations** for the design, methodologies, and conduct of pivotal clinical investigations of medical devices, contained in 30 regulatory guidance documents, 12 standards from the International Organization for Standardization, and 4 consensus statements from academic research consortia. Levels of evidence and types of investigations were addressed in 4 guidance documents, but as a hierarchy in only two, from the US FDA and the Australian TGA.

Thirdly, **preparation of a proposal for a "hierarchy"** was conducted within Task 2.2 of the CORE–MD project as a collaboration between clinical trialists in Uppsala University, and epidemiological and methodological experts in the Clinical Trials Unit of Oxford University. Task members included the leaders of the IDEAL Collaboration and the Good Clinical Trials Collaborative. This group strongly recommended that more randomised controlled trials of new high-risk medical devices should be conducted, before they are approved; essential principles were elaborated in a scientific manuscript on "Large simple trials" intended to simplify the *conduct* of RCTs without reducing their quality. The key features are randomization, and efficient management and data collection achieved by exploiting existing electronic platforms and databases (including registries), which can make studies inclusive and affordable, and able to provide results that are widely generalizable to real-life clinical practice.

A simple proposal for a logical sequence of clinical study designs for new high-risk medical devices was prepared in graphical format and presented for discussion to CORE–MD consortium partners at a Project Board meeting; separately to members of the international Advisory Board of CORE–MD; and then also to European medical device regulators at a Workshop with the Clinical Investigation and Evaluation Working Group (CIE) of the Medical Device Coordination Group (MDCG).

Finally, the draft recommendations were considered in detail at a CORE–MD Workshop held in Brussels on 21st November 2023, to which all consortium members were invited. The consensus view was that it is difficult and probably unhelpful to recommend a single hierarchy, because the most appropriate methodology for a clinical study will vary according to the stage of development of a device and the context of its use. As well as general proposals, two further tables were drafted to show optimal methodologies for the design and conduct of clinical investigations to be used for regulatory submissions, when a device is an innovation for a serious condition for which there is no effective

alternative, or it is a device to treat an orphan indication or disease; and when a new device will be offered as an alternative to existing devices. In the case of innovative and orphan devices, the CORE–MD consortium recommends that any approval on the basis of limited evidence (i.e. with some persisting uncertainty about outcomes) would need to be balanced by mandatory post-market clinical studies.

These recommendations summarise important principles that determine the quality of study design and conduct. A primary requirement must be full transparency of objectives, methodology, inclusion criteria, interventions, analyses, and outcomes. To reach valid conclusions, a trial must be sufficiently powered for the primary outcome measure, with high completeness of follow-up and sufficient duration to detect relevant safety concerns. Risk of bias should be minimized by the study design. The optimal design is a randomised controlled trial (RCT) with at least 80% power at $p<0.05$ significance level to detect a realistic effect on a patient-relevant outcome; with comparison of the device versus a state-of-the-art alternative device (or effective medical treatment) or a sham intervention; with independent determination of validated endpoints or events by blinded observers; and with complete follow-up. A high proportion of patients should be recruited within the population where the approved device will be used.

The tables list alternative designs that are recommended for clinical studies of high-risk devices during 4 stages of their development and introduction into clinical practice (see pages 18 – 20); the stages are called **initial** (referring to first-in-human and preliminary clinical studies); **early** (relating to the assessment of performance, safety, and positive benefit–risk ratio, preparing for later trials powered for efficacy); **definitive (pivotal**) (for confirmation of efficacy for clinical outcomes, and further demonstration of safety); and **long-term (post-market)** (when the objectives are long-term monitoring of device performance and safety, in comparison against alternatives).

This report includes detailed appendices that summarise relevant guidance.

The recommendations from this CORE–MD task have been presented to EU regulators.

1 Introduction

1.1 Background to this CORE–MD recommendation

The EU Horizon 2020 call for a Coordination and Support Action, that was answered by the CORE–MD consortium, stated that there is a need for methodologies to generate improved clinical evidence which can be addressed "by developing and promoting methodological approaches adapted to the specificities of high-risk medical devices".[1] It requested in particular that the consortium should carry out a review of the currently used clinical investigation designs for the evaluation of such devices, and provide a hierarchy of these approaches.

Detailed systematic reviews were thus performed of the clinical evidence and study designs used for high-risk medical devices in cardiology and cardiac surgery [1], orthopaedic surgery [2], diabetic medicine [3], and paediatric practice [4]. They demonstrated that little information from clinical trials is publicly available at the time that new high-risk devices are awarded a certificate of conformity (i.e. approved) in the EU, and secondly that a minority of studies in all fields are randomised trials.

A survey by the CORE–MD consortium (within Task 4.3) confirmed that European medical device regulators and notified body assessors identify a need for methodologies to evaluate specific high-risk devices and to assess benefit-risk ratio and thresholds for acceptability [5]. Clinicians serving in Expert Panels or prescribing high-risk devices also mentioned clinical study designs, their advantages and disadvantages and choice of comparators (standard of care vs. sham vs. placebo), as topics on which they wish to have further guidance and education.

This consensus recommendation builds on insights from all those studies. It should be considered together with CORE–MD advice concerning methodologies that are appropriate for the clinical evaluation of high-risk medical devices for use in infants and children [6].

1.2 Preparation for this CORE–MD Recommendation

1.2.1 Study design recommendations in guidance documents for high-risk medical devices

Task 1.4 of CORE–MD performed a very detailed review of recommendations for the design, methodologies, and conduct of pivotal clinical investigations of medical devices.[2] The study selected 30 regulatory guidance documents, 12 standards from the International Organization for Standardization, and 4 consensus statements from academic research consortia.[3]

[1] SC1-HCO-18-2020: Developing methodological approaches for improved clinical investigation and evaluation of high-risk medical devices.

[2] Protocol published at https://osf.io/w9b5k/

[3] See Deliverable 1.6.

Levels of evidence and types of investigations were addressed in 4 guidance documents, but as a hierarchy in only two. Those documents, from the US FDA and the Australian TGA, are summarised below in Appendices 3 and 4 and then compared in Appendix 5.

Choice of study design was discussed in 8 reports. In comparison, more commonly addressed topics were PICO [Patient, population / Intervention / Comparison / Outcomes], in 18 guidance documents, and how studies are reported and recorded, in 17.

The review confirmed that at the time of sampling for the study (November 2021), there had not yet been any detailed recommendations from EU regulators on which clinical study methodology is most appropriate for investigations of high-risk medical devices at each stage in their life cycle. It is noted, however, that many guidance documents have been published more recently by the Medical Device Coordination Group of EU regulators that is chaired by the European Commission, and new guidance on clinical evaluation is in preparation. Some earlier and less specific, but still related, guidance from the EU is summarised in Appendix 6.

1.2.2 New designs for randomised clinical trials and studies of high-risk medical devices

Task 2.2 within the CORE–MD project was conducted as a collaboration between cardiovascular clinical trialists in Uppsala University, and epidemiological and methodological experts in the Clinical Trials Unit of Oxford University. Members of the study group included the leaders of the IDEAL Collaboration and of the Good Clinical Trials Collaborative (GCTC). The study designs considered appropriate during each stage of the IDEAL framework are shown in Appendix 7, and the contents of the recent GCTC recommendations are outlined in Appendix 8.

A workshop was held in Oxford on 9th November 2022, leading to a scientific manuscript on the essential principles of simplifying the conduct of clinical trials, that has been submitted for publication[4]. The paper draws on experience from running the RECOVERY platform trial of pharmacological therapies for COVID–19 (in Oxford), and TASTE as the first registry-based trial of a high-risk medical device (in Uppsala). Embedding a trial within a medical device registry or other electronic platform can accelerate recruitment, increase efficiency, and reduce costs. The CORE–MD consortium strongly recommends that more randomised controlled trials of new high-risk medical devices should be conducted, before they are approved.

A second output from this task was the preparation of a proposal for a hierarchy of clinical study designs for new high-risk medical devices. The outline that is summarised in Figure 1was presented for discussion to CORE–MD consortium partners at a Project Board meeting; separately to members of the international Advisory Board of CORE–MD; and then also to European medical device regulators at a Workshop with the Clinical Investigation and Evaluation Working Group (CIE) of the Medical Device Coordination Group (MDCG).

[4] See Deliverable 2.2.

The second and third boxes on the right of Figure 1 recommend a logical sequence of study designs, that if followed would progressively accumulate reliable evidence. The dotted line is intended to show that this scheme does not correspond exactly with the "clinical development stages" in current regulatory frameworks, such as the one illustrated in Table 1, taken from Appendix I of EN ISO standard 14155 [7] that has been harmonised into EU legislation. This makes it difficult to gauge any single step at which regulatory review would be appropriate for all devices.

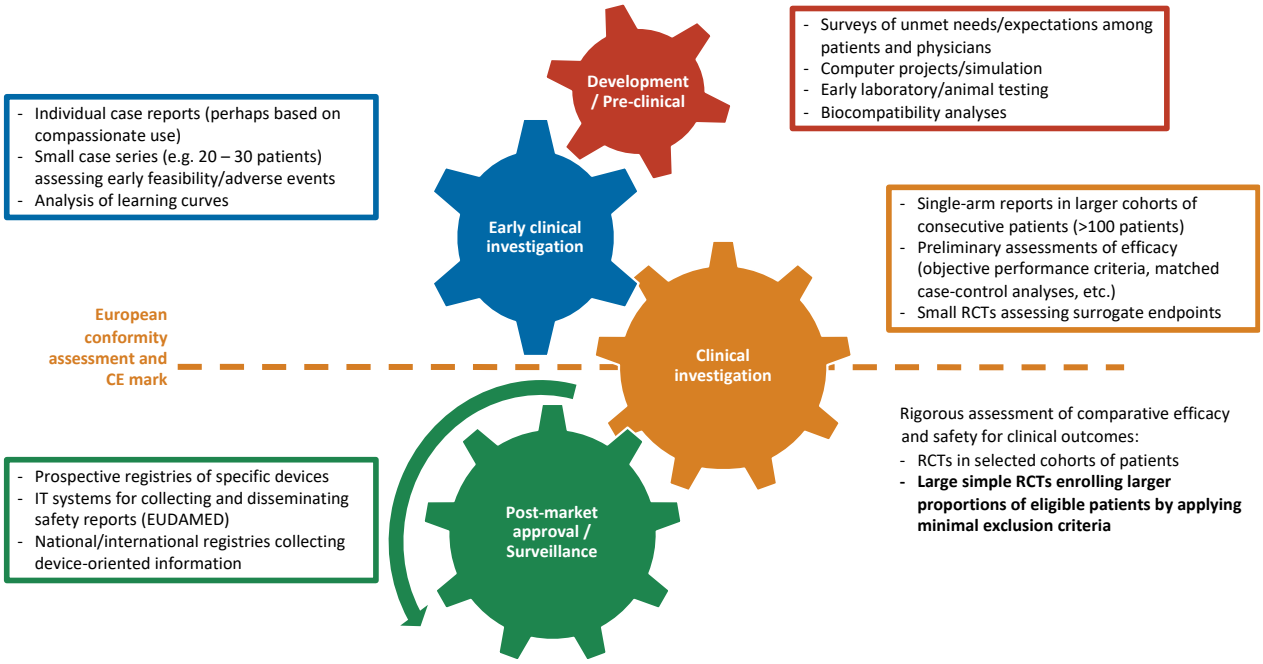

Figure 1. Recommendations on study design from CORE–MD task 2.2

Table 1. Clinical development stages, as recommended by EN ISO standard 14155

Regulatory status	Pre–market		Post–market	
Clinical development stage	Pilot stage	Pivotal stage	Post–market stage	
Type of design	Exploratory or confirmatory	Confirmatory		Observational
Descriptors of clinical investigations	• First in human clinical investigation • Early feasibility clinical investigation • Traditional feasibility clinical investigation	• Pivotal clinical investigation	• Post-market clinical investigation	• Registry • Post-market clinical investigation
Burden to subject	Interventional			Non-Interventional

1.3 Preparation of this report

The specific objective of the task described in this document is "to propose a hierarchy of clinical investigations for high-risk medical devices, ranking trial methodologies and clinical study designs according to the results of CORE–MD work packages 1–3".

In addition to the outputs already summarised, account was taken of the framework for appraising the regulatory utility of real-world evidence (including device registries) that was developed and proposed in Task 3.1 of CORE–MD.[5] Recommendations for clinical study designs to evaluate coronary stents were also relevant; they were prepared in 2015 by members of the European Association of Percutaneous Cardiovascular Interventions (EAPCI) in response to a request from the Clinical Investigation and Evaluation Working Group of the EU Medical Device Coordination Group [8]. That project was led by colleagues who were also partners in CORE–MD.

A CORE–MD Workshop was held in Brussels on 21st November 2023, to consider these recommendations. All consortium members were invited, and participants included members of the Advisory Board and colleagues from notified bodies, EU national regulatory agencies, and the European Commission (all attending in their personal capacity). Preparation of the report was led by consortium members from the European Society of Cardiology (ESC) and the European Federation of National Societies of Orthopaedics and Traumatology (EFORT). Draft conclusions were presented to all

[5] See Deliverable 3.1.

consortium partners at the final meeting of the CORE–MD Project Board on 14th March 2024; and final revisions were then prepared in a series of conference calls between the primary authors, and circulated for final consensus approval.

1.3.1 Guiding principles

The intention has been to prepare general guiding principles, based on the quality of scientific evidence from different study methodologies, unconstrained by current EU guidance documents or regulatory practices. It is recognised that the suggestions contain elements that may require evolution of the EU system of approval of high-risk medical devices. Developments in 'regulatory science' should focus on establishing which essential features are required to minimise risks while permitting patients to benefit from devices that improve clinical outcomes – with the minimum possible administrative burden.

1.3.2 Life-cycle stages

The CORE–MD consortium has recommended methodologies for clinical studies and trials during four phases in the life cycle of devices – called initial, early, definitive, and long-term.

All studies should be rigorous, so that term was not used to describe large-scale RCTs. The term "pivotal" has also not been used, because of variations in how it has been applied in different guidance documents – but it corresponds most closely to "definitive" in the CORE–MD tables.

The following table suggests how these four stages, used in this CORE–MD recommendation, relate to the terms used in EU legislation and guidance, and to the stages described by the IDEAL collaboration:

Table 2. Comparison of phases / stages of clinical evaluation of medical devices

Initial clinical studies	Early clinical studies	Definitive clinical studies	Long-term clinical studies
EU Medical Device Regulation			
• First-in-man • Feasibility • Pilot studies		• Confirmatory pivotal investigations	• Post-market clinical follow-up
ISO 14155			
• First-in-human • Early/feasibility	• Exploratory or confirmatory	• Confirmatory / pivotal clinical investigations	• Post-market clinical investigations
IDEAL Collaboration			
• Idea (First-in-human) (stage 1)	• Exploration (Stage 2b)	• Assessment (Stage 3)	• Long-term study (Stage 4)

• Development (Stage 2a) [e.g. cohort study]	[e.g. multicentre cohort study]	[e.g. randomised controlled trial]	[e.g. high-quality registry]

The most important recommendation in this document from the CORE–MD consortium is that requirements for clinical evidence before approval can be varied for innovative or orphan devices, compared with devices entering an existing market. That implies that 'early' and 'definitive' stages need to be kept separate, without either being linked in all cases to the timing of EU regulatory approval (conformity assessment).

1.4 Structure of this report

Section 1 introduces the objectives and methodology of this task

Section 2 provides the main recommendations, in text as bullet points, and in 3 key tables

Section 3 is a short summary with conclusions

The appendices show extracts or summarise relevant previous regulatory guidance from:

- The FDA in the USA
- The TGA in Australia
- The EU

And from two academic consortia:

- The IDEAL Collaboration
- The Good Clinical Trials Collaborative

2 CORE–MD Recommended study designs for clinical investigations of high-risk medical devices

2.1 Stages and context of clinical investigations

- The EU Medical Device Regulation (MDR) [9] at Annex XIV on Clinical Evaluation (paragraph 1) refers to "exploratory investigations, such as first-in-man studies, feasibility and pilot studies" and to "confirmatory investigations, such as pivotal investigations". Annex XV on Clinical Investigations states (at section 3 on the Clinical Investigation Plan, paragraph 3.6) that the manufacturer will describe the "design of the clinical investigation with evidence of its scientific robustness and validity".
- EU guidance on the clinical evaluation of devices was published in 2016 in the context of the EU medical device directives. MEDDEV 2.7/1 (revision 4) [10] stated in its introduction that "The depth and extent of clinical evaluations should be flexible and appropriate to the nature, intended purpose, and risks of the device in question", and Paragraph 9.3.1.a listed factors that should be evaluated when appraising the study design of pre-market clinical investigations. The document did not indicate any specific requirements for particular types of devices nor did it rank study designs in order of their scientific merit. It will be replaced by new guidance that is being prepared for the MDR.
- The harmonised EU standard on clinical investigation of medical devices (EN ISO 14155) [7] provides a framework for considering the stages of clinical evidence [see appendix 2(b) of this document]. CORE–MD reviews have demonstrated that by placing pivotal investigations only in the pre-market phase, it does not reflect current EU practices. The standard does not indicate optimal study designs at each stage.
- An academic consortium, the IDEAL collaboration, advises 6 stages for evaluating medical devices, called preclinical, idea (first-in-human), development (early clinical studies), exploration (larger, collaborative studies), assessment (RCTs), and long-term (see appendix 7 below) [11].
- EU legislation and guidance does not describe stages of clinical investigations in detail, with specific methodologies for acquiring clinical evidence needed for certification – so this CORE–MD recommendation refers to a simple progression in four stages from 'initial' to 'early' to 'definitive' to 'long-term' clinical studies; associations between these terms and the cited documents are suggested in appendix 2c. The objectives for clinical evaluation and studies at each stage are indicated by the text in italics at the top of each column in the table (page 3).
- The main table in this recommendation summarises major study methodologies in order, as requested by the European Commission (p. 3).
- Two further tables indicate the consensus views of the members of the CORE–MD consortium, on optimal methodologies for the design and conduct of clinical investigations to be used for regulatory submissions, in two particular circumstances:

1. when a device is an innovation for a serious condition for which there is no effective alternative, or it is a device to treat an orphan indication or disease. In either case any approval on the basis of limited evidence (i.e. with some persisting uncertainty about outcomes) would need to be balanced by mandatory post-market clinical studies (see page 4, example 1); and
2. when a new device will be offered as an alternative to existing devices (page 5, example 2).

2.2 Quality of study design and conduct – principles

- A primary requirement for all clinical investigations must be full transparency of objectives, methodology, inclusion criteria, interventions, analyses, and outcomes. They should conform with the FAIR (findable, accessible, interoperable, reusable) principles for sharing data.
- Any recommended 'hierarchy' of designs presupposes that each study will be performed to a high scientific standard. To reach valid conclusions, a trial must be sufficiently powered for the primary outcome measure, with high completeness of follow-up and sufficient duration to detect relevant safety concerns.
- Risk of bias should be minimized by the study design. An open-label study that is underpowered, short in duration, and has high loss to follow-up, will be poor-quality and unreliable even if randomised and controlled. A well-designed and powered observational cohort or registry study with long and complete follow-up, independent adjudication of events, and minimisation or control of possible confounding factors, would then likely be more informative and valuable [12].
- The optimal design is a randomised controlled trial (RCT) with at least 80% power at $p<0.05$ significance level to detect a realistic effect on a patient-relevant outcome; with comparison of the device versus a state-of-the-art alternative device (or effective medical treatment) or a sham intervention; with independent determination of validated end-points or events by blinded observers; and with complete follow-up. A high proportion of patients should be recruited within the population where the approved device will be used.
- Large RCTs should be encouraged but they may be feasible only once the design of a new high-risk device has become stable (mature). Otherwise the appropriate design will depend on the risk class, the stage of development of the device, its context of use, and patients' perception of the balance between benefit and risks. The examples given in the tables indicate preferred designs, but the lists are not exhaustive and the options are not necessarily alternatives.
- Very few *double-blind* RCTs of high-risk devices will be possible.
- A randomised design is more important than blinding.
- The essential features of RCTs have been published by the Good Clinical Trials Collaborative [13].
- Each trial should be registered in advance and its protocol should be published in an open-access database, before recruiting patients.

- Post-market surveillance of high-risk devices should be representative of all patients, and comprehensive (with the recommended target for inclusion in registries >95% of eligible patients [14]), and long-term (e.g. >5 years, to reveal any late adverse effects). Protocols for studying real-world evidence should be publicly registered. Post-market studies should assess the effectiveness of devices according to their intended medical purpose. Results from registries might also support applications to expand approved clinical indications for a device.
- The terms “prospective” and “retrospective” are not applied consistently [15] so use of these descriptors in the tables is limited. Prospective refers to data collection according to a predefined study protocol, after which all data are measured and collected.

2.2.1 Ethical considerations

The Declaration of Helsinki states (as Principle 33) that “The use of placebo, or no treatment, is acceptable in studies where no current proven intervention exists”. It then adds the important qualification that “Where for compelling and scientifically sound methodological reasons the use of placebo is necessary to determine the efficacy or safety of an intervention”, extreme care must be taken to ensure that “the patients who receive placebo or no treatment will not be subject to any risk of serious or irreversible harm”. [16]

There can be a strong placebo benefit from surgical procedures [17] and from the implantation of medical devices. Sham procedures offer a robust methodological approach to minimize the risk of bias due to the “placebo effect”[18]; they may be the only way to confirm real benefit from a high-risk device or conversely to demonstrate that there is no benefit from its use. [19]

These CORE–MD recommendations are intended to encourage investigators to undertake a greater number of randomised trials. When there are other devices already approved for the same clinical indication, then the optimal comparator will be another active device (see Table 3, below). If there is no alternative device, and there is clinical equipoise because of suspicion of a possible significant placebo benefit, then a sham intervention may be appropriate. [20] Randomised trials with well-controlled and ethically approved sham procedures have provided important results concerning cardiovascular devices. [21] Consensus recommendations for interventional trials in a non-life threatening condition such as osteoarthritis advised that it is sometimes appropriate to consider a 3-arm trial, with patients being randomised to active, sham, and no-treatment groups. [22]

Taking these considerations into account, the CORE-MD consortium recommends:

- The use of sham procedures in RCTs of high-risk medical devices should be considered, to strengthen the methodological quality of clinical investigations particularly when the potential for bias due to a placebo effect may be substantial (for example, when assessing long-term effects on symptoms, using patient-reported outcomes).

- A thorough risk-benefit analysis is essential to demonstrate that the knowledge gained from a methodologically rigorous sham-controlled RCT will outweigh any risks associated with the sham procedure (ensuring compliance with the ethical principles of beneficence and non-maleficence).
- Information provided to patients involved in RCTs with a sham procedure should be comprehensive, transparent, and easily accessible.
- Upon completion of the RCT, patients who underwent a sham procedure should be offered the opportunity to cross over to the active treatment, if the results favour the therapeutic intervention.

Further background information and explanatory material are provided in the Appendices.

Table 3. CORE–MD RECOMMENDATIONS FOR STUDY DESIGNS AND METHODOLOGIES

Initial clinical studies	Early clinical studies	Definitive (pivotal) clinical studies *	Long-term (post-market) clinical follow-up study *
First-in-human and preliminary clinical studies: ***All experience to be publicly reported.*** • Case report(s) of first implants or other first use of a new high-risk device.[6] • Observational studies assessing feasibility, safety, and early adverse events.[7]	***Assessment of performance, safety, and positive benefit–risk ratio, preparing for later trials powered for efficacy [8]:*** • Observational study (e.g. single-arm, enrolling >150 consecutive patients), using patient-relevant outcomes [23] and/or validated surrogate end-points [24][25]. • Observational study testing against objective performance criteria (OPCs) [27], with analysis of learning curves. • Case-control or cohort study, assessing differences against another device or current state-of-the-art, and adequately designed to minimise bias.	***Confirmation of efficacy for clinical outcomes [3], and further demonstration of safety:*** • Double-blind RCT, if feasible. • Single-blinded RCT against active comparator [9] – powered for 'superiority'. [10] • 'Assessor-blinded' RCT with sham intervention (if no active comparator available). [11] • Single-blinded RCT (as above) – powered for non-inferiority. • Large multicentre observational study, using OPCs or other validated outcome measures.	***Long-term monitoring of device performance and safety, in comparison against alternatives:*** • 'Large simple' RCT such as a registry-based trial. [12] ** • RCT in enriched cohorts. [13] • Well-designed observational study using a registry, or other real-world source of data, including all devices of the same type, and with results combined through a federated analysis, using appropriate adjustments to minimise bias.

*RCT randomised controlled trial / * Items listed in order of recommended priority / ** Encompasses registry trials, nested trials, etc.*

[6] When use of a new device has been approved for each patient on the basis of an individual humanitarian exemption, data collected retrospectively as a "compassionate-use case series" should be considered insufficient for regulatory approval of that device.

[7] Clinical performance and outcomes documented prospectively in a consecutive series of all patients receiving the device, with a pre-planned common protocol for the method of delivery/use of the device and for all documentation, which may be modified iteratively if necessary in response to initial experience and outcomes.

[8] Against active comparator, 'state-of-the-art' alternative device, or best alternative treatment; and with (co-determined) patient-relevant end-points.

[9] Another device already approved for the same clinical indication, whether of similar or alternative design.

[10] 'Superiority' is a commonly used term, but it refers to a study designed to accept or reject a null hypothesis of no difference (with a two-sided test for significance). See [26].

[11] 'Assessor-blinded' means with independent ascertainment of end-points by blinded observers.

[12] Large trial with minimal exclusion criteria and conduct simplified by embedding study within electronic health record or similar. [See Deliverable 2.2]

[13] If required as a condition on the certificate of conformity, or if a device has been approved with limited evidence because it addresses an important unmet clinical need.

Example 1: Recommendations for clinical investigations of an innovative or orphan medical device

Relevant only when these three conditions are all satisfied:

1. There has been prior determination that the new device can provide for an unmet need (defined independently by experts & patients).
2. The device will be used to diagnose or treat a severe disease or condition.
3. No alternative devices have already been approved for the same indication.

CE

	Initial clinical studies	Early clinical studies	Definitive clinical studies	Long-term clinical evaluation
Preferred designs	• Case report(s) of first implants or other first use. • Planned case series with prospective collection of data.	• Prospective observational study (e.g. single-arm with all consecutive patients). • Comparative study with concurrent matched control subjects.	• RCT versus current 'state of the art', with blinded determination of clinical end-points. [14] • Open-label RCT. • Observational cohort study, with concurrent matched controls.	• Mandatory registry.
Not recommended	• Retrospective review of individual cases.	• Comparative study with historical controls.[15]		• Registry with recruitment <90%.

The shaded column indicates the stage when an EU certificate of conformity / CE mark may be awarded with conditions *for the subsequent collection of definitive evidence, with continued market access thereafter being dependent on confirmation of positive clinical impact.*

[14] The sample size should be sufficient to provide at least 80% power at the two-tailed p value of <0.05 to detect a clinically meaningful effect size; it will depend on the anticipated event rate in the control group (for categorical outcomes such as hospital admission or death) and/or the precision with which the endpoint is measured (for other outcomes such as blood pressure, symptoms, quality of life, etc.). See [28].

[15] But it may sometimes be ethical to compare new life-saving devices only against outcomes in patients for whom no effective options were available.

Example 2: Recommendations for clinical investigations of a new medical device of a well-established type

Relevant for a new high-risk medical device intended to treat an indication for which alternative devices have already been approved:

- Same risk class and type of device (according to EMDN code*) as a well-established technology, but with no claimed equivalence.
- A new device should then be approved in the EU only if it is demonstrated to be at least as safe and effective as the alternative devices.
- The clinical investigations must include head-to-head comparisons with an existing device that represents 'state of the art'.

CE

	Initial clinical studies	Early clinical studies	Definitive clinical studies	Long-term clinical evaluation
Preferred designs	• Case report(s) of first implants or other first use. • Inclusive case series.	• RCT with surrogate end-point(s).[16] • Observational study with objective performance criteria. • Cohort study with matched control subjects.	• RCT against active comparator [17]; powered for demonstration of superiority, or for non-inferiority with a well-justified margin. • RCT with sham control.[18]	• Registry with complete recruitment, recording primary end-points and adverse events.
Not advised	• 'Compassionate use' series.		• RCT with placebo control.[19] • RCT with surrogate end-points. • Single-arm with historical controls (case-control study).	• Registry with incomplete recruitment (<90%).

The shaded column indicates the stage when an EU certificate of conformity / CE mark may be awarded.

[16] Examples could be: for an orthopaedic implant, radiostereometric analysis (RSA) of migration of the device see [29] and for vascular stents and scaffolds, vessel patency determined using intravascular imaging and objective performance criteria [30].

[17] Another device of similar type already approved for the same indication.

[18] Recommended only if the manfacturer offers a well-justified explanation for the lack of access to an active comparator, for their clinical study.

[19] No active comparator or sham intervention available.

Explanatory notes

This CORE–MD recommendation does not include advice on pre-clinical investigations and evaluation of a medical device. The MDR lists requirements, for example in Annex II with reference to Annex XV, Chapter II 4.1 and 4.6. These may include: *in silico* studies; laboratory/bench testing; if relevant, animal testing; and biocompatibility analyses.

The CORE–MD consortium recommends that a logical first step by the developers of a new high-risk medical device will be to commission or conduct a survey of unmet needs and expectations, among patients and physicians.

It is challenging to list all variations of trial designs, since many features can be used in different combinations in an individual trial. The table below gives some options, with those in the clear central columns placed in order of their resulting quality of evidence.

General descriptors of study type	Choices of study design and comparator	Mode of randomisation *or* minimisation of bias	Manner of blinding	Alternative designs	Possible statistical approaches
Prospective / retrospective ***OR*** Randomised (experimental) / observational / secondary analysis [21]	RCT vs. best available and effective alternative treatment: – active comparator, or – medical treatment RCT vs. best control when no effective option available: – sham intervention – placebo Non-randomised observational cohort study vs. other prospective control group Cohort study vs. historical controls	Individual randomisation Cluster randomisation Consecutive enrolment None *or with optimal methods of adjustment:* e.g. Propensity score stratification, or Inverse probability weighting	Double Single Open-label with blinded endpoints None	Factorial Adaptive (feasible only if device has early impact) Cross-over Stepped-wedge etc.	Frequentist Bayesian Win ratio etc.

Further background information and explanatory material are provided in the Appendices.

Summary and conclusions

This document reports the consensus conclusions of the CORE–MD consortium on optimal study designs for the clinical evaluation of high-risk medical devices, according to their stage of development. Separate proposals have been made for medical devices in general; for those used for orphan indications or to satisfy unmet needs for which no alternative devices are available; and for new high-risk devices entering an established market when there are alternative devices within the same type.

According to the circumstances, different levels of clinical evidence might be accepted as the basis for approval, but lesser levels of confidence from smaller pre-market studies might need to be balanced by mandatory post-market clinical investigations. The optimal design will usually be a randomised controlled trial, preferably against an active comparator, but when that is inappropriate or impossible then an observational study can be performed as long as it is adequately powered, incorporates measures to minimise bias, and uses validated outcome measures.

These recommendations will be shared with EU medical device regulators.

References

[1] Siontis GC, Coles B, Häner JD, McGovern L, Bartkowiak J, Coughlan JJ, Spirito A, Galea R, Haeberlin A, Praz F, Tomii D, Melvin T, Frenk A, Byrne RA, Fraser AG, Windecker S; CORE-MD investigators. Quality and transparency of evidence for implantable cardiovascular medical devices assessed by the CORE-MD Consortium. Eur Heart J. 2024; 45: 161–77. https://doi.org/10.1093/eurheartj/ehad567

[2] Lübbeke A, Combescure C, Barea C, Gonzalez AI, Tucker K, Kjærsgaard-Andersen P, Melvin T, Fraser AG, Nelissen R, Smith JA. Clinical investigations to evaluate high-risk orthopaedic devices: a systematic review of the peer-reviewed medical literature. EFORT Open Reviews 2023; 8(11): 781–791. doi: 10.1530/EOR-23-0024. PMID: 37909694. https://eor.bioscientifica.com/view/journals/eor/8/11/EOR-23-0024.xml

[3] Bano A, Künzler J, Wehrli F, Kastrati L, Rivero T, Fraser AG, Stettler S, Hovorka R, Laimer M, Bally L, on behalf of CORE-MD investigators. Clinical evidence for high-risk CE-marked medical devices for diabetes care: a systematic review and meta-analysis. Submitted for publication.

[4] Guerlich K, Patro-Golab B, Dworakowski P, Fraser AG, Kammermeier M, Melvin T, Koletzko B. Evidence from clinical trials on high-risk medical devices in children: a scoping review. Pediatric Research 2023 Sep 28. doi: 10.1038/s41390-023-02819-4. Epub ahead of print. PMID: 37758865. https://www.nature.com/articles/s41390-023-02819-4

[5] Wild C, Ettinger S. Perceived training needs of regulators, Notified Bodies and clinicians for successful implementation of the EU MDR: survey results. Journal of Medical Device Regulation 2023;20(2):45–56. https://www.core-md.eu/wp-content/uploads/2023/06/JMDRMay2023-Training.pdf [full report on the CORE-MD website]

[6] Guerlich K, Patro-Golab B, Barnacle A, Baumann U, Eicken A, Fraser AG, Gruszfeld D, Haas NA, Jonker AH, Kammermeier M, Kenny D, Kolaček S, Lapatto R, Maconochie I, Mader S, McGauran G, Melvin T, Muensterer O, Piscoi P, Romano A, Saxena AK, Schneider DT, Turner MA, Walle JV, Koletzko B; European Academy of Paediatrics. European expert recommendations on clinical investigation and evaluation of high-risk medical devices for children. Acta Paediatrica. 2023; 112: 2440–2448. https://www.core-md.eu/wp-content/uploads/2023/08/Acta-Paediatrica-2023.pdf

[7] ISO 14155:2020. Clinical investigation of medical devices for human subjects – Good clinical practice. https://www.iso.org/standard/71690.html

[8] Byrne RA, Serruys PW, Baumbach A, Escaned J, Fajadet J, James S, Joner M, Oktay S, Jüni P, Kastrati A, Sianos G, Stefanini GG, Wijns W, Windecker S. Report of a European Society of Cardiology–European Association of Percutaneous Cardiovascular Interventions Task Force on the evaluation of coronary stents in Europe: executive summary. Eur Heart J. 2015;36:2608–20.

[9] European Parliament. Regulation (EU) 2017/ 745 of the European Parliament and of the Council - of 5 April 2017 - On medical devices, amending Directive 2001/ 83/ EC, Regulation (EC) No 178/

2002 and Regulation (EC) No 1223/ 2009 and repealing Council Directives 90/ 385/ EEC. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R0745

[10] European Commission. Clinical evaluation: a guide for manufacturers and notified bodies under directives 93/42 and 90/385. MEDDEV 2.7/1 revision 4 June 2016. https://ec.europa.eu/docsroom/documents/17522/attachments/1/translations/

[11] Sedrakyan A, Campbell B, Merino JG, Kuntz R, Hirst A, McCulloch P. IDEAL-D: a rational framework for evaluating and regulating the use of medical devices. BMJ. 2016;353:i2372.

[12] Páez A, Rovers M, Hutchison K, Rogers W, Vasey B, McCulloch P; IDEAL Collaboration. Beyond the RCT: When are randomized trials unnecessary for new therapeutic devices, and what should we do instead? Ann Surg. 2022;275:324–331.

[13] Good Clinical Trials Collaborative. Guidance for good randomized clinical trials. (May 2022) https://www.goodtrials.org/guidance

[14] International Medical Device Regulators Forum. Principles of international system of registries linked to other data sources and tools. IMDRF/REGISTRY WG/N33FINAL:2016. https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-160930-principles-system-registries.pdf

[15] Vandenbroucke JP. Prospective or retrospective: what's in a name? BMJ. 1991;302:249–50.

[16] World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013;310(20):2191–4. Doi: 10.1001/jama.2013.281053.

[17] Louw A, Diener I, Fernández-de-Las-Peñas C, Puentedura EJ. Sham Surgery in Orthopedics: A Systematic Review of the Literature. Pain Med. 2017;18(4):736-750.

[18] Sheldon, R., & Opie-Moran, M. (2017). The Placebo Effect in Cardiology: Understanding and Using It. *The Canadian journal of cardiology*, *33*(12), 1535–1542. https://doi.org/10.1016/j.cjca.2017.09.017

[19] Redberg R. F. (2014). Sham controls in medical device trials. *The New England journal of medicine*, *371*(10), 892–893. https://doi.org/10.1056/NEJMp1406388

[20] Niemansburg SL, van Delden JJ, Dhert WJ, Bredenoord AL. Reconsidering the ethics of sham interventions in an era of emerging technologies. Surgery. 2015 Apr;157(4):801-10.

[21] Byrne RA., Capodanno D., Mahfoud F., et al. Evaluating the importance of sham-controlled trials in the investigation of medical devices in interventional cardiology. EuroIntervention 2018;14(6):708–15. Doi: 10.4244/EIJ-D-18-00481.

[22] Felson, D. T., Redmond, A. C., Chapman, G. J., Smith, T. O., Hamilton, D. F., Jones, R. K., Holt, C. A., Callaghan, M. J., Mason, D. J., Arthritis Research UK Osteoarthritis and Crystal Diseases Clinical Studies Group, & Conaghan, P. G. (2016). Recommendations for the conduct of efficacy trials of treatment devices for osteoarthritis: a report from a working group of the Arthritis Research UK

Osteoarthritis and Crystal Diseases Clinical Studies Group. *Rheumatology (Oxford, England)*, *55*(2), 320–326. https://doi.org/10.1093/rheumatology/kev328

[23] Kersting C, Kneer M, Barzel A. Patient-relevant outcomes: what are we talking about? A scoping review to improve conceptual clarity. BMC Health Serv Res. 2020;20:596.

[24] Christensen R, Ciani O, Manyara AM, Taylor RS. Surrogate endpoints: a key concept in clinical epidemiology. J Clin Epidemiol. 2024 Mar;167:111242.

[25] Ciani O et al. A framework for the definition and interpretation of the use of surrogate endpoints in interventional trials. eClinicalMedicine. 2023;65:102283.

[26] Sackett DL. Superiority trials, noninferiority trials, and prisoners of the 2-sided null hypothesis. ACP J Club. 2004;140(2):A11.

[27] Gressler LE, Marinac-Dabic D, dosReis S, Goodney P, Mullins CD, Shaya F. Creation of objective performance criteria among medical devices. BMJ Surg Interv Health Technol. 2022;4(1):e000106.

[28] van Zwet E, Gelman A, Greenland S, Imbens G, Schwab S, Goodman SN. A new look at p values for randomized clinical trials. NEJM Evid. 2024;3(1):EVIDoa2300003

[29] Sandberg OH, Kärrholm J, Olivecrona H, Röhrl SM, Sköldenberg OG, Brodén C. Computed tomography-based radiostereometric analysis in orthopedic research: practical guidelines. Acta Orthop. 2023;94:33378

[30] Byrne RA, Serruys PW, Baumbach A, Escaned J, Fajadet J, James S, Joner M, Oktay S, Jüni P, Kastrati A, Sianos G, Stefanini GG, Wijns W, Windecker S. Report of a European Society of Cardiology–European Association of Percutaneous Cardiovascular Interventions Task Force on the evaluation of coronary stents in Europe: executive summary. Eur Heart J. 2015;36:2608–20

Appendices

A.1 Recommendations from the Food & Drug Administration (FDA), USA

FDA guidance on pivotal clinical investigations refers in its introduction [20] to:

- the **exploratory clinical stage** in which "the limitations and advantages of the medical device are evaluated. This stage includes first-in-human studies and feasibility studies"
- the **pivotal stage** which "is used to develop the information necessary to evaluate the safety and effectiveness of the device for the identified intended use. It usually consists of one or more pivotal studies" and
- the "**post-market stage** which can include "an additional study or studies for better understanding of device safety, such as rare adverse events and long-term-effectiveness".

The hierarchy of evidence in study designs for medical devices that can be considered for a pre-market approval, derived from the FDA guidance, in descending order is:

1. the randomised double-blinded parallel group study with active or placebo control,
2. the randomised parallel group study with incomplete blinding; at least outcomes evaluators should be blinded, and active or placebo control groups are preferable to control groups without treatment,
3. single-arm studies with historical controls with individual patient data, or
4. single-arm studies with objective performance criteria or performance goals provide the lowest level of evidence.

Definitions of these designs are given in the table on page 13.

Trial designs that the manufacturer should use to describe the level of evidence of the data necessary to support his/her application, as listed in Appendix B of an FDA Guidance document published in 2014 [21], are:

[20] U. S. Food and Drug Administration. Design Considerations for Pivotal Clinical Investigations for Medical Devices. Guidance for industry, clinical investigators, institutional review boards and Food and Drug Administration staff. November 7, 2013. https://www.fda.gov/media/87363/download

[21] The 510(k) Program: Evaluating Substantial Equivalence in Premarket Notifications [510(k)]. Guidance for Industry and Food and Drug Administration Staff Document issued on: July 28, 2014. https://www.fda.gov/media/82395/download

- Randomized, multi-arm, “blinded” study with concurrent sham (placebo) control
- Randomized, multi-arm, “blinded” study with concurrent (“active”) control
- Randomized, multi-arm, un“blinded” study with a control (control that is either active or consists of no treatment)
- Non-randomized study with concurrent (“active”) control
- Single-arm study with patient serving as own control (include designed single-arm crossover)
- Single-arm study with historical control (using patient-level data)
- Single-arm study with literature control (historical control)
- Single-arm study with objective performance criteria
- Single-arm study with performance goals
- *Registry*
- *Observational study*
- *Systematic review (meta-analysis with patient-level data)*
- *Meta-analysis based on summary information only*
- *Literature Summary*
- *Uncertain*

This list comprises study types that were not considered appropriate for a pivotal study (shown above in italics), and it does not include the randomized controlled cross-over design.

Table 4. Definition and classification of study designs for medical devices, according to FDA 2013 and as summarised in Deliverable D1.6

Study design	Description	FDA advantages/disadvantages
Randomised, double-blinded, controlled parallel group clinical study (active or placebo control)	A randomized study is a study in which participants are randomly (i.e. by chance) assigned to one of two or more interventions of a clinical study. Double-blinded indicates that the intervention assignment is not known to the subject or the study staff (including the investigator or any third-party evaluator(s)). Parallel group design means that each subject or sample is assigned only one of the possible treatments being compared. Because a different group of subjects (or samples) is assigned to each treatment, comparisons are made between subject groups. When considering an active control, an important consideration is whether to design the study to demonstrate superiority or non-inferiority.	This study design provides the highest level of assurance that the subject populations in the investigational and control groups are comparable and avoids systematic differences between groups with respect to known and unknown baseline variables that could affect both safety and effectiveness outcomes. However, there are devices for which this design is neither feasible nor practical. Deviation from this study design is especially problematic in situations where there is a possible placebo effect, or when subjective outcome measures are used as study endpoints. Choice of an appropriate active control is based on the current standard of care for the intended subject population. A placebo control is useful if there is thought to be a placebo effect. • It may be challenging to construct a placebo control that appears to function like the investigational device but delivers no therapy. • In some cases, it may be unethical to randomize subjects to a placebo that will provide no known effect.
Randomised, subject as own control, paired clinical study Two-group cross-over design study	In such a study design, the subject could be treated with both the investigational and control interventions at the same time (e.g. side of face). The assignment of intervention is randomised. Another type of such a study design is a two-group cross-over design study, where each subject receives the investigational	This study design is possible when the device effect is only evident locally. It is impossible to evaluate and differentiate systemic safety or effectiveness outcomes when using this study design. The advantage of this study design, when used appropriately, is that the effects of both interventions are measured in the same subject and the variability is smaller so a smaller sample size may be required. With the cross-over design one needs to assume that the effects of the first intervention will not carry over into the second intervention period.

Study design	Description	FDA advantages/disadvantages
	and control interventions sequentially, with a randomly assigned order.	Otherwise, a "wash-out" period may have to be incorporated into the study.
Randomised, non-blinded study with concurrent control (Active, placebo or "no Intervention")	Incomplete blinding refers to instances where the subject, the investigator or the third-party evaluator is not blinded. When no one is blinded, the study is often referred to as an open-label study.	If the subject's assignment to an intervention is not blinded (masked), the behaviour of the subject may be affected by knowledge of the intervention and consequently a bias can be introduced, particularly if a clinical measurement or endpoint is subjective. If the investigator or a third-party evaluator is not blinded (masked) to the intervention assignment, then investigator or evaluator bias can adversely affect the study by influencing the interpretation of clinical outcomes, the performance of surgical implantation of a device, and subsequent clinical decision-making (7.3). If study participants are not blinded, it is very difficult to assess the size of the resulting bias, and it can threaten the scientific validity of an otherwise solid study, even when a truly objective endpoint is used. In instances where blinding of any or all of the study participants (subjects, investigators, evaluators) is not possible, a detailed rationale and explanation of proposed means to address concerns related to bias should be provided to FDA. Choice of a "no intervention" control may present a challenge in recruiting subjects who might receive no intervention or keeping subjects enrolled who were randomized to the "no intervention" control group. • Choice of a "no intervention" control has built-in bias because control group subjects expect to receive no benefit, whereas experimental group subjects expect to receive a benefit. • A "no intervention" control may sometimes be standard of care/best medical management which can provide evidence about any incremental benefit or risk, although the control could vary among the different study centers.

Study design	Description	FDA advantages/disadvantages
Non-randomised study with concurrent control (active or placebo or "no intervention")	In a non-randomised design with a concurrent active control, subjects and investigators are not blinded to the intervention assignment.	Consequently, this study design suffers from all the drawbacks of a randomized, non-blinded study with concurrent control design. In addition, because there is no randomisation and each subject receives only one of the possible interventions, there is a very real possibility of a bias with unknown size due to intervention assignment. This design is generally not recommended since it is as labour intensive as a randomised study, but introduces more biases due to likely differences in the groups, sites, and investigators, including unmeasured, but likely confounding, differences. Even if there appears to be a balance between the two intervention groups for the study overall, there is likely no balance for each participating investigator such that there may be an investigator-by-device interaction, in which the advantage of the investigational device appears to differ by investigator.
Single-group study compared to baseline	Subject's outcomes at baseline compared to outcomes at endpoint evaluations.	Use of baseline outcomes as a comparison for outcome at the endpoint evaluations is inadequate for most therapeutic studies since subjects may improve for reasons unrelated to investigational device (e.g., regression to the mean, placebo effect). It is usually advisable to also have a randomized group with an active or placebo control (or even a "no intervention" control). Such a randomized group in a blinded study will provide a much more stringent control and avoid placebo effect bias as well as temporal bias.
Single-group study with historical control or Information	Historical Control Group: A control group of subjects who were observed prior to the pivotal study. Data collected from this control group is used to compare the performance of the investigational device.	The obvious bias inherent in the use of a historical control is temporal bias, since the groups are not concurrent. The use of a comparator study separated in time can introduce severe and unknown selection bias. Concerns about comparability of groups and that practice of medicine has changed with resultant changes in the expected outcomes. Thus, the disadvantage of this design is that the subject outcomes in a historical control may not be discernible or applicable to the current population being targeted. It may be possible to use a propensity score model to

Study design	Description	FDA advantages/disadvantages
		assess the comparability of the two groups after the current study has been completed. There is no way to assess comparability until the subjects are enrolled and baseline collected and analysed so this approach can be risky. This control presents a significant challenge in addressing the implications of missing data. Sensitivity and missing data analyses may potentially address some concerns associated with bias.
	Objective performance criterion (OPC): A numerical target value derived from historical data from clinical studies and/or registries and may be used by FDA for the comparison of safety or effectiveness endpoints. Performance goal (PG): A numerical value that is considered sufficient by FDA for use as a comparison of the pivotal study results with a safety endpoint, or an effectiveness endpoint.	If a historical control group is not available, the performance of a device may be evaluated through a comparison to a numerical target value, OPC or PG, pertaining to a safety or effectiveness endpoint. Such a study design shares all of the challenges and limitations of comparison to a historical control. In addition, there is no independent way to assess how comparable the current group may be with the historical groups from which the OPC or PG is derived, and it is impossible to quantify the bias. Since there is no control group involved in such studies, comparison to an OPC or PG cannot demonstrate either superiority or non-inferiority.

Note: This table is reproduced from Deliverable D1.6 which is the output of CORE-MD Task 1.4 [Study design recommendations in guidance documents for high-risk medical devices]. It summarizes the definitions, advantages and disadvantages of study types, as described in the FDA guidance document in its sections 7.8, 7.4 Table 1, other parts of section 7, and glossary.

A.2 Recommendations from the Therapeutic Goods Administration (TGA)

The regulatory authority for medical devices in Australia, the Therapeutic Goods Administration (TGA), published comprehensive guidance on study designs in June 2022.[22] It does not provide a hierarchy of evidence, but it refers (at page 30) to the tool for ranking different types of study design according to their levels of evidence, that was published by the Australian National Health and Medical Research Council (NHMRC) in 2009 (but which has now been replaced [23]).

The NHMRC ranked study designs into a hierarchy according to their potential to adequately answer a particular research question, and to the level of bias inherent in the study design. In order, these are:

1. systematic reviews of randomised controlled trials
2. individual randomised controlled trials,
3. pseudo randomised controlled trials,
4. non-randomised comparative trials, and
5. case series.

Description and classification of study designs for medical devices according to NHMRC (reproduced in CORE–MD report [22])

Level of evidence /study design	Description from glossary
I A systematic review of Level II studies	
II Randomised controlled trial	The unit of experimentation (e.g. people, or a cluster of people) is allocated to either an intervention (the factor under study) group or a control group, using a random mechanism (such as a coin toss, random number table, computer-generated random numbers) and the outcomes from each group are compared. Cross-over randomised controlled trials – where the people in the trial receive one intervention and then cross-over to receive the alternate intervention at a point in time – are considered to be the same level of evidence as a randomised controlled trial, although appraisal of these trials would need to be tailored to address the risk of bias specific to cross-over trials.

[22] Australian Government, Department of Health, Therapeutic Goods Administration. Clinical evidence guidelines for medical devices. Version 3.1, June 2022. https://www.tga.gov.au/sites/default/files/clinical-evidence-guidelines-medical-devices.pdf

[23] NHMRC. Guidelines for Guidelines Handbook. www.nhmrc.gov.au/guidelinesforguidelines

Level of evidence /study design	Description from glossary
III-1 Pseudorandomised controlled trial	The unit of experimentation (e.g. people, a cluster of people) is allocated to either an intervention (the factor under study) group or a control group, using a pseudo-random method (such as alternate allocation, allocation by days of the week or odd-even study numbers) and the outcomes from each group are compared
III-2 A comparative study with concurrent controls	
Non-randomised, experimental trial	The unit of experimentation (e.g. People, a cluster of people) is allocated to either an intervention group or a control group, using a non-random method (such as patient or clinician preference/availability) and the outcomes from each group are compared. This can include: (1) a controlled before-and-after study, where outcome measurements are taken before and after the intervention is introduced, and compared at the same time point to outcome measures in the (control) group. (2) an adjusted indirect comparison, where two randomised controlled trials compare different interventions to the same comparator i.e. the placebo or control condition. The outcomes from the two interventions are then compared indirectly.
Cohort study	Outcomes for groups of people observed to be exposed to an intervention, or the factor under study, are compared to outcomes for groups of people not exposed.
Case-control study	People with the outcome or disease (cases) and an appropriate group of controls without the outcome or disease (controls) are selected and information obtained about their previous exposure/ non-exposure to the intervention or factor under study.
Interrupted time series with a control group	Trends in an outcome or disease are measured over multiple time points before and after the intervention is introduced to a group of people, and then compared to the outcomes at the same time points for a group of people that do not receive the intervention.

Level of evidence /study design	Description from glossary
III-3 A comparative study without concurrent controls	
Historical control study	Outcomes for a prospectively collected group of people exposed to the intervention are compared with either (1) the outcomes of people treated at the same institution prior to the introduction of the intervention (i.e. control group/usual care), or (2) the outcomes of a previously published series of people undergoing the alternate or control intervention.
Two or more single arm study	The outcomes of a single series of people receiving an intervention (case series) from two or more studies are compared. Unadjusted indirect comparisons: an unadjusted indirect comparison compares single arms from two or more interventions from two or more separate studies via the use of a common reference ie A versus B and B versus C allows a comparison of A versus C but there is no statistical adjustment for B. Such a simple indirect comparison is unlikely to be reliable (see Song et al 2000).
Interrupted time series without a parallel control group	Trends in an outcome or disease are measured over multiple time points before and after the intervention is introduced to a group of people, and compared (as opposed to being compared to an external control group).
IV Case series with either post-test or pre-test/post-test outcomes	A single group of people exposed to the intervention. Post-test – only outcomes after the intervention are recorded in the series of people, so no comparisons can be made. Pre-test/post-test – measures on an outcome are taken before and after the intervention is introduced to a series of people and are then compared (also known as a 'before- and-after study').

A.3 Comparison of hierarchies of levels of evidence, from FDA and NHMRC

(as summarised in the CORE–MD report [20])

FDA level of evidence [24]	NHMRC level of evidence [25]
Randomized controlled trial	II Randomised controlled trial
Randomized, multi-arm, "blinded" study with concurrent sham (placebo) control	
Randomized, multi-arm, "blinded" study with concurrent ("active") control	
Randomized, multi-arm, un"blinded" study with a control (control that is either active or consists of no treatment)	
Randomized, single-arm study with patient serving as own control (include designed single-arm crossover)	
	III-1 Pseudorandomised controlled trial
Studies with concurrent controls	III-2 Comparative study with concurrent controls
	Non-randomised, experimental studies
	Cohort study *
	Case-control study *
	Interrupted time series with control
Studies with non-concurrent controls	III-3 Comparative study without concurrent control
Single-arm study with historical control (using patient-level data)	Historical control study
Single-arm study with literature control (historical control)	Two or more single arm study (including unadjusted indirect comparisons)
	Interrupted time series without parallel control group
Non-comparative studies	IV Case series with either post-test or pre-test/post-test outcomes
Single-arm study with objective performance criteria	
Single-arm study with performance goals	

[Note that these are observational studies]*

[24] FDA = U. S. Food and Drug Administration. Design considerations for pivotal clinical investigations for medical devices. Guidance for industry, clinical investigators, institutional review boards and Food and Drug Administration staff. 2013. https://www.fda.gov/regulatory-information/search-fda-guidance-documents

[25] NHMRC = National Health and Medical Research Council (Australia). NHMRC levels of evidence and grades for recommendations for developers of guidelines. 2009.

A.4 Recommendations from the European Union

The MDCG guidance document concerning legacy devices (**MDCG 2020-6**) was published in 2020. It provides guidance on clinical data that are considered to provide sufficient clinical evidence for the conformity assessment of legacy devices under the MDR.

Appendix III provides a "**Suggested hierarchy of clinical evidence** for confirmation of conformity with relevant GSPRs [General Safety and Performance Requirements] under the MDR". It states that the suggested hierarchy is "ranked roughly in order from strongest to weakest (some variations may apply dependent on the device, GSPR for which evidence is required, and quality of individual data sources)".

The hierarchy comprises twelve levels of clinical evidence (but note that these are not trial methodologies). It is stated that class III legacy devices and implantable legacy devices which are not well-established technologies (WET) should have sufficient data as a minimum at level four, whereas WET may be able to confirm conformity "via an evaluation of cumulative evidence from additional sources of levels five to twelve". There is no statement whether this minimum level would also apply to devices which do not have a former market authorization. A definition of a WET, which applies to all devices, and not only to legacy devices, is given in section 1.2: "The common features of the devices which are **well-established technologies** are that they all have:

- relatively simple, common and stable designs with little evolution;
- their generic device group has well-known safety and has not been associated with safety issues in the past;
- well-known clinical performance characteristics and their generic device group are standard of care devices where there is little evolution in indications and the state of the art; [and]
- a long history on the market."

The following table reproduces the "hierarchy" in the MDCG guidance:

Suggested hierarchy of clinical evidence for confirmation of conformity with relevant GSPRs under the MDR, from MDCG-2020-6 [26]

The following list was published by the European Commission in a guidance document to clarify what types of clinical evidence could be used for recertification of previously approved medical devices, under the MDR, or to support a new application. Paragraph 4.4.9.2 states that ".. only the first four [levels] are relevant for implantable and class III devices".

Rank	Types of clinical data and evidence	Considerations / comments
1	Results of high quality[1] clinical investigations covering all device variants, indications, patient populations, duration of treatment effect, etc.	This may not feasible or necessary for certain well-established devices with broad indications (e.g Class IIb legacy sutures, which could be used in every conceivable patient population).
2	Results of high quality clinical investigations with some gaps	Gaps must be justified / addressed with other evidence in line with an appropriate risk assessment, and clinical safety, performance, benefit and device claims. Assuming the gaps can be justified, there should be an appropriate PMCF plan to address residual risks. Otherwise, manufacturers shall narrow the intended purpose of the device until sufficient clinical data has also been generated.
3	Outcomes from high quality clinical data collection systems such as registries[2]	Is there sufficient evidence of the quality of the data collected by the registry[3, 4]? Are the devices adequately represented? Are the data appropriately stratified? Are the endpoints appropriate to the safety, performances and endpoints identified in the clinical evaluation plan?
4	Outcomes from studies with potential methodological flaws but where data can still be quantified and acceptability justified[2]	Many literature sources fall into this category, due to limitations such as missing information, publication bias, time lag bias, etc. This applies equally to publications in the peer-reviewed scientific literature. However, for legacy devices where no safety or performance concerns have been identified, these sources can be sufficient for confirmation of conformity to the relevant GSPRs if appropriately appraised and the gaps are identified and handled. High quality surveys may also fall into this category.
Class III legacy devices and implantable legacy devices which are not well-established technologies should have sufficient clinical data as a minimum at level 4. Those devices which are well-established technologies may be able to confirm conformity with the relevant GSPRs via an evaluation of cumulative evidence from additional sources as listed below. Reliance solely on complaints and vigilance is not sufficient.		

[26] Medical Device Coordination Group. MDCG 2020-6 Regulation (EU) 2017/745: Clinical evidence needed for medical devices previously CE marked under Directives 93/42/EEC or 90/385/EEC. A guide for manufacturers and notified bodies2020. [Table 12] https://ec.europa.eu/health/md_sector/new_regulations/guidance_en

5	Equivalence data (reliable / quantifiable)	Equivalence must meet MDR criteria. It is normally expected that manufacturers should gather data on their own devices in the post-market phase, therefore reliance on equivalence should be duly justified, and linked to appropriate PMCF or proactive PMS.
6	Evaluation of state of the art, including evaluation of clinical data from similar devices* as defined in Section 1.2 of this document	This is not considered clinical data under the MDR, but for well-established technologies only can be considered supportive of confirmation of conformity to the relevant GSPRs. Data from similar devices may be also important to establish whether the device under evaluation and similar devices belong to the group of devices considered as "well established technologies" (WET). See section 1.2 in this document for the criteria for WET. Data from similar devices may be used, for example, to demonstrate ubiquity of design, lack of novelty, known safety and performance profile of a generic group of devices, etc.
7	Complaints and vigilance data; curated data	This falls within the definition of clinical data under MDR Article 2(48), but is not generally considered a high quality source of data due to limitations in reporting. It may be useful for identifying safety trends or performance issues. High volume data collected within a robust quality system may provide supportive evidence of device safety.
8	Proactive PMS data, such as that derived from surveys	This falls within the definition of clinical data under MDR Article 2(48), but is not generally considered a high quality source of data due limitations associated with sources of bias and quality of data collection. It may be useful for identifying safety concerns or performance issues.
9	Individual case reports on the subject device	This falls within the definition of clinical data under MDR Article 2(48), but is not considered a high quality source of data due to limitations in generalising findings to a wider patient population, reporting bias, etc. It may provide supportive or illustrative information with respect to specific claims.
10	Compliance to non-clinical elements of common specifications considered relevant to device safety and performance	Common specifications which address clinical investigation or data requirements directly would rank higher in this hierarchy. Common specifications may address clinically relevant endpoints through non-clinical evidence such as mechanical testing for strength and endurance, biological safety, usability, etc.
11	Simulated use / animal / cadaveric testing involving healthcare professionals or other end users	This is not clinical data, but may be considered evidence of confirmation of conformity to relevant GSPRs, particularly in terms of usability, such as for accessories or instruments.
12	Pre-clinical and bench testing / compliance to standards[1]	Pre-clinical and bench testing may address clinically relevant endpoints through non-clinical evidence such as mechanical testing for strength and endurance, biological safety, usability, etc.

1 Refer to data appraisal considerations described in Section 6.3 of this guidance.

2 Please note that the Considerations / Comments listed in point 2 also apply to these studies.

3 http://www.imdrf.org/docs/imdrf/final/technical/imdrf-tech-170316-methodological-principles.pdf

4 http://www.imdrf.org/docs/imdrf/final/technical/imdrf-tech-160930-principles-system-registries.pdf

5 This may be of interest in the case of application of Article 61(10).

* Similar devices = “devices belonging to the same generic device group. The MDR defines this as a set of devices having the same or similar intended purposes or a commonality of technology allowing them to be classified in a generic manner not reflecting specific characteristics” (section 1.2).

GSPR = General safety and performance requirements

A.5 Recommendations from the IDEAL collaboration

The IDEAL collaboration developed recommendations for a logical sequence of investigations to establish the utility of a new surgical procedure.[27] Later these were adapted for studies of medical devices.[28]

The IDEAL Recommendations propose study designs and features specific to the needs of the 6 stages that IDEAL identifies in the device life cycle. It does not rank study methods in order of internal validity, as the FDA hierarchy does, but focuses instead on what questions the study needs to answer. It expresses the key questions for each stage as follows:

IDEAL stages, key questions and main study design recommendations:

Stage 0 **Preclinical evaluation**:

Is the device safe and ready for evaluation patients or human volunteers?

Stage 1 **Idea** (First-in-human study):

What does the device do, how does it work, and how did it perform?

Typical study design recommendation: Comprehensive transparent case report.

Stage 2a **Development** (Early clinical studies):

Have the device itself, its manner of use and its indications reached a stable state?

Typical study design recommendation: Prospective cohort study with sequential reporting of cases.

Stage 2b **Exploration** (Larger, collaborative studies):

Has clinical consensus been reached on the indications and mode of use of the device, its' performance and safety, and the learning curve for high quality delivery?

Typical study design recommendation: Collaborative multicentre prospective cohort study without controls.

Stage 3 **Assessment**:

Is the device superior (or non-inferior) to state-of-the art alternative treatment(s)?

[27] McCulloch P, Altman DG, Campbell WB, Flum DR, Glasziou P, Marshall JC, Nicholl J; Balliol Collaboration; Aronson JK, Barkun JS, Blazeby JM, Boutron IC, Campbell WB, Clavien PA, Cook JA, Ergina PL, Feldman LS, Flum DR, Maddern GJ, Nicholl J, Reeves BC, Seiler CM, Strasberg SM, Meakins JL, Ashby D, Black N, Bunker J, Burton M, Campbell M, Chalkidou K, Chalmers I, de Leval M, Deeks J, Ergina PL, Grant A, Gray M, Greenhalgh R, Jenicek M, Kehoe S, Lilford R, Littlejohns P, Loke Y, Madhock R, McPherson K, Meakins J, Rothwell P, Summerskill B, Taggart D, Tekkis P, Thompson M, Treasure T, Trohler U, Vandenbroucke J. No surgical innovation without evaluation: the IDEAL recommendations. Lancet. 2009;374:1105–12.

[28] Sedrakyan A, Campbell B, Merino JG, Kuntz R, Hirst A, McCulloch P. IDEAL-D: a rational framework for evaluating and regulating the use of medical devices. BMJ. 2016;353:i2372.

Typical study design recommendation: Randomised controlled trial against best alternative treatment.

Stage 4 **Long-term study**:

What variations in the performance and safety of the device emerge over time, and for what reasons?

Typical study design recommendation: High-quality registry.

Key recommendations for research design at each IDEAL phase are shown in the table below [29]. Finally, the IDEAL consortium has also published recommendations for how to report studies at each stage.[30] The relevant checklist was published as the final digital supplement to that paper.[31]

IDEA ***Professional Innovation Database***	DEVELOPMENT ***Prospective Development Studies***	EXPLORATION ***Phase IIS Study***	ASSESSMENT ***Surgical RCT***	LONG TERM MONITORING ***Prospective Registries***
Compulsory reporting of all new innovations Confidential entry allowed to encourage reporting of failed innovations (similar to CHRP system) Hospital or institution to be informed separately as a professional duty	Detailed description of selection criteria Detailed technical description Prospective account of ALL cases consecutively, including those NOT treated with new technique/device Clear STANDARDISED definitions of outcomes reported Description of ALL modifications, and when they were made during the series Registration of PROTOCOL before study starts Use of Statistical Process Control (SPC) methods to evaluate progress	To evaluate technique prospectively and co-operatively To develop a consensus over *definition* of the procedure, *quality* standards and *indications* To gather *data* for *power calculations* To evaluate and monitor *learning curves* To achieve consensus on the *trial question* To develop a *multi-centre randomised trial (RCT)*	RCT – question agreed in Phase IIS Use power calculations from Phase IIS Use learning curve data to decide entry points for clinicians Use Phase IIS consensus to define operation, quality control AND outcome measures Use modified RCTs or recognised alternative if RCT not feasible: Feasibility RCT Expertise-based RCT Cohort multiple RCT Step-wedge design Controlled-interrupted time series	Should monitor indications as well as outcomes SPC used for quality control (Shewart charts, CUSUM, VLAD)

[29] https://www.ideal-collaboration.net/the-ideal-framework/recommendations/

[30] Bilbro NA, Hirst A, Paez A, Vasey B, Pufulete M, Sedrakyan A, McCulloch P; IDEAL Collaboration Reporting Guidelines Working Group. The IDEAL Reporting Guidelines: A Delphi consensus statement stage specific recommendations for reporting the evaluation of surgical innovation. Ann Surg. 2021;273:82–85.

[31] https://journals.lww.com/annalsofsurgery/fulltext/2021/01000/the_ideal_reporting_guidelines__a_delphi_consensus.14.aspx

A.6 Principles, implications and recommendations from the Good Clinical Trials Collaborative [32]

Principle	Implications	Recommendations regarding:
Relevance and utility	Design characteristics of RCTs should be aimed to resolve important uncertainties about the effects of a health intervention.	- Appropriate population - Robust intervention allocation - Adequate size - Blinding and masking of interventions - Adherence to allocated interventions - Completeness of follow-up - Relevant measures of outcomes - Proportionate, efficient and reliable capture of data - Ascertainment of outcomes - Statistical analysis - Assessing beneficial and harmful effects of the intervention - Monitoring emerging information on benefits and harms
Respect of participants	Ethical responsibilities regarding participants, future and current patients, and the public.	- Appropriate communication - Relevant consent - Changing consent - Implications of changing consent - Managing the safety of individual participants - Communication of new information relevant to the intervention
Collaboration and transparency	Practices that contribute to develop trust between all those involved in an RCT and generalize confidence in the RCT ecosystem.	- Working in partnership with people and communities - Collaboration among organizations - Transparency
Appropriateness for their context	Ensuring that a trial is set up to be practicable and produce reliable, actionable results.	- Setting and context - Use of existing resources
Efficiency and management	Competent decision-making and coordinated execution based on goo governance and good trial quality management.	- Competent advice and decision-making - Protecting trial integrity - Planning for success and focusing on issues that matter - Monitoring, auditing and inspection of study quality

[32] https://www.goodtrials.org/guidance

CORE-MD

CORE-MD, Coordinating Research and Evidence for Medical Devices, aims to translate expert scientific and clinical evidence on study designs for evaluating high-risk medical devices into advice for EU regulators.

For more information, visit: www.core-md.eu

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 965246.