



KONINKLIJKE NEDERLANDSE
AKADEMIE VAN WETENSCHAPPEN

EVALUATION OF NEW TECHNOLOGY IN HEALTH CARE

IN NEED OF GUIDANCE FOR RELEVANT EVIDENCE



FORESIGHT STUDY

EVALUATION OF NEW TECHNOLOGY IN HEALTH CARE



© 2014 Royal Netherlands Academy of Arts and Sciences
Some rights reserved.

Usage and distribution of this work is defined in the Creative Commons License, Attribution 3.0 Netherlands. To view a copy of this licence, visit:
<http://www.creativecommons.org/licenses/by/3.0/nl/>

Royal Netherlands Academy of Arts and Sciences
PO Box 19121, NL-1000 GC Amsterdam
T +31 (0)20 551 0700
F +31 (0)20 620 4941
knew@knew.nl
www.knew.nl


pdf available on www.knew.nl

Preferred citation: KNAW (2014). *Evaluation of new technology in health care. In need of guidance for relevant evidence*. Amsterdam, KNAW.

Basic design: Edenspiekermann, Amsterdam
Typesetting: Ellen Bouma, Alkmaar

Illustration cover: Thinkstock/ Wavebreakmedia Ltd

ISBN 978-90-6984-686-6

The paper for this publication complies with the  iso 9706 standard (1994) for permanent paper

EVALUATION OF NEW TECHNOLOGY IN HEALTH CARE

IN NEED OF GUIDANCE FOR RELEVANT EVIDENCE

FOREWORD

Sooner or later, most of us will have some sort of encounter with medical technology for a shorter or longer period of time, either directly or indirectly. The list of medical aids and devices for diagnosis, monitoring or treatment is very long and continues to grow steadily, in step with advances in research and innovation. Medical technology covers a broad spectrum of devices and procedures: from an echo at 20 weeks' pregnancy to an MRI of a fractured knee, and from a pump to administer narcosis or painkillers intravenously to – more dramatically – a pacemaker or a complete hip replacement.

Every medical device must not only be entirely safe, but it must also work as it was meant to. That is true both for the equipment that medical professionals use on our behalf and for the aids that we are increasingly using ourselves, for example diagnostic tests and blood pressure meters. Moreover, when new medical technology becomes available, we should be able to assume that it is at least as good as what the market already offers, and that it also costs less if at all possible. That is no unnecessary luxury if we remember that there are already approximately 500,000 types of medical devices in circulation, from thermometers to surgical robots.

It is quite difficult to evaluate whether a new medical device offers any advantages, and what those advantages are. After all, there is more involved than technical quality and safety. Any evaluation also has to consider the varying usage and user requirements, sector-specific guidelines and legislation. But all the 'ifs, ands and buts' should never throw up insurmountable barriers to the introduction of new medical technologies.

This advisory report was prepared by an Academy foresight committee chaired by Professor Carl Moons. The report aims to offer the various stakeholders guidelines for selecting the research method that best suits the relevant medical device, including postmarketing surveillance. The report makes clear that there are different ways of tackling such evaluations, and that a 'one-size-fits-all' approach is insufficient.

We hope that this foresight study will help stakeholders subject new medical technology to proper testing so that users can depend on safe, good quality, user-friendly diagnostic and therapeutic medical devices.

Hans Clevers

President of the Royal Netherlands Academy of Arts and Sciences

CONTENT

FOREWORD 4

PREVIEW – MAIN POINTS 8

EXECUTIVE SUMMARY 10

1. SCOPE, AIMS AND TARGET READERSHIP 17
 - 1.1 Scope and motivation 17
 - 1.2 Problem definition – medical devices 18
 - 1.3 Aim and target readership 19
 - 1.4 Breadth and limitations of the report 20
 - 1.5 Approach 21
 - 1.6 In this report 21
 - 1.7 Guidance and recommendations per stakeholder 22

2. MEDICAL DEVICES: VARIETY, STAKEHOLDERS, USERS AND CLINICAL EVIDENCE 24
 - 2.1 Definition and indication of variety 24
 - 2.2 Regulatory bodies 25
 - 2.3 Medical device industry 30
 - 2.4 Health care professionals 30
 - 2.5 The user and health policy perspective 31
 - 2.6 Clinical evidence 32
 - 2.7 The medical device life cycle and the stakeholders involved 34

3. STAKEHOLDER VIEWS ON MEDICAL DEVICE INNOVATION AND EVALUATION 37
 - 3.1 General issues per phase of the life cycle of medical devices 38
 - 3.2 Issues related to the clinical evaluation of devices 40

4. CLINICAL RESEARCH ON MEDICAL DEVICES: PRINCIPLES AND APPROACHES 50
 - 4.1 Aim and scope 52
 - 4.2 Overarching principle for device evaluations 52
 - 4.3 Therapeutic devices versus non-therapeutic (diagnostic, screening, monitoring test) devices 53
 - 4.4 Optimal clinical study approach 54

4.5	Deviation from the optimal study approach: no one-size-fits-all	56
4.6	Description of the pathway leading to a device's benefits: a collaborative effort	59
4.7	Direct versus indirect evidence approaches	61
4.8	Summary: key issues to consider in all clinical studies on medical devices	68
5.	COMPLEMENTARY ISSUES	74
6.	CONCLUDING REMARKS	80
APPENDICES I-VII 83		
I	Resolution inaugurating the committee (<i>Instellingsbesluit</i>)	84
II	Committee and secretariat	87
III	Definitions, risk classification, and regulation of medical devices	89
IV	Overview of clinical study approaches for medical devices	101
V	Statistical methods to correct for confounding in non-randomised (direct evidence) study approaches	149
VI	Concise comparison medical devices versus pharmaceuticals	152
VII	Interviews and external consultations	155

PREVIEW – MAIN POINTS

The aim of this report is to provide explicit guidance for research suitable for assessing and inferring the benefits and performance of medical devices, tailored to the various types of devices, their specifics, and the intended contexts, indications and individuals for which they are used.

Establishing the benefits of medical devices poses specific challenges because they are intrinsically diverse in terms of use, users, sector and regulation. All this makes it a complex matter to design and conduct valid research into the merits of medical devices and their use. In light of this diversity, this report considers:

- why there is no one-size-fits-all approach to evaluating the risks, performance and benefits of devices;
- why and how research on the benefits and performance of devices differs between ‘therapeutic’ devices (e.g. pacemakers, nerve stimulators, prostheses) and ‘non-therapeutic’ devices (e.g. diagnostic, monitoring, screening or prognostic tests);
- the key principles of device evaluations;
- the optimal approach for evaluating the benefits of devices;
- alternative research approaches, in view of the specifications and the targeted context, users and individuals/patients of a device;
- why and how device evaluations are enhanced when device developers, manufacturers and end-users (e.g. professionals and targeted individuals/patients) collaborate and describe at an early stage the potential mechanisms/pathways through which, and in whom, device use leads to intended (i.e. benefits) and unintended (i.e. risks) effects on health or health care;
- how knowledge of these ‘working mechanisms’ helps to place evidence taken from multiple studies on a device – e.g. technical, safety and clinical studies – into a ‘linked’ or ‘network of evidence’ perspective;

- why a linked or network of evidence approach is better suited to device evaluations than a 'hierarchy of evidence' approach.

This report is meant to help stakeholders involved in the manufacturing, evaluation, use, and regulation of medical devices choose the proper approach for assessing the health and health care benefits of medical devices, given the device specifics. It gives numerous empirical examples involving a wide variety of medical devices. It should be read against the background of current international policy development, registries regarding the use of medical devices following their market introduction, user involvement, end-user training requirements, and (cost-)effectiveness assessments. This guidance report ultimately aims to protect society and users/end-users against the introduction and use of devices that are 'unsafe' or 'unnecessary'.

EXECUTIVE SUMMARY

Scope

The development, evaluation and introduction of medical innovations and technologies in general and innovative medical devices in particular are considered an increasingly important factor in addressing the grand societal challenge related to health and wellbeing. In view of the rapid technological advances in and the rising cost of health care, society expects new medical devices to have added benefits.

The main topic of this report is how to address the added benefits of medical devices. The overarching principle that actually drives the evaluation and regulation of any health care intervention, including medical devices, is:

To generate and accumulate evidence that the use of a device is not only safe but also has benefits, preferably added benefits beyond existing care, for the health or health care of the intended individuals, patients, professionals or for society at large.

Benefits can include a direct therapeutic effect yielding improved health outcomes for the targeted individuals or users. But it may also include indirect benefits by improving ease of use, facilitating or improving screening or diagnosis of diseases, or reducing the burden on patients or the costs associated with medical care. Establishing the benefits of medical devices poses specific challenges. This report provides tools to support and enhance the clinical research tradition of assessing those benefits.

Aim and targeted readership

The *first aim* of the report is to provide explicit guidance concerning the application of research methods and approaches suitable for assessing the benefits and performance of medical devices. The *second aim* is to tailor this guidance to the various device types and specifics, as well as to the contexts and indications in which they are used, and the individuals in and users by whom they are used.

This guidance is meant for all stakeholders involved in the evaluation, use, and regulation of medical devices, including researchers and professionals in health care, the medical device industry (SMEs in particular), Notified Bodies, health insurance companies, hospital boards, regulatory agencies, funding agencies and medical ethical review committees.

Motivation

In this report, the Royal Netherlands Academy of Arts and Sciences (KNAW) aims to contribute to the European Union's request that its Member States 'take into account that improved research frameworks and criteria are needed to enhance reliability, predictability, speed and transparency in the decision-making on the introduction, use and reimbursement of medical devices'. The specifics of medical devices, however, call for a tailored approach, as they could hamper innovation and economic opportunities in Europe.

Medical Device Variety and Specifics

Chapter 2 of this report offers an overview of the variety of medical devices and associated regulatory aspects. There are many different medical devices (almost 500,000). They are intrinsically diverse in terms of use, users, sector and policy, making it a complex matter to design and conduct valid research into a device's merits.

- *Device use* differs in its aims (e.g. for diagnostic, prognostic, screening, therapeutic, or supportive purposes), durability (e.g. disposable vs. implantable) and mode of action. The need to collect evidence about its (added) benefits is more pressing the higher the risk involved in using a device. However, there is no need to explicitly establish clinical benefits for all medical devices that are used in health care: there is *no one-size-fits-all*.
- Consumers, professional caregivers and health care policymakers may have differing *user perspectives*. Consumers need to be assured that the use of the device is safe and has a positive benefit/risk ratio. Policymakers want the introduction of new, often costly devices to be safe, beneficial *and* cost-effective. Health care professionals share the concerns of both consumers and policymakers; they can, moreover, influence a device's performance, benefits and safety by the way they handle it and by their skills.
- *The market* for medical devices stretches far beyond the professional care setting. Devices are sold by large European multinationals, by SMEs and over the counter. Decentralised health care solutions and widespread use of mobile technologies are opening up innovative ways of increasing patient self-management. The life cycle of medical devices is relatively short (2.5-6 years on average) and in many cases, incremental modifications improve a device over time.

All this contributes to the complexity of research assessing the benefits of device use.

Stakeholders Views

Chapter 3 presents the results of a field survey among stakeholders. All of the stakeholders warned against comparing medical devices with pharmaceuticals and implementing a research paradigm similar to that used in drug approval and reimbursement. However, they also identified critical issues and challenges related to new developments and device evaluation:

- In recently amended EU legislation on medical devices, the emphasis on clinical evaluation has increased but actual guidance on how to design such research is not elaborated. More awareness is needed of the challenge posed by new types of medical devices, such as combination products and self-management products.
- Research into the benefit of a device *after* CE certification to encourage its use and uptake by intended users, suffers from inadequate knowledge of the pros and cons of different research approaches. Frequently, line extensions of existing devices have resulted in numerous small datasets, whereas research commonly takes place in highly controlled settings of experts having undergone extensive user training, which all is very different from regular care.
- There are safety issues related to device use in regular care because devices are commonly applied by untrained professionals. This report does not explicitly provide recommendations for training professionals in new device use, as this subject is addressed by the Dutch Order of Medical Specialists (OMS).
- Registries, including clinical data on using a device in regular care, are crucial for assessing the long-term safety, performance, benefits and cost-effectiveness of devices.
- Health technology assessments (HTA) are hardly ever conducted for medical devices, let alone endorsed. However, in the current era of evidence-based medicine, the benefits and economic considerations of device use need to be taken into account. This should ideally happen in an early phase of new device development.

Research Approaches Tailored to Medical Devices

Given the concerns of the stakeholders, **Chapter 4** focuses on the main principles of research into the performance, benefits and added benefits of medical devices. The chapter provides explicit guidance concerning the pros and cons of different study approaches, given the device specifics and the targeted context, patients and users of a device. This guidance enhances ones understanding of the available evidence, helps in planning subsequent studies, and improves the dissemination, uptake and application of *safe and beneficial devices* by professionals, patients and other stakeholders.

No one-size-fits-all approach

The guidance in Chapter 4 departs from the above overarching principle. It continues to make a distinction between two main categories of devices, therapeutic devices versus non-therapeutic devices. The latter include diagnostic, monitoring, screening or prognostic devices or rather tests. Therapeutic devices usually interfere *directly* with – often targeted – bodily systems and mechanisms. Examples are pacemakers, nerve stimulators, prostheses, breast implants and surgery robots. Therapeutic devices treat specific diseases directly, alleviate specific symptoms or complaints, or improve daily activities. Diagnostic, prognostic, monitoring or screening test devices do not treat or alleviate diseases, symptoms or signs directly, but *indirectly*. Examples include imaging tests, companion diagnostics, laboratory tests, or point-of-care tests. Such devices provide information to users, e.g. professionals or patients, which in turn direct subsequent actions (e.g. therapies or lifestyle changes) that may lead to benefits, e.g. improved health. However, test devices may also be beneficial because they facilitate better therapeutic action by medical drugs (such as companion diagnostics), or because they lead to less invasive, burdensome or costly detection of disorders (such as screening or point-of-care tests). Finally, unlike therapeutic devices, many test devices, e.g. imaging tests, are not intended for just one specific medical condition or indication.

Research approaches to assessing the risks, benefits and performance of both types of devices are markedly different. This is why there is *no one-size-fits-all approach possible* for the evaluation of medical devices.

Pathway of device benefits

The variation in devices is reflected in the many different working pathways or mechanisms through which each device leads to intended (benefits) and unintended (risks) effects on health or health care. Deciding which evidence and research are actually needed can first and foremost be enhanced by device developers, manufacturers and end-users (i.e. targeted professionals and patients) collaborating and describing in detail the potential pathways through which device benefits and risks are likely to arise. This working pathway is ideally defined in the very earliest development stages or even at the conception of a device. A detailed description of the following issues can serve to put evidence from different types of device studies – e.g. technical, safety and clinical studies – into a linked or network of evidence perspective:

- the anticipated technical or analytical capabilities of a device;
- the expected unintended and intended effects in the targeted context;
- in whom these effects are likely to occur, e.g. in the targeted individuals/patients, in the care providers or in health care at large;
- the anticipated mechanisms through which these potential risks and benefits will occur or be achieved in the intended context;

- the existing care in the targeted context and individuals;
- the expected time frame in which potential risks and benefits might occur.

The optimal study approach and alternative strategies

To accumulate evidence that the use of a device is safe and has positive effects on health or health care beyond those achieved by current practice, one could design an ideal study that measures all these aspects directly in the most valid and informative way. **Chapter 4.4** describes the essentials of such a pragmatic or comparative effectiveness trial. Such randomised comparative effectiveness studies are methodologically much more challenging for medical devices than for medical drugs. This is due to the interplay between technical device complexities, user skills and learning-curve issues, all of which influence the benefits and risks of a device and its use. Alternative research approaches are needed to evaluate the performance and benefits of device use.

Three main approaches

To generate evidence that the use of a device has benefits, preferably added benefits beyond existing care, for the intended individuals, professionals or society at large, we can make use of three main study approaches, each with its own merits and vulnerabilities. **Chapter 4.7** provides detailed guidance on these approaches, complemented by numerous examples applied to a variety of medical devices in **Appendix IV**.

1. Studies providing direct evidence of the benefits or added benefits of a device use for health or health care.

Such studies basically address all the issues related to the device use as in practice, the intended context, a comparison strategy, the relevant effects and timing of these effects, and required study size. In addition to the ideal large-scale, long-term, comparative effectiveness trial, there are numerous alternative study approaches that may also provide direct evidence. These include the traditional randomised designs. But they also include more innovative and efficient randomised approaches, as well as non-randomised study approaches that can help generate direct evidence about the long-term performance and benefits of device use.

2. Studies providing indirect evidence of the benefits or added benefits of device use for health or health care, using a *quantitative* linked-evidence approach.

Indirect evidence approaches have in common that they do not directly measure the ultimate health outcomes relevant for the targeted individuals, context or users. Instead, such approaches focus on outcomes measuring intermediate changes along the working pathway of device use. Indirect evidence approaches are very useful in situations where the direct evaluation of device use is absent. Their validity depends

on how well the ‘intermediate’ outcomes relate to long-term health (or health care) outcomes. Linked-evidence modelling approaches link various types of evidence, ranging from technical performance to clinical performance or clinical benefits studies. Quantitative linked-evidence approaches are very relevant for test devices and for devices that are modifications of an existing device. They can quantify to what extent a minor device improvement leads to health benefits for the targeted individuals, context and users, using study results from previous versions of the device.

3. Studies providing indirect evidence on the benefits or added benefits of device use on health or health care, using a *qualitative* linked-evidence approach.

Evidence of a device’s benefits on long-term health or health care outcomes in a specific indication, context or user group, may also be inferred indirectly, for example, when adapting evidence from studies conducted in different individuals, patients or users or from studies across different indications of device use. Evidence gained from technical performance studies, safety studies, studies on related devices or preceding versions of the device, should be linked and put into perspective. This again requires knowing the working pathway through which device use may lead to benefits for health or health care. Inferences about the relevant, long-term health or health care benefits (and risks) of device use taken from such *qualitative* linked-evidence approaches do not have the same validity as the approaches discussed under 1 and 2. Nevertheless, qualitative linked evidence is currently often considered sufficient for market access, and perhaps even for reimbursement decisions.

Complementary issues

Chapter 5 addresses a number of issues that go beyond the explicit guidance for research strategies in Chapter 4, set against the backdrop of European policymaking concerning medical devices. These issues include: registries on the use of medical devices after market access; user involvement; end-user training; and evaluation of the cost-effectiveness (HTA) of devices.

Recommendations for stakeholders

- *For researchers and industry*, explicit guidance for assessing the benefits of device use is presented in Chapter 4 and illustrated across a wide range of medical devices in Appendices IV and V. If this guidance is properly introduced at an early stage of device development, a wealth of suitable research approaches becomes available.
- *For regulatory agencies*, including Notified Bodies, Competent Authorities, medical ethical review committees, and health research funders, Chapters 2 and 3 and Appendix III discuss general regulatory issues. These agencies will also benefit from the general research guidance described in Chapter 4 when judging device research proposals and interpreting the results of such research

- *For health care professionals and hospital boards*, the guidance presented in Chapter 4 is useful for judging what evidence is actually available and what evidence is still lacking when deciding to purchase and implement a particular device.
- *For health insurance companies and health care policy at large*, Chapters 3 and 5 and Appendices III and VI demonstrate that medical devices have their own intrinsic issues and challenges. At the same time, Chapter 4 illustrates the level of evidence that the various research approaches should produce when judging the potential benefits of a particular device.

In conclusion

This report guides stakeholders involved in the evaluation, use, and regulation of medical devices, towards choosing the proper approach for assessing the benefits and performance of a device for health care, given the specifics of the device. The report also offers guidance on how to interpret existing evidence on the risks (safety), performance, and benefits of device use, and how to put this evidence into perspective so as to make better-informed decisions regarding the introduction, use, and reimbursement of a device. This guidance ultimately aims to protect society, users and end-users against the introduction and use of devices that are ‘unsafe’ or ‘unnecessary’.

1. SCOPE, AIMS AND TARGET READERSHIP

1.1 Scope and motivation

Europe faces major long-term societal challenges in health care, for one thing because it has an ageing population and consequently a growing proportion of the overall population with co-morbidities that call for high-quality and innovative health care systems.¹ The introduction of medical innovations, including innovative medical devices, will be increasingly crucial to efforts to address the challenge of sustained health and wellbeing.² As health is central to people's lives, access to safe products and good quality health care services is one of the fundamental human rights.

In the Netherlands, the Life Sciences & Health sector – one of the country's key economic sectors – encompasses an innovative industry, an excellent knowledge base and a long tradition of collaborative and integrative development between industry, academia, health care professionals and health policymakers. The sector aims to develop effective and cost-effective health solutions and to accelerate their delivery.³ This involves considering the impact – safety and benefits – of the total health solution enabled by innovations, complemented by applying more comprehensive Health Technology Assessment (HTA) approaches. It is therefore crucial that breakthrough medical innovations, including medical devices, demonstrate their clinical and other benefits and performance for health care, as well as their safety.

1 ERAB (2009), *Preparing Europe for a New Renaissance – A Strategic View of the European Research Area*.

2 European Council conclusions on innovation in the medical device sector; 2011/C 202/03.

3 Summary Innovatiecontract Topsector LSH, April 2012.

1.2 Problem definition – medical devices

According to European Union legislation, the introduction and assessment of medical devices depend on a 'suitable, robust, transparent and sustainable regulatory framework for the benefit of European patients, consumers and health care professionals, adapted to the needs of tomorrow'.⁴ 'To create the proper conditions allowing safe and beneficial (innovative) medical devices to be deployed within a well-functioning internal market': that is the motivation for the proposed amendment of the EU's medical device legislation (2012).⁵ The proposals are meant to maintain consumer confidence and adapt to a global market; at the same time, they acknowledge the medical devices sector as a key driver of the EU's economic growth. They place more emphasis on the clinical benefits of medical devices.

What are benefits?

In the current situation of rapid technological advances and spiralling health care costs, society expects new medical devices to produce benefits, and preferably added benefits beyond existing or prevailing care. Benefits can be broadly defined as: direct therapeutic effects on patients; indirect benefits that enhance screening, monitoring, diagnosis or prognosis and treatment response; indirect benefits that improve ease of use by professionals or other end-users; and benefits in terms of reducing costs or the burden on patients associated with medical care. More generally stated, medical devices are beneficial if they provide a good and efficient means to address a medical problem. The challenge for society, however, is how to demonstrate the added benefits of a medical device.

Establishing the benefits of medical devices poses specific challenges when compared to pharmaceuticals. The current regulation of medical devices primarily focuses on safety and product performance. There is a less explicit demand for evidence concerning the benefits for users, targeted individuals or patients, let alone for health care at large. New clinical research may not always be necessary to demonstrate a device's benefits, but when such research is needed to produce evidence of a device's benefits, including performance, explicit guidance is required on how to provide such evidence and on the pros and cons of different research approaches.

4 Communication from the European Commission to the European Parliament, the Council, the European Economic and Social Committee, and the Committee of the Regions (COM(2012)540 final, accompanying two revised regulations).

5 Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on *in vitro* diagnostic medical devices, COM(2012) 541 final; 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) 542 final.

Evaluation principle

This report starts from the overarching principle that actually drives the evaluation and regulation of any health care intervention:

To generate and accumulate evidence that the use of an intervention, including a medical device, is not only safe but also has benefits, preferably added benefits beyond existing care, for the health or health care of the intended individuals, patients, professionals or for society at large.

The main reason for this report is the current lack of guidance on methods or designs that can help collect evidence allowing for the assessment of the **benefits (or added benefits)** of medical devices, which in turn would justify their introduction and use in practice and public investment in health care as well as guide reimbursement decisions.

Academy request

In view of the foregoing, the Royal Netherlands Academy of Arts and Sciences (KNAW) decided to initiate an advisory project and installed a committee to develop a strategy that would increase awareness of methodological considerations and provide guidance concerning the **research requirements** for evaluating medical technology, with a focus on medical devices. The guidance covers existing and novel research approaches to facilitating rapid production of the necessary evidence. The resolution inaugurating the committee is given in **Appendix I**.

1.3 Aim and target readership

In this report, the Academy aims to respond to the EU's request that its Member States '*take into account that improved research frameworks and criteria are needed to enhance reliability, predictability, speed and transparency in the decision-making on the introduction (and reimbursement) of medical devices*'.⁶ To this end, the Academy's Council for Medical Sciences has assumed responsibility for building on its independent scientific knowledge base. In addition, the research infrastructure for the scientific evaluation of medical innovations and technologies is crucial, as concluded by European⁷ and Dutch (Academy)⁸ institutions. More specifically:

- the **first aim** of this report is to provide explicit guidance concerning the application of research methods and approaches suitable for assessing the clinical

⁶ European Council conclusions on innovation in the medical device sector; 2011/C 202/03.

⁷ The European Clinical Research Infrastructures Network (ECRIN) is a sustainable, not-for-profit infrastructure supporting multinational clinical research projects in Europe.

⁸ Briefadvies KNAW aan de minister van VWS over versterking infrastructuur klinisch onderzoek (2010). (<http://www.knaw.nl/smartsite.dws?lang=NL&id=26103&pub=20101058>). The NFU and ZonMw endorsed the request for a more sustainable clinical research infrastructure.

- performance and benefits or added benefits of medical devices;
- the **second aim** is to tailor this research guidance to the various device types and specifications as well as the contexts in which and the individuals by whom they are used.

Stakeholders addressed

This guidance intends to support current national and international medical device research and review guidelines and planned developments pertaining to this matter, aligned with other activities in this area. This guidance thus addresses and is meant for various stakeholders and parties involved in the evaluation, use, and regulation of medical devices, both in the Netherlands and abroad. These include:

- Researchers and professionals in health care;
- Medical device industry, and SMEs in particular;
- Notified Bodies (e.g. Dekra);
- Health insurance companies;
- Hospital boards;
- Regulatory agencies both in the Netherlands (e.g. Health Insurance Board) and abroad (e.g. European Medicines Agency);
- Funding agencies in the Netherlands (e.g. ZonMw) and abroad (EU, Horizon 2020);
- Medical ethical review committees (MERC).

1.4 Breadth and limitations of the report

The committee has made the following decisions concerning the breadth of the advisory report:

- The report covers all types of medical devices for therapy, diagnosis, prognosis, screening and monitoring, thus including active implantable devices and *in vitro* diagnostic devices.
- The report covers the whole life cycle of a medical device.
- The report intends to make explicit which type of clinical research approaches and methods are needed to assess the clinical performance and benefits of devices in view of the type of device and intended context. The report does **not** provide a standard clinical research approach (one-size-fits-all approach) suitable for all medical devices, but rather tailored to the types and intended context and use of devices.
- The report focuses on guidance to support the evaluation of the clinical performance and benefits of devices.
- The report does **not** specifically address novel Health Technology Assessment (HTA) approaches, other than referring to existing activities in that field.
- The report does **not** intend to change existing rules for market access or reimbursement decisions, but only provides guidance to allow for informed clinical and policy decision-making.

Finally, the committee has explored whether the burden of proof that is required for the effectiveness and safety of pharmaceuticals could be applied to non-pharmaceutical medical technologies. It concluded that in view of current legislation, and in view of the variety of medical devices available, such a discussion should take place in the appropriate policy organisations. Accordingly, the report does **not necessarily** depart from a 'hierarchy of evidence' approach, but rather promotes a 'network' or 'linked-evidence' approach. To this end, the report also considers how to use cumulative datasets, and how to align data collected pre-CE and post-CE approval, in order to increase the value of data used to infer the benefits or added benefits and performance of a device. In addition, the report discusses parameters for registration that will help build clinical evidence.

1.5 Approach

The committee began by holding several meetings to discuss at length the scope, breadth and limitations of the report, as well as the definitions of benefit, added benefit, performance, clinical value, and the aim and target readership of this guidance. The discussions were followed by input from stakeholders in the fields of policy, research and industry. The input was obtained at several different points by means of a field survey, interviews, a parallel session at the KIVI-NIRIA Technology and Care conference on 11 October 2012,⁹ and by inviting experts to various committee meetings. A list of all interviewees and contributors is provided in Appendix VII. Finally, based on the discussions and a review of the methodological literature focusing on medical devices, the committee formulated the guidance for research approaches to assessing the benefits or added benefits of such devices.

1.6 In this report

This report has surveyed the potential approaches that can be used to assess the health care benefits or added benefits, including performance, of a medical device. Chapter 2 gives a concise overview of the variety of medical devices available and the consequences this has for research into benefits. Appendix III explains the EU and USA legislation on medical devices in more detail. Chapter 3 summarises what stakeholders believe are the most important issues and challenges posed by medical device evaluations. Chapter 4 provides an exhaustive overview of and explicit guidance for research approaches that can help demonstrate and infer the benefits and added benefits of medical devices for health and health care, given the type, intended context, indication and use of the device. This guidance is illustrated by numerous empirical examples involving a wide variety of medical devices, detailed in Appendices IV and V. Chapter 5 provides final remarks and recommendations.

9 <http://www.kiviniiria.net/a/PAG000009587/>.

1.7 Guidance and recommendations per stakeholder

The guidance in this report is meant for various stakeholders and parties involved in the evaluation, use, and regulation of medical devices. Specifically:

- For **researchers and industry, Chapter 4 and Appendices IV and V offer** explicit guidance for assessing the benefits of device use and the key principles involved. The guidance is explicitly illustrated by numerous real-life examples across the wide range of medical devices, given a device's specifics, its intended indication, users and targeted individuals or patients. The report distinguishes between research aimed at providing direct, quantitative evidence of the added health or health care benefits of device use and research aimed at providing indirect evidence of a device's added benefits, which can be furnished quantitatively or qualitatively. Knowledge of the pathway by which a particular device's risks and benefits are addressed, and in whom, is extremely important. If these key principles and pathways are properly addressed at the earliest possible stage of device development, then a wealth of suitable research approaches is available. Chapter 4 also provides a **list of key issues** that need to be considered by researchers and industry in any study concerning the health and health care benefits of devices.
- For **regulatory agencies, including Notified Bodies, Competent Authorities, medical ethical review committees, and health research funders**, the general regulatory issues are discussed in **Chapters 2 and 3** and **Appendix III**. These regulatory agencies may also benefit from the general research guidance given in **Chapter 4** when judging research proposals on devices and interpreting the results of such research. Notably, it is useful for regulatory stakeholders to have a knowledge of: the pathways or mechanisms through which the risks and benefits of a particular device can be addressed and in/for whom; the key issues of device research; and the type of evidence that is required in connection with the regulatory requirements. This applies not only to national but also to international regulatory bodies.
- For **health care professionals and hospital boards, Chapter 4** offers guidance on ascertaining what evidence is available and what evidence is still lacking when making decisions on purchasing and implementing a particular device. This knowledge, plus a knowledge of the pathways or mechanisms through which the risks and benefits of a particular device are addressed and in/for whom, will increase their awareness and motivate them to ask manufacturers for such information, and not simply introduce and use devices without such information.
- For **health insurance companies and health care policy** at large, **Chapters 2, 3 and 5**, and **Appendices III and VI** demonstrate that medical devices have their own intrinsic issues and challenges. At the same time, Chapter 4 illustrates the degree of evidence required in the different research approaches when judging the

potential health benefits of a particular device. These approaches can be incorporated into formal cost-effectiveness or health technology assessments (HTA), tailored to the medical device concerned. Only such HTA approaches make it possible to assess the risks and benefits of introducing a health care device into society at large.

2. MEDICAL DEVICES: VARIETY, STAKEHOLDERS, USERS AND CLINICAL EVIDENCE

This chapter summarises the variety of medical devices available and the relevant issues involved in evaluating them. The benefits of such devices are described from the perspective of the intended user and patient. The chapter further discusses the context in which devices are used in health care. It also introduces the medical devices industry and briefly reviews the regulatory requirements. The chapter ends with a description of the different phases in the medical device life cycle.

2.1 Definition and indication of variety

Definition

The committee adopted the European Commission definition of a medical device: *'any instrument, apparatus, appliance, software, material or other article, whether used alone or in combination, including the software intended by its manufacturer to be used specifically for diagnostic and/or therapeutic purposes and necessary for its proper application, intended by the manufacturer to be used for human beings'* (Directive 93/42/EC).

See **Appendix III** for more details on medical device definitions and legislation. In addition to diagnostic and therapeutic purposes, we explicitly note that medical devices can also be used for screening, prognostics, and disease or therapy monitoring. This definition actually encompasses a very broad range of medical devices. While maintaining this breadth, it goes without saying that the guidance for clinical research approaches provided in this report does not truly cover *all* medical technology or innovations, but focuses on those used for therapy, diagnosis, prognosis, screening or monitoring.

Variety of medical devices

There are many different types of devices in different classes, ranging from medical implants and medical aids to *in vitro* diagnostic tests and medical imaging, to mention only a few. Of the almost 500,000 different medical devices, the majority are relatively simple, e.g. a disposable syringe or an ear thermometer. Various devices are complex, however, and reflect the latest advances in medical technology, for example new imaging equipment, various heart, vessel, bone and joint implants, and advanced point-of-care lab tests. Medical devices are thus intrinsically diverse: their lifespan varies widely, with extremes ranging from a few seconds for disposable devices to several decades for some implantable devices and medical equipment. Some medical devices have expiration dates, whereas other long-lived individual equipment may undergo replacement of components.

Variety in medical device use

Medical devices are used for a variety of diagnostic, prognostic, screening, monitoring and therapeutic indications. Consequently, there are differences in the level of risk and in the regulatory systems used to manage those risks; in the manufacturing costs and sale prices; in the standards and in the nomenclature systems; and in the various approaches used to determine their safety, performance, and benefits.¹⁰ The devices market stretches far beyond the professional care relationship. Devices are also sold over the counter. They may be developed for use in laboratories, in first aid kits, in kindergartens and in homes for the elderly. In the coming years, the market for medical devices will become even more crowded. Companies are preparing a range of self-testing devices, and major European multinationals such as DSM, Philips and Siemens are also investing a great deal of effort in health care devices and other types of medical support, often intended for primary and home care and for patient self-management. The question is how to deal with the wide variety of existing medical devices, which is set to increase, perhaps in ways that we cannot anticipate.

2.2 Regulatory bodies

The variety of medical devices available and the growing number of technological innovations are also having a significant impact on the methodologies and types of study needed to collect evidence in order to ensure the introduction not only of 'safe' but also of 'useful' and 'beneficial' devices in health care. **Box 2.1** lists some key definitions and concepts of medical device research.

Legislation imposes an obligation to assess the balance between risk and performance, where performance means the ability of a device to achieve its intended

10 *Medical devices: managing the mismatch: an outcome of the priority medical devices project* (2010), World Health Organisation. ISBN 978 92 4 156404 5.

purpose (or use) as claimed by the manufacturer. Performance is thus not necessarily clinical performance, let alone clinical or health/health care benefits. Clinical evaluation is, moreover, not needed for all medical devices. The variety of different devices makes it difficult if not impossible to define a common (one-size-fits-all) framework for clinical investigations evaluating the risks and performance of a device (and its use), let alone to define its benefits or added benefits for health care.

BOX 2.1 TERMINOLOGY, CONCEPTS AND PRINCIPLES USED FOR MEDICAL DEVICES

The definitions given below are derived from the Global Harmonization Task Force (GHTF) and its follow-up initiative, the International Medical Devices Regulators Forum documents. There are no restrictions on reproducing this information; however, it should be noted that reproduction does not convey or represent an endorsement of any kind by the GHTF.

Clinical evidence – The information or clinical data that supports the scientific validity and performance (analytical performance and, where applicable, clinical performance) of a device when used as intended by the manufacturer.

Analytical/Technical Performance – The ability of a device to correctly detect or measure a particular analyte.

Clinical evaluation – The assessment and analysis of clinical data pertaining to a medical device to verify the clinical safety and performance of the device when used as intended by the manufacturer. Clinical evaluation is an ongoing process; information about clinical safety and performance (e.g. adverse event reports, results from any further clinical investigations, published literature, etc.) should be monitored routinely by the manufacturer once the device is available on the market and the benefits and risks reassessed in light of this additional information.

Clinical investigation (synonymous with ‘clinical trial’ and ‘clinical study’) – Any systematic investigation or study in or on one or more human subjects, undertaken to assess the safety and/or performance of a device.

Clinical utility – A concept generally used for diagnostic devices that refers to the likelihood of the test significantly improving the health outcomes of the targeted individuals. In other words, the capacity of the test to rule in and/or rule out the disorder of interest, and to facilitate a decision to adopt or to reject a subsequent (e.g. therapeutic) action. Clinical utility is an increasingly common concept in health care, but one that lacks an agreed formal definition or conceptualisation. The term is commonly used as a synonym for clinical effectiveness.

Clinical performance – The ability of a device to yield results that are correlated with a particular clinical condition or a physiological state in accordance with a target population and intended user. Interventional clinical performance is where the test results may influence patient management decisions and/or may be used to guide treatment.

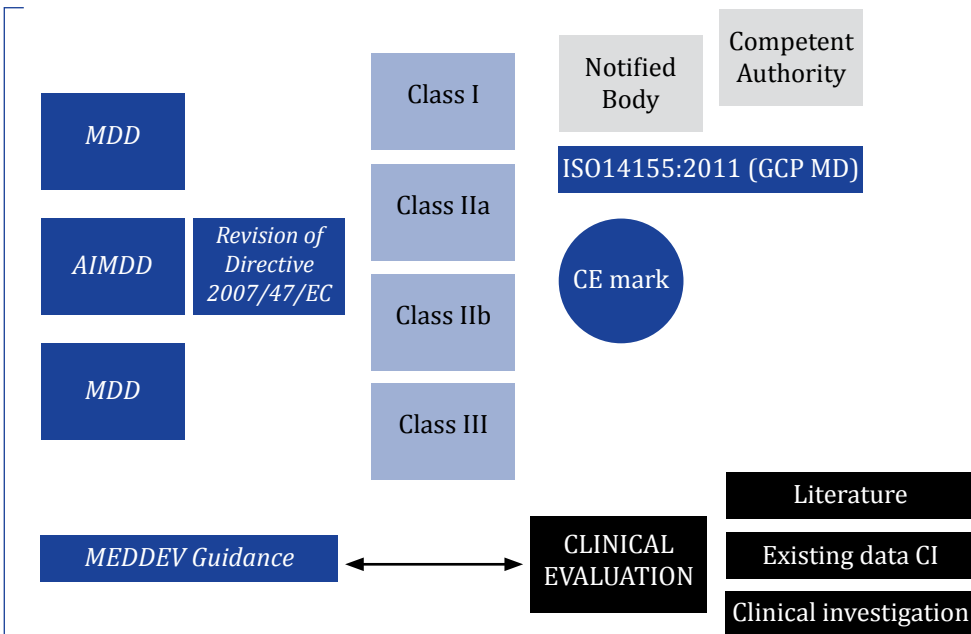


Figure 2.1 Diagram of EU legislation concerning medical devices.

Blue boxes at the left represent the directives covering medical devices (MD, including In vitro devices or IVD, and active implantable medical devices or AIMD), and the Guidance indicating how to interpret the legislation (MEDDEV Guidance). In a lighter color, the four risk categories are depicted from low-risk (class I) to high-risk (class III) devices. To obtain market access (i.e. a CE mark in blue circle), a decision from the relevant authorities (in grey) is required, which depends on compliance with the directives. The higher the risk, the more requirements there are with regard to safety, performance, and clinical evaluation (in black). For lower risk devices, clinical evaluation may depend on the literature or existing data; for higher risk devices, new clinical investigations are requested.

EU legislation on medical devices in brief

A diagram of medical device regulatory requirements is presented in Figure 2.1. More detailed explanations on the medical device regulatory framework are available in **Appendix III**.

EU legislation concerning medical devices is subject to constant improvement, with more emphasis on clinical evaluation and on obtaining data through clinical investigations.¹¹ The relevant directives and MEDDEV guidance documents clearly state **when** investigations are needed to obtain sufficient evidence on clinical benefit, but not **how**

¹¹ See <http://ec.europa.eu/health/medical-devices/documents/guidelines/> for an overview of all guidance documents. See <http://ec.europa.eu/enterprise/policies/european-standards/harmonised-standards/medical-devices/> for harmonized standards for industry.

this should be done; that is the motivation for this report.¹² All this, however, does not necessarily mean that the directives are properly enforced. It is up to the Notified Bodies and Competent Authorities to demand new clinical investigations or not. The Working Group on Clinical Investigation and Evaluation (CIE) welcomes methodological contributions and a rationale for clinical research on medical devices.¹³ The fact that it is rather difficult to establish the clinical benefit of device use poses even more challenges when it comes to assessing the relative or added benefits (effectiveness) of a device compared with another device or with care as usual.

Notified Bodies in the EU

When the regulatory process is more stringent, the Notified Body must implement and impose these stricter requirements. Notified Bodies in Europe are formally autonomous from the Competent Authorities. In Europe, they have been organised to discuss the European guidelines.¹⁴ The Notified Bodies take a somewhat technical approach, focusing primarily on device safety and technical performance. One major problem, however, is that the quality of the Notified Bodies in the Member States varies widely, affecting the quality of the decision-making process across countries. In addition, the Notified Bodies' internal procedures are often not very transparent. It is easier to access the market in some EU countries than in others, and industry can choose their own Notified Body. The EU is currently addressing the uneven quality of the Notified Bodies.

USA legislation

In the case of breakthrough medical device innovations, the regulatory requirements in the USA are stricter than in the EU. In the USA, a Pre Market Approval (PMA) for new devices is required. But the impression is that in practice the equivalence argument is also used regularly to replace a PMA by the more flexible 510(k) procedure.¹⁵ There is a tendency for industry to choose Europe for innovative new devices so that they can enter the market faster, whereas the USA 510(k) procedure is used more frequently for line extensions (modifications) of existing devices, building upon equivalence arguments.¹⁶ For more on the differences between EU and USA device legislation, see **Appendix III**.

12 MEDDEV 2.7.1, see also Annex 3.

13 http://ec.europa.eu/health/medical-devices/dialogue-parties/working-groups/index_en.htm.

14 The European Association for Medical Devices of Notified Bodies; <http://www.team-nb.org/>.

15 There is a fundamental difference between the PMA and 510(k) pathways. In a PMA review, the FDA determines whether the device is reasonably safe and effective for its intended use. In a 510(k) review, the FDA determines whether the device is substantially equivalent to another device whose safety and effectiveness may never have been assessed.

16 *Congressional Research Service Report for Congress: FDA Regulation of Medical Devices*. J.A. Johnson, June 25, 2012.

Global Harmonisation

Although there has been progress in the worldwide standardisation of medical device regulation, the process is not finished and differences between the USA and Europe remain. The Global Harmonization Task Force (GHTF), and its follow-up initiative the International Medical Devices Regulatory Forum (IMDRF),¹⁷ will continue collaborating with the WHO on harmonising regulatory practices, which will eventually be implemented in national legislation. They recognised that the first step towards international harmonisation in the regulation and convergence of data requirements for medical devices is to arrive at a common terminology, concepts and principles. The GHTF's five study groups, of which Study Group 5 focused on clinical safety/performance, created several guidance documents providing key definitions and concepts for medical devices (see also **Box 2.1**).

Postmarketing surveillance

Postmarketing surveillance (PMS) is obligatory in both EU and USA legislation and should include a follow-up involving broader user populations to further build datasets on the device. Manufacturers are responsible for PMS, but the Notified Bodies should supervise, in collaboration with the Competent Authorities. Our interviewees felt that PMS supervision could improve in individual Member States and in the registries (e.g. Eudamed) (see **Chapter 3**). Registries could help amass the knowledge needed to properly review the larger follow-up studies, as we will discuss in **Chapter 5**.

Reimbursement of devices in the Netherlands

Reimbursement of medical devices in a Dutch hospital setting is part of an agreement covering 'hospitals buying' and 'industry delivering'. Hospitals sometimes do purchase devices even when the evidence for clinical benefit, let alone cost-effectiveness, is insufficient (see **Chapter 3** for example). **Appendix III** provides more background on the Dutch reimbursement system. Rising health care costs have revealed an apparent need for evidence-based medicine, cost-effectiveness, and general health technology assessment. This is likely to continue in the coming years, largely due to the continuous advances in medical technology rather than increasing life expectancy.¹⁸

17 <http://www.imdrf.org/>.

18 CPB Netherlands Bureau for Economic Policy Analysis (www.cpb.nl), published 21 March 2013.

2.3 Medical device industry

Given the variety of different medical devices available, the medical devices market is equally diverse, making it difficult to provide a general description. The market dynamics differ from those of the pharmaceuticals market, particularly with regard to the life cycle of medical devices and the innovation processes (see also **Appendix VI**). From conception to obsolescence, many medical devices have relatively short product cycles of 2.5 years on average (ranging to up to 6 or 7 years for imaging devices). Often, they undergo incremental modifications over time, renewing the cycle. Improved versions of a basic prototype may have even shorter commercial life cycles. Patents also play a different role than in the pharmaceutical sector: the focus is more on freedom to operate than exclusivity. Generally, many patents are needed to protect all the innovations in one medical device. But there are also many devices for which patents are not critical. The return on investment is based on quick market access, and the increase in the number of new patents applications coincides with the advent of new business models (like pay per use).

The medical devices industry (both large companies and SMEs) often collaborates closely with medical doctors in hospitals who are involved in clinical studies. But similarly to the pharmaceutical sector, when a ground-breaking new device enters the market, safety and performance requirements are crucial; clinical research is compulsory, and it is expensive to build up dossiers of clinical research data. The medical device sector in Europe comprises around 18,000 SMEs.¹⁹ A large number of new and innovative medical devices are thus expected to enter the market. Globally, SMEs comprise about 80% of the market.²⁰

2.4 Health care professionals

The development of new medical devices and their subsequent introduction into the health care system is often the result of close collaboration between developers and health care professionals. Depending on the type of device, health care professionals are frequently involved in different stages of the medical devices life cycle in the following ways (see also Figure 2.2):

- a. as co-developers, in the device development phase. With regard to clinical research involving a prototype, this requires the approval of the local medical ethical review committee (MERC) as well as of hospital instrument services.
- b. as researchers, participating in the clinical evaluation/implementation phase. This takes place in the context of medical scientific research that is often (partly)

¹⁹ Council of the European Union; Council conclusions on innovation in the medical device sector; 3095th EMPLOYMENT, SOCIAL POLICY, HEALTH and CONSUMER AFFAIRS Council meeting- Health issues – Luxembourg, 6 June 2011.

²⁰ <http://www.eucomed.org/newsroom/66/57/Enabling-European-medical-technology-SMEs-to-innovate?cntnt01homepage=yes&cntnt01origid=103>.

sponsored by industry.

- c. as users of a new device in regular health care. This type of use cannot be compared to device use in a research setting (under b), which usually involves well-trained researchers, professionals and a support team.

Knowledge about possible alternative clinical research approaches to assessing the health benefits and added benefits of medical devices is relevant for all health care professionals and researchers. It is equally relevant for the committees (MERCs, but also funding agencies and regulatory agencies) that review device research proposals and device research results to know about the potential research approaches to a given device and what evidence is actually produced by which research approach. This was part of the motivation for the present report.

User interference, learning curve aspects, and user risks are particularly relevant in the case of inexperienced use in regular care. To tackle this issue, the Dutch Order of Medical Specialists (OMS) is collaborating with the Dutch Health Insurance Board (CvZ) to prepare a guideline on the safe introduction of new medical devices in regular health care (see also **Chapter 5**). In this report, we therefore do not address these types of risks in detail.

2.5 The user and health policy perspective

The user perspective is important when assessing the performance of a device, largely because many, if not most, devices have no *direct* therapeutic use or benefit for the targeted patients or individuals (as we will explain in detail in **Chapter 4.3**). These notably include devices used for diagnostic, screening, monitoring or prognostic purposes. Such devices commonly consist of tests that generate information, which in turn directs further clinical, therapeutic or lifestyle actions that subsequently may change patients' health. The health effects are thus *indirect*. Such devices sometimes provide abstract information, which first needs to be interpreted by professionals before decisions can be made and the health consequences for the patient can be inferred. Interpretation and subsequent decision-making depend not only on the properties (e.g. performance) of the device, but also on the judgment or expertise of the professional, who takes other parameters into account as well.

Consumers and professional perspectives

Medical devices are used by:

1. the *patient or consumer*;
2. the *professional or caregiver*.

The user perspective is not always clear and this impairs any assessment of risk and performance and hence the risk/benefit ratio. For example, consumers may influence the risk/benefit ratio by their use of and adherence (or non-adherence) to the device. If a medical doctor is the user, there is a difference one who is trained and

one who is not trained. Furthermore, there may be differences between professional and non-professional (e.g. patient) user perspectives for the same device.²¹ Another relevant factor in this regard is that devices are used in different settings: in hospitals by experts; in decentralised settings by professionals; or at home by professionals or patients/clients.

Eventually, the individuals or patients in or for whom the device is to be used need to be assured that such use will have a positive risk/benefit ratio. There should be evidence of the positive (intended) and negative (unintended) effects of the device in the specific context of the individuals in which the device is to be used. In addition, the device should – ultimately – have benefits beyond prevailing care.

The health policy perspective

Today, with efforts being made to contain health care costs on the one hand and introduce new, often costly medical devices on the other, it is important that market access and reimbursement are provided for those medical devices that are (relatively) safe and – ultimately – lead to improvement of patients' or clients' health or of health care at large. The more invasive, burdensome and perhaps costly the devices are, the greater the need to provide evidence that the device has a positive risk/benefit ratio. In other words, there should be a balance between added benefit and extra burden, risks or monetary costs. From a monetary cost perspective, a less advanced, cheaper device that carries a similar risk and offers a similar performance or benefits may be acceptable as well.

In summary, the direct or indirect benefits of a medical device, the type of setting, user and his or her experience in handling a device, the regular adaptation of previous versions, and the short life cycle all go to create a dynamic process that increases the complexity of clinical research design.

2.6 Clinical evidence

The committee takes the view that regardless of the type of device, evidence of some form of benefit to be gained from using the device should be addressed. The committee has therefore introduced (in **Chapter 1.2**) the overarching principle that actually drives the evaluation and regulation of any health care intervention, including medical devices: *To generate and accumulate evidence that the use of a device is not only safe but also has benefits, preferably added benefits beyond existing care, for the health or health care of the intended individuals, patients, professionals or for society at large.* How much evidence, however, and what form it should take may differ according to

21 Hauber, A.B, et al. (2009). Older Americans' risk-benefit preferences for modifying the course of Alzheimer disease. *Alzheimer Dis Assoc Disord*, 23:23-32; Nigel K. Arden, N.K, et al. (2012). How do physicians weigh benefits and risks associated with treatments in patients with osteoarthritis in the United Kingdom? *J.Rheum.*, DOI:10.3899.

the device's risk classification and its life cycle phase: is it a breakthrough device that is ready for a first-in-man study, or is it the fifteenth version of a device that already has a CE marking and whose precursors are regularly being used in health care? Regardless of the phase, there are three crucial aspects to bear in mind when designing research into the benefit or added benefit of devices.

1. *Safety*: In essence, the need to collect evidence is more important for *devices associated with a higher risk* during device use. The EU concept of risk classification (see **Appendix III**) depends on the intended use of the device and distinguishes four classes of risks, ranging from low to high. A risk category is linked to specific requirements for manufacturers to demonstrate safety and performance. Safety according to the Directive (MDD 93/42/EEC) means that the clinical condition or the safety of patients, or the safety and health of users, or, where applicable, other persons, is not compromised. Furthermore, it means that any risks that may be associated with the devices intended use constitute acceptable risks when weighed against the benefits for the patient and are compatible with a high level of protection of health and safety.

2. The *performance* and thus benefit of medical devices is heavily influenced by the intended use and *the mode of action*. Medical devices produce mainly local and physical effects on the body. They may contain single or multiple active components, and both single (e.g. disposables) and multiple uses by one user (e.g. a blood glucose self-monitoring device) of the same device are possible. In fact, there are often multiple users of the same device (e.g. most imaging devices). Cessation of use can be simple for external devices and temporary implants, but it can be complex as well (e.g. for long-term implants).

3. *User interference*. As stated above, the safety and performance of a device depends not only on the device itself but also on how it is used. The user interface of a medical device is often not direct (device–patient), but in many cases involves an indirect or intermediary (device–operator–patient) situation. There is often a ‘learning curve’ associated with the use of medical devices, particularly in the case of complex high-tech devices, and a need for technical training and support. This may have a major impact on the benefits for the intended patients or individuals or their health outcomes. User interference becomes even more crucial when a medical device is applied in a wider population than it was originally evaluated in, emphasising the need for extensive postmarketing surveillance. In addition, the health benefits or outcomes are influenced by the fact that heterogeneous populations are tested. There is an inherent need to take both user interference and heterogeneous populations into account when designing research and judging the benefits of the device.

Added health benefit of a medical device

As we explain in **Chapter 4**, the term ‘added health benefit’ can be interpreted in a number of ways but it is commonly accepted to mean that the device demonstrates an improvement in a certain parameter of health compared to some alternative management or intervention, e.g. no intervention or the prevailing intervention. Added health benefit may also include non-inferiority to a known effective intervention but with less burdensome or more efficient care. In other cases, added benefit includes a clear improvement in the time required to detect disease, leading to earlier effective intervention, which in turn leads to an improvement in the patient’s symptoms or quality of life.²²

Prominent questions concerning the investigation of a device’s added benefit are:

- Which study results or evidence needs to be produced given the type of device and its intended use?
- Which study designs may be used, and what are their pros and cons?
- Which patients or clinical context and indication may need to be studied?
- Which types of health benefits or outcomes need to be studied?
- Which type of data analysis is recommended?
- Which study results or evidence needs to be produced?
- When and how should these studies be set up in view of the life cycle and the type of medical device (e.g. diagnostic, screening or other type)?

All these questions are discussed and illustrated in detail in **Chapter 4**. The basic concept of evaluation is similar to that of pharmaceuticals, but in practice medical devices have a larger number of variables (see **Appendix VI**).

2.7 The medical device life cycle and the stakeholders involved

We can divide the medical device life cycle into four discrete phases. Each of these phases has specific characteristics that have consequences for the way in which the added clinical benefit can or cannot be determined. Figure 2.2 presents an overview of the phases, the stakeholders involved and the key requirements.

Phase 1: From idea to product

Phase 2: Experimental use in practice

Phase 3: Uptake in regular care

Phase 4: Long-term implementation and reimbursement

The transition between the first and the second phase is the CE certification allowing market access (see Appendix III). The transition between the second and the third phase is the gradual dissemination of devices into regular care, imposing new issues

²² End points in cancer clinical trials and the drug approval process, Richard L. Schilsky. *Clin Cancer Res* April 2002 8; 935.

with respect to safety, performance and benefits. The essential difference is that phase 2 is research-based, whereas phase 3 is implementation-based. Long-term implementation of a device in regular care and its reimbursement often depend on the assessment of its cost-effectiveness.

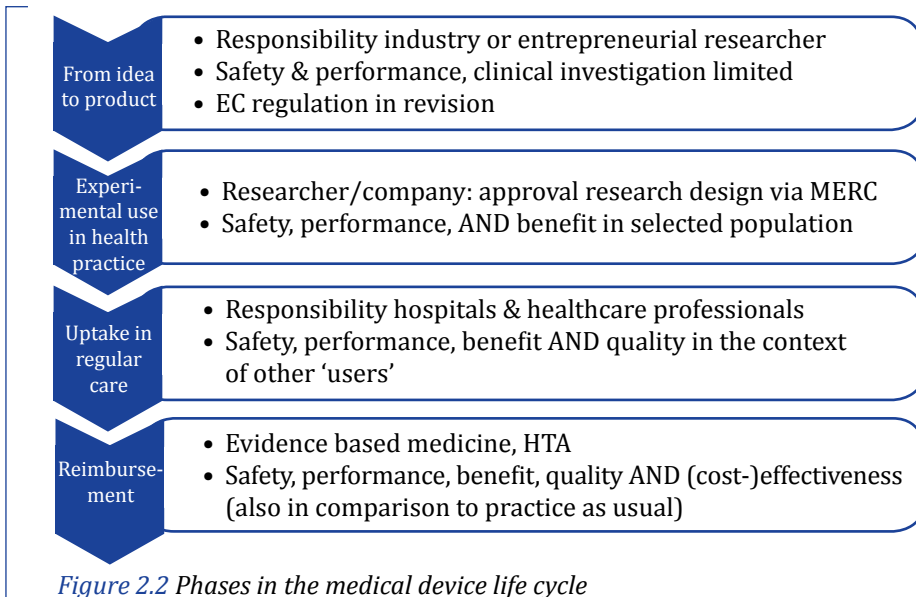


Figure 2.2 Phases in the medical device life cycle

Each of the phases involves different stakeholders and different needs. In phase 1, industry, researchers and, eventually, the Notified Bodies and Competent Authorities are involved. The medical ethical review committees (MERCs) and professionals may also be involved when investigating a device for its safety and performance in a new clinical investigation.

In phase 2, health care professionals, researchers and industry are directly involved; indirectly, the stakeholders are the medical ethical review committees, hospital instrument committees, and health research funders as representatives of intermediary organisations.

Phase 3 sees an increasing number of health care professionals, institutions and boards getting involved, such as health insurance boards and companies, hospital boards, health care professional organisations, and health research funding agencies. Industry continues to play a role in introducing a new device into regular practice.

In phase 4, the stakeholders include health care quality institutes, health insurance boards, health insurance companies, hospital boards, funding agencies as well as various other regulatory and advisory agencies, for example policy advisors who advise on developing user protocols and guidelines and advisory agencies for reimbursement decisions.

Regulatory approval is not sufficient for the successful implementation of a new device in health care. In the current era of evidence-based medicine, the benefits and

indeed health care-related economic considerations need to be taken into account. They should ideally be included from the early phase of a new device's development. More comprehensive health technology assessment approaches should cover all phases to consider the new device's total impact on health care.

3. STAKEHOLDER VIEWS ON MEDICAL DEVICE INNOVATION AND EVALUATION

In this chapter we consider the issues and challenges raised during our field survey by various stakeholders who are involved in developing, evaluating, introducing and using medical devices. The stakeholders can be roughly divided into four categories that are closely related: industry, health care professionals (e.g. medical doctors and nursing staff), regulators, and targeted patients and clients (users). Their perspectives were presented in **Chapter 2**. The field investigation took the form of interviews, invitations to participate in committee meetings, and feedback from the representatives of the various parties at a health technology conference. A list of the participants involved in the field investigation is provided in **Appendix VII**.

In this chapter, we begin (in **Section 3.1**) by discussing general points raised by the stakeholders concerning the evaluation of medical devices. These general points are summarised in the context of the device's life cycle. **Section 3.2** discusses more issues and challenges raised by the stakeholders with regard to establishing evidence of the performance and added clinical or other benefit of medical devices. The list of issues is not exhaustive.

Both general and specific issues and challenges must be taken into account when providing guidance for clinical investigations of medical devices. The field survey indicated that a one-size-fits-all approach for clinical investigations of medical devices is not feasible: the variety of medical devices available and their market sector require a flexible approach to clinical investigation.

What do we know about clinical research involving medical devices?

Given the increasingly stringent clinical requirements for medical devices, both in the EU (amendment of directives) and in the USA (FDA), we are witnessing a clear

tendency towards more transparency in clinical device research. The website of the US National Institutes of Health (NIH) provides a register of clinical studies (both closed and on-going), including clinical studies of devices, conducted worldwide (see **Box 3.1**).²³ The majority of clinical trials registered involve pharmaceutical products. In the Netherlands, clinical studies are registered in the Dutch trial register.²⁴ **Appendix IV** gives numerous examples from these registries to illustrate each clinical research approach (discussed in **Chapter 4**) that can be used in evaluating a device's benefits or added benefits and performance. In this chapter, we give some of the examples described in Appendix IV to show the limitations and potential of clinical research involving medical devices.

Box 3.1

ClinicalTrials.gov currently lists 143,826 studies with locations in 184 countries. Of these, around 10% (14,473 studies) involve medical devices (using the key word 'device'), and a roughly similar amount involves surgical procedures. Eighty per cent of the studies registered are interventional, whereas 20% are observational. The majority of these studies have been or are being performed in the US (6145) and Europe (4523).

The *Dutch Trial Register* lists 4,159 studies. Of these, 35 (<1%) involve a device (using the key word 'device'). Other key words relevant for medical devices are: imaging (in 108 studies); hip (76); monitoring (70); surgical (57); stent (54); diagnostic (47); implant (31); biomarker (23); prognostic (14); robot and in vitro (8 studies each).

3.1 General issues per phase of the life cycle of medical devices

This section summarises general issues and bottlenecks per phase of a device's life cycle (**Chapter 2.7**; see **Figure 2.2**) that were addressed by various stakeholders.

Phase 1 From idea to product

- There are regulatory guidelines and harmonised standards for assessing safety and performance and sufficient awareness that compliance with these is required for CE approval.²⁵ There is also a growing emphasis on clinical evaluations in all the regulatory guidelines, but no explicit guidance on designing, conducting and analysing clinical investigations for manufacturers, researchers, policymakers and health care professionals.

23 <http://clinicaltrials.gov/ct2/resources/trends>, accessed 18 April 2013.

24 <http://www.trialregister.nl/trialreg/index.asp>, accessed 24 January 2014.

25 See <http://ec.europa.eu/health/medical-devices/documents/guidelines/> for an overview of all guidance documents. See <http://ec.europa.eu/enterprise/policies/european-standards/harmonised-standards/medical-devices/> for harmonized standards for industry.

- Knowledge about useful, alternative designs or approaches to clinical research is relevant for all health care professionals and device researchers. This was also the reason behind a recent report on alternatives for randomised clinical trials in research into the effect of therapeutic interventions, which was commissioned by the Netherlands Organisation for Health Research and Development (ZonMw).²⁶ Chapter 4 and Appendix IV build on this overview.
- In many cases of CE certification, a clinical study is not necessary. Companies use lab or benchmark tests whenever possible by claiming that a device that has been altered incrementally does not aim to address a different context, users or patients than the original device. This is often considered sufficient by the Notified Bodies and other regulatory agencies, although the concept of equivalence (see 3.2 and Appendix III) is not always used appropriately, according to the Dutch Health Care Inspectorate (IGZ).²⁷
- The quality of the Notified Bodies in Europe differs from one country to another, and they do not focus specifically on clinical evaluations, hampering enforcement of the changing legislation.
- More awareness is needed of the challenge involved in studying new types of medical devices for which the existing clinical research paradigm is less suitable. Examples are: combination products, self-management products, AMTPs. For these types of medical devices, a 'network of evidence' approach may be more appropriate.

Phase 2 Experimental use in practice

- In the Netherlands, the local medical ethical review committee (MERC) and instrumentation departments of hospitals frequently request a CE certificate for the clinical study plan to guarantee the safety of the product. The majority of MERC requests for medical devices already have a CE certificate. This may mean that clinical investigations have been carried out outside the Netherlands as part of the CE certification process, or that CE certification focuses primarily on performance rather than on the clinical benefits of device use.
- Clinical research (after CE certification) on the benefits of a device meant to encourage its use and uptake suffers from inadequate knowledge of the possible research approaches, each with its pros and cons.
- Having a large number of incremental changes/line extensions/new versions of an existing device results in many small datasets, each of them inadequate as statistical evidence.
- Clinical studies of devices by health care professionals often take place in the

²⁶ *Alternatieven voor Randomized Clinical Trials in onderzoek naar de effectiviteit van interventies.* M. Langendam, L. Hooft, R. Scholten, P. Bossuyt. Dutch Cochrane Centre (DCC), Academisch Medisch Centrum, Amsterdam, February 2013.

²⁷ *Inspectie voor de gezondheidszorg, May 2013. Metaal-op-metaal-heupimplantaten. De keten voor de kwaliteitsborging van medische hulpmiddelen moet beter functioneren.*

context of scientific research that is sponsored (at least in part) by industry. In these situations, the health professionals usually receive extensive training in how to handle a device safely, and the conditions under which the intervention is used are standardised as well. For devices in cardiology (stents and pacemakers: both class III devices, see **Appendix III**), rehabilitation care and orthopaedics, clinical research is fairly large-scale. The main channel of communication for this type of clinical research is through scientific publications.

Phase 3 Uptake in regular care

- There are no systematic registries for postmarketing surveillance and long-term benefit assessment of devices.
- Uptake of devices in regular care is hampered by a new set of safety issues related to their use by untrained professionals. Inexperience in handling a device creates another type of risk and may lower the expected benefits, negatively influencing the benefit/risk ratio of a device. User interference and learning curve aspects are particularly relevant in the case of inexperienced use in regular care. To tackle this issue, the Dutch Order of Medical Specialists (OMS) is collaborating with the Dutch Quality Institute and the Health Insurance Board on a guideline for the safe introduction of new medical devices in regular health care (see also **Chapter 5.1**). We therefore do not address these risks in detail in this report.

Phase 4 Long-term implementation and reimbursement

- Lack of transparency and openness about regulatory practices between countries.
- Clinical benefit is usually discussed at the point where reimbursement (and thus cost-effectiveness) is being considered, and not when the device is being introduced.
- In response to the rising cost of health care and society's need for devices that are not only safe but also beneficial (especially when they are expensive and/or high-risk), calls are growing for a framework of research approaches to assess the benefits and added benefits of devices and for health technology assessment (HTA) models.

3.2 Issues related to the clinical evaluation of devices

Table 3.1 lists the issues or challenges mentioned by the various stakeholders, related to the clinical evaluation of medical devices. Setting up a traditional randomised clinical trial (see **Appendix IV**) of a medical device is considered complicated because, e.g.:

- for many devices (e.g. prostheses, robots and implants), it is not possible to perform double-blind randomised studies;

- the benefits often vary considerably due to user interference and expertise;
- there are often multiple effect and safety outcomes to be defined;
- multiple outcomes are not of equal significance.

Accordingly, it is often hard to formally assess the risk/benefit ratio of a device using the traditional intervention research methodology used for pharmaceutical products. Below, we describe the issues listed in Table 3.1 in more detail without necessarily providing a solution for each issue.

1. No specific medical indication

The intended application of a medical device is often broad and not limited to a specific medical indication. A good example is imaging devices (CT, PET, SPECT, or MRI scanners) that are used for diagnosis and monitoring of a wide range of medical conditions, e.g. in cancer and cardiology. However, when an improved version of the device shows more detail, the consequences for patient management and thus for patients' health are not always straightforward. Is it always beneficial to see more, or does it lead to over-diagnosis and overtreatment?

Table 3.1 Issues and challenges in the clinical evaluation of medical devices, addressed by the stakeholders and illustrated by examples of medical devices.

Issue or challenge	Medical device example	Clinical research example
1. No specific medical indication: general purpose devices	Imaging devices – phantom images	Postmarketing outcome study (Appendix IV C1.1)
2. User interference; long-term adverse effects	Hip implants	Device-based registries (see Appendix IV B2.2.1)
3. Co-development diagnostic marker and pharmaceutical	Companion diagnostics	Herceptin and many more to follow
4. Equivalence and incremental improvement	Therapeutic (surgical) support device	Case series, controlled before and after (Appendix IV B1.2)
5. Combination products: Medical device and pharmaceutical in one product	Inhaler devices for asthma and COPD	Adaptive trial design (Appendix IV A4)
6. Self-management products and online monitoring	Mobile phone health applications	Guidance for Mobile Medical Applications (FDA)
7. Expensive devices and cost-effectiveness	da Vinci robots	Large-scale follow-up studies
8. The balance between regulation and speed of introduction	Stem cells (Advanced Medicinal Therapeutic Products)	Conditional market approval in combination with pivotal trials
9. Health Technology Assessment (HTA)	HTA International Policy Forum and Dutch Innovative Medical Devices Initiative (IMDI)	NICE (UK)

On the one hand, it may not be sensible to require new clinical investigations for each new version or small adaptation of an existing imaging device. A small increase or adaptation of such a device may only need a few calibration or recalibration studies and tests by imaging phantoms, or a comparison with computer-generated results

followed by subsequent inter-observer and intra-observer variation studies.²⁸ On the other hand, when the device improvements are so major that it seems necessary to carry out a clinical investigation in order to demonstrate that the use of a new imaging device is beneficial for patient management and ideally for patients' health, the question is what type of research will show these benefits. Clearly, from a health perspective, it is relevant to show that the benefits of using the new imaging device are at least equal to the benefits of existing care.

Example C1.1 in Appendix 4 shows a single device study to assess the benefit (defined here in terms of percentage of accurate diagnosis) of PET-FDG in determining regional nodal spread of breast cancer (ClinicalTrials.gov/ NCT00201942) by comparing it to the more burdensome and cumbersome reference standard involving the histologic examination of all excised (sentinel and non-sentinel) axillary lymph nodes.

2. The role of user interference and long-term adverse effects

When the user of a medical device is a professional, user interference may influence the outcome of device use. The more experienced and trained the professional, the less influence the user will have on the potential benefits and risks of device use. One example of the influence of user interference on device benefits and risks is total hip replacement surgery. Hip-replacement devices are classified in EU risk category IIB (see **Appendix III**), which means that they do not need to be formally evaluated in patients before entering the market. However, long-term follow-up studies recently revealed the metal-to-metal hip implants wear out more quickly and caused adverse effects when not precisely positioned.²⁹ Adverse effects were also reported for specific metal-to-metal hip devices that were not user-related, and this device was eventually recalled.³⁰ Interestingly, all this could not have been shown by any clinical investigation, which would generally lack a long-term follow up of ten to twenty years. This illustrates that the influence of users on the benefits and risks of device use need to be carefully addressed in the clinical evaluation of devices, pre-market and post-market.

Discussions of the adverse effects of total hip replacement overshadow the fact that this intervention has also been called one of the most successful operations of the 20th century and that it is currently performed on one million patients each year worldwide (see footnote 29). In the Netherlands, the Health Care Inspectorate has therefore investigated how to improve the quality chain of medical devices in general and those

28 Das IJ, Cao M, Cheng CW, Mistic V, Scheuring K, Schüle E, Johnstone PA. A quality assurance phantom for electronic portal imaging devices. *J Appl Clin Med Phys*. 2011 Feb 2;12(2):3350.

29 A series of papers in *BMJ* 21 May 2011, vol. 342:1115-1130. In this issue of *BMJ*, serious concerns were raised about the regulation of high-risk medical devices and how well they are tested before they come to market, taking the example of metal-to-metal hip implants, while at the same time mentioning the success of these interventions.

30 Cohen, D. Out of joint. *BMJ* (2011) 342: 1116-1122; d3009. In August 2010, DePuy Orthopaedics issued a voluntary recall of its ASR™ XL Acetabular Hip System and DePuy ASR™ Hip Resurfacing System.

of metal-to-metal hip replacements in particular.³¹ Internationally, the International Consortium of Orthopaedic Registries (ICOR) is a US FDA-sponsored initiative consisting of a public-private partnership, with over 30 registries participating worldwide since 2011.³² The aim is to facilitate and enhance inter-registry collaboration on total hip replacements by providing a supportive infrastructure and the development of a distributed data network that uses innovative approaches to analyse the available postmarketing surveillance data.

3. Companion diagnostics

Companion diagnostic devices constitute a specific category of diagnostic devices. A companion diagnostic device can be an *in vitro* diagnostic device that provides information essential for the safe and effective use of a corresponding therapeutic product. Such devices are aimed at aiding the development and use of disease-specific pharmaceuticals. Biomarkers, molecular imaging, microdosing and pharmacogenetics are companion diagnostics approaches. They help optimise therapy and drug therapy options and early detection and treatment of disease and disease prevention. They also may assist in studying specific targets for drug delivery, in reducing the time and costs involved in drug development, and in eventually enhancing the intended effects and decreasing the unintended effects of a drug. Incorporating new diagnostic technology helps us understand disease processes and as well as patient diversity and predisposition, thereby offering an opportunity to change the current drug discovery and development paradigm. One example of a medicinal product that uses a companion diagnostic is Herceptin (HER2 breast cancer).

There has been a boom in new companion diagnostic devices, as shown by the 7% growth (between 2005 and 2010) in the global *in vitro* diagnostics market and the 40% increase in funding (in 2010 compared to 2009) raised for diagnostic companies.³³ Companion diagnostics devices' access to the market is also regulated by the medical device directives (IVDD or MDD). Their development is often combined with that of a therapeutic drug, organised directly by pharmaceutical companies.

4. Equivalence and incremental improvement

Equivalence refers to a new device (or parts of a new device) that either continues the development of an existing version or is similar in type to an existing device. Equivalence claims in the EU are made explicit in guidance documents.³⁴ There are nine criteria for assessing equivalence: four technical, four clinical and one biological (**Appendix III**). If there is information available for six of the nine criteria, for

31 Metaal op metal heupimplantaten: De keten voor de kwaliteitsborging van medische hulpmiddelen moet beter functioneren. Inspectie voor de gezondheidszorg, Utrecht May 2013 (in Dutch only).

32 <http://www.icor-initiative.org/>.

33 Personalized Medicine and Companion Diagnostics, DutchCC workshop, 15 November 2011.

34 MEDDEV 2.7.1, see also Annex 3.

example, then new clinical investigations can be requested focusing on the missing three criteria. In the case of equivalence being claimed for a new device, the question is how to obtain sufficient evidence of the benefits of the new device for health and health care. Therapeutic support devices used by professionals in a hospital setting, such as staplers for closing an incision or new catheters, can enter the market with a CE mark without a new clinical investigation. The evaluation can be conducted using existing data and comparison of the above nine criteria with equivalent devices or a previous version of the same device. Once on the market, these devices are sometimes tested in small studies that often are extended over time. An example of such a study is given in Appendix 4 (B1.2): a before-after study shows how to investigate the benefits of pancreatic stump closure using a reinforced staple line. The study is a second series of four years of experience using this technique in a larger sample, following the first small group.³⁵ The technique is straightforward and results in reduced morbidity and cost. In addition to such studies, the use of registry data on longer-term follow-up or linking datasets of different versions of the device are other possibilities for evaluating its benefits and risks.

5. Combination products

Combination products consist of a medical device and a pharmaceutical joined into one product, e.g. inhaler devices for asthma and COPD. The primary mode of action (or function) of the product dictates how its market access is regulated, so a combination product can never be both a medicinal product and possess CE marking as a device. Appendix IV (A4) describes an example of an adaptive trial in healthy volunteers intended to optimise the rotacap formulation and ROTHALER device for delivery of Fluticasone Propionate/Salmeterol in COPD patients (ClinicalTrials.gov/NCT01540708). A key element of research on 'adaptive' features is that changes in design or analyses are guided by an examination of the accumulated data at interim points in the trial. Adaptive designs, time series, and linked evidence or synthesis of data are more flexible research strategies than the conventional comparative randomised trials, which may be particularly useful for the evaluation of medical devices.³⁶

6. Self-management products and online monitoring

The future of health care delivery is connected to decentralised health care solutions, e.g. the continuous monitoring of vital signs (e.g. HbA1c monitoring in patients with

35 Wallace CL, Georgakis GV, Eisenberg DP, Macaulay WP, Jimenez RE. *Conn Med*. 2013 Apr;77(4):205-10.

36 Bluhm R. From hierarchy to network: a richer view of evidence for evidence based medicine. *Perspective Biol Med* 2005;vol 8:535-547; Walach H., Falkenberg T, Fonnebo V, Lewith G., Jonas W.B. Circular instead of hierarchical: methodological principles for the evaluation of complex interventions. *BMC Med Res Methodol* 2006;vol 6:6-29; Van de Wetering G., Steuten L.M.G., Birgelen von C., Adang E.M.M., IJzerman M.J. Early Bayesian modeling of a potassium lab-on-a-chip for monitoring of heart failure patients at increased risk of hyperkalaemia. *Technological Forecasting and Social Change*. 2012 Sep;79(7):1268-79.

diabetes) outside the hospital and the possibility of supervising the functioning of devices remotely. The widespread adoption and use of mobile technologies is also opening the door to new and innovative ways to improve health and health care delivery. Increasingly, these innovations are becoming available for home care and for patient self-management. They represent a whole new type of health care intervention, where the benefit is related to user and patient expectations. However, such technology also challenges us to define or redefine the risks and benefits associated with their use. It involves taking a different approach to introduction, supervision and training. It remains unclear what type of clinical investigation is suitable for such technology. The FDA recently published a guidance for Mobile Medical Applications (September 2013).³⁷ According to industry estimates, 500 million smartphone users worldwide will be using a health care application by 2015, and by 2018, 50% of the more than 3.4 billion smartphone and tablet users will have downloaded mobile health applications. These users include health care professionals, consumers, and patients.³⁸

7. Expensive devices and cost-effectiveness

There should be a balance between added benefit and extra costs; from a cost-effectiveness perspective, a less advanced but cheaper device that is as accurate as or equivalent to the existing care or device can be acceptable. However, new devices are often relatively expensive and are used for conditions for which they have not necessarily been shown to be beneficial. One example is the da Vinci Surgical System, a robotic surgical system for complex surgery using a minimally invasive approach, controlled by a surgeon from a console.³⁹ Da Vinci robots operate in several thousands of hospitals worldwide (more than 2,585 da Vinci Systems have been installed in over 2,025 hospitals worldwide at a cost of around USD 2 million).⁴⁰ The FDA cleared the da Vinci Surgical System for adult and paediatric use in several laparoscopic surgical procedures (e.g. urologic and gynaecologic). Nevertheless, there is criticism that the system is difficult for users to learn and that it has not been shown to be more effective than traditional laparoscopic surgery. A recent study evaluated outcomes in more than a quarter of a million people comparing either laparoscopic or robotically assisted hysterectomies in 441 hospitals between 2007 and 2010. Both methods are minimally invasive and involve smaller incisions than open abdominal surgery. No overall

37 <http://www.fda.gov/medicaldevices/productsandmedicalprocedures/connectedhealth/mobilemedicalapplications/default.htm>.

38 <http://www.research2guidance.com/500m-people-will-be-using-healthcare-mobile-applications-in-2015/>.

39 The Slow Rise of the Robot Surgeon. *MIT Technology Review*. 24 March 2010; http://www.intuitivesurgical.com/products/davinci_surgical_system/.

40 http://www.intuitivesurgical.com/products/products_faq.html#sthash.f1YfH8q2.dpuf.

difference in complication rates was seen between the two groups.⁴¹ There was a big difference in costs, however. Robotically assisted hysterectomies cost on average about a third more than laparoscopic surgery.

8. The balance between regulation and speed of introduction

Stricter regulation of clinical investigation poses a dilemma: on the one hand, there are considerations of safety and of the benefits for the targeted patients or users; on the other hand, there is the potential delay in introducing better technologies. More generally, asking for more extensive evidence of the benefits, let alone added benefits, of medical devices may increase development costs. This has been recognised by the European Council. A Council meeting document explicitly states: *'An efficient system of market access for new products that **satisfy the highest possible safety standards** is crucial in satisfying the needs of patients today and tomorrow. There is certainly a scope for improving the current system. Interventions should directly address weaknesses in the system, **without creating additional barriers for innovation and availability for patients**. Serious incidents underscore the need for post-market follow-up in evaluating the safety of medical devices as they are used in the field.'*^{42 43}

Awareness of such barriers can be seen in the field of cell therapy (subject to the Advanced Medicinal Therapeutic Products regulation of pharmaceuticals in Europe⁴⁴). While there have been no mesenchymal stem cell (MSC)-based medicine authorisations in the EU, three MSC products have received marketing approval in other regions since 2011. South Korea is leading with two MSC products registered and the first authorisation granted in 2011.⁴⁵ The authorisation seems to be linked to a conditional market approval procedure within South Korea's regulatory framework that allows

41 Jason D. Wright, MD; Cande V. Ananth, PhD, MPH; Sharyn N. Lewin, MD; William M. Burke, MD; Yu-Shiang Lu, MS; Alfred I. Neugut, MD, PhD; Thomas J. Herzog, MD; Dawn L. Hershman, MD. Robotically Assisted vs Laparoscopic Hysterectomy Among Women With Benign Gynecologic Disease. *JAMA*. 2013;309(7):689-698. doi:10.1001/jama.2013.186.

42 Council of the European Union, meeting document DS1360/1/13, 22 May 2013 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) No178/2002 and Regulation (EC no 1223/2-009).

43 EU Council conclusions on innovation in the medical device sector; 2011/C 202/03

44 http://ec.europa.eu/health/human-use/advanced-therapies/index_en.htm; Gene therapy, somatic cell therapy and tissue engineering are Advanced Therapy Medicinal Products. They are included in the medicinal product regulation (Directive 2001/83/EC) and further in Regulation (EC) No 1394/2007.

45 Janis Ancans. Cell therapy medicinal product regulatory framework in Europe and its application for MSC-based therapy development. *Front Immunol*. 2012; 3: 253. Published online 2012 August 14. doi: 10.3389/fimmu.2012.00253; PMID: PMC3418507.

commercial sales in certain instances while pivotal trials are still underway.⁴⁶ A similar regulatory decision has been adopted in Canada.⁴⁷ The results of clinical studies of MSC-based therapy clinical trials in the past ten years have led to the conclusion that MSC applications are safe and feasible. However, the benefits of the device itself often could not be convincingly demonstrated in single studies, as the therapies also advanced along with the development of the MSC products.⁴⁸ This is also illustrated by the absence of MSC-based products in the European market.

Adaptive licensing, e.g. conditional approvals, could be introduced based on incremental learning in circumstances of acknowledged uncertainty, with iterative phases of post-market data gathering and regulatory re-evaluations.⁴⁹ Adaptive licensing requires a different approach from the standardised dichotomous unapproved/approved CE paradigm, however. Initiatives are under way at the European Medicines Agency aimed at facilitating MSC-based medicinal product development and authorisation in the EU. It is important for the European Commission to monitor the trend towards (omics-based) personalised treatments closely and to reflect on how it can support this development by regulatory and non-regulatory means.

9. Health technology assessment (HTA)

With regard to HTA and reimbursement discussions, the Health Technology Assessment International (HTAi) Policy Forum supports and promotes ongoing discussions and scientific advances in this field. The most recent deliberations (2013) deal with HTA and value, but definitions of value vary depending on the stakeholder view (patient, societal, health system, industry).⁵⁰ However, all participating HTA experts agree that patient health and benefits are central, and that patient health improvement

46 Cartistem has become the world's first allogenic, off-the-shelf MSC-based product (umbilical cord blood (UCB)-derived MSCs for the treatment of traumatic and degenerative osteoarthritis). In 2011 the Korean company FCB PharmiCell received Korean FDA approval for commercial sale of HeartiCellgram (autologous bone marrow-derived MSC) indicated for post-acute myocardial infarction treatment.

47 The Osiris Therapeutics Inc. product Prochymal consists of allogenic MSCs. The company was granted an authorisation for the treatment of acute graft-vs-host disease (GvHD) in children under Health Canada's Notice of Compliance with conditions (NOC/c) in May 2012.

48 Wang S., Qu X., Zhao R. C. (2012) Clinical applications of mesenchymal stem cells. *J. Hematol. Oncol.* 5: 19.

49 Eichler H. G., Oye K., Baird L. G., Abadie E., Brown J., Drum C. L., Ferguson J., Garner S., Honig P., Hukkelhoven M., Lim J. C., Lim R., Lumpkin M. M., Neil G., O'Rourke B., Pezalla E., Shoda D., Seyfert-Margolis V., Sigal E. V., Sobotka J., Tan D., Unger T. F., Hirsch G. (2012). Adaptive licensing: taking the next step in the evolution of drug approval. *Clin. Pharmacol. Ther.* 91 426–437.

50 Chris Henshall and Tara Schuller, on behalf of the HTAi Policy Forum. HEALTH TECHNOLOGY ASSESSMENT, VALUE-BASED DECISION MAKING, AND INNOVATION. *International Journal of Technology Assessment in Health Care*, 29:3 (2013), Cambridge University Press 2013 doi:10.1017/S0266462313000378.

Background Paper: *HTA and Value: Assessing value, making value-based decisions, and sustaining innovation*. HTAi Policy Forum Meeting 2013, 3-5 February 2013, Barcelona, Spain.

is considered a core element of value. Wider elements of value include other benefits for patients (e.g. less burdensome, improved time to diagnosis, early detection), benefits for caregivers, and benefits for health care and society in general. Although value is often defined in economic terms, there are divergent views about whether or not value is proportional to cost, i.e. cost-effectiveness thresholds. Value measurements (both core and wider) can be quantitative and qualitative. There are different approaches that health systems take to value-based decision-making.

The Dutch IMDI (Innovative Medical Devices Initiative, part of the Dutch Life Sciences & Health key economic sector) offers another way to approach the HTA of medical devices. IMDI is investigating whether it is possible to establish potential clinical use and added benefit in a very early stage of device development. It has defined added value along three lines, which takes a system approach instead of an individual patient approach:

- Added value in health care/public health, which is further defined as health gains, quality of life, safety, self-reliance, and effect on the number of people working in health care.
- Added value in industry/economy. The Ministry of Health in the Netherlands does not want to limit entrepreneurship in the health care industry by imposing stricter controls. Its view is that innovation in health care would benefit from social innovation (such as self-management, informal care, community welfare services) or innovation in the organisation of health care, and not so much from regulation.
- Added value in research.

In the UK, the National Institute for Health and Care Excellence (NICE) provides guidance to the government on the clinical benefits and cost-effectiveness of selected new and established technologies. The Institute provided a guidance document with an overview of the principles and methods of health technology assessment, describing key principles of appraisal methodology.⁵¹ However, despite continuous development of technology appraisal methodology, there remain many areas of controversy and uncertainty. See also **Chapter 5**.

Summary

Based on this field survey, it is the opinion of the committee that, in every phase of a medical device's life cycle, it would be of major value to provide more explicit guidance on how to perform studies to assess the benefit or added benefit of the device and which type of studies are possible and needed given the type of device and the intended context and users.

⁵¹ Guide to the methods of technology appraisal 2013, National Institute for Health and Care Excellence (NICE). <http://publications.nice.org.uk/guide-to-the-methods-of-technology-appraisal-2013-pmg9>.

Chapter 4 therefore offers an overview of research approaches for collecting evidence that is not limited to clinical investigations needed for CE marking, but also include guidance on conducting clinical research with CE-certified devices. Chapter 4 aims to provide detailed guidance on investigating the different clinical benefits of medical device use while acknowledging the specific nature and market of medical devices. Improved guidance in assessing the clinical benefit or added benefit of medical devices will lead to more clinical and statistically relevant data that can be used to inform CE-marking, uptake and reimbursement discussions.

4. CLINICAL RESEARCH ON MEDICAL DEVICES: PRINCIPLES AND APPROACHES

A variety of different approaches are available for evaluating the safety, technical/analytical performance, and clinical benefits/utility/effectiveness (including clinical performance) of a medical device (**see also Box 2.1**).^{52,53,54} The field survey (**Chapter 3**) clearly showed that studies meant to substantiate the clinical benefits or effectiveness – for example for a patient or user – of a device still raise questions and challenges among device manufactures, researchers, care providers, Notified Bodies and policymakers. The field survey also indicated that there generally is a good understanding of the types of studies required to demonstrate the safety of medical devices as well as their technical and analytical performance.

This chapter therefore does not address approaches for evaluating a device's safety or technical/analytical performance; instead, it provides explicit guidance as to which study approach allows for inferences about the clinical benefits/utility/effectiveness (or added benefits) – including clinical performance – of a device and to what extent given the device specifics and indications. The purpose is to enhance our understanding of the evidence on a device, help the parties involved plan future studies (if needed), and improve the dissemination, uptake and application of safe and beneficial devices. We discuss the essentials of numerous research approaches, and how these can inform us about the likelihood that a device will indeed provide the intended benefits for health or health care when introduced in regular care. We show that, unlike pharmaceuticals, there is **no one-size-fits-all** approach in the realm of medical device evaluations, simply because of the wide variety of medical devices available.

52 Clinical Evidence – Key Definitions and Concepts Study Group 5 Final Document SG5/N1R8, May 2007. Retrieval from: <http://www.imdrf.org/docs/ghf/final/sg5/technical-docs/ghf-sg5-n1r8-clinical-evaluation-key-definitions-070501.pdf>.

53 <http://www.cdc.gov/genomics/gtesting/ACCE/>.

54 Smart A. A multi-dimensional model of clinical utility. *Int J Qual Health Care*, 2006; 18 (5): 377-382. doi: 10.1093/intqhc/mzl034.

Outline of this chapter

The chapter begins by describing its aim and scope (**Section 4.1**), followed by an explanation of the general principle for assessing the performance and benefits or added benefits of medical devices (**Section 4.2**). In **Section 4.3** we discuss the overall distinction between therapeutic and non-therapeutic (i.e. diagnostic, screening, monitoring or prognostic) devices or test devices. Therapeutic devices commonly intervene with bodily systems and thus may directly benefit the health of targeted users or patients. Diagnostic, screening, monitoring or prognostic tests yield information that in turn can lead to better, earlier, or enhanced interventions. Hence, such devices have different benefits and improve the health of targeted users or individuals indirectly. We continue by outlining the optimal research approach for assessing the benefits or added benefits of devices, and why this is the case (**Section 4.4**). In **Section 4.5** we discuss why and when deviations from this optimal research approach may be indicated. We also stress how helpful it is to have a mental picture of the theoretical optimal research approach when designing or choosing alternative approaches, with a view to the validity of such research. In **Section 4.6** we describe why it is important to define, as early as possible, the pathway or pathways through which the device will lead to intended and unintended effects on health or health care. We also provide guidance as to which elements should be included in such pathway descriptions, and explain that this is a collaborative effort among device developers, manufacturers and end-users (e.g. medical professionals and targeted individuals or patients). **Section 4.7** then shows the absence of a one-size-fits-all approach. It describes the three main study approaches to generating evidence about the benefits or added benefits of device use for the intended individuals, professionals or society at large. These include 1) studies providing direct evidence and quantification of the benefits/added benefits of device use on health or health care (**Section 4.7.1**); 2) studies providing indirect evidence of the benefits/added benefits of device use on health or health care using a quantitative linked-evidence approach (**Section 4.7.2**); 3) studies providing indirect evidence of the benefits/added benefits of device use on health or health care using a qualitative linked-evidence approach (**Section 4.7.3**). For each approach, we discuss which types of devices can be covered and in which situations, and how each approach has its merits and vulnerability to specific forms of bias. Explicit study designs and their pros and cons are given for each approach and illustrated with real-life empirical examples from a wide range of devices (**in Appendix IV**). Finally, **Section 4.8** provides an *overall summary* of the chapter with a list of key issues that should be considered by researchers studying the clinical benefits, including clinical performance, of devices. **Appendix V** addresses overarching statistical recommendations.

This overview and guidance are not based on a formal, systematic literature review, but rather survey and summarise the existing methodological literature on assessing the performance and benefits/effectiveness of medical devices, supplemented by the

expert knowledge of the committee members and interviewees. In relation to this, our guidance on choosing research approaches (given the type, context and intended use of a device) is not set in stone. It should be seen as a recommendation and option for all stakeholders involved in assessing medical devices; it is the start rather than the end of a process and a discussion among the stakeholders concerning research methods for medical devices.

4.1 Aim and scope

This chapter and accompanying Appendices IV and V discuss and illustrate the need, design, execution and interpretation of available approaches to investigating a device's clinical benefits/added benefits or effectiveness, in view of the wide variety of medical devices available. We illustrate that such variety means there can be no *one-size-fits-all* approach. Knowledge of the pros and cons of these different approaches will improve the quality and understanding of device evaluations. In the current era of evidence-based medicine and accountability, this will increase the uptake and use of devices by end-users (patients and care providers), and ultimately result in better health or health care for those for whom the device is intended.

This chapter does not aim to:

- change any government or legal rules or obligations for market access or uptake in health insurance systems;
- prescribe at which point in a device's life cycle – e.g. before or after market access – which studies must be performed;
- prescribe which study approaches should specifically be applied for IVDD, AIMDD, or MDD devices, or for which risk class. For this we refer to the existing international and national guidelines;
- describe methods for evaluating the cost-effectiveness of devices. Some stakeholders may be specifically interested in methods for such evaluations or in which cost-effectiveness thresholds should be used for reimbursement decisions. Such methods are rare, however, and are still under development even for drug evaluations, let alone for device evaluations. Moreover, cost-effectiveness-threshold decisions require a political rather than a methodological debate. We briefly address the challenges of cost-effectiveness device evaluations in **Chapter 5**.

4.2 Overarching principle for device evaluations

Medical devices vary greatly in complexity and application and thus in their potential unintended (safety) and intended (benefits) effectiveness. Accordingly and perhaps arguably, unlike medical drugs, there is *no unique* optimal study approach to investigating the risks and benefits of a device and its use. Compared to pharmaceuticals, it is much more difficult to describe concisely which research methods are applicable

in which instance. Many factors play a role in choosing the ‘optimal study approach’, including:

- phase of the device’s life cycle;
- type of device;
- working mechanism through which a device leads to risks, benefits or less burdensome care;
- intended medical context or indication;
- intended users;
- prevailing care in the intended context.

Nevertheless, as stated in **Chapter 1 (and 2)**, the **overarching principle** that drives the evaluation and regulation of medical drugs, devices or any other health care intervention, is the same:

To generate and accumulate evidence that the use of an intervention is not only safe but also has benefits, preferably added benefits beyond existing care, for the health or health care of the intended individuals, patients, professionals or for society at large.

As we stated earlier, this principle is the starting point of this report, the aim being to ensure as much as possible the introduction and use of medical devices that are ‘safe’ and/or ‘necessary’.

4.3 Therapeutic devices versus non-therapeutic (diagnostic, screening, monitoring test) devices

The above principle applies to all types of medical devices, but we can distinguish two main categories: therapeutic devices and non-therapeutic devices, including diagnostic, monitoring, screening or prognostic tests. The benefits of both types of devices and how these benefits are achieved are markedly different. As we show in **Sections 4.5-4.7**, this is in part why there is *no one-size-fits-all approach* to evaluating the benefits and risks of medical devices.

Therapeutic devices are devices that interfere *directly* with – often targeted – bodily systems and mechanisms. Examples are pacemakers, nerve stimulators, prostheses, breast implants and medical robots. Therapeutic devices are beneficial because they treat specific diseases (e.g. heart rhythm disorders), alleviate specific symptoms or complaints (e.g. pain or tremors), or improve daily activities or quality of life *directly*. Due to their direct interference and thus effects on targeted patients, evidence of the benefits or added benefits of therapeutic devices is much easier to obtain directly from a single, properly designed, conducted and analysed clinical study, using a variety of study approaches. This may include randomised and non-randomised study approaches. This will be explained in detail in **Sections 4.5-4.7**.

Diagnostic, prognostic, monitoring or screening devices are devices or rather **tests** that do not by themselves treat or alleviate diseases, symptoms or signs directly, but rather do so *indirectly*. Examples of such test or information-generating devices

include imaging tests, companion diagnostics, laboratory tests, point-of-care or bedside tests. Diagnostic, prognostic, monitoring or screening tests provide information to the users, i.e. professionals, patients or other individuals, which in turn indicates subsequent actions or interventions, such as therapies or lifestyle changes. These interventions or actions may subsequently lead to benefits in terms of improved health outcomes of individuals or patients. Diagnostic, prognostic, monitoring or screening tests are information-generating devices rather than therapeutic devices, and therefore benefits the health of the targeted patients or individuals indirectly.

Test devices may also be beneficial by improving the therapeutic actions of medical drugs; examples include companion diagnostics, molecular imaging devices or imaging devices to guide surgery. A new diagnostic test may be beneficial because it means less invasive detection of disorders, avoiding additional unnecessary testing that is more invasive or costly. A new screening or monitoring test that leads to early detection makes it possible to administer the appropriate treatment at an early stage.

We note that test devices can still be invasive to varying extents (e.g. angiographic procedures or imaging techniques requiring radioactive contrast agents), but they are often minimally invasive (point-of-care tests requiring blood) or even non-invasive (ultrasound techniques). Safety aspects related to these devices are usually related directly to the extent of their invasiveness.

Finally, unlike therapeutic devices, many of the diagnostic, prognostic, monitoring or screening devices have a general purpose (such as imaging techniques, blood analysers, point-of-care tests). They are often not intended for one specific medical condition or indication.

All of the above has implications for the type of evidence and study approaches that can be used to infer the benefits/added benefits of diagnostic, screening, monitoring or prognostic tests for the targeted individuals/patients. This evidence may still come from a single clinical (randomised or non-randomised) study. But these devices may also profit from *network or linked-evidence approaches*. This will be explained in detail in **Sections 4.5-4.7**.

4.4 Optimal clinical study approach

Given the common goal of medical intervention evaluation, i.e. to accumulate evidence that the intended use of a device (therapeutic or non-therapeutic) is safe and has positive effects on health or health care beyond what is achieved by current practice, it should be possible to design a single study that measures all these aspects directly in the most valid and informative way. This is in fact the same optimal study approach used to evaluate the effectiveness of medical drugs. The study should:

investigate the device and its use similarly to how it will be used in practice;

- investigate the device in the same clinical context and with the indications as in practice;
- use the device by/for the same targeted individuals (e.g. professionals and

- patients) as in practice;
- compare the device with the prevailing alternative care. Ideally, this would be a randomised comparison: the targeted individuals would be allocated randomly to either the device use or a comparative care strategy;
 - in case of diagnostic, monitoring, screening or prognostic test devices, study their use in combination with any therapeutic management actions indicated by the test/test information;
 - measures all outcomes/endpoints relevant for the targeted individuals, health care professionals, and ideally for society at large. This would include, for example, any unintended outcomes, intended health outcomes (for patient and users), burden and ease of device use, speed of administering therapeutic decisions (for non-therapeutic devices), and even costs of the device use;
 - be sufficiently long to investigate the long-term health care effects of device use;
 - be designed and executed such that it produces valid results, i.e. minimises the risk of bias;
 - be sufficiently large to obtain precise estimates of the safety and health benefits of device use.

Applying all these main criteria would lead to the design of a large-scale, long-term **pragmatic or comparative effectiveness randomised trial**. For more details on this research approach, see **Appendix IV, part A**.⁵⁵ The two comparison groups are created randomly or by chance splitting. In the index group, the new device is used (in combination with subsequent therapeutic actions in the case of diagnostic, screening, monitoring or prognostic test devices), with the prevailing management being applied in the comparison (control) group. Provided they are large enough, then, the two groups are ‘the same’ except for the device under study. Any observed differences between the two groups in terms of benefits (and risks) of whatever type can then be ascribed to the difference in management, and thus to the device use.

Randomised comparative effectiveness studies compare the use of the device under evaluation – combined with subsequent actions in the case of non-therapeutic devices – directly with the best alternative strategy in the right population, measuring all relevant outcomes (for patients, users, and health care) in the short and long-term, and should generate the most *direct and valid evidence* as to whether the device use will indeed produce the intended relevant benefits, at an acceptable level of safety, as compared to prevailing care. This approach can even address the cost-effectiveness of device use as compared to the alternative care.

55 Clarke J., Wentz R. Pragmatic approach is effective in evidence based health care. *Bmj* 2000;321:566-7; Helms P.J., ‘Real world’ pragmatic clinical trials: what are they and what do they tell us? *Pediatr Allergy Immunol* 2002;13:4-9; Hunink M.G., Krestin G.P. Study design for concurrent development, assessment, and implementation of new diagnostic imaging technology. *Radiology* 2002;222:604-14; MacRae K.D. Pragmatic versus explanatory trials. *Int J Technol Assess Health Care* 1989;5:333-9; Roland M., Torgerson D.J., What are pragmatic trials? *Bmj* 1998;316(7127):285.

4.5 Deviation from the optimal study approach: no one-size-fits-all

In the case of medical devices, both therapeutic and non-therapeutic, such pragmatic randomised trials present a greater methodological challenge than in the case of medical drugs. For example, in clinical studies of devices there is often an intricate interplay between the technical complexities of the device, its user (often specific skills are needed), and ‘user/interpretation learning-curve effects’, compromising intervention adherence and sometimes even the possibility of randomised allocation (see also **Section 4.7.1. and Appendix IV, part A**). Indeed, the benefits and risks of a device often depend not only on the device itself but are a result of that complex interplay.

Non-randomised or observational studies of therapeutic devices are even more challenging, however. Selecting the proper study subjects, choosing a comparison group, adjusting for other influential factors, and addressing learning curve issues all require the conscious selection of study design, conduct and analysis plan. For example, when existing clinical databases or registries are used to obtain evidence of the potential benefits and risks of device use in practice, relevant information about other influences is often not – or only partially – available. This compromises any valid inferences about its true benefits and risks (see **Section 4.7.1. and Appendix IV, part B**). Moreover, for diagnostic, screening, prognostic or monitoring devices, i.e. devices without a direct therapeutic effect on patients’ outcomes, a single randomised comparative study may even not be indicated or may be very cumbersome.⁵⁶ Such devices may profit from a network or linked-evidence approach instead (see **Sections 4.7.2 and 4.7.3, and Appendix IV, parts C+D**). Clearly, there is *no one-size-fits-all* approach to evaluating the benefits of medical devices.

4.5.1 Reminder of optimal clinical study approach and overarching principle

The challenge is thus to identify the type of evidence and study approach that is required, given the specifics of the device (e.g. therapeutic or information-generating, invasive or non-invasive, requiring user interference/interpretation or not), its intended context, intended indication, and targeted individuals. However, we strongly recommend *maintaining a mental picture of the optimal randomised comparative effectiveness study when designing or choosing any alternative research approaches, to*

56 Bossuyt P.P.M., Lijmer J.G., Mol B.W., Randomised comparisons of medical tests: sometimes invalid, not always efficient. *Lancet* 2000;356:1844-1847; Lord S.J., Irwig L., Simes R.J., When is measuring sensitivity and specificity sufficient to evaluate a diagnostic test, and when do we need randomized trials? *Ann Intern Med* 2006;144:850-5; Biesheuvel C.J., Grobbee D.E., Moons K.G., Distraction from randomisation in diagnostic research. *Ann Epidemiol* 2006;16:540-4; Lijmer J.G., Bossuyt P.M., Various randomized designs can be used to evaluate medical tests. *J Clin Epidemiol* 2009;62:364-73.

allow for valid inferences about the benefits/added benefits of device use for health or health care.

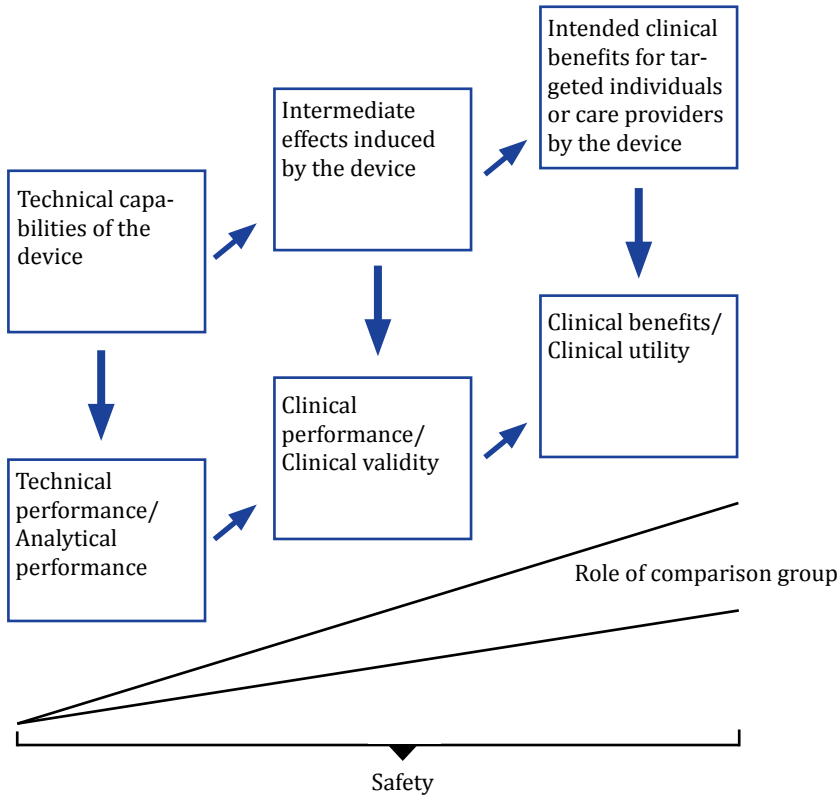


Figure 4.1 Pathway through which devices may eventually lead to benefits/added benefits for health or health care, and how each step is linked to a specific approach. The pathway ranges from the technical or analytical performance of a device through its intermediate changes or effects on health or health care outcomes to the final benefits for patients, care providers or society at large.

Indirect versus direct evidence

To obtain *direct evidence* of the health or health care benefits of a device, alternative study approaches may include simpler or more tailored randomised studies and non-randomised comparative studies. But there are also approaches that generate *indirect evidence* of the device's intended health or health care benefits. These include single study approaches followed by some kind of network or linked-evidence approach. For some devices (e.g. devices that are simply a minor modification of an existing device), a non-comparative or perhaps even only technical performance study can be justified using a network or linked-evidence approach. For other devices, direct evidence obtained in randomised comparative effectiveness studies or at least

a non-randomised comparative study may be required. This will be discussed and illustrated in **Section 4.7**.

A clear description of the pathway (see also **Section 4.6**) through which a device may ultimately affect health or health care outcomes, and which type, provides relevant clues for designing the following proper study in the life cycle of a device (see **Figure 4.1**). Studies focusing on earlier steps in this pathway generate indirect evidence of the benefits of a device for health or health care. This indirectness may relate to various sources, including:

- *outcomes*, e.g. short-term/intermediate rather than long-term outcomes, diagnostic accuracy outcomes rather than patient-relevant outcomes, clinical performance rather than clinical utility outcomes, care provider rather than patient outcomes;
- *use of the device*, e.g. by experts in specialised care rather than less experienced users in regular care;
- *type of device*, e.g. a new diagnostic device may lead to better or earlier detection of a disorder, whereas its benefits for improved therapeutic decision-making, let alone for improved patient health, has not been proven as yet;
- *individuals studied*, e.g. investigating a device in a sample of healthy individuals rather than in a sample taken from the intended context and patient population.

Bias versus precision

In addition to the directness or indirectness of a study on the health care benefits of a device, there are two other *key dimensions of quality* to consider in clinical research on devices (**Figure 4.2**). These are:

- *risk of bias (invalidity)*;
- *precision* of study estimates.

Every type of clinical study is vulnerable to specific *biases*, which we discuss in detail in **Section 4.7** and in the corresponding **Appendix IV**. Shortcomings in study design or execution can produce incorrect or invalid results and thus incorrect inferences about clinical performance, validity, benefits or utility. Minimising such bias is the aim of every clinical study. Generally speaking, the more one deviates from the optimal study approach, the more challenging it is to draw valid inferences about the benefits/added benefits of a device for health or health care. Awareness of this deviation is a major step forward, however. **Section 4.7** shows how to make the best possible inferences about a device's benefits or added benefits for health care when deviating from the optimal study approach.

Imprecision arises when study results are based on a small study sample. Imprecision leads to uncertainty about the study results. Proper statistical analysis can describe the uncertainty in study results by using confidence intervals.

Figure 4.2 integrates the overarching principle and pathway through which devices may lead to benefits or added benefits for health or health care, and the three main dimensions of quality for evidence obtained from device evaluations (indirectness of evidence; risk of bias; precision of estimates).

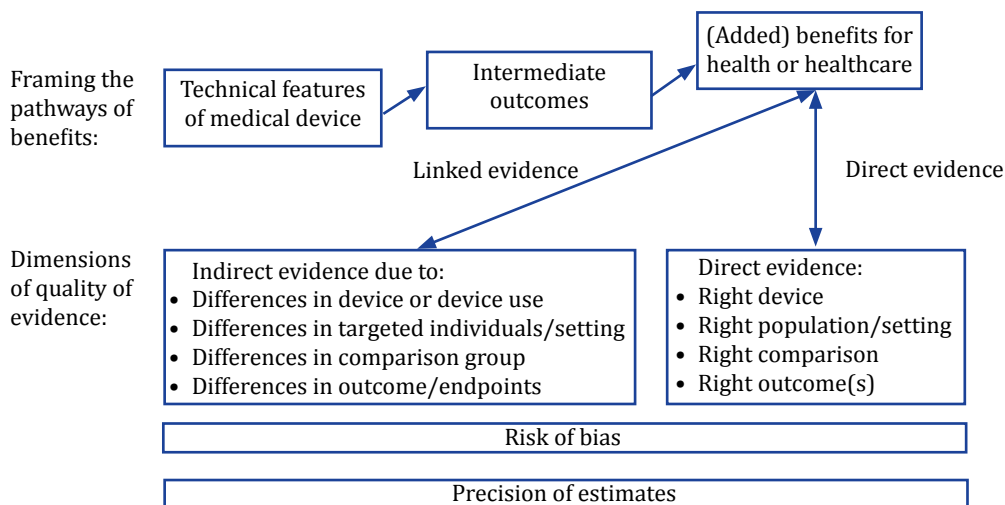


Figure 4.2 Relationship between the pathway through which devices may lead to benefits/added benefits for health or health care and the three dimensions of quality for evidence (indirectness of evidence, risk of bias, precision of estimates).

The framework outlined in **Figure 4.2** is further discussed in the remaining sections. **Section 4.6** provides guidance by explaining how and why we must describe the pathway through which a device leads to benefits (and risks). **Section 4.7** then describes various research approaches for generating evidence of the benefits of device use for the targeted individuals (patients or professionals). Here we focus on two main issues: the directness or indirectness of the evidence and the risks of bias (internal invalidity). Precision of study results (estimates) is a more straightforward concept for which the discussion in Appendix V suffices.

Appendix IV provides a detailed overview of the many different study designs that are available for investigating the intended health benefits of a device, either therapeutic, diagnostic, prognostic, screening or monitoring devices. The empirical examples are meant to serve as a source of inspiration for both researchers and manufacturers.

4.6 Description of the pathway leading to a device's benefits: a collaborative effort

The variety of devices available is reflected in the variety of pathways through which each device leads to intended and unintended effects on health or health care – e.g. directly or indirectly. It would be easier to decide which evidence and studies are needed to make valid inferences about the benefits or added benefits and risks of a device if device developers, manufacturers and end-users (i.e. medical professionals and targeted individuals or patients) were to collaborate and describe in detail the potential pathways through which benefits and risks are likely to occur. Such working pathways or mechanisms address:

- the anticipated technical or analytical capabilities of a device;
- the unintended and intended outcomes or effects expected in the targeted context;
- where these effects are likely to occur, e.g. in the targeted individuals/patients, in the care providers, other parties, or health care at large (e.g. cost-effectiveness);
- the anticipated mechanisms through which these potential risks and benefits will occur or be achieved in the intended context;
- existing care in the targeted context and individuals;
- the expected time frame in which potential risks and benefits might occur.

4.6.1 Why we must describe the pathways leading to device benefits

The working pathway is ideally defined in a very early stage of a device's development, or even at its conception. When manufacturers/developers provide information on the potential technical capabilities of a device and when the targeted end-users indicate what is needed in the intended medical context and what can be obtained with the prevailing care, the process of designing and – where necessary – altering devices and their anticipated studies becomes easier.

Having a detailed description of the working pathway at an early stage is particularly useful in situations where the device poses a major risk or is expensive to use. This is certainly true for invasive therapeutic devices such as pacemakers, nerve stimulators, breast implants, prostheses or medical robots, but it also holds for invasive and non-invasive diagnostic, prognostic, monitoring or screening devices, such as clinical chemistry tests to detect specific biomarkers, imaging tests with or without radioactive labelled contrast agents, and electrophysiology tests. This is particularly useful for the latter type of devices because their effects on health or health care are more indirect. For example, they affect health or health care outcomes indirectly by diagnosing a disease earlier or in an less burdensome manner, or through early detection that allows for more timely and effective treatment.

A detailed description of the above issues can also be used to consider evidence taken from different consecutive studies – e.g. technical, safety and clinical studies – in a more linked or network of evidence perspective. For instance, if technical performance studies fail to provide evidence for the intended technical capabilities of the device, further studies are unnecessary. If safety and technical performance studies of a new version (modification) of an existing device show that its safety and technical performance is similar to the preceding version but that it is less burdensome or cheaper to use, then subsequent studies may not be needed.

In summary, a clear description of the pathways through which a device generates potential intended and unintended health or health care effects, for whom and in what time frame, has the following merits:

- it promotes the logical build-up of a *network of evidence* leading to an integrated portfolio of evidence for a device;

- it provides a transparent and common starting point for designing the next optimal study;
- it leads to improved acceptance and uptake of a device by the medical profession, targeted individuals, and other related parties, certainly in the current era of evidence-based medicine, accountability and patient rights, and ever-decreasing health care resources.

4.7 Direct versus indirect evidence approaches

There are direct and indirect methods for generating evidence that a *therapeutic or non-therapeutic device* has benefits for health or health care (preferably added benefits beyond existing care) and for the intended individuals, professionals or society at large. Based on the level of indirectness, we can distinguish three main research approaches:

1. Studies providing **direct evidence and quantification** of the benefits or added benefits of device use for health or health care (**Section 4.7.1**).
2. Studies providing **indirect evidence** of the benefits or added benefits of device use for health or health care, via a **quantitative** linked-evidence approach (**Section 4.7.2**).
3. Studies providing **indirect evidence** of the benefits or added benefits of device use for health or health care, via a **qualitative** linked-evidence approach (**Section 4.7.3**).

Each approach has its own merits, and the vulnerability to specific forms of biases also varies across approaches. Below we discuss the essentials of these three research approaches, and how they can be informative about the likelihood that a device, whether therapeutic or non-therapeutic, will have the intended benefits on health or care when introduced in regular care. In this discussion we will be mindful of the expected pathways through which benefits (and risks) are generated (**Section 4.6**). In **Appendix IV**, we provide a detailed overview of the specific study types or designs that may be used in each of these three research approaches, along with their pros and cons (directness & risk of bias), and we illustrate each one with an example of a trial used for medical device evaluation.

4.7.1 Direct evidence and quantification of the benefits or added benefits of a device

As discussed above, a study aimed at providing direct evidence of a device's benefits for the health outcomes of the targeted individuals or for health care at large should basically address issues concerning its use in practice, i.e. the intended context, comparison strategy, relevant effects, time and size. In **Appendix IV (part A)**, as well as in **Section 4.4**, we discuss why this randomised pragmatic or comparative effectiveness design is the optimal, and most valid (lowest risk of bias), study design for a *direct*

evidence approach. This applies both to therapeutic and non-therapeutic devices. However, as discussed, this optimal single study approach – in which a device is studied in a strictly random manner against the best current care in the right population, measuring directly all relevant outcomes over a sufficiently long period of time – is not always feasible or possible

We consequently provide numerous alternative study approaches in **Appendix IV** that also produce the desired direct evidence of the benefits or added benefits of device use for the relevant and long-term health outcomes of the targeted individuals in the intended medical context. These designs may include alternative randomised designs, such as cluster (rather than patient level), factorial or cross-over studies, but they may also include more modern and efficient designs such as adaptive or N=1 trials.⁵⁷ These alternative study types are often as valid as the randomised pragmatic, comparative effectiveness trial, only less cumbersome and costly. **Appendix IV, part A** describes the essentials of each of these study types, including their pros and cons, and when they may be suitable in device evaluations.

There are many non-randomised study approaches that may also contribute to generating *direct* evidence about the long-term, comparative benefits or effectiveness of the relevant outcomes of a new device. These range from quasi-randomised (or quasi-experimental) studies to controlled before-after studies and cohort or case-control studies.⁵⁸ These non-randomised studies are more prone to bias due to differences in demographic and clinical characteristics between the groups being compared. Fortunately, there are various approaches to controlling or adjusting for such biases (see **Appendix V**), and these may be possible alternatives to clinical device studies. In **Appendix IV, part B** we survey all these possible alternative, non-randomised designs, their advantages and disadvantages, and the possibility of minimizing the risk of bias in each design.

All the study approaches mentioned above can be applied to all types of devices, whether *therapeutic devices or diagnostic, prognostic, screening or monitoring test devices*. In **Appendix IV, part C3**, however, we provide explicit empirical examples in which the added health or health care benefits of test devices are studied directly.

57 Pocock SJ. *Clinical Trials: a practical approach*: John Wiley & Sons, 1984; Friedman L.M., Furberg C.D., DeMets D.L., *Fundamentals of Clinical Trials*. 4th ed: Springer, 2010; Meinert CL. *Clinical Trials*. 2nd ed: Oxford University Press, 2012; Grobbee D.E., Hoes A.W., *Clinical Epidemiology – Principles, Methods and Applications for Clinical Research*. London: Jones and Bartlett Publishers, 2009; Rothman K.J., Greenland S., *Modern epidemiology*. Second ed. Philadelphia: Lincot-Raven Publishers, 1998.

58 Grobbee D.E., Hoes A.W., *Clinical Epidemiology – Principles, Methods and Applications for Clinical Research*. London: Jones and Bartlett Publishers, 2009; Hennekens C.E., Buring J.E., *Epidemiology in Medicine*. Boston/Toronto: Little, Brown & Co, 1987; Rothman KJ, Greenland S. *Modern epidemiology*. Second ed. Philadelphia: Lincot-Raven Publishers, 1998; Sackett D.L., Haynes R.B., Tugwell P. *Clinical epidemiology; a basic science for clinical medicine*. Boston: Little, Brown & Co, 1985; Rosenbaum PR. *Design of Observational Studies*. 2nd ed: Springer, 2002.

4.7.2 Indirect or linked evidence approach: Quantitative

Many clinical studies of devices do not measure the device's effects on the ultimate health outcomes for the targeted individuals or health care at large directly. Instead, they focus on measuring intermediate effects/outcomes along the working pathway of a device (see **Figures 4.1 and 4.2 and Section 4.5**). However, it is possible to undertake an explicit – quantitative – exercise that translates the device's effects on intermediate effects/outcomes into its effects on relevant (long-term) health or health care outcomes.

Therapeutic devices: short-term and surrogate endpoints

Device studies frequently focus on relatively the short-term effects or benefits of device use, rather than on the more relevant long-term outcomes. For example, the benefits of a new cardiac synchronisation device for exercise tolerance or quality of life improvement are studied in the first three months after implantation, rather than after 12 or 24 months.

Another, related, issue is the decision to measure surrogate or intermediate outcomes rather than patient-relevant (often called 'hard') endpoints. For example, restoring blood flow by inserting a new stent, measured directly after implantation, does not mean that it reduces cardiovascular events down the road. A nerve stimulator may reduce tremor in Parkinson patients, but not the ability to pick up things with their hands or to walk without falling, thus improving their quality of life.

The key issue when choosing short-term and intermediate outcomes is how closely they correlate with long-term and relevant health outcomes. Short-term and intermediate outcomes are more valuable if previous studies have repeatedly found a close relationship between them and long-term and relevant outcomes. All the study approaches discussed in **Section 4.7.1** (direct evidence approaches) and in **Appendix IV parts A and B** can also be used to study the benefits of device use for short-term or intermediate/surrogate health outcomes, with the same pros and cons. The main advantage of using short-term or surrogate/intermediate outcomes is that they often require a smaller sample size, shorter follow-up and thus smaller budgets. However, it must be acknowledged that the effects or benefits of the device for the desired longer term and/or patient relevant endpoints are unobserved in such studies. End-users (patients, care providers) as well as health care policymakers are usually more interested in the impact (and safety) of devices in the longer term and in participant-relevant outcomes, such as quality of life or improvement in daily activities. This is particularly so for implantable devices that are implanted for a longer period of time, such as breast implants, artificial hips, joints or heart valves, stents, pacemakers, and nerve stimulators.

Linked-evidence method. There are, however, ways to link evidence from different device studies quantitatively in order to investigate the benefits of a device for

long-term, relevant outcomes (and the safety of that device).⁵⁹

- First, standard study approaches as discussed in Section 4.7.1 (**Appendix IV, parts A and B**) have documented the benefits of device use on short-term and/or surrogate outcomes.
- Second, there are other studies that have quantified the association between short-term and long-term outcomes or between surrogate outcomes and participant relevant outcomes.
- Third, using decision or Markov modelling approaches, one can (fairly) easy link both types of evidence, and actually quantify the benefits (and risks) of device use on the long-term and relevant health outcomes.
- Fourth, these linked-evidence models can include various sensitivity analyses, for example accounting for the insecurities involving in using various types of evidence taken from different sources (studies).

In **Appendix IV, part D** we provide various examples of a *quantitative linked-evidence approach* for many different types of devices.

Diagnostic, screening, prognostic or monitoring test devices

A special case of using intermediate outcomes applies to information-generating devices such as diagnostic, screening, prognostic, and monitoring tests. As noted in **Section 4.3**, their effects on an individual's health outcomes are usually determined via the management or treatments that are initiated based on the information they provide.⁶⁰ Studies involving such devices usually begin by examining their predictive, diagnostic or screening accuracy and establishing the relationship between the device results and the presence/absence of a certain disease or condition. Such studies indicate to what extent the information provided and interpreted by the new device accurately predicts or detects the presence/absence of a specific disorder. Here, the pure predictive accuracy of the device is at stake, and is sometimes compared to the predictive accuracy of another, competing device. Despite the obvious relevance of such studies, predictive accuracy is still an intermediate endpoint: good predictive accuracy does not guarantee improved therapeutic management, let alone improved health outcomes at later stage.

59 Bluhm R. From hierarchy to network: a richer view of evidence for evidence based medicine. *Perspective Biol Med* 2005;vol 8:535-547; Walach H., Falkenberg T., Fonnebo V., Lewith G., Jonas W.B., Circular instead of hierarchical: methodological principles for the evaluation of complex interventions. *BMC Med Res Methodol* 2006;vol 6:6-29; Merlin T., Lehman S., Hiller J.E., Ryan P., The 'linked evidence approach' to assess medical tests: a critical analysis. *International journal of technology assessment in health care* 2013;29(3):343-50; Schaafsma J.D., Van der Graaf Y., Rinkel G.J., Buskens E. Decision analysis to complete diagnostic research by closing the gap between test characteristics and cost-effectiveness. *J Clin Epidemiol* 2009;62(12):1248-52.

60 Lord SJ, Irwig L, Simes RJ. When is measuring sensitivity and specificity sufficient to evaluate a diagnostic test, and when do we need randomized trials? *Ann Intern Med* 2006;144(11): 850-5; Bossuyt P.M., McCaffery K. Additional Patient Outcomes and Pathways in Evaluations of Testing. *Med Decis Making* 2009; Bossuyt P.M., Reitsma J.B., Linnet K., Moons K.G., Beyond diagnostic accuracy: the clinical utility of diagnostic tests. *Clin Chem* 2012;58(12):1636-43.

Appendix IV, part C provides an overview of the possible approaches to studying the predictive accuracy of diagnostic, screening, monitoring or prognostic test devices, their essentials and their pros and cons.

Linked-evidence method. Another option is to apply a similar *quantitative linked-evidence approach* as discussed above that combines evidence from these predictive accuracy studies with evidence from therapeutic studies – preferably randomised – in order to quantify the potential benefits of a diagnostic, screening, prognostic or monitoring device for relevant, long-term health outcomes.⁶¹ A key issue is whether the additional individuals identified by the diagnostic, screening or prognostic test will benefit from the therapy in the same way as the traditional individuals in the existing therapeutic studies.⁶²

In **Appendix IV, part D** we provide empirical examples of a *quantitative linked-evidence approach* that quantifies the benefits of a test device for long-term, relevant health outcomes.

4.7.3 Qualitative linked-evidence approach

As discussed in **Section 4.5**, indirect evidence that a device has benefits for long-term, relevant health outcomes can also come from other sources than those discussed in **Section 4.7.2**. For example:

- **Differences in device use** (e.g. by experts in specialised care rather than less experienced users in regular care);
- **Differences in the individuals studied** (e.g. investigating a device in a sample of healthy individuals rather than in a sample from the intended context and patients);
- **Different indications** (e.g. investigating a device for a specific medical context or indication and make inferences beyond this context);
- **Differences in the device itself** (e.g. a device is a modified version of a previous device and only a technical performance study is conducted, without any clinical study or evaluation of the newer device).

In such situations it is clearly much harder to apply a quantitative linked-evidence approach. In many cases, we do not know what kind of effects occur when a device is used, implanted or interpreted in regular care by less experienced users, or to what

61 Merlin T, Lehman S., Hiller J.E., Ryan P, The ‘linked evidence approach’ to assess medical tests: a critical analysis. *International journal of technology assessment in health care* 2013;29(3):343-50; Schaafsma J.D., Van der Graaf Y., Rinkel G.J., Buskens E., Decision analysis to complete diagnostic research by closing the gap between test characteristics and cost-effectiveness. *J Clin Epidemiol* 2009;62(12):1248-52; Koffijberg H., Van Zaane B., Moons K.G., From accuracy to patient outcome and cost-effectiveness evaluations of diagnostic tests and biomarkers: an exemplary modelling study. *BMC Med Res Methodol* 2013;13:12.

62 Ferrante di Ruffano L., Hyde C.J., McCaffery K.J., Bossuyt P.M., Deeks J.J., Assessing the value of diagnostic tests: a framework for designing and evaluating trials. *Bmj* 2012;344:e686.

extent these effects are intended and unintended. Similarly, we often do not know to what extent the benefits and risks change when a device studied in a specific sample of individuals (e.g. in patients with a specific type of cancer) or for a specific medical indication (e.g. stents for coronary arteries) is used for other types of patients (another type of cancer) or other indications (e.g. in the carotid arteries). Indeed, many devices can be used for several medical indications, which certainly applies to various diagnostic, prognostic and screening devices, such as imaging tests. Moreover, many devices are merely slightly modified versions or ‘copies’ of existing devices, with no testing beyond technical performance studies.

Quantitative linked-evidence approaches may not be easy in any of these situations. However, the use of sensitivity and scenario analysis may help assess the impact of device use on health or health care, providing a good indication of the need for additional studies. When this is not feasible, then some kind of *qualitative linked-evidence approach* is indicated to evaluate the risks and benefits for health or health care.⁶³ Here, evidence concerning the device’s use – e.g. from technical performance studies (**Figure 4.1**) – and evidence of its safety taken from other studies or from studies on closely related devices – as well as evidence from studies on the device’s health benefits should be combined and viewed from this perspective. A nice *example* is the recent study on the safety and benefits of stents for treating renal artery stenosis.⁶⁴ Knowing the pathway through which benefits and risks of device use arise (**Section 4.6**) can be extremely helpful in this context.

We must note that inferences about the relevant and long-term health benefits (and risks) of device use derived from qualitative linked-evidence approaches do not have the same validity as the direct quantitative approaches discussed in **Sections 4.7.1 and 4.7.2**. Nevertheless, qualitative linked evidence is currently often considered sufficient for market access, and perhaps even for reimbursement decisions.

Summary of linked-evidence approaches

Linked-evidence approaches offer an alternative in situations where direct evaluation of a device’s added benefits or effects on long-term, patient-relevant, ‘hard’ health outcomes is difficult or impossible. Their validity depends on how well the ‘intermediate’ outcomes predict the long-term, relevant health or health care outcomes. Linked-evidence approaches can be used to test therapeutic and information-generating devices. They may be valuable specifically for devices that are in fact *modifications of an*

63 Pope C., Mays N., *Qualitative research in health care*. 3rd ed: Blackwell Publishing, 2013; Bossuyt P.M., Reitsma J.B., Linnert K., Moons K.G., Beyond diagnostic accuracy: the clinical utility of diagnostic tests. *Clin Chem* 2012;58(12):1636-43; Bluhm R. From hierarchy to network: a richer view of evidence for evidence based medicine. *Perspective Biol Med* 2005;vol 8:535-547; Walach H., Falkenberg T., Fonnebo V., Lewith G., Jonas W.B., Circular instead of hierarchical: methodological principles for the evaluation of complex interventions. *BMC Med Res Methodol* 2006;vol 6:6-29.

64 Postma C.T., Spiering W., Kroon A.A., Renal artery stenosis: do not toss the stent aside. *Ned Tijdschr Geneesk.* 2013;157:A6350

existing device: they can be used to infer to what extent a minor device improvement will lead to benefits for the health or health care of the targeted individuals, context and users, by combining results from studies (technical and clinical) on the previous version or versions of the device.

The literature provides a wealth of examples of linked-evidence approaches applied to medical devices in general, and to diagnostics in particular. The UK Health Technology Assessment programme, supervised by the NHS and NICE, offers various examples over the past ten years. We cite a few.⁶⁵ Most linked-evidence approaches use general templates methods, generally based on similar approaches for drug evaluations.⁶⁶ There are also general guides on the use of linked-evidence approaches

-
- 65 Riemsma R., Al M., Corro Ramos I., Deshpande S., Armstrong N., Lee Y.C., et al., SeHCAT [tauroselcholic (selenium-75) acid] for the investigation of bile acid malabsorption and measurement of bile acid pool loss: a systematic review and cost-effectiveness analysis. *Health Technol Assess* 2013;17(61); Meads C., Sutton A., Malysiak S., Kowalska M., Zapalska A., Rogozinska E., et al. Sentinel lymph node status in vulval cancer: systematic reviews of test accuracy and decision-analytic model-based economic evaluation. *Health Technol Assess* 2013;17(60); Lord J., Willis S., Eatock J., Tappenden P., Trapero-Bertran M., Miners A., et al., Economic modelling of diagnostic and treatment pathways in National Institute for Health and Care Excellence clinical guidelines: the Modelling Algorithm Pathways in Guidelines (MAPGuide) project. *Health Technol Assess* 2013;17(58); Waugh N., Cummins E., Royle P., Kandala N.B., Shyangdan D., Arasaradnam R., et al. Faecal calprotectin testing for differentiating amongst inflammatory and non-inflammatory bowel diseases: systematic review and economic evaluation. *Health Technol Assess* 2013;17(55); Sutcliffe P., Connock M., Pulikottil-Jacob R., Kandala N.B., Suri G., Gurung T., et al. Clinical effectiveness and cost-effectiveness of second- and third-generation left ventricular assist devices as either bridge to transplant or alternative to transplant for adults eligible for heart transplantation: systematic review and cost-effectiveness model. *Health Technol Assess* 2013;17(53); Ward S., Scope A., Rafia R., Pandor A., Harnan S., Evans P., et al. Gene expression profiling and expanded immunohistochemistry tests to guide the use of adjuvant chemotherapy in breast cancer management: a systematic review and cost-effectiveness analysis. *Health Technol Assess* 2013;17(44); Simpson E., Stevenson M., Scope A., Poku E., Minton J., Evans P., Echocardiography in newly diagnosed atrial fibrillation patients: a systematic review and economic evaluation. *Health Technol Assess* 2013;17(36); Orlando R., Pennant M., Rooney S., Khogali S., Bayliss S., Hassan A., et al., Cost-effectiveness of transcatheter aortic valve implantation (TAVI) for aortic stenosis in patients who are high risk or contraindicated for surgery: a model-based economic evaluation. *Health Technol Assess* 2013;17(33); Mowatt G., Scotland G., Boachie C., Cruickshank M., Ford J., Fraser C., et al., The diagnostic accuracy and cost-effectiveness of magnetic resonance spectroscopy and enhanced magnetic resonance imaging techniques in aiding the localisation of prostate abnormalities for biopsy: a systematic review and economic evaluation. *Health Technol Assess* 2013;17(20); Meads C., Auguste P., Davenport C., Malysiak S., Sundar S., Kowalska M., et al., Positron emission tomography/computerised tomography imaging in detecting and managing recurrent cervical cancer: systematic review of evidence, elicitation of subjective probabilities and economic modelling. *Health Technol Assess* 2013;17(12)
- 66 Mushlin A.I., Rushlin H.S., Callahan M.A., Costeffectiveness of diagnostic tests. *Lancet* 2001;358:1353-55; A. Briggs, M. Sculpher, K. Claxton. *Decision Modelling for Health Economic Evaluation*. Oxford: OUP 2006. Pp. 256; Hunink M., Glasziou P., Siegel J., et al., *Decision making in health and medicine. Integrating evidence and values*. Cambridge: Cambridge University Press, 2001.

specifically for medical (diagnostic, screening, monitoring and prognostic) tests, obtained from organisations such as NICE, AHRQ and EUNetHTA.⁶⁷ However, the prevailing view that evidence linkage for medical devices merely requires the application of general linked-evidence principles has recently been challenged.⁶⁸ Much work is therefore being carried out on developing linked-evidence methods.

4.8 Summary: key issues to consider in all clinical studies on medical devices

This section *summarises the main concepts of this chapter* by providing a list of key issues that need to be considered by all researchers studying the clinical benefits, including clinical performance, of devices, whether therapeutic or non-therapeutic. These issues should be addressed regardless of whether the study approach is direct or indirect. They are also summarised in **Table 4.1**.

Device category/type – It is helpful to first specify the type of the device under study: a *therapeutic device* or a *test (diagnostic, prognostic, monitoring or screening) device* (**Section 4.3**). The intended health or health care benefits of the latter type are achieved indirectly because appropriate or more timely therapeutic management is subsequently provided, or because potentially harmful or more burdensome or costly additional management is avoided. However, undergoing an invasive diagnostic, screening, monitoring or prognostic test can still be harmful and may cause side-effects that negatively affect the benefit/risk ratio.

Pathway of benefits and aim of planned evaluation – Defining the pathways through which device use may lead to benefits and risks and determining the device's evaluation phase indicates, directly, what evidence should be produced in the next phase. This may range from evidence concerning the device's safety and technical performance to its clinical performance and clinical utility (**see Figures 4.1 and 4.2**).

67 <http://www.nice.org.uk/media/A0B/97/DAPManualFINAL.pdf>; Agency for Healthcare Research and Quality. Methods Guide for Medical Test Reviews [posted November 2010]. Rockville, MD. Available at: <http://effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayProduct&productID=>; <http://www.eunethta.eu/sites/5026.fedimbo.belgium.be/files/HTA%20Core%20Model%20for%20Diagnostic%20Technologies%201.0r.pdf>.

68 Lord J., Willis S., Eatock J., Tappenden P., Trapero-Bertran M., Miners A., et al., Economic modelling of diagnostic and treatment pathways in National Institute for Health and Care Excellence clinical guidelines: the Modelling Algorithm Pathways in Guidelines (MAPGuide) project. *Health Technol Assess* 2013;17(58); Laking G., Lord J., Fischer A., The economics of diagnosis. *Health Econ* 2006 Oct;15(10):1109-20; Sutton A.J., Cooper N.J., Goodacre S., Stevenson M., Integration of Meta-analysis and Economic Decision Modeling for Evaluating Diagnostic Tests. *Med Decis Making* 2008 28: 650. DOI: 10.1177/0272989X08324036; Velasco G.M., Mangiapane S., Surrogate outcomes in health technology assessment: an international comparison. *Int J Technol Assess Health Care* 2009 Jul;25(3):315-22. doi: 10.1017/S0266462309990213.

Specification of the device pathway of effects and its current evaluation phase offers direct guidance on the aim and design of the next study. Each planned study should describe the current state (phase) of evidence, why the next study is needed, and what additional evidence is lacking. This should be linked to the ‘working pathway’ of the device (**Section 4.6**) by indicating what step in the pathway will be evaluated. Focused research aims or questions and an explanation of the evidence still lacking are critical starting points for discussing the next device evaluation stage. Appropriate choices about the type of study design, the intended population, context, use and outcomes to be studied (**Sections 4.4 and 4.5**) all depend on the underlying research aims and questions. In the context of diagnostic, prognostic or monitoring test devices, there is often a sequence of prior tests still being used in regular practice. Specifying the intended role and place of the new test device in this sequence of tests is therefore crucial for designing the next study and how to generate the most informative evidence concerning the device’s benefits.

Targeted individuals & setting – The clinical benefit or performance of a device clearly depends on the medical setting and targeted users (care professionals) and/or patients. Hence, it is necessary to determine whether the *indications for use and diversity of the targeted individuals* in the planned study – both defined by the inclusion/exclusion criteria – are as similar as possible to the setting, targeted patients and/or users for which the device is intended. A device use may improve the health outcomes of the targeted patients or individuals directly, but it may also result in less burdensome care with comparable safety and health outcomes for the patient, more timely or efficient care by professionals, or more cost-effective care for society at large (**Section 4.3**). Examples include the use of a less invasive laparoscopic surgery device that offers the same level of safety and same effectiveness as conventional surgery, or the use of a less invasive bed-side diagnostic device that has the same diagnostic performance as the prevailing but more burdensome laboratory or advanced test. A device may also be beneficial because it improves the health of care professionals, for example the use of patient lifts that assist in lifting and turning patients in bed so that nursing personnel can avoid back complaints.

Other devices address care providers and society at large. Examples include faster and thus less costly diagnostic analysers in clinical laboratories that offer the same diagnostic performance as prevailing analysers, and thus safety and health effects for patients similar to those of prevailing care. Such new devices lead to more efficient or cost-effective health care and are thus relevant for care providers, health policymakers and indeed society at large.

Table 4.1 Key factors to consider in all clinical studies evaluating a medical device.

Key factors	Description	Implications
Type of device	Therapeutic device or test device	The benefits of a test device are achieved indirectly (informing subsequent health care actions), whereas a therapeutic device should have a direct effect. This difference has major implications for decisions related to device design, outcomes, study subjects and costs. Details about research on both types of devices are given in Appendix IV, part A, B and C.
Pathway of benefits	Specification of the pathway through which a device will generate beneficial effects	Researchers can use this pathway to indicate what type of evidence their study will produce. When evaluating devices, it is often efficient to start with more simple forms of evaluation. Pathway specification greatly improves the build-up of a portfolio or network of evidence.
Targeted individuals/setting	Indications for use and targeted individuals in the planned study	The clinical benefit or performance of a device clearly depends on the targeted medical context, individuals and users. Ideally, the targeted context, individuals and users in the planned study are as similar as possible to the 'real world' situation in which the device will be used.
Device use	The level of difficulty of device use, the user's required level of expertise, and the need for training and level of training required are all key factors that may influence the benefits and risks of the device (and its use).	Ideally, the device and the way it is used in a clinical study are the same as they will be in the 'real world' situation.
Comparison group	The evaluation of medical devices is in essence a comparative process: the new intervention is compared with prevailing care.	There are several design, conduct and analysis options for generating results based on a fair comparison between groups. The details of numerous research strategies, with their pros and cons, are given in Appendix IV.
Outcome/end-points	A variety of outcomes can be addressed in a device study. These range from: in and by whom the outcomes are measured, the degree of (in)directness of the outcomes, the subjectivity of the outcome measurement, and the timing of that measurement.	The selection of outcomes to be addressed in a device study is crucial and should be in line with the type of device, the (expected) pathway of effects, the device evaluation phase, and the type of evidence a study is meant to generate. Details about possible research approaches and the outcomes addressed in each approach are given in Appendix IV.

Device use and user interference – The device and the way it is used in a clinical study is ideally the same as will be in the 'real world'. The *level of difficulty of the device, the end-user's required level of expertise, the need for training and the level of training required by the end-user* are key factors that may influence benefits and risks.

However, evidence concerning the benefits (and risks) of devices is often generated in an environment quite different from that of real-world use. For example, many clinical studies of complex devices are conducted in an environment staffed by highly skilled and trained operators. Often there is a learning curve associated with the use of such devices, and the results of such studies cannot simply be generalised to expected benefits and risks in regular practice. Besides a detailed description of the pathways through effects (intended or unintended) may occur, we therefore also need to know the conditions for designing a proper clinical evaluation, for example potential user variability and learning effects on use or interpretation. If the benefits (and risks) of the device depend on how it is used or how its results are interpreted, and are thus susceptible to user variation, this should be explicitly defined.

In clinical investigations, it is advisable to apply *the same conditions – training, expertise of end-users, etc.* – as might be encountered in the targeted context and populations. Any deviations from this might compromise the performance/benefits of the device when used in practice. Clinical studies can encompass this by including more users (to allow for inter-user variation assessment), or by including users other than highly experienced ones who have been specifically trained for that purpose (in order to better reflect regular practice), and by having the same user use or interpret the device (e.g. an imaging test) multiple times (in order to assess intra-user variation and learning curve effects).

Finally, we wish to reiterate that users can be care providers (such as radiologists in the case of imaging tests or clinical chemists in the case of laboratory tests), but also individuals or patients (e.g. point-of-care tests to detect specific markers in body fluids such as the pregnancy test, but also automated blood-pressure measurement devices).

Comparison strategy & device allocation – Clinical benefits and performance – including improved or less burdensome diagnosis, prognosis, screening or monitoring – are relative and not absolute features. This in fact applies to any health care intervention. To draw valid inferences about the benefits, performance or accuracy of a device, it is important to compare observed benefits, performance or accuracy in the device group with those observed in a comparison (control) group.

This comparison is ideally involves a *direct evidence approach* (**Sections 4.5 and 4.7.1**). The control group individuals may undergo/use an earlier version of the same device, a competing device or any other alternative care, including no intervention. Ideally, the *comparative strategy will offer a realistic alternative*. Sham (placebo) interventions are less suitable for device studies. They may only be used if they meet ethical standards, and then only if there is thought to be a placebo effect.⁶⁹ Ideally, the two groups being compared are created by *random (chance) splitting*. Accordingly, provided the groups are large enough, it is more likely that any observed differences

69 Kirkley A., Birmingham T.B., Litchfield R.B., Giffin J.R., Willits K.R., Wong C.J., Feagan B.G., Donner A., Griffin S.H., D'Ascanio L.M., Pope J.E., Fowler P.J., A randomized trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med.* 2008 ;359:1097-107.

in benefits, performance, accuracy, risks, etc., can be assigned to the difference in management, and thus to the device use. Participants can also serve as controls for themselves (e.g. left vs. right hand). This may be indicated when the effects of management with the new device and of control management are local. A control group may also include *non-randomised, concurrent controls*, for example individuals who are as similar as possible to the individuals in the new device group but managed or treated at another centre or in another country that has not used the new device. Controls can also be *non-concurrent*, such as the use of *historical control groups*. For example, one may compare observed benefits, accuracy or performance in a group of individuals at the same or another hospital before the new device was used (control group) to the same outcome parameters after it was introduced.

The challenge in all non-randomised comparisons is to ensure that the groups being compared are comparable on all other aspects that might influence health outcomes, e.g. similar characteristics and additional care. Discussions about whether or not to include a comparison in a study and how to choose or create these comparison groups are *the same, whether the devices are therapeutic or test or information-generating*. However, in direct comparative studies of test devices it is often possible to use *cross-over designs* in which each subject is administered the new and alternative tests (see **Appendix IV, part C**).

It is possible to estimate the benefits, accuracy or performance of a new device as compared to alternative strategies indirectly using a *linked-evidence approach* (see **Sections 4.5 and 4.7.2 and 4.7.3**). The observed benefits of a device on intermediate, short-term or test accuracy outcomes are extrapolated to long-term effects on the relevant individual or patient outcomes using a quantitative modelling approach. This is described in detail in **Section 4.7.2**. The linked-evidence approach can also involve qualitative modelling. Here, evidence concerning device use (e.g. taken from technical performance studies), evidence concerning its safety (taken from other studies or from studies on closely related devices) and evidence taken from studies on the device's health benefits for the relevant outcomes are combined and placed into an overall working pathway perspective. Inferences about the relevant, and long-term health benefits (and risks) of a device obtained by means of qualitative linked-evidence approaches do not have the same validity as direct quantitative approaches.

Outcomes/endpoints – the choice of health outcome parameters (endpoints) is dictated by the expected capabilities and working pathway of device use. This involves addressing the following issues.

- *Which outcomes may be expected* by whom: the patient, care provider (e.g. ease of use, less burdensome or more efficient to use), others (e.g. hospital boards, policy-makers, society at large).
- *The type of outcomes*: patient-reported endpoints (e.g. perceived pain or quality of life), endpoints requiring subjective interpretation (e.g. lesions on imaging, epileptic foci on EEGs), or objective endpoints (all-cause mortality, myocardial infarction)

based on predefined increase in blood parameters).

- *The timing of outcome measurement*: short term or long term, in other words how much time should pass between device application and measuring the outcome in the participants. This depends on whether there might be a rebound effect, i.e. whether a short-term effect will diminish after 12 or 24 months.
- *Surrogate/intermediate versus hard/patient-relevant outcomes*. This choice is again guided by the pathway through which effects might occur.
- *Composite endpoints*, i.e. a pre-specified combination of more than one endpoint, are sometimes used. This choice should be based on the assumption that the effect of the intervention on each of the components will be similar, and is also of similar importance.
- A special instance of intermediate outcomes concerns information-generating devices such as diagnostic, monitoring, prognostic, and screening tests. Such devices are often only studied for their *predictive accuracy*, which in fact can be considered an intermediate outcome between technical performance outcomes and clinical effectiveness outcomes (**Sections 4.3 and 4.5**).
- Regardless of the types of outcomes, outcomes are *measured systematically* using standardised assessment tools in each patient (for patient-relevant endpoints) or for each care provider (for care provider outcomes).

5. COMPLEMENTARY ISSUES

Although the committee does not intend to prescribe when to use which type of study approach in which situation or for which type of device, there are a number of complementary issues that go beyond general guidance for the development of evidence strategies from a methodological point of view, as discussed in the previous chapter. These issues are discussed below. They should be seen against the background of European policymaking on medical devices, and are referenced here in order to improve or highlight the context of medical device evaluation in particular cases. In September 2012, the EU published an amended version of the MDD directive, following consultations. The new version is meant to tighten up the clinical research requirement, combines the MDD and AIMDD into one, and will be turned into a Regulation (2015), meaning that the EU guidelines must be implemented in national law (effective 2020).⁷⁰

New clinical data or time series ('off label' use)

Assuming that a medical device has a specific indication or targeted population, the question remains whether new clinical investigations are needed for each new potential application, context or target population. It is critical in this context to identify when a new device is an incremental development, e.g. the equivalent of a previous version or a similar device, whether it is a new device, or an existing device for a totally new indication or intended medical context. The equivalence concept is key in this discussion (**see Appendix III**).

⁷⁰ http://ec.europa.eu/health/medical-devices/documents/revision/index_en.htm and http://ec.europa.eu/health/medical-devices/documents/interpretative-documents/index_en.htm.

Randomised or non-randomised comparative studies are generally time-consuming. Modifications (e.g. to enhance safety, patient comfort, ease of use) of low-risk technologies that already have been available for many years and have been properly characterised from a safety, performance and even benefits perspective might not require new clinical evidence derived from a newly designed and conducted comparative study. Bench and/or animal testing may be sufficient to substantiate the benefits (performance or effectiveness) of the enhanced device. For modifications of high-risk and some medium-risk devices, however, new clinical studies – in targeted individuals – are basically required. This requirement is even stricter in the amended directives. What approach and which specific designs are useful depend on the type of information needed, as discussed at length in **Chapter 4** and **Appendix IV**. For example, historical controls involving a previous version of the device can be used as a comparison group.

Linked evidence, notably quantitative, approaches (combining the clinical data of small user groups and building evidence by combining data and stratifying for specific parameters, see **Section 4.7.2**) are useful because it is not feasible to require entirely new studies for every new target group. Instead, subsequent time series can be collected. This is only possible if the research parameters and possible bias factors are well described and if the details of the study are easily accessible for others.

The committee recommends further study of linked-evidence approaches for medical device evaluations, as discussed in Sections 4.7.2 and 4.7.3. Appraisal of consecutive clinical studies may build on a more holistic perspective, with a more flexible strategy to select and synthesise data, an approach known as a network of evidence as opposed to a hierarchy of evidence.⁷¹

Registration of the use of medical devices – information in databases

Registration systems and routine care databases used to obtain high-quality data that provide a nuanced understanding of safety, performance and clinical benefit in

71 Lord J., Willis S., Eatock J., Tappenden P., Trapero-Bertran M., Miners A., et al., Economic modelling of diagnostic and treatment pathways in National Institute for Health and Care Excellence clinical guidelines: the Modelling Algorithm Pathways in Guidelines (MAPGuide) project. *Health Technol Assess* 2013;17(58); Laking G., Lord J., Fischer A., The economics of diagnosis. *Health Econ* 2006 Oct;15(10):1109-20; Sutton A.J., Cooper N.J., Goodacre S., Stevenson M., Integration of Meta-analysis and Economic Decision Modeling for Evaluating Diagnostic Tests. *Med Decis Making* 2008 28: 650. DOI: 10.1177/0272989X08324036; Velasco GM, Mangiapane S. *Surrogate outcomes in health technology assessment: an international comparison.*; A. Briggs, M. Sculpher, K. Claxton. Decision Modelling for Health Economic Evaluation. Oxford: OUP 2006. Pp. 256; Hunink M., Glasziou P., Siegel J., et al., *Decision making in health and medicine. Integrating evidence and values.* Cambridge: Cambridge University Press, 2001. IJzerman M.J., Early Bayesian modeling of a potassium lab-on-a-chip for monitoring of heart failure patients at increased risk of hyperkalaemia. *Technological Forecasting and Social Change*. 2012 Sep;79(7):1268–79.

a broad-based population are difficult and costly. These registrations are, however, crucial to assessing the long-term safety, performance and effectiveness of medical devices, especially in the case of calamities. The committee wishes to emphasise the relevance of such registrations, as well as the need to comply with regulatory post-marketing surveillance requirements, given the need for continuous monitoring of risks and benefits of medical devices.⁷² Here, unique device identification in the digital health infrastructure may prove of value.

In this respect, the EU is also taking steps: the Global Medical Device Nomenclature (GMDN), now supported by the GMDN Maintenance Agency,⁷³ has developed the Nomenclature into what it is today: a comprehensive, regularly updated web-based nomenclature accessible to manufacturers for a licence fee. Closer public-private partnerships in Europe are needed to combine and connect existing registries and to analyse such data in-depth. In addition, these registries require high-quality data for proper analysis. Such databases can combine all available data taken from technical studies with all clinical investigation and long-term use data.

The committee realises that the responsibility for lifetime registries of medical devices is best placed at the user end, and that this requires transparency from manufacturers and regulators.

User involvement

User needs, user contexts, user outcomes and user evaluation are all aspects of a more user-centred world. One of the principles of device development and evaluation is the early and continuous focus on users. The modes of involvement range from informative, through consultative to participative.⁷⁴ User involvement is needed at an early stage of medical device development in order to avoid risks and support benefits. In assessing the benefit/risk ratio, patient-reported outcome measures (PROMs) can also support the quality of care delivered to patients from a patient perspective.⁷⁵

⁷² A European databank for medical devices has been developed under the name Eudamed (under provision of the medical device directives). It is a secure web-based portal acting as a central repository for information exchange between national Competent Authorities and the Commission and is not publicly accessible. Eudamed use is obligatory since May 2011. The aim of Eudamed is to strengthen market surveillance and transparency in the field of medical devices. Eudamed contains (among others) data obtained in accordance with the vigilance procedure and also data on clinical investigations.

⁷³ <http://www.gmdnagency.com/>.

⁷⁴ Kujala S., User involvement: a review of benefits and challenges. *Behaviour & Information Technology*; 2003, vol 22(1): 1-16. DOI: 10.1080/0144929021000055530.

⁷⁵ Crawford B., Mathias S.D., Sullivan E., Basic principles of patient-reported outcomes research. In: *Therapeutic and diagnostics device outcomes research*; 2011, ISPOR. ISBN 9780974328928; IJzerman M.J., Steuten L.M.G., Early Assessment of Medical Technologies to Inform Product Development and Market Access. *Appl Health Econ Health Policy*. 2011 Sep;9(5):331-47.

User preferences are also relevant in benefit-risk appraisal. Usually the regulator defines what is acceptable, whereas risk tolerance will vary among patients, in turn affecting individual patient decisions as to whether the risks are acceptable in exchange for a probable benefit. Some patients are willing to take on a very high risk to achieve a small benefit, whereas others are more risk averse. The basic principles for what constitutes a meaningful benefit and risk may continue to evolve.⁷⁶

Finally, users operate autonomously and are increasingly able to buy tests ‘over the counter’. Devices, for instance a genetic test or a blood pressure meter, are offered or sold on the internet and the test results are delivered without any advice or interpretation by a health care professional. A growing number of remote monitoring and mobile phone applications are becoming available for homecare and for patient self-management. They represent a whole new type of health care intervention whose benefits are related to user and patient expectations. However, such technology also challenges us to define/redefine the risks and benefits associated with their use. This publicly accessible, *direct-to-consumer* segment beyond any medical control may be one of the fastest-growing parts of the market in the near future. Should these devices nevertheless meet the same standards as those developed for use by health care professionals?

The committee stresses that these ‘modern’ device use methods should be taken into account in any further guidance for research into clinical performance and benefit.

End-user training and follow-up on safety issues

User interference is a crucial factor in the introduction of medical devices and their use throughout the entire device life cycle, not only when accessing the market through well-trained and properly supported key opinion leaders, but also when device use is being extended to other hospitals and users. Untrained use may seriously affect device safety and hence negatively impact the device’s risk/benefit ratio. The Dutch Order of Medical Specialists (OMS) and the Dutch Health Insurance Board (CvZ) therefore defined guidelines for the safe introduction and use by professionals of new medical innovations in regular care. By acknowledging the responsibility of health care professionals and their institutions for the safety of their patients, OMS and CvZ wish to implement a system of back-testing for unintended events. The guideline (“Leidraad”) recommends the need for responsible implementation to prevent such risks, including extensive end-user training with medical devices and prospective risk analysis in order to maximise the benefit and minimise the *risk when used in regular practice*.

76 Hauber, A.B, et al. (2009). Older Americans’ risk-benefit preferences for modifying the course of Alzheimer disease. *Alzheimer Dis Assoc Disord*, 23:23-32; Nigel K. Arden, N.K, et al. (2012). How do physicians weigh benefits and risks associated with treatments in patients with osteoarthritis in the United Kingdom? *J.Rheum*, DOI:10.3899.

When institutions or professionals cannot indicate that they have followed these guidelines, the Health Care Inspectorate (IGZ) and/or CvZ may take action (e.g. regarding reimbursement). The new guidelines for the safe introduction of medical devices in regular care are currently being discussed with OMS and medical research societies. The guidelines will be presented in 2014.⁷⁷ Of particular interest is that the guidance also describes how long-term follow up could be organised.

The committee endorses the OMS/CvZ plan for guidelines for the safe introduction of medical devices in regular care, as they will support the follow-up of medical device use and help build and enhance linking evidence.

Evaluation of cost-effectiveness of devices

Cost-effectiveness analysis aims to provide evidence at the level of the health care system and builds upon modelling and standardisation. The international standards are more or less established for pharmaceuticals. Standards of research are comparable for medical devices, yet their application and implementation in comparable phases of development and as part of reimbursement decisions have not received the same level of support, i.e. have not yet become standard policy. Standards will inevitably follow further advances in health technology assessment, but for medical devices they may need to overcome many challenges first.⁷⁸

The cost-effectiveness of an individual device depends on our knowing the costs and the device's effectiveness. Both knowledge sets are unstable, so the result of the equation is volatile as well, and not yet widely applicable. Costs are not always stable because prices are negotiable at the level of the institution. Effectiveness is very often not established in terms of health outcomes, although there is a growing demand for patient-reported outcomes, and surrogate endpoints are non-informative as primary endpoints.

For example, the National Institute for Health and Care Excellence (NICE,) provides guidance to the NHS in the UK on the clinical and cost-effectiveness of selected new and established technologies. The Institute issues a guidance document with an overview of the principles and methods of health technology assessment, describing

77 *Leidraad Introductie van Nieuwe Interventies in de Klinische Praktijk*; Draft December 2013. Kennisinstituut van medische Specialisten, op initiatief van de orde van Medisch Specialisten en College (OMS) voor Zorgverzekeringen (CvZ).

78 Schaafsma J.D., Van der Graaf Y., Rinkel G.J., Buskens E., Decision analysis to complete diagnostic research by closing the gap between test characteristics and cost-effectiveness. *J Clin Epidemiol* 2009;62(12):1248-52; Koffijberg H., Van Zaane B., Moons K.G., From accuracy to patient outcome and cost-effectiveness evaluations of diagnostic tests and biomarkers: an exemplary modelling study. *BMC Med Res Methodol* 2013;13:12.

key principles of appraisal methodology.⁷⁹ However, despite ongoing advances in the methodology of technology appraisal, areas of controversy and uncertainty remain. Methods of cost-effectiveness analysis are no exception in terms of ongoing development.⁸⁰ Thus, there will be contributions to our understanding and further development of HTA knowledge and expertise both the in the Netherlands and internationally (e.g. the HTA international policy forum, HTAi).

The committee realises that innovative approaches and further research and guidance are needed with respect to the cost-effectiveness assessment of medical devices.

79 Guide to the methods of technology appraisal 2013, National Institute for Health and Care Excellence (NICE). <http://publications.nice.org.uk/guide-to-the-methods-of-technology-appraisal-2013-pmg9>.

80 Van de Wetering G., Steuten L.M.G., Birgelen von C., Adang E.M.M., IJzerman M.J., Early Bayesian modeling of a potassium lab-on-a-chip for monitoring of heart failure patients at increased risk of hyperkalaemia. *Technological Forecasting and Social Change*. 2012 Sep;79(7):1268-79; IJzerman M.J., Steuten L.M.G., Early Assessment of Medical Technologies to Inform Product Development and Market Access. *Appl Health Econ Health Policy*. 2011 Sep;9(5):331-47.

6. CONCLUDING REMARKS

To enable well-informed health care and reimbursement decisions, research is required to provide evidence of the risks (safety), performance, benefits (effectiveness) and even cost-effectiveness of any health care intervention, including medical devices. As applies to all medical interventions, medical devices are more likely to be used when it is clear that their risk/benefit ratio is more favourable than – or at least equal to – current care. Accordingly, this involves studying medical devices in a valid scientific manner.

The committee has discussed the issues presented in this report at length over a period of almost two years. During that time, it addressed many viewpoints, ideas, comparisons, policy directives, frameworks and designs. It was not easy to find a strategy that could build on the committee’s combined scientific expertise in clinical research methods without interfering with the current state of affairs and developments in the regulatory, clinical, academic and industrial sectors. The committee explicitly decided to focus on offering guidance for research into the added performance and benefits of medical devices, as stakeholders felt that this was notably lacking in current research guidelines for medical devices.

The guidance in this report supports the design, conduct, analysis and reporting of suitable, robust, transparent and sustainable device research in order to build evidence of the risks (safety), performance and benefits of medical devices, thereby helping to meet the needs of tomorrow. This report can help stakeholders involved in the development, use, evaluation and regulation of medical devices choose the proper approach when evaluating the benefits and performance of medical devices for health care, given the device specifics. The report stresses that there is no one-size-fits-all approach possible, given the wide diversity of medical devices available and their specific characteristics. The report also helps to interpret existing evidence of the risks,

performance, and benefits of a device use and put this evidence in perspective so as to make more informed decisions regarding the introduction, use, and reimbursement of a device. The committee trusts that this report is useful for national and international stakeholders, including:

- researchers and professionals in health care;
- medical device industry, and SMEs in particular;
- Notified Bodies;
- health insurance companies;
- hospital boards;
- regulatory agencies;
- funding agencies;
- medical ethical review committees.

The committee appreciates that the guidance in this report is not set in stone. It should be regarded as a set of recommendations and options for all stakeholders involved in assessing or using medical devices. The guidance ultimately aims to protect society and users/end-users against the introduction and use of devices that are 'unsafe' or 'unnecessary'.

APPENDICES I-VII

APPENDIX I

RESOLUTION INAUGURATING THE COMMITTEE (*INSTELLINGSBESLUIT*)



KONINKLIJKE NEDERLANDSE
AKADEMIE VAN WETENSCHAPPEN

RESOLUTION INAUGURATING THE COMMITTEE ON THE INTRODUCTION OF TECHNOLOGY IN HEALTHCARE

The Board of the Royal Netherlands Academy of Arts and Sciences (KNAW), *given Section 8 of the Regulations governing the Academy [Reglement van de KNAW], considering that the current methods of clinical research with medical innovation technologies make testing the effectiveness (in terms of better patient outcomes) of the technologies very difficult, and that consequently, on the one hand, rapid implementation in clinical practice is not possible and, on the other, technologies are implemented too quickly; considering that the Netherlands has very sound expertise in the field of clinical research and has therefore built up an excellent international reputation, considering that the Netherlands can export its ideas on new methods of design, performance, and analysis of clinical research with innovative technologies; and considering that the Netherlands now has the right momentum for medical technology; resolves, at the proposal of the Council for Medical Sciences, to appoint an advisory committee on the 'Introduction of Technology in the Care Sector', hereinafter referred to as 'the Committee'.*

Section 1. Task of the Committee

The task of the Committee, within an exploratory process, is to increase awareness of the methodology for research in connection with the introduction of medical technologies ('devices') in the healthcare sector. This development of technologies has advanced very rapidly in recent years and will continue to do so in the years ahead. In this task, the term 'technology' refers to diagnostic (e.g. fMRI, electrophysiology), prognostic (e.g. biomarkers, including genetic markers), and therapeutic (i.e. non-pharmaceutical) technologies (e.g. pacemakers, stents). The Committee will make recommendations for ethically responsible research methods for the introduction of such innovative technologies in the healthcare sector. This will include, for example, new designs and analysis methods, different endpoints, and the performance of relatively small-scale studies that must be able to fuel one another quickly. The recommendations will be substantiated by methodological and statistical knowledge. The Committee will set out its findings in an exploratory report.

The Committee will consider the following questions:

- To what extent is it possible and desirable to require that the burden of proof that is considered necessary, on the basis of legislation and regulations, for determining the effectiveness and safety of a pharmaceutical should also apply to non-pharmaceutical medical technologies ('devices')?

- What is the current state of practice and knowledge regarding the methodology of clinical research in connection with innovative technologies ('devices')?
- How can improvements in the methodology of clinical research in connection with medical technologies ('devices') improve introduction?

The advisory report is intended in the first instance for the professional groups (doctors and researchers) and developers of the new technologies, but it is also intended for evaluation and review committees, such as the NWO (ZonMw), CCMO and MRECs. In the past, researchers and industries have to a large extent been trained to adhere to classical methods of diagnostic, prognostic and pharmaceutical research, which are not in all cases sufficient to quantify the true clinical value (i.e. effectiveness) of technologies, or are not feasible in practice in the light of the developments outlined above, and are sometimes not even ethically desirable. It is extremely important to instigate sophisticated discussion within the professional groups concerned of other research methods for innovative technologies.

Given that the problem is a global one, the advisory report will be drawn up in English.

The Committee will ensure that its advisory report can be submitted to the Board before 30 June 2012, after which the peer review process will take place (see Section 3).

Section 2. Composition and Term*

The following persons have been appointed as members of the Committee:

- Prof. A. van den Berg (Professor of Microfluidics & Nanofluidics, University of Twente) **
- Prof. P.M.M. Bossuyt (Professor of Clinical Epidemiology, Amsterdam Academic Medical Centre)
- Prof. E. Buskens (Professor of Medical Technology Assessment, University Medical Centre Groningen)
- Prof. J. Kievit (Professor of Clinical Decision Analysis and surgeon, Leiden University Medical Centre)
- Prof. E. van Leeuwen (Professor of Medical Ethics, Radboud University Nijmegen Medical Centre)
- Prof. K.G.M. Moons (Professor of Clinical Epidemiology, Utrecht University Medical Centre), chair
- Prof. W.J. Niessen (Professor of Biomedical Imaging, Erasmus Medical Centre)
- Prof. T. Stijnen (Professor of Medical Statistics, Leiden University Medical Centre)
- Prof. E.G.E. de Vries (Professor of Medical Oncology and oncologist, University Medical Centre Groningen)

The Committee's term will conclude on 1 December 2012.*

The Committee will receive support from the Academy's Staff Department in accordance with the instructions of the Director General.

* Later the term was extended to 2014.

** Early 2012 prof. Van den Berg was replaced by prof. M.J. IJzerman.

Section 3. Quality management

Prior to being appointed, the members of the Committee have taken note of the Preamble to the Advisory Committee Statement of Interests [Belangenverklaring Adviescommissie] and filled in and returned the KNAW Advisory Committee Statement of Interests Form [Formulier Belangenverklaring Adviescommissie KNAW].

The peer review policy is described in the Policy Framework for Quality Assurance in Advisory Reports [Beleidskader Kwaliteitsborging Adviezen]. There will be no deviation from this policy .

Section 4. Follow-up and Communication

The Committee will follow up and provide for communication concerning its findings. During the process, the Committee will also devote considerable attention to communication with stakeholders (see Section 1).

Section 5. Costs and Remuneration

Pursuant to Section 18(2) of the Regulations governing the Academy, the members of the Committee will be compensated for the travel costs that they incur.

Section 6. Confidentiality

The Committee will observe confidentiality in respect of all information that becomes known to them in the context of the implementation of this resolution and that can be considered to be of a confidential nature.

Adopted in Amsterdam by the Board of the Royal Netherlands Academy of Arts and Sciences on 17 January 2011.

On behalf of the Academy Board,

Dr K.H. Chang
Director General of the Royal Netherlands Academy of Arts and Sciences

APPENDIX II

COMMITTEE AND SECRETARIAT

Prof. K. Moons	University Medical Center Utrecht (Chair)
Dr I. Meijer	Center for Science and Technology Studies (CWTS), University Leiden (secretary)
Prof. P. Bossuyt	Academic Medical Center Amsterdam
Prof. E. Buskens	University Medical Center Groningen
Prof. M. IJzerman	University of Twente. Early 2012 he succeeded prof. Van den Berg as committeemember
Prof. J. Kievit	Leiden University Medical Center
Prof. E. van Leeuwen	University Medical Center St Radboud Nijmegen
Prof. W. Niessen	Erasmus Medical Center Rotterdam
Prof. T. Stijnen	Leiden University Medical Center
Prof. E. de Vries	University Medical Center Groningen

Technical support

Dr L. Hooft	Dutch Cochrane Center
Dr J. Reitsma	University Medical Center Utrecht

APPENDIX III

DEFINITIONS, RISK CLASSIFICATION, AND REGULATION OF MEDICAL DEVICES

This appendix provides information on the regulatory requirements for medical device innovations and the organisation of market access, which is regulated at international level.

Definition of medical device

The term *medical device* covers a multitude of different items. A few are complex and reflect the latest advances in medical technology, e.g. new imaging equipment, various implants in heart, vessels, bones and joints, and advanced point-of-care lab tests. Most, however, are relatively simple, e.g. thermometers, latex gloves and crutches, to mention only a few. There are also numerous ‘me-too’ devices, i.e. different suppliers producing similar devices that offer the same performance. The medical device world is thus diverse in its nature, applications and user categories. Diversity is also evident in the risks involved and in the regulatory systems used to manage those risks; in the manufacturing costs and sales prices; in the standards and nomenclature systems; and in the various approaches used to determine the effectiveness, cost-effectiveness, and safety of devices.¹

The committee decided to take the European Commission definition of a medical device as its starting point. The Commission defines a medical device as:

‘any instrument, apparatus, appliance, software, material or other article, whether used alone or in combination, including the software intended by its manufacturer to be used specifically for diagnostic and/or therapeutic purposes and necessary for its proper application, intended by the manufacturer to be used for human beings’ (Directive 2007/47/EC).

Devices are to be used for the purpose of:

- *Diagnosis, prevention, monitoring, treatment or alleviation of disease.*
- *Diagnosis, monitoring, treatment, alleviation of or compensation for an injury or handicap.*

1 Priority Medical Devices, WHO, 2010.

- *Investigation, replacement or modification of the anatomy or of a physiological process*
- *Control of conception*

This includes devices that ‘do 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’. The latter refers to products combining medical devices and drugs, in which the main mode of action is device-related.

In fact, Directive 2007/47/EC is the technical revision of the three main directives constituting the core legal framework for medical devices. Compliance with Directive 2007/47/EC became mandatory for manufacturers, Notified Bodies and Competent Authorities on 21 March 2010. The three directives aim to ensure a high level of protection for human health and safety and the proper functioning of the single market. The three main directives are:

- Directive 93/42/EEC regarding medical devices (MDD). The MDD covers all medical devices that are not regulated by the more specific AIMD Directive or IVD Directive (see Box III.1)
- Directive 90/385/EEC regarding active implantable medical devices (AIMDD). Active Implantable Medical Devices are devices relying for their functioning on a source of electrical energy or any source of power other than that directly generated by the human body or gravity which are intended to be totally or partially introduced, surgically or medically, into the human body or by medical intervention into a natural orifice, and which are intended to remain after the procedure. A pacemaker is an example of an AIMD.
- Directive 98/79/EC regarding *in vitro* diagnostic medical devices (IVDD). An *in vitro* diagnostic medical device (IVD) means any medical device which is a reagent, reagent product, calibrator, control material, kit, instrument, apparatus, equipment, or system, whether used alone or in combination, intended by the manufacturer to be used *in vitro* for the examination of specimens, including blood and tissue donations, derived from the human body, solely or principally for the purpose of providing information: on a physiological or pathological state; or on a congenital abnormality; or to determine the safety and compatibility with potential recipients; or to monitor therapeutic measures.

BOX III.1: THE FOLLOWING DEVICES ARE REGULATED BY THE MDD 1993

- Hospital equipment (e.g. anaesthetic instruments, heart-lung machines, x-ray machines, and surgical instruments)
- Dental equipment (e.g. chair, UV-light apparatus, implants)
- Audiometric devices (e.g. measuring instruments, hearing aids)
- Ophthalmic devices (e.g. diagnostic instruments, glasses and contact lenses)
- Prostheses (both implantable and non-implantable)
- Devices for disabled or impaired persons (e.g. wheelchairs, and rehabilitation devices)
- Devices for single use (e.g. contraceptive rubbers).

The latest revision of these directives began recently with the publication of two Commission proposals (26 September 2012) for two new regulations to replace the existing three directives. The new Medical Devices Regulation will integrate the MDD and the AIMDD into a single regulation; and the IVDD will be replaced by the In vitro diagnostic medical devices regulation. The proposed changes (covered in both proposals) include greater transparency, stricter requirements on traceability in the supply chain, more scrupulous designation and auditing of notified bodies, additional pre-market scrutiny for higher risk devices, more clinical evidence required for higher risk (see below) and implantable devices, and the introduction of central reporting of serious incidents to a central database. More structured post-marketing surveillance activities will also be required.

Risk Classification of medical devices

One key difference across the many different types of medical devices is the wide range of risks associated with their use. These risks are particularly relevant when a request for market access is filed for new devices. The risk classification system is different from that for pharmaceuticals. The EU concept of risk classification depends on the intended use (as described by the manufacturer) and distinguishes four classes of risks, ranging from low risk to high risk (see box below). This classification system is outlined in Annex IX to Directive 93/42/EEC. The classification is based on the risk for patients if the medical device were to fail and is set out in a set of rules. The rules are broad statements relating to situations, functions, parts of the body treated, properties, etc.² They differentiate between numerous different categories in order to offer more flexibility and more easily assess the risk of new developments in medical technology.

- Class I – generally regarded as low risk
- Class IIa – generally regarded as medium risk
- Class IIb – generally regarded as medium risk
- Class III – generally regarded as high risk

The classification is not a list of products, which would require constant updating. The rules are designed to minimise any possible ambiguity. The manufacturer can assign each new device, or part of a (updated) device, to a risk category, which in turn is connected to specific requirements such as how manufacturer are to demonstrate a device's performance. The rules work well, with disagreements about risk

² European Commission, DG Health and Consumer; Directorate B, Unit B2 'Cosmetics and medical devices'. MEDDEV 2.4/1 Rev 9; June 2010. Guidance document – Classification of medical devices Guidelines relating to the application of the Council Directives 93/42/EEC on Medical Devices.

classification arising for only ten or so of the 500,000 different devices. The main criterion that distinguishes the different classes is whether clinical evaluation requires (new) clinical investigations, or whether it can refer to the literature or previous clinical investigations with equivalent devices. Risk classification is particularly relevant for *new* classes of medical devices. It is generally not critical for improvements to a device that is based on the same working principles.

The following concepts are used in the legislation governing access to the medical devices market (see also Box 2.1 in Chapter 2, main text). These concepts are not directly comparable to the concepts used for pharmaceuticals, such as efficacy, effectiveness and safety.

- Clinical evidence – Information or clinical data that supports the scientific validity and performance (analytical performance and, where applicable, clinical performance) of the device when used as intended by the manufacturer;
- Analytical/Technical Performance – The ability of a device to correctly detect or measure a particular analyte;
- Clinical evaluation – The assessment and analysis of clinical data pertaining to a medical device to verify the clinical safety and performance of the device when used as intended by the manufacturer. Clinical evaluation is an ongoing process; information about clinical safety and performance (e.g. adverse event reports, results from any further clinical investigations, published literature etc.) should be monitored routinely by the manufacturer once the device is available on the market and the benefits and risks reassessed in light of this additional information;
- Clinical investigation (synonymous with ‘clinical trial’ and ‘clinical study’) – Any systematic investigation or study in or on one or more human subjects, undertaken to assess the safety and/or performance of a device;
- Clinical utility – A concept generally used for diagnostic devices that refers to the likeliness of the test significantly improving the health outcomes of the targeted individuals. In other words, the capacity of the test to rule in and/or out the disorder of interest, and to facilitate a decision to adopt or to reject a subsequent (e.g. therapeutic) action. Clinical utility is an increasingly common concept in health care, but one that lacks an agreed formal definition or conceptualisation. The term is commonly used as a synonym for clinical effectiveness;
- Clinical Performance – The ability of a device to yield results that are correlated with a particular clinical condition or a physiological state in accordance with a target population and intended user. Interventional clinical performance is where the test results may influence patient management decisions and/or may be used to guide treatment.

In the next section we describe the way the regulatory system is organised in the EU and in the USA. This is the extended version of the concise regulatory requirements presented in Chapter 2 in combination with a diagram. To aid the reader and provide an overview, the diagram is reproduced here (see Figure III.1).

Regulations for market access of medical devices

We begin by comparing the EC regulatory system to that of the USA and then discuss international harmonisation activities.

EU regulatory system

According to the Medical Devices Directives, the government of each Member State is required to appoint a **Competent Authority** (in the Netherlands: the Ministry of Health, Welfare and Sport and/or the Health Care Inspectorate/IGZ) responsible for medical devices. The Competent Authority (CA) acts on behalf of the government of the Member State to ensure that the requirements of the Medical Device Directives are transposed into National Law and are applied.

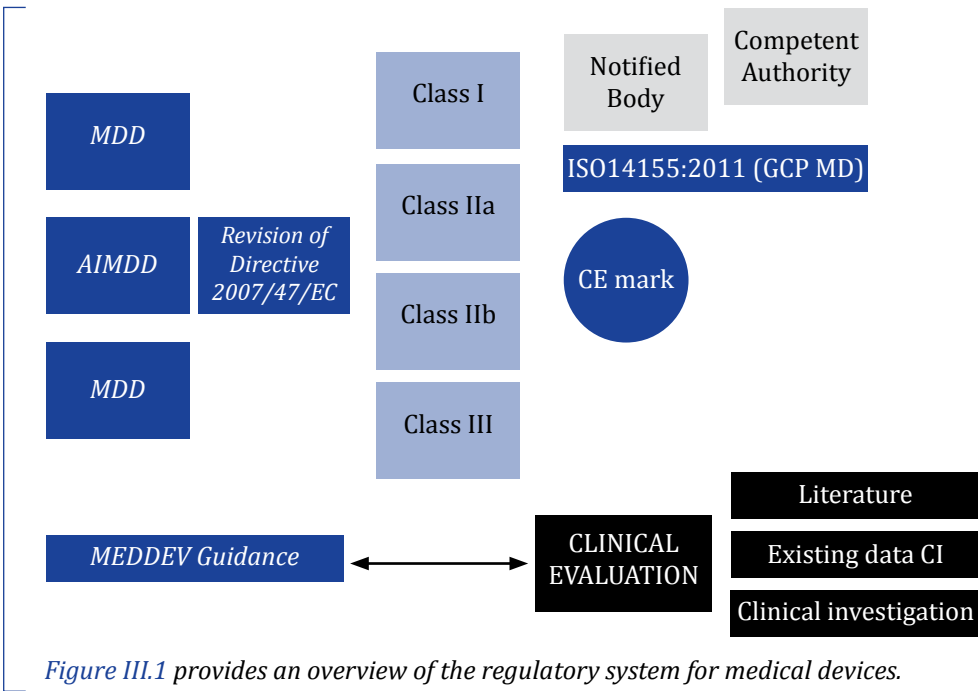


Figure III.1 provides an overview of the regulatory system for medical devices.

The authorisation of medical devices (i.e. market access) is guaranteed by a **Declaration of Conformity**. This declaration is issued by the manufacturer itself, but for products in Class IIa, IIb or III, it must be verified by a **Certificate of Conformity** issued by a **Notified Body** (in the Netherlands: DEKRA, formerly KEMA). A Notified Body is a public or private organisation that has been certified to validate the device’s compliance with the European Directive. Medical devices assigned to class I (provided that they do not need to be sterilised or are not used to measure a function) can be put on the market after self-certification. Manufacturers can choose any Notified Body for Conformity assessment.

When the Notified Body has approved the Declaration of Conformity, the manufacturer is allowed to identify the medical devices with a **CE (Conformité Européenne) mark**. CE marking indicates that the product meets the essential requirements of the applicable EU directives (such as safety, health, environmental protection requirements). CE marking is compulsory for all medical devices and to sell or use a device without it is an economic offence, but CE marking is not compulsory for a device that is still subject to clinical research, or for a personally customised product. For a new device subject to clinical research, an investigational medical device dossier (IMDD, available from the CCMO website) can be given to the medical ethical committees for approval. However, in the Netherlands the medical ethical committees prefer a device to have a CE marking before they approve clinical research, and the technical instrumentation services at hospitals also play an active role in this case. Use of a device beyond its intended purpose is considered off-label use.

By and large, then, CE marking refers to the technical quality and safety of the product, and compliance with the directives, including risk analysis. Medical devices must not only be safe, however, but also function both medically and technically as the manufacturer ‘intended’. This is regulated through the ISO system. Of particular interest is the recent NEN-EN-ISO 14155:2011 international standard, which provides a ‘Good Clinical Practice for medical devices for human use’. ISO 14155:2011 specifies general requirements intended to ‘protect the rights, safety and well-being of human subjects, ensures the scientific conduct of the clinical investigation and the credibility of the results, define the responsibilities of the sponsor and principal investigator, and assist sponsors, investigators, ethics committees, regulatory authorities and other bodies involved in the conformity assessment of medical devices’, in essentially the same way as the ICH GCP applies to medicinal products. ISO 14155:2011 does not apply to *in vitro* diagnostic medical devices. This ISO standard describes the design of clinical investigations (A6) and statistical considerations (A7) in largely general terms.

Clinical evaluation

The Medical Devices directives, as well as the Guidance documents³ that accompany the directives, describe the processes and essential requirements in terms of safety and **clinical evaluation** in some detail. One of the essential requirements in the EU is to perform a clinical evaluation (see Medical Device Directive Art. 15 and Annex X). Performing a clinical evaluation basically means looking at all the available clinical data and assessing the safety and performance. Clinical data are derived from:

- a critical evaluation of the relevant scientific *literature* currently available and relating to the safety, performance, design characteristics and intended purpose of the device;
- or a critical evaluation of the results of all clinical investigations made;

³ e.g. MEDDEV 2.7/4 Guideline on Clinical Investigation: A Guide for Manufacturers and Notified Bodies.

- or a critical evaluation of the combined clinical data provided.

The clinical evaluation and its documentation must be actively updated by data obtained from the postmarketing surveillance. The update should include not only data on the manufacturer's own medical device but also data on other similar medical devices (regardless of their manufacturer). A clinical evaluation report describes the data sources, assesses their suitability, and presents an evaluation and conclusions.

When the available data are not sufficient to draw valid conclusions about the device's safety and performance, additional **clinical investigations are required**. The objectives of such clinical investigations are:

- to verify that, under normal conditions of use, the performance of the devices conform to those intended by the manufacturer, and
- to determine any undesirable side-effects, under normal conditions of use, and assess whether they constitute risks when weighed against the intended performance of the device.

Studies must be conducted in accordance with Article 15, Annex VIII and Annex X of the EU MDD. Annex VIII describes what should be submitted to the Competent Authority before the start of the study. It also describes what documentation is needed for inspection purposes. Pre-approval from a CA is required.

Guidance from the Medicines and Healthcare Regulatory Agency (MHRA, UK) states that clinical investigations should probably be carried out if: 1) it is a Class III device, or an implantable device; 2) it is a new design or new application; 3) it contains new materials; 4) the device is modified in such a way that it is potentially significantly affecting clinical safety and performance; and 5) sufficient data do not exist. A clinical investigation is probably not needed (according to the MHRA) if: 1) it is a low-risk device; 2) sufficient data exist; and 3) *it is an equivalent product*.

The medical devices regulatory process relies heavily on the concept of **equivalence**, which refers to the fact that data from equivalent devices can be used to support the safety and/or performance of the device in question. In order to be equivalent, devices should have the same intended use and need to be comparable with respect to their technical and biological characteristics. These characteristics should be similar to such an extent that there would be no clinically significant difference in the devices' performance and safety. The nine criteria for equivalence depend on:

The *intended use* relating to:

- the clinical condition being treated,
- the severity and stage of disease,
- the site of application to/in the body and
- the patient population;

The *technical characteristics* relate to:

- the design, specifications, physiochemical properties including energy intensity,
- deployment methods,

- critical performance requirements,
- principles of operation and conditions of use.

Biological characteristics relate to biocompatibility of materials in contact with the same body fluids/tissues.

It is the responsibility of the Notified Bodies to judge whether equivalence has been sufficiently shown by the manufacturer. Based on supposed equivalence, a manufacturer can build a case for something new on existing data and literature. It is possible to show equivalence for some of the criteria; in that case additional investigations are only required for the non-equivalent elements of the device.

To summarise, the critical issue is that Good Clinical Practice for medical devices is not supported by a specific process description of how to collect new clinical data for every new product. Altogether, the regulatory authorities have neither a basic rationale or conceptual framework nor methodological requirements for the clinical data needed to provide sufficient evidence of added clinical benefit. This is because the range of device products is much larger than in the pharmaceuticals sector and because the purpose and role of a device are often low risk, which requires less clinical evidence because manufacturers are allowed to rely on equivalence. But this is case-dependent and the more high-risk a device is, the more clinical investigations are required. In addition, when a new device enters the market and has a profile similar to another, less expensive, device meant for a similar purpose, the more comparative effectiveness studies are needed. In the case of medical devices, the critical judgment is in the hands of the Notified Bodies, which often take a more technical view than the more clinically based European Medicines Agency appraisal. The Notified Bodies are under pressure because they are uneven in quality.

US regulatory system

Although there are many similarities between the regulatory process in the United States and that in the European Union, there are some important differences that impact the time and cost associated with the introduction of a new medical device. This is particularly true with regard to the organisation involved in approval, the criteria for approval, and the local institutional review board's position and negotiation options. In the USA, the definition and classification of medical devices is slightly different than in the EU, although Class III devices are the highest-risk ones in the USA as well. In addition the USA has a different procedure (see below). The most notable difference is that the Center for Devices and Radiological Health (CDRH), part of the Food and Drug Administration (FDA), is responsible for granting market access instead of non-governmental notified bodies regulating the approval and post-approval process.

For a *new* medical device, first clinical use is perceived as a key milestone.

1. Clinical testing of an unapproved significant-risk medical device requires FDA approval in the form of an *Investigational Device Exemption (IDE)*. The IDE application provides the FDA with information on device design and qualification, as well as on the study protocol. An IDE may also be required for studies in which an approved device is used for a purpose distinct from its approved indication. This is typically the case when a trial is sponsored by a company for the purpose of expanding a device's indication or making significant changes in the instructions for use.
2. Following FDA approval to initiate clinical studies, subsequent review is required by the *institutional review board (IRB)* at the clinical site.
3. The regulatory approval process by the FDA's CDRH is as follows for the different risk categories. Note that risk classification in the USA is not exactly the same as in the EC:
 - 3a. Class I (lowest risk) devices are subject to *general controls*, which are published standards pertaining to labelling, manufacturing, postmarketing surveillance, and reporting. Formal FDA review is not required for most class I devices before their market introduction. This correlates with the EU self-certification process.
 - 3b. Class II devices are those higher-risk devices for which general controls alone are insufficient to provide reasonable assurance of safety and effectiveness. They require *special controls* that may include performance standards, design controls, and postmarketing surveillance programmes. In addition, most class II devices require *FDA clearance of a premarket notification application (PMA or 510[k])* before the device may be marketed. In the 510(k) application, the medical device manufacturer must provide data to demonstrate that the new device is 'substantially equivalent' to a legally marketed device. Although substantial equivalence can usually be demonstrated on the basis of bench and animal testing alone, approximately 10% of 510(k) applications include clinical data. In other words, the concept of equivalence is a dominant principle in the USA as well.
 - 3c. Class III devices (such as heart valves and coronary artery stents) are judged to pose the highest potential risk. Most class III devices require *FDA approval of a PMA* before they can be legally marketed. Approval of the PMA generally requires clinical data demonstrating reasonable assurance that the device is safe and effective in the target population.

The PMA process typically involves a series of studies starting with first clinical use and culminating in a multicentre, prospective (randomised) control trial (pivotal trial). The complexity and extent of the clinical testing programme is dictated by the nature of the device and its proposed use. The clinical study programme is developed by the company in conjunction with clinician investigators, all in close collaboration with FDA/CDRH. Even when the FDA has authorised early US clinical trials, clinical

testing must pass significant additional hurdles at each clinical site in terms of IRB and contract approvals. This difference has a profound impact on the size and scope of the clinical studies for regulatory approval.⁴ When few data on existing standards are available, the FDA typically requires randomised rather than single-arm studies, in which the new device is compared against concurrent controls treated with current best medical practice. That comparison may be powered to show that the new treatment is superior to prior approaches, or that it is non-inferior (equivalent or better) compared with a previously approved treatment/device.

For new high-risk devices, US regulatory requirements are more extensive and require more time and resources than those of other countries. It is estimated that obtaining FDA approval to initiate clinical studies in the United States adds 3 to 6 months to the process of device development. In addition, subsequent review by the institutional review board (IRB) at the clinical site can add an additional 3 to 6 months to this timeline. Because of these factors, most initial clinical device testing has shifted to outside the United States. It is estimated that more than 75% of first clinical use of cardiovascular device testing now takes place outside the United States. However, high-profile new devices that require new clinical data for approval are the exception rather than the rule. Most devices currently in testing are similar to well-characterised approved devices. When FDA/CDRH has substantial data on the device class metrics, comparisons may be made to historical data or objective performance criteria. The vast majority of device clinical trials are case series that carefully document product performance.

To summarise, the USA is generally more stringent and the innovation climate in the EU is more favourable in the case of new and innovative devices, but for subsequent versions of previously approved devices, USA practice is less stringent. The 510K procedure is currently under review and the requirements for *in vitro* diagnostic tests (high risk in the USA, medium risk in the EU) have been circumvented by the application of 'lab developed tests', which do not require extensive clinical data.

Harmonisation

The reason that the EU and the USA (and the rest of the world, for that matter) differ with regard to medical devices is that the international harmonisation of legislation has not been concluded. One example of the lack of harmonisation is the ISO 14155 (2011),⁵ which has not been accepted by the FDA; it currently works with CFR 820 but may accept ISO 14155 in the near future. However, regulatory authorities and industry from all over the world joined forces in 1992 to respond to the growing need for

4 Medical Device Development: from Prototype to regulatory approval. Kaplan et al. (2004). *Circulation* 109: 3068-3072.

5 ISO 14155 (2011) Clinical investigation of medical devices for human subjects – Good Clinical Practice.

international harmonisation by creating the Global Harmonization Task Force (GHTF),⁶ which worked on definitions and concepts for medical devices. Since its inception, the GHTF has consisted of the representatives of five founding members (Australia, Japan, USA, EU, Canada) grouped into three geographical areas – Europe, Asia-Pacific and North America – each of which actively regulates medical devices using their own unique regulatory framework. Chairmanship of the GHTF is rotated among the regulatory representatives of the five founding members. Japan is the current chair.

The purpose of the GHTF is to encourage convergence in regulatory practices related to ensuring the safety, effectiveness/performance and quality of medical devices, promoting technological innovation and facilitating international trade. The primary way in which this purpose is accomplished is via the publication and dissemination of harmonised documents on basic regulatory practices. These documents, which are developed by five different GHTF Study Groups, provide a model for the regulation of medical devices that can then be adopted/implemented by national regulatory authorities. The GHTF study groups have delivered working documents. The Group on clinical safety and performance has delivered reports on Clinical evidence⁷. The other study groups have documented their agreements. All the guidance documents are discussed in the ASEAN harmonization working party to expand global coverage. GHTF is now in the process of liquidation. In collaboration with the WHO, a new forum is being created: the International Medical Devices Regulatory Forum (IMDRF),⁸ which started in 2011. IMDRF will further continue to develop the documents, to prepare for worldwide implementation in national law at a global scale. The medical device industry has achieved this harmonisation in almost 20 years.

Use in wider populations – Postmarketing surveillance

The realities of logistics, time, and resources restrict the size and duration of most new device trials to 800-1500 patients, thereby limiting the power of these trials to detect events with an occurrence rate of <1%. In addition, the most experienced physicians at medical centres with sufficient research infrastructure and patients to recruit usually carry out the pivotal trials. The need to extrapolate the results to the real-world post approval is equally or even more pressing than in the case of pharmaceuticals (because of user interference), since the degree of generalizability to a wider population is much smaller. This cannot be captured in randomised clinical trials (RCT), but requires follow-up studies. There is therefore a need for stricter postmarketing surveillance data that are accurate and timely, both in the US and Europe.

Several systems for obtaining high-quality data are available that provide a nuanced understanding of safety, performance and clinical benefit in a broad-based

6 <http://www.ghtf.org/>.

7 Key definitions and concepts, June 07 (N1); Clinical evaluation, June 07 (N2); Clinical investigations April 10 (N3).

8 <http://www.imdrf.org/>.

population. These are difficult and costly. In the US there is:

- The Medical Device reporting (MDR) system (the EU has a similar system). The MDR system relies mainly on manufacturers, importers and user facilities complying with mandatory reporting requirements. It is easy and relatively inexpensive for monitoring, but not suitable for detecting rare (but serious) adverse events. This is due to underreporting and inadequate characterisation of the patient population at risk.
- Claims-based data drawn from health insurance companies or other third parties that can combine admission and diagnosis-related groupings with adverse events. These databases may be useful but are not designed for postmarketing surveillance. Moreover, there is a substantial delay between reporting and the event, and coverage may be poor.
- Regional databases. These allow for follow-up of in-hospital complications, for example, but are not structured for following outcomes associated with a specific device.
- Multicentre registries (disease-oriented). They may examine practice patterns and the outcomes of procedures in specific medical disciplines. The data are limited in that these databases are not designed to track specific devices.
- Device-specific registries. They are often established by manufacturers as part of their postmarketing surveillance programmes. They typically contain larger amounts of data, but numbers frequently remain below the number (>5,000 or so) required to detect and quantify a 1 in 200 adverse event frequency with adequate statistical power.

New approaches would need to do at least the following:

- assign device-related complications, distinguishable from spontaneous events related to disease natural history;
- sample the entire population;
- establish a frequency threshold.

Requirements for reimbursement in The Netherlands

Unlike the authorisation process for medical devices in the EU, which all Member States (the Competent Authority and Notified Body) follow to a greater or lesser extent, the reimbursement procedures for medical devices vary significantly from country to country. The health care policy of a country largely determines whether and how a medical device is eligible for reimbursement. In the Netherlands the health care policy, including the rules and measures for compensation, is developed and controlled by the Ministry of Health, Welfare and Sport. The Dutch reimbursement policy for medical devices is based on diagnosis-related groups (DRGs). The system is called the Diagnosis Treatment Combination (DBC) and it is used for patient classification, cost containment and Health Technology Assessment (HTAs), which quantify the

cost-effectiveness of interventions (including devices). On the basis of established procedures (see below), the Health Insurance Board in The Netherlands (CvZ) assesses whether drugs, medical devices and other devices are part of the basic insurance package. Most medical devices are included (in part) in the basic health insurance package for specific indications. Some reimbursement instruments are available that are based on other than health insurance schemes, such as the Exceptional Medical Expenses Act (AWBZ) and the Social Support Act (WMO).

Medical care must meet at least two criteria: it must be care that ‘professionals tend to offer’ and the care must reflect ‘the state of science and practice’. For some types of care, additional conditions or limitations apply. Care must also be delivered in a way that professionals consider up to professional standards. CvZ usually determines whether this is the case on the basis of guidelines and professional standards. To determine whether care meets ‘state of science and practice’ standards, CvZ adheres to the principles of evidence-based medicine. This method involves the careful, explicit and judicious use of the best available evidence on the effectiveness, cost-effectiveness and safety of the care in question. CvZ’s premise is that there must be a ruling based on medical-scientific data that meet the highest possible standards for evidence. CvZ basically uses only published and peer-reviewed literature in its assessments.

In essence, care must comply with the principles of effectiveness, cost-effectiveness, necessity and feasibility. For the effectiveness principle, CvZ makes a distinction between health-related and welfare-related tools. CvZ has made this distinction to clarify the level of proof required. The *Assessment Framework Tools* report (for out-patient medical devices) is especially elaborating on the principles of necessity and effectiveness.

APPENDIX IV

OVERVIEW OF CLINICAL STUDY APPROACHES FOR MEDICAL DEVICES

The huge variety of medical devices available calls for a flexible but integrated approach to accumulating evidence that a device is safe and will produce relevant benefits for health or health care. There are numerous study approaches available for building a portfolio of evidence for a device. One key difference between these approaches is whether the study will provide 1. direct evidence; 2. indirect quantitative evidence (e.g. short-term outcomes, intermediate/surrogate outcomes, accuracy measures) or 3. indirect qualitative evidence (e.g. results from different patient population or when the device tested differs from the device of interest). Whether a direct or indirect (linked) evidence approach is used, it is still possible to choose from a wide range of different study types or designs that can be used in each of these three research approaches. Each of these study types has its pros and cons (e.g. risks of bias).

In this appendix we describe traditional and novel study types and designs – *everything from traditional randomised studies to novel, more tailored randomised studies, to semi-randomised and non-randomised designs* – which can all be used in medical device evaluations that are aimed at providing evidence of the benefits of a device (or its use) for health or health care. The advantages and drawbacks of each study type are given, and they are illustrated by examples of medical device evaluations taken from many different areas of medicine and many different types of devices. For background literature, see the references cited in the main text, which are complemented by some additional references (without being exhaustive) where deemed necessary.

A. Randomised (direct evidence) study approaches⁹

Randomisation ensures that the groups being compared – one in which the device is used and one (or more) in which not the device but some alternative, comparative management (including placebo or sham management) is used – are created by random or chance splitting. Provided they are large enough, the groups are ‘the same’

⁹ This section builds upon the extensive literature and books written on these topics, some of them referred to in the main text of chapter 4, and on the report *Alternatieven voor Randomized Clinical Trials in onderzoek naar de effectiviteit van interventies*. M. Langendam, L. Hooft, R. Scholten, P. Bossuyt. Dutch Cochrane Centre (DCC), Academisch Medisch Centrum, Amsterdam, februari 2013.

except for the device under study. All other modifying factors – the ‘confounders’ – are equally distributed across the comparison groups. Accordingly, any observed differences in benefits (and risks) between the groups can be assigned with a larger measure of likelihood to the actual difference in management, and thus to the device use; all other modifying (confounding) factors are ruled out due to the randomisation. Provided it is successful, randomisation allows researchers to draw conclusions about true cause-and-effect relationships between the device use and the health or other outcomes.

A randomised clinical trial (RCT) can have varying sample sizes, ranging from small trials to very large (mega) trials.¹⁰ Randomisation can be executed on the individual or patient level, where the usual aim is to compare the effects of device use on patient-related endpoints, intermediate or long term. It can also be executed at the practitioner or end-user level, where the endpoints are user-relevant, such as a difference in ease of use, efficiency, time-to-diagnosis, etc.

Device companies often struggle to fund and organise such trials. Randomised designs, especially the more novel and tailored approaches (see below under B), still can play a role by providing high-level evidence (low risk of bias) for the comparative safety, performance, benefits or effects, and even cost-effectiveness of devices.

A1. Pragmatic or comparative effectiveness randomised trials

The best randomised design for direct evaluation of a medical device’s benefits or added benefits is the parallel randomised pragmatic design. This applies both to therapeutic and test devices, i.e. diagnostic, monitoring, screening or prognostic tests (see Chapter 4.3). Targeted individuals or clusters of individuals are allocated randomly, concurrently and in parallel to either the new intervention or the comparative (control) strategy. A large-scale, long-term **pragmatic or comparative effectiveness randomised trial** compares the use of the device in question directly with the best alternative care in the right population, measuring all relevant outcomes over the long term and with use as it would be in everyday practice. Of course in the case of diagnostic, screening or monitoring devices, the therapeutic actions dictated by the results of such devices are also part of the intervention.

Such a design would produce the most *direct and valid evidence* of whether the device will indeed produce the intended relevant health benefits, at an acceptable level of safety, as compared to prevailing care. Even the cost-effectiveness of the device use can be addressed. This is true both for devices that interfere with bodily systems in order to treat or alleviate specific health conditions (therapeutic devices) and for test devices such as diagnostic, prognostic, monitoring or screening tests that rather generate information based on which therapeutic actions are administered (see Chapter 4.3, main text).

¹⁰ Charlton B.G., Mega-trials: methodological issues and clinical implications. *J R Coll Physicians Lond* 1995;29:96-100.

Pragmatic randomised trials are a specific and – for device evaluations – arguably the best form of a parallel RCT. The key feature of the pragmatic trial is that the comparison is not a placebo intervention (see below) but an alternative intervention, usually best current practice, with no restrictions on their application. Pragmatic trials are designed to evaluate the effectiveness of the index intervention (e.g. device use) under routine practice conditions, as opposed to placebo controlled trials, which test whether a specific intervention-aspect (in drugs this is referred to as the ‘pharmacological agent or substance’) is effective under optimal ‘experimental’ conditions. The pragmatic aspects of a trial may include, among others:

- the use of broad eligibility criteria to specifically select participants with heterogeneous characteristics, conform daily care;
- having flexible strategies for the use of the device;
- include a variety of practitioners with different expertise regarding the device use;
- include a variety of clinical settings;
- assess a variety of clinically meaningful or patient relevant outcomes.

Accordingly, the outcomes of pragmatic randomised trials are considered to have greater relevance for clinical practice and health policy makers.

The main disadvantage of pragmatic trials is that the participants and the practitioners are commonly not blinded, potentially obscuring the net effects of the index device. Their advantage, however, is that results are often directly applicable to daily care. The benefits (and risks) found in such trials are highly generalizable because pragmatic trials usually include participants who will be dealing with the intervention in the real world, recruit patients from a broader range of study sites (not only academic or highly experienced medical centres and professionals), and have outcomes that often represent a full range (not only short-term or intermediate outcomes) of relevant patient and professional health outcomes.

A2. Randomised placebo-controlled trials

A **placebo or sham-controlled parallel randomised design** is the most traditional randomised design, and comes from the pharmaceutical domain. The major advantage of this design is that all possible influences from any source – patient and professional interpretation influences and other confounding factor influences – are controlled for due to the double-blinded, placebo-controlled randomised character of the study. This design is much less straightforward for the evaluation of medical devices than it is for medicines, mainly because devices are part of complex interventions where, for example, double blinding, and thus a full placebo or sham-management control group can be extremely difficult. Moreover, traditional randomised placebo-controlled studies are often conducted in an environment staffed by highly skilled and trained operators and by high-volume medical centres, with very specific patient inclusion and exclusion criteria. As a result, these trials deliver evidence in an environment that can be quite different from real-world, pragmatic use.

In the medical device sector, the comparative strategy often involves alternative, prevailing management or even no management (wait-and-see), see under A1. Sham or placebo-device interventions as comparison exist but are rare in device evaluations. When they do take place, they are almost always for therapeutic devices, and not for diagnostic, screening or monitoring devices.

EXAMPLES OF PRAGMATIC TRIALS USED FOR MEDICAL DEVICE EVALUATION

Title: Clinical evaluation of silicone hydrogel lens wear with a new multipurpose disinfection care product.

Registration number: Not reported.

Source: MEDLINE (through PubMed)

Reference: Zigler L, Cedrone R, Evans D, Helbert-Green C, Shah T. Eye Contact Lens. 2007 Sep;33(5):236-43.

Abstract: **PURPOSE:** To evaluate subjective symptoms and clinical signs in silicone hydrogel contact lens wearers with three different multipurpose solution (MPS) lens care regimens. **METHODS:** In a double-masked, randomized, concurrently controlled study, 233 subjects from 12 clinical sites wore one of two silicone hydrogel lens brands (ACUVUE Advance or Focus NIGHT & DAY) for 1 month on a daily-wear basis supported by a new reconditioning multipurpose disinfecting solution (MPDS) preserved with POLYQUAD and ALDOX, regimen 1 (OPTI-FREE RepleniSH Multi-Purpose Disinfecting Solution), or by one of two MPSs preserved with polyhexamethyl biguanide, regimen 2 (ReNu MultiPlus Multi-Purpose Solution No Rub Formula) or regimen 3 (Complete MoisturePLUS Multi-Purpose Solution). **RESULTS:** Significant differences in favor of regimen 1 were found in subjective responses of subjects wearing ACUVUE Advance lenses. For Focus NIGHT & DAY lens wearers, regimen 1 was associated with significantly less corneal staining (severity [P=0.0019], area [P=0.0077]) than regimen 2 was. The average number of times per day that rewetting drops were used was significantly higher for subjects randomized to regimen 3 than for subjects using regimen 1. **CONCLUSIONS:** The clinical performance of the new MPDS product with silicone hydrogel lenses was generally as good as or better than the two comparative polyhexamethyl biguanide-preserved MPSs. Clinical differences were evident between the products. Practitioners should be aware that MPS product choice for use with silicone hydrogel lenses may lead to different clinical outcomes, particularly in regard to stress on the ocular surface, as evidenced by the corneal staining response.

Title: Safety and efficacy of gravitational shunt valves in patients with idiopathic normal pressure hydrocephalus: a pragmatic, randomised, open label, multicentre trial (SVASONA).

Registration number: ISRCTN51046698

Source: MEDLINE (through PubMed)

Reference: Lemcke J, Meier U, Müller C, Fritsch MJ, Kehler U, Langer N, Kiefer M, Eymann R, Schuhmann MU, Speil A, Weber F, Remenez V, Rohde V, Ludwig HC, Stengel D. *J Neurol Neurosurg Psychiatry*. 2013 Aug;84(8):850-7. doi: 10.1136/jnnp-2012-303936. Epub

2013 Mar 1.

Abstract: OBJECTIVES: To investigate whether gravitational valves reduce the risk of overdrainage complications compared with programmable valves in ventriculoperitoneal (VP) shunt surgery for idiopathic normal pressure hydrocephalus (iNPH). BACKGROUND: Patients with iNPH may benefit from VP shunting but are prone to overdrainage complications during posture changes. Gravitational valves with tantalum balls are considered to reduce the risk of overdrainage but their clinical effectiveness is unclear. METHODS: We conducted a pragmatic, randomised, multicentre trial comparing gravitational with non-gravitational programmable valves in patients with iNPH eligible for VP shunting. The primary endpoint was any clinical or radiological sign (headache, nausea, vomiting, subdural effusion or slit ventricle) of overdrainage 6 months after randomisation. We also assessed disease specific instruments (Black and Kiefer Scale) and Physical and Mental Component Scores of the Short Form 12 (SF-12) generic health questionnaire. RESULTS: We enrolled 145 patients (mean (SD) age 71.9 (6.9) years), 137 of whom were available for endpoint analysis. After 6 months, 29 patients in the standard and five patients in the gravitational shunt group developed overdrainage (risk difference -36%, 95% CI -49% to -23%; $p < 0.001$). This difference exceeded predetermined stopping rules and resulted in premature discontinuation of patient recruitment. Disease specific outcome scales did not differ between the groups although there was a significant advantage of the gravitational device in the SF-12 Mental Component Scores at the 6 and 12 month visits. CONCLUSIONS: Implanting a gravitational rather than another type of valve will avoid one additional overdrainage complication in about every third patient undergoing VP shunting for iNPH.

EXAMPLES OF PLACEBO-CONTROLLED RCTS USED FOR MEDICAL DEVICE EVALUATION

Title: A novel nasal expiratory positive airway pressure (EPAP) device for the treatment of obstructive sleep apnea: a randomized controlled trial.

Registration number: NCT00772044

Source: MEDLINE (through PubMed)

Reference: Berry RB, Kryger MH, Massie CA. Sleep. 2011 Apr 1;34(4):479-85.

Abstract: STUDY OBJECTIVES: Investigate the efficacy of a novel nasal expiratory positive airway pressure (EPAP) device as a treatment for obstructive sleep apnea (OSA). DESIGN: A prospective, multicenter, sham-controlled, parallel-group, randomized, double-blind clinical trial. SETTING: 19 sites including both academic and private sleep disorder centers. PATIENTS: Obstructive sleep apnea with a pre-study AHI ≥ 10 /hour. INTERVENTIONS: Treatment with a nasal EPAP device (N=127) or similar appearing sham device (N=123) for 3 months. Polysomnography (PSG) was performed on 2 non-consecutive nights (random order: device-on, device-off) at week 1 and after 3 months of treatment. Analysis of an intention to treat group (ITT) (patients completing week 1 PSGs) (EPAP N=119, sham N=110) was performed. MEASUREMENTS AND RESULTS: At week 1, the median AHI value (device-on versus device-off) was significantly lower with EPAP (5.0 versus 13.8 events/h, $P < 0.0001$) but not sham (11.6 versus 11.1 events/h, $P = \text{NS}$); the decrease in the AHI (median) was greater (-52.7% vs. -7.3%, $P < 0.0001$) for the ITT group.

At month 3, the percentage decrease in the AHI was 42.7% (EPAP) and 10.1% (sham), $P < 0.0001$. Over 3 months of EPAP treatment the Epworth Sleepiness Scale decreased (9.9 ± 4.7 to 7.2 ± 4.2 , $P < 0.0001$), and the median percentage of reported nights used (entire night) was 88.2%. **CONCLUSIONS:** The nasal EPAP device significantly reduced the AHI and improved subjective daytime sleepiness compared to the sham treatment in patients with mild to severe OSA with excellent adherence.

Title: Evaluating the Effect of Tooth Cleaning Devices on Oral Health

Registration number: NCT01250769

Source: ClinicalTrials.gov

Reference: Not published (yet).

Protocol: This is a study to evaluate the safety and the efficacy of tooth and interproximal cleaning modalities on oral health. Study Design: Allocation: Randomized; Endpoint Classification: Safety/Efficacy Study; Intervention Model: Parallel Assignment; Masking: Single Blind (Outcomes Assessor); Primary Purpose: Treatment. Condition: Dental Plaque. Intervention: Device: Manual Toothbrush; Device: Interproximal Cleaning Device. Primary Outcome Measures: Residual Protein Concentration [Time Frame: 14 days] [Designated as safety issue: No], Residual protein concentration of interproximal plaque samples. Secondary Outcome Measures: Residual Protein Concentration [Time Frame: 28 days] [Designated as safety issue: No], Residual protein concentration of interproximal plaque samples, Modified Gingival Index [Time Frame: 14 days] [Designated as safety issue: No], Gingival inflammation evaluation on an ordinal scale of 0 to 4 (0 is best; 4 is worst), Modified Gingival Index [Time Frame: 28 days] [Designated as safety issue: No], Gingival inflammation evaluation on an ordinal scale of 0 to 4 (0 is best; 4 is worst), Gingival Bleeding Index [Time Frame: 14 days] [Designated as safety issue: No], Gingival bleeding evaluation using an ordinal scale of 0 to 3; (0 is best; 3 is worst), Gingival Bleeding Index [Time Frame: 28 days] [Designated as safety issue: No], Gingival bleeding evaluation using a ordinal scale of 0 to 3; (0 is best; 3 is worst).

Enrollment: 170 patients

Study Start Date: April 2010

Study Completion Date: June 2010 (Final data collection date for primary outcome measure)

A3. Cross-over trials

In addition to RCTs involving parallel groups, pragmatic or placebo-controlled, there are other approaches in which subjects can serve as self-controls. In a randomised controlled **paired or cross-over design** (in contrast with parallel comparison), participants still undergo randomisation but also serve as their own control. They can retrieve the device (e.g. nerve stimulator) and the alternative treatment simultaneously, with the new intervention being assigned to, for example, the left leg or arm and the control treatment to the right leg or arm. In a cross-over design, the participants receive both interventions, although consecutively in a randomly assigned order. This

design is possible when the target disease/condition is chronic and regresses – in due time – to the original baseline status when the first assigned intervention ceases. Cross-over trials are also recommended when the new device and control intervention have relatively local effects that do not overlap after a ‘washout period’.

Paired or cross-over approaches have two major advantages over parallel designs. First, they are less susceptible to imbalances between other modifying factors (confounders) because each patient serves as his or her own control. Second, these trials are very efficient and require smaller sample sizes because they produce within-participant comparisons (whereas parallel designs produce between-participants comparisons).¹¹

Cross-over trials are sometimes infeasible, however, and they have their disadvantages. As we said, this approach is generally performed in patients with chronic or incurable diseases. For curative treatments or rapidly changing (instable) conditions, cross-over trials may be infeasible or unethical. The effects of both interventions should also have rapid onset, be of short duration and be administered after a sufficiently long ‘washout period’ to avoid ‘carry-over effects’ between the two intervention periods. In practice, however, planning a sufficiently long wash-out period requires expert knowledge of the treatment dynamics, which are often unknown. Related to this problem is the issue of ‘order effects’, because the order in which treatments are administered may affect the endpoints under study.

EXAMPLE OF A CROSS OVER TRIAL USED FOR MEDICAL DEVICE EVALUATION

Title: A comparison of the Supreme laryngeal mask airway with the Proseal laryngeal mask airway in anesthetized paralyzed adult patients: a randomized crossover study.

Registration number: Not reported.

Source: MEDLINE (through PubMed)

Reference: Tham HM, Tan SM, Woon KL, Zhao YD. *Can J Anaesth.* 2010 Jul;57(7):672-8. doi: 10.1007/s12630-010-9312-6. Epub 2010 Apr 22.

Abstract: **PURPOSE:** The Supreme laryngeal mask airway (SLMA) is a new single-use advanced form of the Proseal laryngeal mask airway (PLMA). This study tested the hypothesis that the SLMA is equally as effective as the PLMA as a supraglottic ventilatory device in anesthetized paralyzed adult patients. **METHODS:** Size 4 SLMAs and PLMAs were compared in a randomized crossover study involving 60 patients aged 21-75 yr and American Society of Anesthesiologists physical status I and II. Once the patients were anesthetized and paralyzed, the SLMA and the PLMA were inserted into each patient in random order. The primary outcome measure was the laryngeal seal pressure (LSP) at an intracuff pressure of 60 cm H₂O. Secondary outcome measures included the ease of inserting the laryngeal mask airway devices (LMADs) and the fiberoptic position of the airway tube. **RESULTS:** There was no statistically significant difference in LSP between the

11 Louis T.A., Lavori P.W., Bailar J.C. III, Polansky M. ‘Crossover and self-controlled designs in clinical research’. In: Bailar J.C. III, Mosteller F, eds. *Medical uses of statistics*, 2nd edn. Boston, MA: New England Medical Journal Publications, 1992:83-104.

SLMA and the PLMA. The mean LSP was 19.6 +/- 5.8 cm H(2)O and 20.9 +/- 6.7 cm H(2)O for the SLMA and the PLMA, respectively. There was a similarity between the SLMA and the PLMA regarding the number of attempts required and the duration for insertion. However, fiberoptic positioning was better with the PLMA than with the SLMA ($P < 0.0001$). CONCLUSION: The clinical performance of the SLMA as a ventilatory device is comparable with that of the PLMA, as illustrated by the similar LSPs. The inferior position of the SLMA airway tube compared with that of the PLMA does not affect its ease of ventilation.

A4. Adaptive trials

This novel approach is defined as ‘a clinical study design that uses accumulating data to decide how to modify aspects of the study as it continues, without undermining the validity and integrity of the trial.’¹² The randomisation ratio is changed during the course of the trial; these changes should not be ad hoc but by design. In general, there are two main approaches: monitoring the balance of baseline covariates in the randomisation during the continuing trial; and a response-adaptive treatment allocation. An example of the latter is that, based on observed benefits (or absence of benefits) in particular subgroups in the first series of randomised subjects, one continues the randomisation (or discontinues the trial) in those subgroups only. Another version of an adaptive design is a *sequential trial*. In this approach, the number of participants is not pre-specified. Instead, participants are recruited until intervention differences are observed (or can be dismissed). These trials are designed with the idea that the accumulated evidence at – pre-specified – interim analyses is sufficient to draw appropriate inferences about the benefits (and risks) of the device. Particularly in these adaptive trials, the use of *Bayesian methods* has been on the rise; that has also been true of clinical trials of medical devices over the past decade, as it may offer many advantages when it comes to evaluating more invasive, implantable medical devices.¹³ In Bayesian approaches, estimates of the benefits (and risks) are made based on prior information, which is complemented by and weighted against empirical observations taken from new studies. Accordingly, inferences about the benefits and risks of a device are not based solely on the empirical studies but on the weighted evidence of previous knowledge and new evidence. Prior information can come from previous studies or informative registries.

The major advantage of adaptive trial designs is that they can suffice with considerably smaller sample sizes and are very efficient compared to typical randomised trials

12 U.S. Food and Drug Administration: Draft Guidance for Industry: adaptive design clinical trials for drugs and biologics. <http://www.fda.gov/downloads/DrugsGuidanceComplianceRegulatoryInformation/Guidances/UCM201790.pdf>.

13 Bonangelino P. et al., ‘Bayesian Approaches in Medical Device Clinical Trials: A Discussion with Examples in the Regulatory Setting’. *Journal of Biopharmaceutical Statistics* 2011; 21:5, 938-953; Pibouleau L, Chevret S. Bayesian statistical method was underused despite its advantages in the assessment of implantable medical devices. *Journal of Clinical Epidemiology* 64 (2011) 270e279.

with fixed numbers of participants, thereby shortening the device evaluation process. As for Bayesian modelling, the combined use of existing data could further improve the efficiency of the evaluation process by potentially decreasing costs and shortening the timelines for device evaluation. Adaptive designs also have greater flexibility. Another advantage is the possibility of recalculating and adjusting the required sample sizes, with more real data (beyond assumptions) on differences in benefits (and risks) and variation between the groups while the trial is ongoing. Finally, there is the built-in possibility of stopping a trial prematurely in the presence of unexpected risks, or even in the clear presence (or absence) of major benefits.

A disadvantage of these designs is that not all outcomes or devices lend themselves to early discontinuation. Interim analyses may not be feasible when the primary endpoint cannot be measured relatively soon after start of the trial. Trials may also be discontinued prematurely in error. Moreover, sufficient experience using a superior device may be lacking if the trial is discontinued prematurely in the presence of clear benefits and the device is subsequently introduced to the medical profession. Finally, this design requires trained statisticians with specialist knowledge to conduct the analyses and interpret the results.

EXAMPLES OF ADAPTIVE TRIAL DESIGNS USED FOR MEDICAL DEVICE EVALUATION

Title: A Single Centre Study in Healthy Volunteers to Optimise the Rotacap Formulation and ROTAHALER Device for Delivery of Fluticasone Propionate/Salmeterol

Registration number: NCT01540708

Source: ClinicalTrials.gov

Reference: Not published (yet)

Protocol: The purpose of this study is to optimise the device and/or formulation of the Fluticasone propionate (FP)/salmeterol (SALM) unit dose powder inhaler (Rotahaler) to achieve drug delivery characteristics comparable to the Fluticasone propionate /salmeterol DISKUS inhaler. The indication is asthma and chronic obstructive pulmonary disease. The study is an open-label, randomised, cross-over, single centre study in healthy volunteers and will be conducted in a maximum of 3 parts, A, B and C. The design is adaptive and pharmacokinetic (PK) data analysis follows each part to enable a decision on whether progression to the subsequent parts is required. Part A of the study will test an alternative version of the Rotahaler with a low airflow resistance. The study will then test one or more of the following options depending on the outcome of part A. If progressed, part B will test modified Rotacap formulations including: (1) modified blend formulation, (2) reduced capsule fill weights, (3) different capsule types. Part B will also test other versions of the Rotahaler with intermediate airflow resistance. Part C will test the lower strength FP/salmeterol (100/50 mcg or lower) and/or a new unit dose DPI device (BUDI). A total of 36 subjects will be enrolled in each part to ensure 32 complete. In each cross-over arm, subjects will be administered 7 doses (3.5 days bid) with PK sampling following administration of the 7th dose. A three-day minimum wash-out period will separate each cross-over arm.

Enrolment: 36

Start date: January 2012

Study completion date: September 2012

Title: Prospective evaluation of elastic restraint to lessen the effects of heart failure (PEERLESS-HF) trial.

Reference: Costanzo MR, Ivanhoe RJ, Kao A, Anand IS, Bank A, Boehmer J, Demarco T, Hergert CM, Holcomb RG, Maybaum S, Sun B, Vassiliades TA Jr, Rayburn BK, Abraham WT. *J Card Fail.* 2012 Jun;18(6):446-58.

Registration number: Not reported

Source: MEDLINE (through PubMed)

Abstract: BACKGROUND: Left ventricular (LV) remodeling predicts poor outcomes in heart failure (HF) patients. The HeartNet(®) cardiac restraint device (Paracor Medical Inc., Sunnyvale, CA) may reduce LV remodeling and improve functional capacity, quality of life, and outcomes in HF patients. To evaluate the safety and efficacy of the HeartNet Ventricular Support System in HF patients receiving optimal medical therapy. METHODS AND RESULTS: Prospective, randomized, controlled, multicenter trial in patients with symptomatic HF and LV ejection fraction $\leq 35\%$ on optimal medical and device therapy. The primary efficacy end points were changes in peak VO(2), 6-minute walk (6MW) distance, and Minnesota Living with Heart Failure (MLWHF) quality of life score at 6 months. The primary safety end point was all-cause mortality at 12 months. Because the planned adaptive interim analysis of the first 122 subjects with a completed 6-month follow-up indicated futility to reach the peak VO(2) end point, trial enrollment was suspended. Hence, the results on the 96 treatment and 114 control subjects are reported. Groups were similar at baseline. At 6 months, responder frequency for a prespecified improvement was similar between groups for peak VO(2) ($P = .502$) and MLWHF score ($P = .184$) but borderline higher for improvement in 6MW distance in the treatment compared with the control group (33 [38%] vs. 25 [25%]; $P = .044$). At 6 months, the treatment group had a significantly greater improvement in Kansas City Cardiomyopathy Questionnaire (KCCQ) ($P < .001$) and decrease in LV mass ($P = .032$), LV end-diastolic diameter ($P = .015$), LV end-systolic diameter ($P = .032$), and LV end-diastolic volume ($P = .031$) as compared with controls. At 12 months, all-cause mortality and responder rates were similar in the 2 groups. Success rate for the HeartNet implantation was 99%. CONCLUSION: Enrollment in the trial was stopped because an interim analysis showed futility of reaching the peak VO(2) end point. However, because of the device safety and favorable signals for LV remodeling and quality of life, further investigation of this device is warranted.

TWO EXAMPLES OF SUCCESSFUL USE OF BAYESIAN METHODS IN DEVICE TRIALS ARE*:

TRANSCAN (<http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfTopic/pma/pma.cfm?num=p970033>). Prior information was used to incorporate results from previous

studies, resulting in a reduced sample size for demonstration of effectiveness.

INTERFIX (<http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfTopic/pma/pma.cfm?num=p970015>). An interim analysis was performed; based on Bayesian predictive modeling of the future success rate, the trial was stopped early. No prior information was used.

**FDA: Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials. Document issued on: February 5, 2010*

A5. Cluster trials

In cluster trials, aimed at quantifying the effects of device use on patient outcomes, randomisation takes place at a higher level; with practitioners, general practices, hospitals, villages or families being randomised to the index device and alternative intervention or management rather than individual participants.¹⁴

Motivation for performing a cluster trial may be to avoid contamination. This can happen – for example – when evaluating the effects of new diagnostic or monitoring devices on treatment decisions. The physician may learn from patients randomised to the index device group. Subsequent control patients, in whom the device is not used, may be treated differently based on experiences in similar patients from the index group. Another advantages of cluster-randomised trials over individually randomised trials include the ability to study interventions that cannot be randomly assigned to individuals, as discussed above.

The main disadvantage, compared with individually randomised trials, is that cluster randomised trials are more complex to design and require more complex analysis because, for example, two different units of measurement (the cluster and the patient) are used. In addition, this approach commonly requires a larger number of participants to obtain enough statistical power. Finally, selection bias can occur because informed patient consent is requested after randomisation of the clusters, such that specific patients may selectively refuse (or consent) to participate.

EXAMPLES OF CLUSTER TRIALS USED FOR THE EVALUATION OF MEDICAL DEVICES

Title: Integrated community case management of fever in children under five using rapid diagnostic tests and respiratory rate counting: a multi-country cluster randomized trial.

Registration number: NCT00720811.

Source: MEDLINE (through PubMed)

Reference: Mukanga D, Tiono AB, Anyorigiya T, Källander K, Konaté AT, Oduro AR, Tibenderana JK, Amenga-Etego L, Sirima SB, Cousens S, Barnish G, Pagnoni F. *Am J Trop Med Hyg.* 2012 Nov;87(5 Suppl):21-9.

Abstract: Evidence on the impact of using diagnostic tests in community case management

14 Puffer S. et al., 'Evidence for risk of bias in cluster randomised trials: review of recent trials published in three general medical journals'. *BMJ* 2003; 327: 785–9.

of febrile children is limited. This effectiveness trial conducted in Burkina Faso, Ghana, and Uganda, compared a diagnostic and treatment package for malaria and pneumonia with presumptive treatment with anti-malarial drugs; artemisinin combination therapy (ACT). We enrolled 4,216 febrile children between 4 and 59 months of age in 2009-2010. Compliance with the malaria rapid diagnostic test (RDT) results was high in the intervention arm across the three countries, with only 4.9% (17 of 344) of RDT-negative children prescribed an ACT. Antibiotic overuse was more common: 0.9% (4 of 446) in Uganda, 38.5% (114 of 296) in Burkina Faso, and 44.6% (197 of 442) in Ghana. Fever clearance was high in both intervention and control arms at both Day 3 (97.8% versus 96.9%, $P = 0.17$) and Day 7 (99.2% versus 98.8%, $P = 0.17$). The use of diagnostic tests limits overuse of ACTs. Its impact on antibiotic overuse and on fever clearance is uncertain.

Title: Field evaluation of permethrin long-lasting insecticide treated nets (Olyset®) for malaria control in an endemic area, southeast of Iran.

Registration number: Not reported.

Source: MEDLINE (through PubMed)

Reference: Soleimani-Ahmadi M, Vatandoost H, Shaeghi M, Raeisi A, Abedi F, Eshraghian MR, Madani A, Safari R, Oshaghi MA, Abtahi M, Hajjaran H. *Acta Trop.* 2012 Sep;123(3):146-53. Epub 2012 May 9.

Abstract: Long lasting insecticide treated nets (LLINs) have been advocated as an effective tool for prevention and control of malaria. Olyset net was the first LLINs which became commercially available and obtained WHO approval. According to the national strategic plan on evaluation of Olyset net, a field trial was conducted to determine the efficacy of these nets against malaria vectors in an endemic area in the southeast of Iran. Fourteen villages with similar topographical and epidemiological situations were selected and randomly assigned to two clusters of the study: Olyset net and untreated net. Distribution of nets was carried out to cover 100% of the population in Olyset net and untreated net cluster. Anopheline mosquitoes were collected monthly using different WHO standard methods in both areas to determine their abundance, feeding pattern and resting behaviour. Human blood index was determined using ELISA test. Additionally, Olyset nets were evaluated for their biological activity using WHO cone bioassay test by susceptible colony of *Anopheles stephensi* (Beech strain) and then for insecticide residues by employing high performance thin layer chromatography. Malaria incidence was measured by passive and active case detection from all study population. In total 2115 adult anopheline mosquitoes were collected and identified using morphological characters. They comprised of seven species: *Anopheles dthali* (Liston), *A. culicifacies* (Giles), *A. stephensi* (Liston), *A. superpictus* (Grassi), *A. fluviatilis* (James), *A. moghulensis* (Christophers) and *A. turkhudi* (Liston). *A. dthali*, *A. culicifacies* and *A. stephensi* were most prevalent species in both areas. In the Olyset net study area, there was a significant reduction of 41.1%, 54.4%, 59.39% and 64.1% in the indoor-resting density of *A. culicifacies*, *A. stephensi*, *A. dthali* and *A. superpictus*, respectively, with an overall reduction of 39.3% in total mosquitoes in

comparison with untreated net area. A significant reduction was also observed in human blood index of vector species in the Olyset net villages. Bioefficacy test results of Olyset nets showed that the median knockdown time was 1.48 and 3.25min, while the average mortality rate was 100% and 72.3%±7.07 in baseline and after 1 year of intervention, respectively. The average permethrin content reached to 68.31% (683.1mg/m²) of the initial insecticide dose of 937±21.69mg/m² (nearly 1000mg/m²) at the end of intervention. Malaria incidence was reduced by 96.6% and 64.8% in the village with Olyset nets and in the villages with untreated nets, respectively. During intervention period, there was a reduction of 93.2% in malaria incidence in Olyset net area as compared to the untreated area. This study indicated that Olyset nets have a major impact on malaria vectors and disease burden; therefore it could be recommended as an effective personal protection tool for malaria control in malarious areas.

A6. Factorial designs

In a factorial randomised design, each participant is randomly assigned to a group that receives a particular combination of interventions or non-interventions.¹⁵

The advantages of this approach over parallel comparison are its ability to efficiently investigate multiple interventions and detect interaction between different treatments. It also requires fewer patients and is less costly than conducting multiple parallel group trials.

The main disadvantage is that most actual trials are underpowered to detect significant treatment interactions because they are often designed on the assumption that interaction will be absent and, additionally, that interactions can only be determined at the end of the study. If such an effect is present, analysing the results by ignoring the other treatment assignments becomes invalid. It follows that one of the main advantages, i.e. the smaller required sample size, is also invalidated.

EXAMPLE OF A FACTORIAL DESIGN USED FOR THE EVALUATION OF MEDICAL DEVICES

Title: A randomized controlled trial comparing the effects of counseling and alarm device on HAART adherence and virologic outcomes.

Registration number: NCT00273780

Source: MEDLINE (through PubMed)

Reference: Chung MH, Richardson BA, Tapia K, Benki-Nugent S, Kiarie JN, Simoni JM, Overbaugh J, Attwa M, John-Stewart GC. PLoS Med. 2011 Mar;8(3):e1000422. doi: 10.1371/journal.pmed.1000422. Epub 2011 Mar 1.

Abstract: **BACKGROUND:** Behavioral interventions that promote adherence to antiretroviral medications may decrease HIV treatment failure. Antiretroviral treatment programs in sub-Saharan Africa confront increasing financial constraints to provide comprehensive

15 Daniel B. Efficient Designs: Factorial Randomized Trials. *J Bone Joint Surg Am*, 2012 Jul 18;94(Suppl 1(E)):34-38.

HIV care, which include adherence interventions. This study compared the impact of counseling and use of an alarm device on adherence and biological outcomes in a resource-limited setting. **METHODS AND FINDINGS:** A randomized controlled, factorial designed trial was conducted in Nairobi, Kenya. Antiretroviral-naïve individuals initiating free highly active antiretroviral therapy (HAART) in the form of fixed-dose combination pills (d4T, 3TC, and nevirapine) were randomized to one of four arms: counseling (three counseling sessions around HAART initiation), alarm (pocket electronic pill reminder carried for 6 months), counseling plus alarm, and neither counseling nor alarm. Participants were followed for 18 months after HAART initiation. Primary study endpoints included plasma HIV-1 RNA and CD4 count every 6 months, mortality, and adherence measured by monthly pill count. Between May 2006 and September 2008, 400 individuals were enrolled, 362 initiated HAART, and 310 completed follow-up. Participants who received counseling were 29% less likely to have monthly adherence <80% (hazard ratio [HR]=0.71; 95% confidence interval [CI] 0.49-1.01; p=0.055) and 59% less likely to experience viral failure (HIV-1 RNA \geq 5,000 copies/ml) (HR 0.41; 95% CI 0.21-0.81; p=0.01) compared to those who received no counseling. There was no significant impact of using an alarm on poor adherence (HR 0.93; 95% CI 0.65-1.32; p=0.7) or viral failure (HR 0.99; 95% CI 0.53-1.84; p=1.0) compared to those who did not use an alarm. Neither counseling nor alarm was significantly associated with mortality or rate of immune reconstitution.

CONCLUSIONS: Intensive early adherence counseling at HAART initiation resulted in sustained, significant impact on adherence and virologic treatment failure during 18-month follow-up, while use of an alarm device had no effect. As antiretroviral treatment clinics expand to meet an increasing demand for HIV care in sub-Saharan Africa, adherence counseling should be implemented to decrease the development of treatment failure and spread of resistant HIV.

A7. Preferences trials

Preferences trials are variations of RCTs in which participants' preferences – i.e. whether or not to receive the new device or the control treatment – are taken into account.¹⁶ The primary purpose is to minimise refusals at the recruitment stage owing to the reluctance of some participants to be randomised. The best-known preferences trial is probably the comprehensive cohort design. Here, all patients who meet the eligibility criteria can be recruited regardless of their informed consent to randomisation. This trial includes at least one group in which the (eligible, but non-randomised) participants are allowed to choose their own preferred treatment from several options offered. The second group is designed as a common parallel RCT, with participants being randomly allocated to either the treatment or the control.

Other well-known versions of preferences trials are the Zelen's design or the

16 Kao L.S. et al., Clinical Research Methodology I: Introduction to Randomized Trials. *J Am Coll Surg*. 2008 February; 206(2): 361–369. Silverman W.A., Altman D.G., Patients' preferences and randomised trials. *Lancet* 1996;347:171-4.

Wennberg's design. In the Zelen's design, participants are randomised to either the treatment (new device) or control group before giving informed consent, instead of prior to randomisation. In this design, participants that give their informed consent are certain of receiving the new device. Participants who do not consent (for other than privacy reasons) can receive the control treatment – often care as usual – in the clinical trial. In the Wennberg's design, participants are randomised to either a preference group or a randomisation group. Participants in the randomisation group are assigned to either the new device or the control treatment, and patients in the preference group are offered their treatment of choice.

The main advantage of these designs is that eligible patients who normally would refuse to participate in trials, either because they do not want to be randomised to placebo or because they have strong preferences for a particular intervention, will now participate in clinical research. Often patients in a randomised trial only represent a small proportion of the patients eligible for the trial in the real world. This design allows the applicability of trial results to be assessed by comparing the RCT group with the cohort of patients who met the eligibility criteria but did not consent to randomisation. On average, the outcomes of the patients who normally would not participate in randomised trials and those who participate and have strong preferences may differ.

A disadvantage of this design is that the trials must be open, thereby introducing the possibility of a 'placebo' effect (performance bias), and that the statistical power of the study may be affected if a high proportion of participants choose to receive the same treatment.

EXAMPLES OF PREFERENCE TRIALS USED FOR THE EVALUATION OF MEDICAL DEVICES

Title: Effectiveness of nurse delivered endoscopy: findings from randomised multi-institution nurse endoscopy trial (MINuET).

Reference: Williams J, Russell I, Durai D, Cheung WY, Farrin A, Bloor K, Coulton S, Richardson G. *BMJ*. 2009 Feb 10;338:b231. doi: 10.1136/bmj.b231.

Registration number: ISRCTN82765705.

Source: MEDLINE (through PubMed)

Abstract: **OBJECTIVE:** To compare the clinical effectiveness of doctors and nurses in undertaking upper and lower gastrointestinal endoscopy. **DESIGN:** Pragmatic trial with Zelen's randomisation before consent to minimise distortion of existing practice. **SETTING:** 23 hospitals in the United Kingdom. In six hospitals, nurses undertook both upper and lower gastrointestinal endoscopy, yielding a total of 29 centres. **PARTICIPANTS:** 67 doctors and 30 nurses. Of 4964 potentially eligible patients, we randomised 4128 (83%) and recruited 1888 (38%) from July 2002 to June 2003. **INTERVENTIONS:** Diagnostic upper gastrointestinal endoscopy and flexible sigmoidoscopy, undertaken with or without sedation, with the standard preparation, techniques, and protocols of participating hospitals. After referral for either procedure, patients were randomised between doctors and

nurses. **MAIN OUTCOME MEASURES:** Gastrointestinal symptom rating questionnaire (primary outcome), gastrointestinal endoscopy satisfaction questionnaire and state-trait anxiety inventory (all analysed by intention to treat); immediate and delayed complications; quality of examination and corresponding report; patients' preferences for operator; and new diagnoses at one year (all analysed according to who carried out the procedure). **RESULTS:** There was no significant difference between groups in outcome at one day, one month, or one year after endoscopy, except that patients were more satisfied with nurses after one day. Nurses were also more thorough than doctors in examining the stomach and oesophagus. While quality of life scores were slightly better in patients the doctor group, this was not statistically significant. **CONCLUSIONS:** Diagnostic endoscopy can be undertaken safely and effectively by nurses.

Title: Designs for mechanical circulatory support device studies.

Reference: Neaton JD, Normand SL, Gelijns A, Starling RC, Mann DL, Konstam MA. *J Card Fail.* 2007 Feb;13(1):63-74.

Registration number: not reported

Source: MEDLINE (through PubMed)

Abstract: **BACKGROUND:** There is increased interest in mechanical circulatory support devices (MCSDs), such as implantable left ventricular assist devices (LVADs), as 'destination' therapy for patients with advanced heart failure. Because patient availability to evaluate these devices is limited and randomized trials have been slow in enrolling patients, a workshop was convened to consider designs for MCSD development including alternatives to randomized trials. **METHODS AND RESULTS:** A workshop was jointly planned by the Heart Failure Society of America and the US Food and Drug Administration and was convened in March 2006. One of the panels was asked to review different designs for evaluating new MCSDs. Randomized trials have many advantages over studies with no controls or with nonrandomized concurrent or historical controls. These advantages include the elimination of bias in the assignment of treatments and the balancing, on average, of known and unknown baseline covariates that influence response. These advantages of randomisation are particularly important for studies in which the treatments may not differ from one another by a large amount (eg, a head-to-head study of an approved LVAD with a new LVAD). However, researchers have found it difficult to recruit patients to randomized studies because the number of clinical sites that can carry out the studies is not large. Also, there is a reluctance to randomize patients when the control device is considered technologically inferior. Thus ways of improving the design of randomized trials were discussed, and the advantages and disadvantages of alternative designs were considered. **CONCLUSIONS:** The panel concluded that designs should include a randomized component. Randomized designs might be improved by allowing the control device to be chosen before randomisation, by first conducting smaller vanguard studies, and by allowing crossovers in trials with optimal medical management controls. With use of data from completed trials, other databases, and registries, alternative designs that include both a randomized component (eg, 2:1 allocation for new device versus control) and a

nonrandomized component (eg, concurrent nonrandomized control, historical control, or a comprehensive cohort design) should be evaluated. This will require partnerships among academic, government, and industry scientists.

A8. N=1 trials

Basically, N=1 trials (or individual patient trials) are cross-over trials in which one participant receives both the experimental and the control interventions; this single case study is the entire trial.¹⁷ Usually, the number of treatments is not specified in advance, so that the clinician and the patient can decide to stop when they observe important differences between the interventions.

The major advantage of this design is that it can be very effective in confirming causality between an achieved effect and the administered treatment. This is very useful when it is not clear whether a treatment will help a particular patient and it can guide clinical decisions, for example when there is little evidence from trials supporting which treatment to use in the event of rare diseases or in a particular patient population (children, elderly). In addition, the costs of the N=1 trial are considerably lower than those of other randomised designs.

The main disadvantages are that these trials provide individual outcomes rather than generalizable results, and that they are time-consuming.

EXAMPLES AN N=1 TRIAL USED FOR THE EVALUATION OF MEDICAL DEVICES

Title: High AC/A accommodative esotropia strabismus treated with contact lenses: a single case design (N=1) study.

Registration number: Not reported.

Reference: Gonzalez E, Barra F, Sanchez I, Antona B, Barrio A. Binocul Vis Strabismus Q. 2007;22(2):90-5.

Abstract: **PURPOSE:** The purpose was to determine the efficacy of two types of contact lenses (spherical disposable and aplanatic) as treatment in a patient with esotropia with a high Accommodative Convergence/Accommodation Ratio (AC/A). Due to the possibility of the appearance of accommodative insufficiency in this kind of patient, (i.e., following many years of bifocal glasses use), the elimination of the plus addition lens is advisable. Nevertheless, in some patients, this change leads to the appearance of a residual angle of esodeviation in near vision. It was expected that monofocal aplanatic contact lenses could achieve, due to their optical characteristics, an accurate and orthotropic binocular alignment, without aggravating an undesirable manifestation of the accommodative insufficiency. **METHODS:** An experimental design of a single case (N=1) was used in which

17 Guyatt G.H., Sackett D., Taylor D.W., Chong J., Roberts R.S., Pugsley S. Determining optimal therapy – randomized trials in individual patients. *N Engl J Med* 1986;314:889-92; Guyatt G.H., et al., The n-of-1 randomized controlled trial: clinical usefulness. Our three-year experience. *Ann. Intern. Med.* 1990; 112 (4): 293–9.

the subject acted as his own control. With bifocal glasses the subject displayed stability in his binocular and accommodative system at every distance of vision for the past three years. We compared the efficacy of two different types of hydrophilic contact lenses to control the angle of deviation, both at distance and at near vision. **RESULTS:** Neither of the two contact lenses produced the results of stability and the correct binocular alignment that had been achieved with bifocal glasses. This subject experienced a worse manifest esodeviation in distance vision with aplanatic lenses than with the disposable ones. **CONCLUSIONS:** These monofocal contact lenses did not create acceptable binocular alignment and stability in a subject with a high AC/A accommodative esotropia.

B. Non-randomised (direct evidence) study approaches

Non-randomised studies (NRS) may be important alternative approaches to studying the benefits of devices. However, ensuring proper selection of the subjects and a comparison group, adjusting for other influential factors, and addressing learning curve issues all require even more fine-tuning of the specific study design, conduct and analysis plan. In the absence of randomisation, conclusions as to whether an observed difference in benefits between the two groups is indeed due to the device are compromised (e.g. higher risk of bias). The main challenge of NRS in device evaluations is to correct for such biases – technically known as correcting for confounding factors – and thus to reach valid (unbiased) conclusions about the benefits of device use. There are various ways to address this confounding in NRS for devices (see Appendix V).¹⁸

Non-randomised studies can be divided into semi- or quasi-experimental trials (experimental is used here as a synonym for randomisation) and observational studies:

- Quasi-experimental studies are very similar to RCTs, but they specifically lack the element of random assignment (and thus concealed allocation) of participants to groups.¹⁹ Instead, quasi-experimental research allows the investigator to assign the participants to either the device or the control group using other allocation methods, including date of birth, the number of the hospital record or the day of invitation. These allocation methods are not regarded as ‘purely random’; these trials are therefore called quasi-experimental or pseudo-randomised trials. Ill luck may also make them more susceptible to (confounding) bias than RCTs and they may therefore require one of the adjustment methods described in Appendix V.
- Observational studies (comparative and non-comparative) use neither randomisation nor any pseudo-randomisation. Study participants are only observed on their received interventions (e.g. device use versus other intervention) and their

18 Cousens S. et al., Alternatives to randomisation in the evaluation of public-health interventions: statistical analysis and causal inference. *J Epidem Community Health* 2011;65:576-581; McNamee R., Regression modelling and other methods to control confounding. *Occup Environ Med* 2005;62:500-506 doi:10.1136/oem.2002.001115

19 Grimshaw J., Campbell M., Eccles M., Steen N., Experimental and quasi-experimental designs for evaluating guideline implementation strategies. *Fam Pract.* 2000 Feb;17 Suppl 1:S11-6.

outcomes. In general these designs can be divided into (retrospective or prospective) comparative and (retrospective or prospective) non-comparative observational studies.

B1. Quasi-experimental studies

B1.1 Non-randomised controlled study

This is an experimental study in which people are allocated to different interventions using one of the semi-randomisation methods described above, e.g. the number of the hospital record, the day of hospital admission, the day of invitation, etc.

Advantages of the quasi-experimental approach are that the results are more representative for routine clinical practice, such designs are easier to set up, and to some extent the allocation is still random and there is thus less potential for (confounding) bias compared to other NRS. This design can be used when randomisation is not feasible or in longitudinal research that involves a longer time horizon.

B1.2 Controlled before-and-after study

This approach involves making and measuring observations before and after implementing an intervention, both in a group that receives the device and in a control group that does not. The control group usually has similar characteristics (matching) to the group of participants receiving the (new) device. Outcomes of interest are measured in both groups before the intervention is introduced and again after the intervention has been introduced.

Even though these two designs use some form of pseudo-randomisation or well-matched controls, their main disadvantage is that any differences in outcomes may still be explained (in part) by other differences between the control and device use group. To correct for this, one of statistical methods described in Appendix V can and should be used.

EXAMPLE OF A CONTROLLED BEFORE-AFTER STUDY USED FOR THE EVALUATION OF MEDICAL DEVICES

Title: Further experience with pancreatic stump closure using a reinforced staple line.

Reference: Wallace CL, Georgakis GV, Eisenberg DP, Macaulay WP, Jimenez RE. *Conn Med.* 2013 Apr;77(4):205-10.

Registration number: not reported

Source: MEDLINE (through PubMed)

Abstract: **BACKGROUND:** We previously demonstrated that pancreatic transection with a reinforced staple line results in significantly lower fistula rates than when stapling without reinforcement. (*J Gastrointest Surg.* 2007;11:345-349). Criticism of this initial study focused on the small size of the treated group (N = 13). We report four more years of experience with this technique with a larger sample size. **METHODS:** This was a before-after trial. Patients included had distal pancreatectomies with stapled stump closure. The main intervention analyzed was staple-line reinforcement with Seamguard. The experimental

group consisted of a consecutive series of stapled pancreatectomies with reinforcement performed from 2005 to 2010. The control group was a consecutive series of stapled pancreatectomies without reinforcement performed between 2003 and 2005 (previously published). The main outcome measure was pancreatic fistula. **RESULTS:** 54 patients were included; 36 in the experimental group and 18 in the control group. Mean age was 62; 50% were males. The most common diagnoses were adenocarcinoma (31%), cystic neoplasm (24%), and neuroendocrine tumor (22%). There were no mortalities. Postoperative pancreatic leak rate was 39% in the control group, and 8% in the experimental group ($P = 0.01$). Seven of ten patients with leak required additional drain placement. Development of pancreatic leak resulted in prolonged hospital stays (12 vs eight days, $P < 0.007$). **CONCLUSION:** We demonstrate sustained success of reinforced stapling for pancreatic stump closure. Our technique is straightforward and results in reduced morbidity and cost. Our results suggest that surgical drains may not be needed when this technique is applied.

B1.3 Interrupted time series study (ITS)

This approach is a kind of a before-after study and observes a group of targeted individuals (and their outcomes) at multiple time points before the index intervention (i.e. device) is used or introduced in daily care; after the index intervention is introduced and used, the observations are compared to those of another group of individuals – preferably in the same setting or location as the control group (the ‘interruption’/‘pseudo-randomisation by time’). The design attempts to detect whether the intervention has had an effect significantly greater than any underlying trend over time: the multiple time points before the intervention allow the underlying trend to be estimated, the multiple time points after the intervention allow the intervention effect to be estimated. Any differences may be attributed to the intervention, accounted for by a time trend.

This approach measures the effect of an intervention when randomisation or identification of an appropriate concurrent control group is impractical.

A disadvantage of any time series study – i.e. when control and index group are split by time – is that the measured effect observed in the index, the device-use period, can be the effect of other events or changes in care occurring at the same time as the study intervention, or other differences between the two groups that occurred over time. To correct for this, one of the above methods should be used. Another limitation is that it is often difficult to collect sufficient data points during both the control and index periods, unless such studies are carefully planned, designed and conducted, or conducted in practices where comprehensive routine data sources or registries are available to allow for proper adjustment of other influential (confounding) factors. Many published interrupted time series have not been analysed correctly, so that the

effects of interventions are frequently overestimated.²⁰

EXAMPLE OF AN INTERRUPTED TIME SERIES STUDY USED FOR THE EVALUATION OF MEDICAL DEVICES

Title: Helmet legislation and admissions to hospital for cycling related head injuries in Canadian provinces and territories: interrupted time series analysis.

Reference: Dennis J, Ramsay T, Turgeon AF, Zarychanski R. *BMJ*. 2013 May 14;346:f2674. doi: 10.1136/bmj.f2674.

Reference number: Not reported

Source: MEDLINE (through PubMed)

Abstract: **OBJECTIVE:** To investigate the association between helmet legislation and admissions to hospital for cycling related head injuries among young people and adults in Canada. **DESIGN:** Interrupted time series analysis using data from the National Trauma Registry Minimum Data Set. **SETTING:** Canadian provinces and territories; between 1994 and 2003, six of 10 provinces implemented helmet legislation. **PARTICIPANTS:** All admissions (n=66,716) to acute care hospitals in Canada owing to cycling related injury between 1994 and 2008. **MAIN OUTCOME MEASURE:** Rate of admissions to hospital for cycling related head injuries before and after the implementation of provincial helmet legislation. **RESULTS:** Between 1994 and 2008, 66,716 hospital admissions were for cycling related injuries in Canada. Between 1994 and 2003, the rate of head injuries among young people decreased by 54.0% (95% confidence interval 48.2% to 59.8%) in provinces with helmet legislation compared with 33.1% (23.3% to 42.9%) in provinces and territories without legislation. Among adults, the rate of head injuries decreased by 26.0% (16.0% to 36.3%) in provinces with legislation but remained constant in provinces and territories without legislation. After taking baseline trends into consideration, however, we were unable to detect an independent effect of legislation on the rate of hospital admissions for cycling related head injuries. **CONCLUSIONS:** Reductions in the rates of admissions to hospital for cycling related head injuries were greater in provinces with helmet legislation, but injury rates were already decreasing before the implementation of legislation and the rate of decline was not appreciably altered on introduction of legislation. While helmets reduce the risk of head injuries and we encourage their use, in the Canadian context of existing safety campaigns, improvements to the cycling infrastructure, and the passive uptake of helmets, the incremental contribution of provincial helmet legislation to reduce hospital admissions for head injuries seems to have been minimal.

20 Grilli R, Freemantle N., Minozzi S., Domenighetti G., Finer D., Impact of Mass Media on Health Services Utilisation (Cochrane Review). In *The Cochrane Library, Issue 3* Oxford: Update Software; 1998.

B2. Observational study approaches: comparative and non-comparative studies

B2.1 Comparative observational studies

Cohort study – In a cohort study on the benefits of a specific intervention (device), a defined group of people with common characteristics is tracked over time in order to examine associations between different interventions that they have received and subsequent outcomes. Some in the cohort have received or used the device under evaluation and others have not; the observed outcomes of the two are then compared.

In a **prospective cohort study (B2.1.1)**, the investigator recruits participants before any intervention has started or been used and tracks them into the future to observe their outcomes. A predefined comparison of outcomes can be made between interventions, e.g. a sub-cohort that received/used the device versus a sub-cohort that did not receive/use the device but some other form of care, including no care at all.

The design is a valid alternative in situations where randomisation is not ethical. The prospective manner allows the investigator to pre-specify and collect all necessary data, e.g. on the outcomes and other influential (confounding) factors. However, a major disadvantage is that any differences in observed outcomes can be explained by other differences between the two sub-cohorts. This is even more so than for the quasi-experimental designs. Hence, one of the correction methods (see above) always need to be applied in order to isolate the effects of the device as much as possible from any confounding effects. As we said, however, the investigator has full control over which data to collect so as to ensure that all potentially known influential factors are defined and measured in all participants. This is in contrast to retrospective cohort studies (see below).

EXAMPLES OF PROSPECTIVE COHORT STUDY USED FOR THE EVALUATION OF MEDICAL DEVICES

Title: Transversus abdominis plane (TAP) catheters inserted under direct vision in the donor site following free DIEP and MS-TRAM breast reconstruction: a prospective cohort study of 45 patients.

Reference: Zhong T, Wong KW, Cheng H, Ojha M, Srinivas C, McCluskey SA, Clarke H, Jacks L, Hofer SO. *J Plast Reconstr Aesthet Surg.* 2013 Mar;66(3):329-36. doi: 10.1016/j.bjps.2012.09.034. Epub 2012 Nov 9.

Registration number: not reported

Source: MEDLINE (through PubMed)

Abstract: **INTRODUCTION:** The transversus abdominis plane (TAP) block is a peripheral nerve block of T6-L1 intercostal nerves of the abdominal wall. The purpose of this study was to evaluate the usefulness of intermittent TAP blockade for the first two postoperative days following free muscle sparing-transverse rectus abdominis muscle (MS-TRAM) or deep inferior epigastric perforator (DIEP) flap reconstruction of the breast. Therapeutic-Level II evidence. **MATERIAL AND METHODS:** This prospective cohort consisted of 45 consecutive patients who underwent DIEP or MS-TRAM free-flap breast reconstruction.

Intra-operatively, a multi-orifice epidural catheter was inserted under direct vision into the TAP. Ten millilitres of 0.25% bupivacaine was injected into each TAP catheter every 12 h until removal on day 3. The control group consisted of 80 consecutive patients who underwent free MS-TRAM or DIEP free-flap breast reconstructions by the same two surgeons without TAP block. Postoperatively, both groups had patient-controlled analgesia (PCA) and the primary outcome was intravenous (IV) PCA opioid consumption in the first 48 h. **RESULTS:** There were no complications associated with using TAP catheters. The 48-h PCA-delivered opioid requirement was significantly less ($p < 0.001$) in the TAP block group (17.10 ± 17.23 mg IV morphine equivalent) compared to the control group (48.44 ± 39.53 mg). **CONCLUSION:** Intermittent delivery of bupivacaine through the TAP block significantly reduced postoperative parenteral opioid requirements following free MS-TRAM or DIEP flap reconstruction of the breast. This is the first report of the TAP block being inserted under direct vision to provide postoperative analgesia at the abdominal flap donor site following microsurgical breast reconstruction.

Title: Risk of revision for fixed versus mobile-bearing primary total knee replacements.

Reference: Namba RS, Inacio MC, Paxton EW, Ake CF, Wang C, Gross TP, Marinac-Dabic D, Sedrakyan A. *J Bone Joint Surg Am.* 2012 Nov 7;94(21):1929-35. doi: 10.2106/JBJS.K.01363.

Registration number: not reported

Source: MEDLINE (through PubMed)

Abstract: **BACKGROUND:** Mobile-bearing total knee arthroplasty prostheses were developed to reduce wear and revision rates; however, these benefits remain unproven. The purposes of this study were to compare the short-term survivorship and to determine risk factors for revision of mobile-bearing and fixed-bearing total knee replacements. **METHODS:** A prospective cohort study of primary total knee arthroplasties performed from 2001 to 2009 was conducted with use of a community total joint replacement registry. Patient characteristics and procedure details were identified. Cox regression models were used. Bearing type was investigated as a risk factor for revision while adjusted for other risk factors such as age, American Society of Anesthesiologists (ASA) score, body mass index, sex, race, diagnosis, bilateral procedures, cruciate-retaining versus posterior-stabilized components, surgical approach, fixation, patellar resurfacing, hospital and surgeon volumes, and fellowship training. **RESULTS:** The study cohort consisted of 47,339 total knee arthroplasties, with 62.6% of the procedures in women. Fixed bearings were used in 41,908 knees (88.5%) and mobile bearings in 4830 (10.2%). Rotating-platform designs were used in all mobile-bearing total knee arthroplasties (3112 had a Rotating-Platform Press-Fit Condylar posterior-stabilized design; 1053, a Low Contact Stress [LCS] design; and 665, a Rotating-Platform Press-Fit Condylar cruciate-retaining design). Patients who received fixed-bearing total knee arthroplasty systems were older (mean age, 68.1 years) than those who received mobile-bearing total knee arthroplasty systems (mean age, 62.2 years); the difference was significant ($p < 0.001$). Overall, 515 knees (1.1%) were revised for reasons other than infection. The survival rate was 97.8% (95% confidence interval

[CI, 97.4% to 98.0%) at 6.7 years. The adjusted risk of aseptic revision for the LCS total knee replacements was 2.01 times (95% CI, 1.41 to 2.86) higher than that for fixed-bearing total knee replacements ($p < 0.001$). There was no significant revision risk for the other mobile-bearing total knee arthroplasty systems. There was no association with surgeon and hospital case volumes and the risk of revision total knee arthroplasty. **CONCLUSIONS:** Our study suggests the benefit of potential long-term wear reduction with the LCS implant may not be realized in a community-based setting, where a variety of surgical skills, surgical experience, and diverse patient demographic factors may affect early outcomes.

A specific form of prospective cohort study is the **prospective unpaired before-after study (B2.1.2)**. Here, participants act as self-controls or self-comparisons. Their outcomes are first measured before they receive/use the device, and once again after the device has been received/used (*non-randomised (unpaired) cross-over trial*). An advantage over the two-group cohort design is that smaller sample sizes are often needed because they produce within-participant comparisons (whereas parallel designs produce between-participant comparisons). But as with paired (randomised) cross-over trials, the problem with using participants as their own controls is that they may improve for unknown reasons unrelated to the device (e.g. regression to the mean or placebo effect). This design is also only possible when the target condition is chronic, and non-progressive. For rapidly changing (instable) conditions, either self-limiting or progressive, such designs are infeasible, as the true effects of the device can hardly be isolated from the natural course.

A **retrospective cohort study (B2.1.3)** uses an existing cohort (e.g. database from a previous study or existing registry) and selects *from the same cohort* a sub-cohort of participants that had received the intervention in the past and another sub-cohort that had not. Thus the use of the device and comparative interventions had both taken place in the past, relative to the starting point of the study. Then the outcomes of both groups – measured after the device has been used – are taken from that same database or perhaps even measured by the investigator at the present time.

The advantage of this cohort design over the prospective variants is that it is more efficient as it makes use of data collected earlier. It does require that the device and target condition were properly understood and used in the past, have not changed much over time, and also that the comparison group or current best practice has not changed much over time either.

The disadvantage is that, as in all observational studies, confounding correction methods must be applied to draw the most valid possible inference about whether the device has in fact caused the differences in outcomes observed between the two sub-cohorts. However, unlike the prospective cohort study designs, the use of earlier data means that the investigator has no control over the data and must rely on data that were collected previously and thus on data availability. Important confounding factors could therefore have been missed or not collected, making it impossible to make valid inferences about the benefits of the device. Moreover, existing databases or registries often lack data, which can also lead to bias.

EXAMPLE OF A RETROSPECTIVE COHORT STUDY USED FOR THE EVALUATION OF MEDICAL DEVICES

Title: A Retrospective Study Evaluating the Use of Permacol Surgical Implant in the Repair of Abdominal Wall Defects.

Registration number: NCT01214252

Source: ClinicalTrials.gov

Reference: Not published (yet).

Protocol: Purpose: To evaluate short, mid and long term clinical outcomes associated with the use of Permacol in the treatment of abdominal wall defects. Study Population: Patients who have undergone surgical repair of their abdominal wall defects with Permacol Surgical Implants with at least 12 months(-30days)follow-up. Inclusion Criteria:•Equal or over 18 years of age; •Had undergone surgical repair or reconstruction of abdominal wall defects, ventral hernias or incisional hernias using Permacol Surgical Implant; •At least 12 months of follow-up post date of surgery (-30 days); •Undergone open or laparoscopic repairs. Exclusion Criteria: •Had undergone inguinal, parastomal, diaphragmatic or paraesophageal/hiatal hernia repair; •Any prior use of Permacol in abdominal wall repair.

Enrollment: 472

Start date: October 2010

Study completion date: January 2012

In a non-randomised **historically controlled study (B2.1.4)**, a group of participants who have received/used the device is compared – often with prospective data collection – with a group of individuals who have not received/used the device but some form of alternative management. This control group can be taken from existing registries, or from a previous cohort or trial conducted in similar individuals. This is a common design for device evaluations and it is more efficient than a prospective cohort study. In addition, it shares the same advantages as a retrospective cohort study (**B2.1.3**).

Unfortunately, it also shares the same disadvantages as the retrospective cohort study, and may in fact have one more disadvantage, because the index (device) group and control group are by definition not selected from the same underlying total cohort. The risk of incomparability between the two groups is thus much larger. This incomparability can be due to changes over time (device group very different from the historical control group; they may not have received current best practice, and therefore have other outcomes by definition); not all relevant confounding factors have been documented in the historical control group; and the methods used for outcome evaluation might not be similar or appropriate in both groups.

EXAMPLE OF A HISTORICALLY CONTROLLED STUDY USED FOR THE EVALUATION OF MEDICAL DEVICES

Title: Comparison Study of Two Different Surgical Clips During Laparoscopic Urologic Surgery

Registration number: NCT01008709

Source: ClinicalTrials.gov

Reference: Not published (yet)

Protocol: Purpose. Intracorporeal suturing and knot tying during robotic prostatectomy and laparoscopic and robotic renal surgery have historically been considered the most technically challenging and time consuming aspects of these procedures. With improved operative technique as well as the use of innovative surgical devices, vascular control during these surgeries is often less cumbersome as compared with traditional techniques. Current standard methods of hemostasis include the use of clips, of which the most popular design is the Hemolock, a locking, nonabsorbable plastic clip, or the use of the very expensive endomechanical stapler. Unfortunately while they are associated with time savings in the operating room, there is a great deal of disposable costs associated with these various devices as well as a not insignificant device malfunction rate reported in the literature. The aim of this case-controlled study is to evaluate the Aesculap U-clip device compared to our current technique of vascular control using the Teleflex Hemolock clip device during minimally invasive genitourinary surgery.

Enrolment: 20

Start date: October 2009

Study completion date: October 2012

The **case-control study (B2.1.5)** compares people with a specific outcome of interest ('cases') with people from the same source population but without that outcome ('controls'), and then studies differences in the interventions they have undergone. Unlike cohort studies, where subjects are selected for having or not having undergone/used the device, here the subjects are selected for having or not having the outcome of interest.

This design is particularly useful for investigating the development (causes) of rare outcomes. It can be a useful for assessing the association of a particular event – often a particular rare outcome or unintended effect – with a specific device use.

A disadvantage is that case-control studies are generally considered to have a larger risk of bias than cohort studies, mainly because data collection on device use in the past is often retrospective, leading to incomplete data collection. As in the case of retrospective cohort studies, information on many confounders is also often missing. Moreover, data re-collection to determine previous exposure to possible causes of the relevant outcomes may differ depending on whether the subjects had or had not experienced the outcome under study (recall bias).

B2.2 Non-comparative observational studies

Case series (uncontrolled longitudinal study; B2.2.1) are descriptive studies in which observations are made on a series of individuals, usually all receiving the same intervention, before and after the intervention but with no control group. It can be retrospective or prospective, consecutive or non-consecutive, and usually involves a smaller

number of patients than case-control studies or randomised controlled trials.²¹

A disadvantage of case series is that they may be confounded by selection bias, which limits the strength of conclusions concerning the causality of correlations observed.

EXAMPLES OF CASE-SERIES USED FOR THE EVALUATION OF MEDICAL DEVICES

Title: Cardiovascular implantable electronic device replacement infections and prevention: results from the REPLACE Registry.

Reference: Uslan DZ, Gleva MJ, Warren DK, Mela T, Chung MK, Gottipaty V, Borge R, Dan D, Shinn T, Mitchell K, Holcomb RG, Poole JE. *Pacing Clin Electrophysiol.* 2012 Jan;35(1):81-7. doi: 10.1111/j.1540-8159.2011.03257.x. Epub 2011 Nov 11.

Registration number: NCT00395447

Source: MEDLINE (through PubMed)

Abstract: **BACKGROUND:** Infection following cardiovascular implantable electronic device (CIED) replacement is a serious complication, and rates of infection have increased. Analysis of procedural and clinical data from device replacement procedures collected by the REPLACE Registry may provide insights into infection prevention strategies and outcomes.

METHODS: We prospectively evaluated procedural complications in patients undergoing CIED replacement over 6 months from 72 U.S. sites. Major and minor infections were predefined and adjudicated by an independent blinded clinical events committee. Data regarding infection prevention strategies and infectious outcomes were analyzed for their potential relationships.

RESULTS: A total of 1,744 patients were included in REPLACE. All patients received preoperative intravenous antibiotics and 68.7% received postoperative systemic antibiotic therapy. CIED infection developed in 22 patients (1.3%), of which 14 cases were major (0.8%, 95% confidence interval [CI] 0.4%-1.3%) and eight were minor (0.5%, 95% CI 0.2%-0.9%). Patients with infections were more likely to have had postoperative hematomas (five of 22 [22.7%] vs 17 of 1,722 [0.98%], $P = 0.002$). Participating sites experiencing infection rates $>5\%$ were more likely to use povidone-iodine for topical antisepsis, had lower implantation volume, and had patients with higher Charlson Comorbidity Index (2.79 vs 2.32, 95% CI for difference 0.08-0.86, $P = 0.019$).

CONCLUSIONS: In this multicenter prospective study with 6 months of follow-up, infections associated with CIED replacements were surprisingly infrequent, possibly due to the use of preoperative antibiotics. Patients with infections were more likely to have had a postoperative hematoma, and sites with higher infection rates had sicker patients and lower overall procedural volume.

21 Dekkers O.M., Egger M., Altman D.G., Vandenbroucke J.P., Distinguishing case series from cohort studies. *Ann Intern Med.* 2012 Jan 3;156(1 Pt 1):37-40. doi: 10.7326/0003-4819-156-1-201201030-00006.

Title: Clinical Evaluation of the Interlace Medical Hysteroscopic Morcellator.

Reference: Not published (yet).

Registration number: NCT01026805

Source: ClinicalTrials.gov

Protocol: Purpose. This study assesses the effectiveness of intrauterine fibroid and polyp removal using the Interlace Medical 1st generation hysteroscopic morcellator device based on a retrospective review of medical records of women who have been treated with the device. Case series: 11 women previously receiving hysteroscopic myomectomy or polypectomy using the hysteroscopic morcellator device.

Enrollment: 11

Start date: February 2009

Study completion date: March 2009

Device-based registries (B2.2.2) can be useful not only for long-term surveillance and postmarketing but also for evaluating medical devices outcomes. The main advantage is that registry data reflects the use of devices in routine practice (in the real-world setting) over time. Because registries contain data on large numbers of participants, subgroup analyses are possible. Moreover, by linking device exposures and long-term outcomes, registries permit follow-up that can span decades.

Special challenges related to the construct (and use) of a medical device registry include the need for unique identification of devices and the need for follow-up to collect all relevant information on outcomes, patient factors, device factors, user interface, experience and learning curves, device modifications, and the ability to combine all multiple components within a device.

EXAMPLES OF REGISTRIES USED FOR THE EVALUATION OF MEDICAL DEVICES

Title: The international registry infrastructure for cardiovascular device evaluation and surveillance.

Reference: Marinac-Dabic D, Holmes DR. JAMA. 2013 Jul 17;310(3):257-9. doi: 10.1001/jama.2013.7133 (http://www.ncbi.nlm.nih.gov/pubmed?term=Sedrakyan%20A%5BAuthor%5D&cauthor=true&cauthor_uid=23860981)

Source: MEDLINE (through PubMed)

Abstract: Since the creation of the US Food and Drug Administration (FDA) Medical Device Epidemiology Network (MDEpiNet),¹ there has been an increasing FDA commitment to support the development of a global medical device research and surveillance infrastructure. The FDA's new postmarket surveillance plan² strengthens this commitment and highlights the importance of national and international registries, and multi-stakeholder involvement for ensuring this commitment is fulfilled. This Viewpoint summarizes the potential for development of an International Consortium of Cardiovascular Registries, modeled on a consortium established for orthopedic devices, as an initial project in the realm of cardiovascular devices with the focus on transcatheter aortic valve replacement (TAVR).

Summary: Modeling after the orthopedic ICOR initiative, a global consortium of cardiovascular device registries has great potential to improve the public health, facilitate and strengthen regulatory processes, and advance clinical practice using innovative approaches. The exploration of the new International Consortium of Cardiovascular Registries focused on the novel technology of TAVR has multiple potential scalable positive outcomes for a larger cardiovascular initiative. The initiative may improve collaboration of the different stakeholders and enhance efficiency of registries and could facilitate the evaluation of the safety and efficacy of new devices and approaches, thereby reaching the goal of harnessing the global knowledge.

Title: Kaiser Permanente implant registries benefit patient safety, quality improvement, cost-effectiveness.

Reference: Paxton EW, Kiley ML, Love R, Barber TC, Funahashi TT, Inacio MC. *Jt Comm J Qual Patient Saf.* 2013 Jun;39(6):246-52.

Source: MEDLINE (through PubMed)

Abstract BACKGROUND: In response to the increased volume, risk, and cost of medical devices, in 2001 Kaiser Permanente (KP) developed implant registries to enhance patient safety and quality, and to evaluate cost-effectiveness.

METHODS: Using an integrated electronic health record system, administrative databases, and other institutional databases, orthopedic, cardiology, and vascular implant registries were developed in 2001, 2006, and 2011, respectively. These registries monitor patients, implants, clinical practices, and surgical outcomes for KP's 9 million members. Critical to registry success is surgeon leadership and engagement; each geographical region has a surgeon champion who provides feedback on registry initiatives and disseminates registry findings. **RESULTS:** The registries enhance patient safety by providing a variety of clinical decision tools such as risk calculators, quality reports, risk-adjusted medical center reports, summaries of surgeon data, and infection control reports to registry stakeholders. The registries are used to immediately identify patients with recalled devices, evaluate new and established device technology, and identify outlier implants. The registries contribute to cost-effectiveness initiatives through collaboration with sourcing and contracting groups and confirming adherence to device formulary guidelines. Research studies based on registry data have directly influenced clinical best practices. **CONCLUSIONS:** Registries are important tools to evaluate longitudinal device performance and safety, study the clinical indications for and outcomes of device implantation, respond promptly to recalls and advisories, and contribute to the overall high quality of care of our patients

Title: Comparison of 21-gauge and 22-gauge aspiration needle in endobronchial ultrasound-guided transbronchial needle aspiration: results of the American College of Chest Physicians Quality Improvement Registry, Education, and Evaluation Registry.

Reference: Yarmus LB, Akulian J, Lechtzin N, Yasin F, Kamdar B, Ernst A, Ost DE, Ray C, Greenhill SR, Jimenez CA, Filner J, Feller-Kopman D; American College of Chest Physicians Quality Improvement Registry, Education, and Evaluation (AQuIRE). *Chest.* 2013

Apr;143(4):1036-43. doi: 10.1378/chest.12-1205.

Source: MEDLINE (through PubMed)

Abstract: **BACKGROUND:** Endobronchial ultrasound-guided transbronchial needle aspiration (EBUS-TBNA) is a minimally invasive procedure originally performed using a 22-gauge (22G) needle. A recently introduced 21-gauge (21G) needle may improve the diagnostic yield and sample adequacy of EBUS-TBNA, but prior smaller studies have shown conflicting results. To our knowledge, this is the largest study undertaken to date to determine whether the 21G needle adds diagnostic benefit. **METHODS:** We retrospectively evaluated the results of 1,299 patients from the American College of Chest Physicians Quality Improvement Registry, Education, and Evaluation (AQUIRE) Diagnostic Registry who underwent EBUS-TBNA between February 2009 and September 2010 at six centers throughout the United States. Data collection included patient demographics, sample adequacy, and diagnostic yield. Analysis consisted of univariate and multivariate hierarchical logistic regression comparing diagnostic yield and sample adequacy of EBUS-TBNA specimens by needle gauge. **RESULTS:** A total of 1,235 patients met inclusion criteria. Sample adequacy was obtained in 94.9% of the 22G needle group and in 94.6% of the 21G needle group ($P = .81$). A diagnosis was made in 51.4% of the 22G and 51.3% of the 21G groups ($P = .98$). Multivariate hierarchical logistic regression showed no statistical difference in sample adequacy or diagnostic yield between the two groups. The presence of rapid onsite cytologic evaluation was associated with significantly fewer needle passes per procedure when using the 21G needle ($P < .001$). **CONCLUSIONS:** There is no difference in specimen adequacy or diagnostic yield between the 21G and 22G needle groups. EBUS-TBNA in conjunction with rapid onsite cytologic evaluation and a 21G needle is associated with fewer needle passes compared with a 22G needle.

Performance goals (B2.2.3) – In a performance goal study, participants are enrolled through databases, registers or surveys only to receive the device under investigation. The performance of a device can be evaluated by comparing it to a numerical target value that pertains to a safety or effectiveness endpoint. Developing performance goals requires good historical data or publications, and a-priori agreement on who should develop the goals in question.

A performance goal study has limitations comparable to that of a non-randomised historically controlled study. In addition, since there is no control group, such studies cannot demonstrate either superiority or non-inferiority to another management, device or intervention.

EXAMPLE OF A PERFORMANCE GOAL STUDY USED FOR THE EVALUATION OF MEDICAL DEVICES

Title: GORE® Embolic Filter in Carotid Stenting for High Risk Surgical Subjects (EMBOLDEN)

Reference: Not published (yet)

Registration number: NCT00766493

Source: ClinicalTrials.gov

Protocol: Purpose: Compare the 30-day safety and efficacy of the GORE® Embolic Filter used in conjunction with FDA-approved carotid stents to a performance goal obtained from carotid stent studies utilising distal embolic protection.

Enrollment: 250

Start Date: January 2009

Study Completion Date: July 2010

C. Study approaches to evaluating diagnostic, screening, monitoring and prognostic test devices

As described in Chapter 4.3 main text, non-therapeutic devices include diagnostic, screening and prognostic test devices. Test devices commonly have no direct therapeutic effects on the health outcomes of the intended population; instead, they are tests that generate information which in turn directs health care professionals (or sometimes patients themselves – depending on the type of device) to guide clinical management (see Figure 1). The health of the targeted individuals will thus improve indirectly because they receive more appropriate or earlier treatment, and because they can avoid potentially harmful additional tests.

Researchers of such devices should consider carefully whether they will focus on the results generated by the device often referred to as (predictive) accuracy studies – or also measure the down-stream consequences of test use, for example changes in the management of health care professionals or even in the health effects induced by combining the use of the device and follow-up treatment (device test-treatment evaluations).

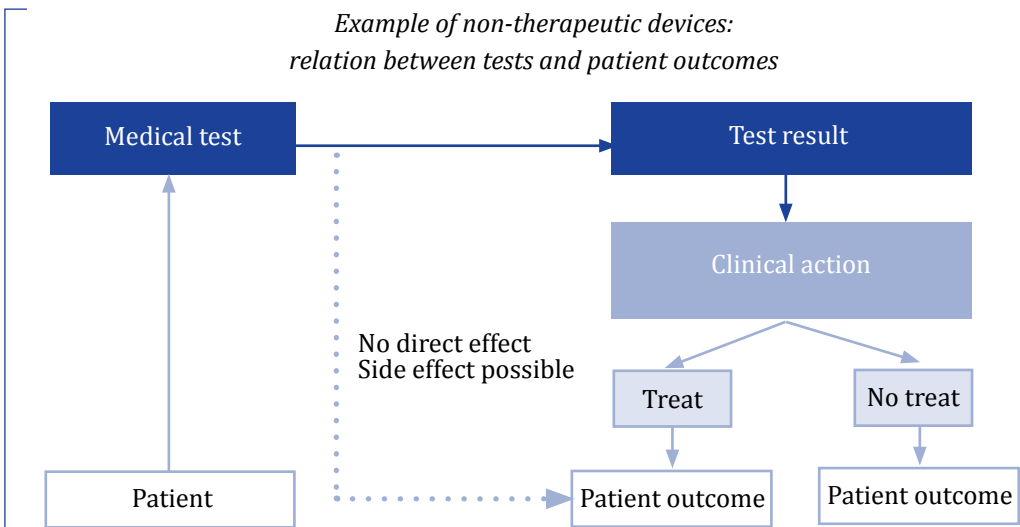


Figure 1. Possible working pathways of diagnostic, screening, monitoring, prognostic test devices.

C1. Diagnostic and screening device accuracy designs

Diagnostic and screening test accuracy studies aim to measure the relationship between the results of the test and the presence or absence of a certain disease or condition (e.g. disease or any other health state) of interest *at the same time*. This leads to a cross-sectional (though time) design where subjects will undergo both the index device (test) under evaluation and the reference standard. The reference standard is the best available method – usually more invasive, burdensome or costly than the index device – for determining whether a subject in fact (‘truly’) has or does not have the condition of interest. Because the interest is in the cross-sectional relationship, the device and the reference standard should ideally be performed within a short time interval. Sometimes it is not possible or ethical to establish the presence or absence of a target condition at the moment when the index device is used. Follow-up can then be used as a means of confirming or refuting the true presence or absence of the target condition. This is frequently applied in studies on the accuracy of screening devices for detecting early stages of a certain disease, e.g. breast cancer. Several accuracy measures can be used to express the agreement between the index device/test under evaluation and the true presence or absence of disease, for example sensitivity & specificity, positive and negative predictive values, likelihood ratios, and the diagnostic odds ratio.

Test accuracy studies are vulnerable to some specific types of biases, as described at length in the literature (see e.g. references in Chapter 4).²² For example, in diagnostic accuracy studies, the individuals of interest are often defined by their presenting symptoms, signs and setting. Using patients already diagnosed with the condition or disease of interest and a separate group of control subjects without the disease can lead to overoptimistic measures of accuracy, because patients with the disease are more typical or advanced cases of the target condition. Also, the selected control group can be atypical, for instance when a healthy control group has been used. Such case-control designs (see above) using healthy controls may indeed be used to examine whether there a device test holds any promise. However, the results of such designs and the accuracy (agreement) found between the index device and the reference test cannot simply be generalised to the device accuracy that would be found when the device is used in the targeted clinical population in routine practice. For diagnostic and screening test accuracy studies, it is crucial to think about the targeted individuals and setting beforehand (see Chapter 4.5). Another key quality item of such studies is that all patients will undergo the same (best available) reference standard. Finally, the results of the index diagnostic or screening test device should be determined or interpreted without knowing the outcome of the reference standard; otherwise, there is the risk of producing overoptimistic estimates of accuracy for the device under evaluation.

²² Bossuyt P.M., Irwig L., Craig J., Glasziou P., Comparative accuracy: assessing new tests against existing diagnostic pathways. *BMJ*. 2006 May 6;332(7549):1089-92.

C1.1 Single (non-comparative) device test accuracy studies

The aim in single device test accuracy studies is to determine the accuracy of a specific device test (see Figure 2). This leads to a cross-sectional diagnostic or screening device study where the targeted individuals (ideally a consecutive series of such individuals) undergo both the index device test and the reference standard. A key quality item of such studies is that all patients will undergo the same optimal reference standard. The results of the index test should be determined or interpreted without knowing the outcome of the reference standard; otherwise, there is the risk of producing overoptimistic estimates of accuracy.

Single device test accuracy studies are not suitable for determining whether a specific test leads to improved accuracy beyond what is achieved with available test results, or which order of testing is to be preferred (see C1.2); neither are they suitable for determining whether the use of the diagnostic or screening device will lead to actual (additional) health benefits in individuals (see C.3).

EXAMPLE OF A SINGLE DEVICE TEST ACCURAY STUDY (1)

Title: A Prospective Study to Determine the Role of 2-[18F]Fluoro-2-Deoxy-D-Glucose (FDG)Positron Emission Tomography (PET)in the Assessment of Regional Nodal Spread of Disease in Breast Cancer Patients

Registration number: NCT00201942

Source: ClinicalTrials.gov

Protocol: Patients will have histologic confirmation of invasive breast cancer and will have a FDG-PET scan prior to axillary node assessment. All patients will have a sentinel node biopsy if any sentinel nodes can be located. Patients with a positive sentinel node will have an axillary node dissection. The results of the PET will be compared to the reference standard of histologic examination of all excised (sentinel and non-sentinel) axillary lymph nodes which will be referred to as axillary node assessment. Sensitivity, specificity, positive and negative predictive values for PET-FDG will be determined.

EXAMPLE OF A SINGLE DEVICE TEST ACCURAY STUDY (2)

Title: Diagnostic accuracy of plasma glial fibrillary acidic protein for differentiating intracerebral hemorrhage and cerebral ischemia in patients with symptoms of acute stroke.

Reference: Foerch C, Niessner M, Back T, Bauerle M, De Marchis GM, Ferbert A, Grehl H, Hamann GF, Jacobs A, Kastrup A, Klimpe S, Palm F, Thomalla G, Worthmann H, Sitzer M; BE FAST Study Group. Clin Chem. 2012 Jan;58(1):237-45. doi: 10.1373/clinchem.2011.172676. Epub 2011 Nov 28.

Source: MEDLINE (through PubMed)

Abstract: **BACKGROUND:** Glial fibrillary acidic protein (GFAP) is a biomarker candidate indicative of intracerebral hemorrhage (ICH) in patients with symptoms of acute stroke. GFAP is released rapidly in the presence of expanding intracerebral bleeding, whereas a more gradual release occurs in ischemic stroke. In this study the diagnostic accuracy of plasma GFAP was determined in a prospective multicenter approach.

METHODS: Within a 1-year recruitment period, patients suspected of having acute (symptom onset <4.5 h before admission) hemispheric stroke were prospectively included into the study in 14 stroke centers in Germany and Switzerland. A blood sample was collected at admission, and plasma GFAP was measured by use of an electrochemiluminometric immunoassay. The final diagnosis, established at hospital discharge, was classified as ICH, ischemic stroke, or stroke mimic.

RESULTS: The study included 205 patients (39 ICH, 163 ischemic stroke, 3 stroke mimic). GFAP concentrations were increased in patients with ICH compared with patients with ischemic stroke [median (interquartile range) 1.91 µg/L (0.41-17.66) vs 0.08 µg/L (0.02-0.14), $P < 0.001$]. Diagnostic accuracy of GFAP for differentiating ICH from ischemic stroke and stroke mimic was high [area under the curve 0.915 (95% CI 0.847-0.982), $P < 0.001$]. A GFAP cutoff of 0.29 µg/L provided diagnostic sensitivity of 84.2% and diagnostic specificity of 96.3% for differentiating ICH from ischemic stroke and stroke mimic.

CONCLUSIONS: Plasma GFAP analysis performed within 4.5 h of symptom onset can differentiate ICH and ischemic stroke. Studies are needed to evaluate a GFAP point-of-care system that may help optimize the prehospital triage and management of patients with symptoms of acute stroke.

C1.2 Comparative test accuracy studies

In medical practice, professionals are commonly interested in the accuracy of a new test device for a specific target condition as compared to the accuracy of other existing tests, rather than the accuracy of a single test device. Comparing tests within the same study – where all the subjects undergo both the new and existing test devices and the same reference standard – is likely to produce more valid results than comparing accuracy estimates taken from different studies, each evaluated as a separate test.²³ In the latter situation, many factors other than the use of a different device test may be responsible for a difference in accuracy, for example differences in the setting, population, observers or users of the device, and reference standard protocol. Comparative test device accuracy studies are thus similar to the cross-over design in therapeutic designs (see above).

Comparative test accuracy studies also allow us to study the accuracy of combinations of diagnostic tests, whether a new device can lead to improved classification or accuracy for patients beyond what current tests have already achieved, in which order different tests may be conducted, and at which point in time a specific new test device may be best suited. Often, the focus in such comparative studies is on assessing whether a certain test adds information to previous test results in terms of improved accuracy or classification.

Comparative accuracy designs are attractive in any test device research (diagnostic, screening or prognostic tests) because it is often easy to meet the requirements that the first index test does not influence the result of the second index test (due to

23 Takwoingi Y, Leeflang M.M., Deeks J.J., Empirical evidence of the importance of comparative studies of diagnostic test accuracy. *Ann Intern Med.* 2013 Apr 2;158(7):544-54.

learning effects, for example) and that the targeted disease (stage) should not change between the first and second index test. If these requirements are violated or if it is too burdensome for subjects to receive all index test devices, subjects can be randomised to receive either one of the test devices under evaluation. However, cross-over designs are more efficient as they require only half the number of patients (or even fewer) than needed in a parallel randomised design. An additional benefit of cross-over comparative test studies is that such studies reveal whether index tests tend to make errors in the same subjects (correlated errors). The different test accuracy designs are outlined in Figure 2.

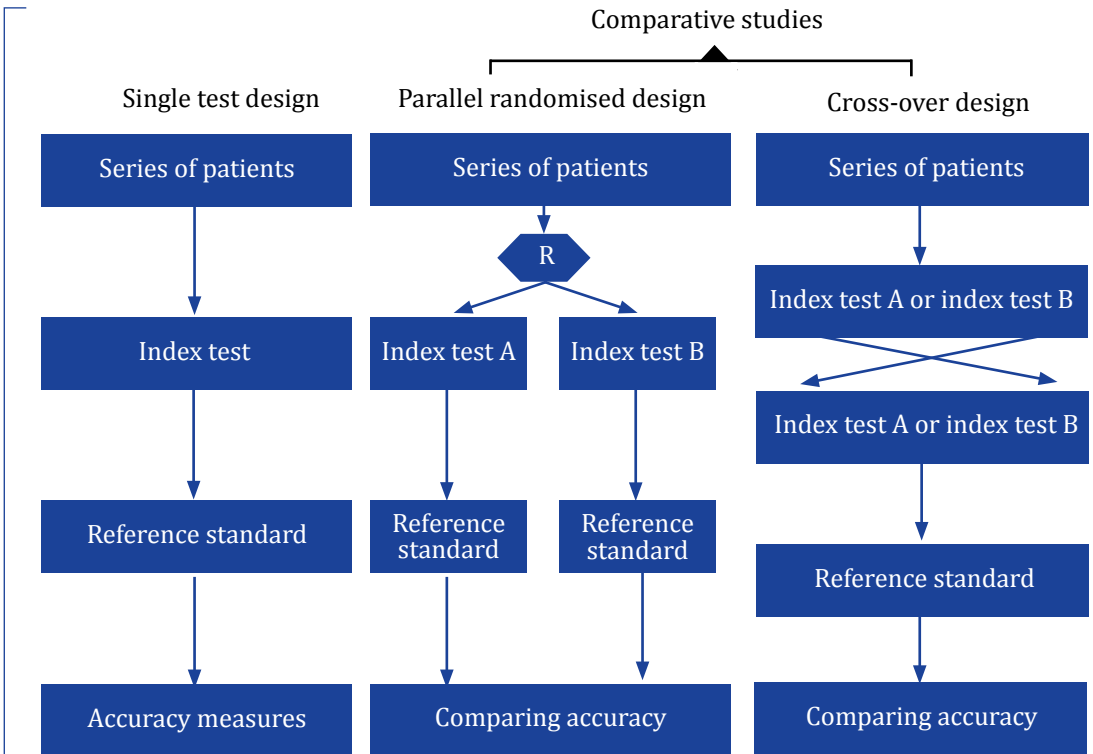


Figure 2. Single test and comparative diagnostic or screening test accuracy designs

EXAMPLE OF A COMPARATIVE DEVICE TEST ACCURACY STUDY (CROSS-OVER DESIGN)

Title: Comparative study of automated breast 3-d ultrasound and handheld B-mode ultrasound for differentiation of benign and malignant breast masses.

Reference: Chen L, Chen Y, Diao XH, Fang L, Pang Y, Cheng AQ, Li WP, Wang Y. *Ultrasound Med Biol.* 2013 Oct;39(10):1735-42. doi: 10.1016/j.ultrasmedbio.2013.04.003. Epub 2013 Jul 9.

Source: MEDLINE (through PubMed)

Abstract: The automated breast volume scanner (ABVS) represents a new technology for diagnosing breast masses. In this study, a total of 219 breast masses in 175 patients underwent both conventional handheld B-mode ultrasound (HHUS) and ABVS examinations, and the differences in the diagnostic values of the two modalities for benign and malignant breast masses were compared with the final pathologic findings. In addition, the diagnostic accuracy for breast masses with features including retraction phenomenon and hyperechoic rim in the coronal plane of the ABVS was evaluated. There were no differences between the ABVS and HHUS in terms of sensitivity (92.5% vs. 88.0%), specificity (86.2% vs. 87.5%), accuracy (88.1% vs. 87.2%), false-positive rate (13.8% vs. 12.5%), false-negative rate (11.8% vs. 7.5%), positive predictive value (74.7% vs. 75.6%) and negative predictive value (96.3% vs. 94.3%) ($p > 0.05$ for all). However, there were significant differences between the malignant and benign masses with respect to retraction phenomenon and hyperechoic rim in the coronal plane of the ABVS. For retraction phenomenon, both the specificity and positive predictive value of a malignant diagnosis reached 100%, and the accuracy and false-positive rate were 96.8% and 0, respectively; for the hyperechoic rim, the specificity, negative predictive value and accuracy of a benign diagnosis were 92.8%, 95.3% and 95.9%, respectively.

Conclusion: Overall, ABVS is a promising modality for the clinical diagnosis of breast masses with retraction phenomenon and hyperechoic rim in the coronal plane, although the ABVS and HHUS do not differ in diagnostic accuracy for the differentiation of malignant or benign breast masses.

EXAMPLE OF A COMPARATIVE DEVICE TEST ACCURACY STUDY (PARALLEL RANDOMISED DESIGN)

Title: Acetic acid compared with i-scan imaging for detecting Barrett's esophagus: a randomized, comparative trial.

Reference: Hoffman A, Korczynski O, Tresch A, Hansen T, Rahman F, Goetz M, Murthy S, Galle PR, Kiesslich R. *Gastrointest Endosc.* 2013 Aug 13. pii: S0016-5107(13)02129-9. doi: 10.1016/j.gie.2013.07.013. [Epub ahead of print]

Source: MEDLINE (through PubMed)

Abstract: BACKGROUND: Traditional surveillance in patients with Barrett's esophagus (BE) has relied on random biopsies. Targeted biopsies that use advanced imaging modalities may significantly improve detection of specialized columnar epithelium (SCE).

OBJECTIVE: We compared the efficacy of targeted biopsies that used i-scan or acetic acid to random biopsies in the detection of SCE.

DESIGN: Patients with visible columnar lined epithelium or known BE were randomized at a 1:1 ratio to undergo acetic acid application or i-scan with targeted biopsies.

SETTING: Targeted biopsies were performed based on surface architecture according to the Guelrud classification followed by 4-quadrant biopsies.

PATIENTS: A total of 95 patients were randomized.

INTERVENTION: A total of 46 patients underwent acetic acid staining, and 49 underwent i-scan imaging. Random biopsies were performed in 86 patients.

MAIN OUTCOME MEASUREMENTS: The primary outcome was the yield of SCE as

confirmed by histologic assessment.

RESULTS: The diagnostic yield for SCE was significantly higher with targeted biopsies than with random biopsies in both groups combined (63% vs 24%; $P = .0001$). The yield of targeted biopsies was significantly greater with both i-scan (66% vs 21%; $P = .009$) and acetic acid (57% vs 26%; $P = .012$) technologies and did not differ between these groups. The accuracy for predicting SCE was 96% ($k = .92$) for i-scan and 86% ($k = .70$) for acetic acid analysis.

LIMITATIONS: No dysplastic lesions were found.

CONCLUSION: The i-scan or acetic acid-guided biopsies have a significantly higher diagnostic yield for identifying SCE, with significantly fewer biopsies, as compared with a protocol of random biopsies. Acetic acid and i-scan showed comparable results diagnosing SCE in our study. (Clinical trial registration number: NCT01442506.).

C2. Prognostic (and monitoring) device accuracy studies

In accuracy studies for prognostic and monitoring test devices (or biomarkers), the interest is in the relationship between the information from the index test device and the occurrence of an health outcome (or health event) in the future – in other words, how information from the index test device can predict a future health outcome. Such outcomes can be an objective event such as a particular disease progression or recurrence, a new event, or a treatment response, but can also include more subjective measures such as decreased pain experience or improved quality of life. The future may be measured in hours or days (e.g. devices to measure blood loss during surgery in order to predict the need for postoperative blood transfusion), weeks, months or even years.

The time sequence between the index test device and outcome occurrence requires a different type of study design than diagnostic or screening device accuracy studies. The biggest difference between diagnostic and screening device accuracy studies and prognostic device accuracy studies is time. More specifically, the time interval between the index test result and the outcome of interest. In prognostic test accuracy studies, subjects must undergo the test device and are tracked for a certain amount of time to establish whether or not they develop the targeted study outcome(s) in the future. This time may range from hours, days, weeks, months or even years, depending on the working mechanism of the device. Such prognostic test device study thus have to follow a longitudinal design, such as a prospective or retrospective observational cohort study (see above). Also, data from randomised studies may be used to evaluate the predictive accuracy of prognostic test devices.

In studies of prognostic test devices, the population of interest can be patients diagnosed with a particular condition for whom a particular future health outcome is to be predicted, but may also involve predicting whether healthy individuals will develop diabetes type 2, for example. The overall question in such studies is whether the information provided by the test device is helpful in labelling or stratifying

patients by their risk of developing a certain outcome in the future. Study individuals or patients should again be representatives of the intended spectrum of individuals (patients) and not only a specific small subgroup. Important choices must be made about which relevant health outcomes to predict and thus observe in a study, and what the relevant duration of the follow-up is. Studies with longer follow-ups are usually more relevant for patients, professionals and society at large, but they are also more time-consuming, expensive, and run a higher risk of losing the study individuals. The accuracy of a prognostic device can be expressed in terms of how well the device predicts actual observed outcomes and how well it differentiates between those who do or do not experience the health outcome.

C2.1 Single prognostic device studies

Like single diagnostic and screening test device accuracy studies, prognostic test devices can also be evaluated individually (in isolation) for their predictive prognostic accuracy. Such studies often apply the classic cohort design, with follow-up of the subjects using the prognostic test device until the predefined time period for developing the health outcome. One important quality item of any prognostic device study, whether single or comparative (see next section), is the completeness of the follow-up; ideally, all the study follow up all the subjects for the predefined time period. Like single diagnostic and screening test accuracy studies (C1.1), the main result of a single prognostic test device study is the strength of the association between the test device information and the occurrence of the health outcome of interest. And like single diagnostic and screening test accuracy studies, single prognostic test accuracy studies are not suitable for determining whether a new prognostic device has added predictive value beyond the accuracy already obtained by existing prognostic tests or factors; they are also not suitable for identifying the order in which prognostic devices can best be used (see C.2.2) or for determining whether actual use of the prognostic device or use of the diagnostic or screening device will lead to actual (or added) health benefits in individuals (see C.3).

EXAMPLE OF A COMPARATIVE DEVICE TEST ACCURACY STUDY (PARALLEL RANDOMISED DESIGN)

Title: Acetic acid compared with i-scan imaging for detecting Barrett's esophagus: a randomized, comparative trial.

Reference: Hoffman A, Korczynski O, Tresch A, Hansen T, Rahman F, Goetz M, Murthy S, Galle PR, Kiesslich R. *Gastrointest Endosc.* 2013 Aug 13. pii: S0016-5107(13)02129-9. doi: 10.1016/j.gie.2013.07.013. [Epub ahead of print]

Source: MEDLINE (through PubMed)

Abstract: **BACKGROUND:** Traditional surveillance in patients with Barrett's esophagus (BE) has relied on random biopsies. Targeted biopsies that use advanced imaging modalities may significantly improve detection of specialized columnar epithelium (SCE).

OBJECTIVE: We compared the efficacy of targeted biopsies that used i-scan or acetic acid

to random biopsies in the detection of SCE.

DESIGN: Patients with visible columnar lined epithelium or known BE were randomized at a 1:1 ratio to undergo acetic acid application or i-scan with targeted biopsies.

SETTING: Targeted biopsies were performed based on surface architecture according to the Guelrud classification followed by 4-quadrant biopsies.

PATIENTS: A total of 95 patients were randomized.

INTERVENTION: A total of 46 patients underwent acetic acid staining, and 49 underwent i-scan imaging. Random biopsies were performed in 86 patients.

MAIN OUTCOME MEASUREMENTS: The primary outcome was the yield of SCE as confirmed by histologic assessment.

RESULTS: The diagnostic yield for SCE was significantly higher with targeted biopsies than with random biopsies in both groups combined (63% vs 24%; $P = .0001$). The yield of targeted biopsies was significantly greater with both i-scan (66% vs 21%; $P = .009$) and acetic acid (57% vs 26%; $P = .012$) technologies and did not differ between these groups. The accuracy for predicting SCE was 96% ($k = .92$) for i-scan and 86% ($k = .70$) for acetic acid analysis.

LIMITATIONS: No dysplastic lesions were found.

CONCLUSION: The i-scan or acetic acid-guided biopsies have a significantly higher diagnostic yield for identifying SCE, with significantly fewer biopsies, as compared with a protocol of random biopsies. Acetic acid and i-scan showed comparable results diagnosing SCE in our study. (Clinical trial registration number: NCT01442506.).

EXAMPLE OF A SINGLE PROGNOSTIC DEVICE STUDY (2)

Title: High-sensitivity troponin T is a prognostic marker for patients with aortic stenosis after valve replacement surgery.

Reference: Saito T, Hojo Y, Hirose M, Ikemoto T, Katsuki T, Kario K. *J Cardiol*. 2013 May;61(5):342-7. doi: 10.1016/j.jjcc.2013.01.005. Epub 2013 Mar 6.

Source: MEDLINE (through PubMed)

Abstract: **BACKGROUND:** Aortic stenosis (AS) is recognized as a cause of sudden cardiac death. Recently, the measurement of high-sensitivity troponin T (hs-TnT) has become possible. Several studies have clarified that hs-TnT is a marker to indicate mortality of cardiovascular diseases.

OBJECTIVES: To examine whether hs-TnT can be used as a prognostic marker to predict the operative outcome of AS.

METHODS: We enrolled 60 patients with AS (mean age= 68.7 ± 9.6 years, male/female=30/30). Cardiac catheterisation and echocardiography were performed to evaluate the severity of AS. Aortic valve replacement surgery was performed in all patients. We defined major adverse cardiac events (MACE) as composite events of heart failure, fatal arrhythmia, and all causes of death.

RESULTS: We followed up the patients for 922 ± 800 days. Mean left ventricular ejection fraction was $60.0 \pm 1.8\%$. Mean aortic valve area was 0.61 ± 0.03 cm². MACE occurred in 11 patients (18%), including 5 sudden cardiac deaths. We divided the patients into

three groups based on the percentile of the plasma levels of hs-TnT. Kaplan-Meier curve revealed a statistically significant difference in MACE rate among the groups (log-rank test, $\chi(2)=13.0$, $p=0.002$). We conducted a Cox proportional hazard analysis with a model including age, sex, estimated glomerular filtration rate, and hs-TnT tertile as explanatory variables to predict MACE. We found that hs-TnT tertile to be a significant factor to predict MACE (hazard ratio: 3.71, $p=0.03$).

CONCLUSIONS: hs-TnT can be a prognostic marker for patients with AS after valve replacement surgery.

C2.2 Comparative prognostic device studies

As in the case of comparative diagnostic and screening test accuracy studies (C1.2), professionals in medical practice are generally interested in the predictive accuracy of any new prognostic device as compared to or beyond the accuracy of existing prognostic test devices. Comparing prognostic tests devices within the same study – where all the subjects undergo the existing and new test devices – is likely to produce more valid results than comparing the accuracy estimates from different studies, each one evaluating a separate prognostic test device. As discussed under C.1.2, in the latter case many factors other than the use of a different device test may be responsible for a difference in accuracy, for example differences in the setting, population, and observers or users of the device.

Like comparative diagnostic or screening test device studies, comparative prognostic device studies allow us to study whether a new device improves prediction beyond what current tests already achieve, what the predictive accuracy is of combinations of different prognostic devices, in which order different prognostic tests should be conducted, and at which time point a specific new prognostic test device may be most suitable.

The advantages of comparative diagnostic and screening test accuracy studies (see C.1.2) also apply for prognostic comparative accuracy studies.

EXAMPLE OF A COMPARATIVE PROGNOSTIC DEVICE STUDY (ADDED VALUE STUDY)

Title: Does coronary CT angiography improve risk stratification over coronary calcium scoring in symptomatic patients with suspected coronary artery disease? Results from the prospective multicenter international CONFIRM registry.

Reference: Al-Mallah MH, Qureshi W, Lin FY, Achenbach S, Berman DS, Budoff MJ, Callister TQ, Chang HJ, Cademartiri F, Chinnaiyan K, Chow BJ, Cheng VY, Delago A, Gomez M, Hadamitzky M, Hausleiter J, Kaufmann PA, Leipsic J, Maffei E, Raff G, Shaw LJ, Villines TC, Cury RC, Feuchtner G, Plank F, Kim YJ, Dunning AM, Min JK. *Eur Heart J Cardiovasc Imaging*. 2013 Aug 21. [Epub ahead of print]

Source: MEDLINE (through PubMed)

Abstract: AIMS: The prognostic value of coronary artery calcium (CAC) scoring is well established and has been suggested for use to exclude significant coronary artery disease

(CAD) for symptomatic individuals with CAD. Contrast-enhanced coronary computed tomographic angiography (CCTA) is an alternative modality that enables direct visualisation of coronary stenosis severity, extent, and distribution. Whether CCTA findings of CAD add an incremental prognostic value over CAC in symptomatic individuals has not been extensively studied.

METHODS AND RESULTS: We prospectively identified symptomatic patients with suspected but without known CAD who underwent both CAC and CCTA. Symptoms were defined by the presence of chest pain or dyspnoea, and pre-test likelihood of obstructive CAD was assessed by the method of Diamond and Forrester (D-F). CAC was measured by the method of Agatston. CCTAs were graded for obstructive CAD (>70% stenosis); and CAD plaque burden, distribution, and location. Plaque burden was determined by a segment stenosis score (SSS), which reflects the number of coronary segments with plaque, weighted for stenosis severity. Plaque distribution was established by a segment-involvement score (SIS), which reflects the number of segments with plaque irrespective of stenosis severity. Finally, a modified Duke prognostic index-accounting for stenosis severity, plaque distribution, and plaque location-was calculated. Nested Cox proportional hazard models for a composite endpoint of all-cause mortality and non-fatal myocardial infarction (D/MI) were employed to assess the incremental prognostic value of CCTA over CAC. A total of 8627 symptomatic patients (50% men, age 56 ± 12 years) followed for 25 months (interquartile range 17-40 months) comprised the study cohort. By CAC, 4860 (56%) and 713 (8.3%) patients had no evident calcium or a score of >400, respectively. By CCTA, 4294 (49.8%) and 749 (8.7%) had normal coronary arteries or obstructive CAD, respectively. At follow-up, 150 patients experienced D/MI. CAC improved discrimination beyond D-F and clinical variables (area under the receiver-operator characteristic curve 0.781 vs. 0.788, $P = 0.004$). When added sequentially to D-F, clinical variables, and CAC, all CCTA measures of CAD improved discrimination of patients at risk for D/MI: obstructive CAD (0.82, $P < 0.001$), SSS (0.81, $P < 0.001$), SIS (0.81, $P = 0.003$), and Duke CAD prognostic index (0.82, $P < 0.0001$).

CONCLUSION: In symptomatic patients with suspected CAD, CCTA adds incremental discriminatory power over CAC for discrimination of individuals at risk of death or MI.

C3. Clinical utility of diagnostic, screening, monitoring and prognostic test devices: direct evidence

The clinical utility (impact or effectiveness) of a test device is commonly defined as the degree to which actual use of the test device will lead to the intended benefits, ideally added benefits, in the relevant health outcomes of the targeted patients and setting. This is similar to therapeutic devices, as discussed in Chapter 4.2 and 4.3, and above. The main difference between a clinical utility study of a test device versus the same study of a therapeutic device is that the latter may impact health outcomes directly, either in the short term or the long term, whereas test devices do this indirectly. Test devices provide information or results that change or guide therapeutic decisions and

management, which in turn lead to benefits or added benefits. Hence, when studying the added benefits of test devices for the relevant health outcomes in the targeted individuals and context, we not only study the test device but the test device in combination with subsequent treatments/management – also known as test-treatment studies (see also Figure 3, left panel).

The study designs discussed above under parts A and B can thus also be used – subject to the same pros and cons – to evaluate the clinical benefits or added benefits of using a certain test device combined with the treatment of choice. Here as well, a randomised design – notably a comparative effectiveness or pragmatic randomised design – is the preferred choice of design. Depending on how the test device is used, however, the type of randomisation may vary, as described in various papers (see references in Chapter 4). Also, the results of a test device and the subsequent therapeutic management decisions must be clearly linked. Without this, the results of a trial become more difficult to interpret. The link is needed when using a traditional randomised design, a novel randomised design (both part A) or a non-randomised design (part B). All the issues mentioned in part A and B apply equally when evaluating the clinical utility of non-therapeutic devices.

EXAMPLE OF A CLINICAL UTILITY STUDY TO DOCUMENT THE IMPACT OF USING A TEST DEVICE ON PATIENT OUTCOMES

Title: Cardiac computed tomography guided treatment strategy in patients with recent acute-onset chest pain: Results from the randomised, controlled trial: CArdiac cT in the treatment of acute CHest pain (CATCH).

Reference: Linde JJ, Kofoed KF, Sørgaard M, Kelbæk H, Jensen GB, Nielsen WB, Hove JD. *nt J Cardiol.* 2013 Aug 14. pii: S0167-5273(13)01545-3. doi: 10.1016/j.ijcard.2013.08.020. [Epub ahead of print]

Source: MEDLINE (through PubMed)

Abstract: **OBJECTIVES:** In patients admitted on suspicion of acute coronary syndrome, with normal electrocardiogram and troponines, we evaluated the clinical impact of a Coronary CT angiography (CCTA)-strategy on referral rate for invasive coronary angiography (ICA), detection of significant coronary stenoses (positive predictive value [PPV]) and subsequent revascularisations, as compared to a function-based strategy (standard care). Secondly we assessed intermediate term clinical events.

METHODS AND RESULTS: We randomised 600 patients to a CCTA-guided strategy (299 patients) or standard care (301 patients). In the CCTA-guided group referral for ICA required a coronary stenosis >70% or >50% in the left main, and for intermediate stenoses (50-70%), a stress test was used. A significant stenosis on ICA was defined as a stenosis $\geq 70\%$ or reduced FFR ≤ 0.75 in intermediate stenoses (50-70%). Referral rate for ICA was 17% with CCTA vs. 12% with standard care ($p=0.1$). ICA confirmed significant coronary artery stenoses in 12% vs. 4% ($p=0.001$), and 10% vs. 4% were subsequently revascularised ($p=0.005$). PPV for the detection of significant stenoses was 71% with CCTA vs 36% with standard care ($p=0.001$). Clinical events (cardiac death, myocardial infarction,

unstable angina pectoris, revascularisation and readmission for chest pain), during 120 days of follow-up, were recorded in 8 patients (3%) in the CCTA-guided group vs. 15 patients (5%) in the standard care group ($p=0.1$).

CONCLUSION: In patients with recent acute-onset chest pain, a CCTA-guided diagnostic strategy improves PPV for the detection of significant coronary stenoses, and increases the frequency of revascularisations, when compared to a conventional functional approach.

D. Linked-evidence approaches for both therapeutic and non-therapeutic (test) devices

In the main text, we described how linked-evidence approaches can be used to evaluate the clinical benefits or added benefits for the relevant, long-term health outcomes in the targeted individuals and context of medical devices. That is the case for both therapeutic and test devices. Whether quantitative or qualitative in nature (see main text), the linked-evidence approach is often a key component when building evidence for the clinical benefits of devices. In linked-evidence approaches, the results of studies on the impact of device use on short-term outcomes or surrogate/intermediate outcomes, or of test accuracy studies can be linked to the results of other studies – e.g. long-term therapeutic studies on the relevant health outcomes – in order to make inferences about the long-term effectiveness (and cost-effectiveness) of a medical device. Figure 3 below shows the difference between linked-evidence and direct evidence approaches for evaluating test devices, notably diagnostic and screening devices.

Linked-evidence approaches, preferably quantitative, may be justified if there is clear evidence that, for example, short-term or surrogate/intermediate outcomes are indeed associated with clinically meaningful outcomes – in other words, if there is sufficient confidence that indirect outcomes can predict clinical meaningful outcomes. For test devices (Figure 3) for which only a test accuracy study has been performed, a linked-evidence approach may be justified if evidence from longitudinal studies has shown that treatment of the condition of interest has a positive impact on relevant, long-term health outcomes in the same targeted individuals as those in which the new test device was studied.

Bear in mind that a linked-evidence approach is basically inferior to a direct evidence approach. For details about these approaches, we refer to Chapter 4, main text. Here we provide a few examples of linked-evidence approaches of different kinds (inferring from short-term outcomes to long-term outcomes, from surrogate to 'hard', relevant outcomes, and from test accuracy parameters to relevant health outcomes and even cost-effectiveness) and from different types of devices (therapeutic, diagnostic, and screening devices).

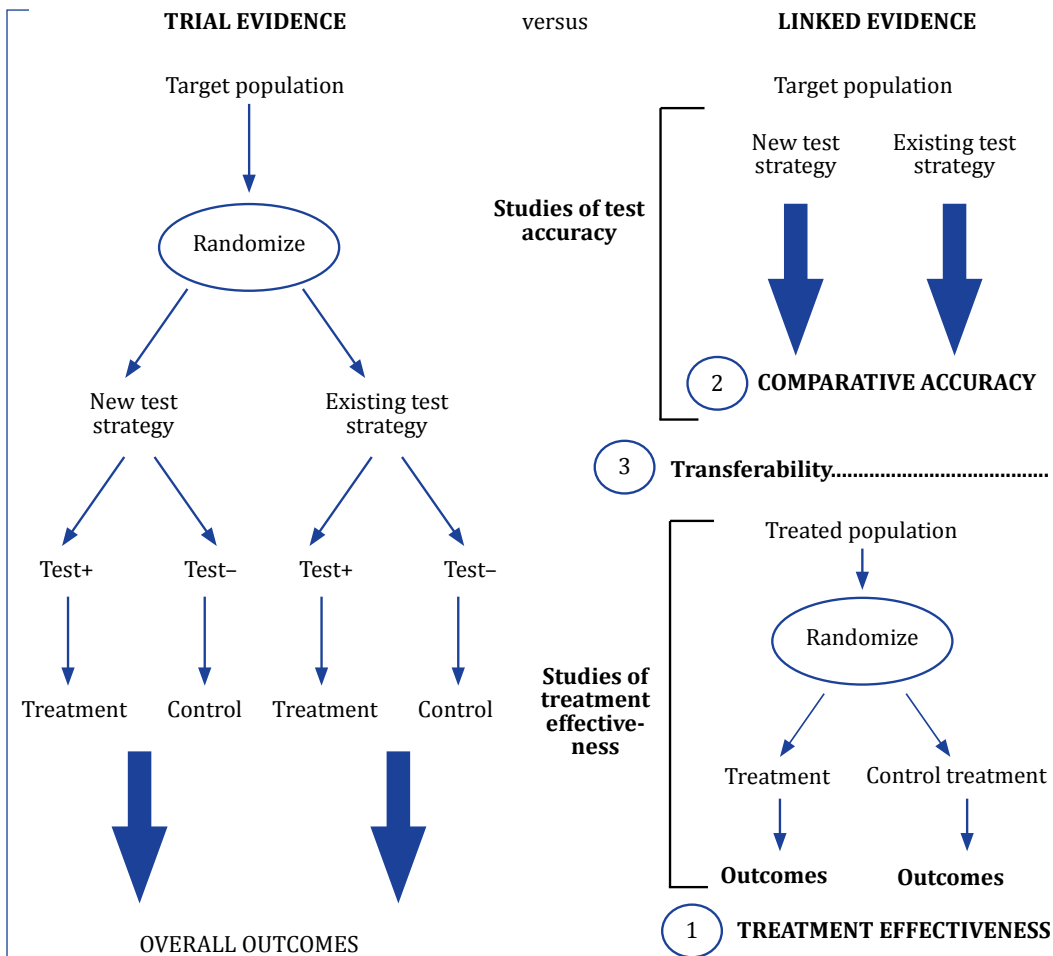


Figure 3. The use of direct evidence compared with linked evidence in the evaluation of test devices²⁴

EXAMPLES LINKED EVIDENCE APPROACHES USED FOR THE EVALUATION OF MEDICAL DEVICES

Title: The long-term cost-effectiveness of cardiac resynchronisation therapy with or without an implantable cardioverter-defibrillator.

Reference: Yao G, Freemantle N, Calvert MJ, Bryan S, Daubert JC, Cleland JG. *Eur Heart J*. 2007 Jan;28(1):42-51. Epub 2006 Nov 16.

Source: MEDLINE (through PubMed)

Abstract: AIMS: Cardiac resynchronisation therapy (CRT-P) is an effective treatment for

24 Lord S., Ghersi D., Simes J., Irwig L., (2005) *Medical Services Advisory Committee: Guidelines for the assessment of diagnostic technologies*. Canberra: Commonwealth of Australia). Note that the right panel – the linked-evidence approach – may be quantitative or qualitative in nature.

patients with heart failure and cardiac dyssynchrony with moderate or severe symptoms despite pharmacological therapy. The addition of an implantable cardioverter-defibrillator (ICD) function may further reduce the risk of sudden death. We assessed the cost-effectiveness of CRT-P compared with medical therapy (MT) alone, and the cost-effectiveness of CRT-ICD + MT compared with CRT-P + MT, on incremental cost per quality adjusted life year (QALY) and life year using data from two landmark clinical trials. **METHODS AND RESULTS:** A Markov model with Monte Carlo simulation to assess costs, life years, and QALYs associated with CRT (+/- ICD) and MT in patients with heart failure and cardiac dyssynchrony, on the basis of a UK healthcare perspective was constructed. NYHA class distribution and transitions, associated health utilities, rates and cause of hospitalisation and death were estimated from individual patient data from the CARDiac RESynchronisation in Heart Failure (CARE-HF trial). The estimated additional benefit on survival of an ICD was based on results from COMPANION. The base case analysis used 10 000 individual life-time simulations assuming a battery life of 6 years for CRT-P and 7 years for CRT-ICD. From a life-time perspective in a 65-year-old patient, the incremental cost-effectiveness of CRT-P compared with MT is 7538 euros (95% CI 5325-11,784 euros) per QALY gained and 7011euros (95% CI 5346-10,003 euros) per life year gained. The incremental cost-effectiveness of CRT-ICD compared with CRT-P is 47,909 euros (95% CI 35,703-79,438 euros) per QALY gained, and 35,864 euros (95% CI 26,709-56,353 euros) per life year gained. **CONCLUSION:** Long-term treatment with CRT-P appears cost-effective compared with MT alone. From a life-time perspective, assuming a reasonable life expectancy when receiving effective treatment for heart failure, CRT-ICD may also be considered cost-effective when compared with CRT-P + MT.

Title: The model-based cost-effectiveness analysis of 1-year adjuvant trastuzumab treatment: based on 2-year follow-up HERA trial data.

Reference: Shiroywa T, Fukuda T, Shimozuma K, Ohashi Y, Tsutani K. Breast Cancer Res Treat. 2008 Jun;109(3):559-66. Epub 2007 Jul 28.

Source: MEDLINE (through PubMed)

Abstract: **BACKGROUND:** Several randomized controlled trials have confirmed the usefulness of trastuzumab as an adjuvant therapy for HER2-overexpressed breast cancer patients; however, the costs for 1-year treatment are high. Therefore, we performed an economic analysis regarding the efficient distribution of medical resources. **METHODS:** To analyze the cost-effectiveness for a 1-year adjuvant trastuzumab treatment group compared with the observation group, we constructed a Markov model adopting a 3% per year discount rate for costs and outcomes. The time horizon was 50 years. The perspective was that of health-care payers, as only direct medical costs were calculated. The outcome was measured as life-year gained (LYG) from 2-year follow-up HERA trial data. **RESULTS:** The ICER of the standard setting (5 years efficacy and 50-60 kg patient weight) was JPY 2,600,000 (<euro>17,000) per LYG. The calculation results of other weight class ICER were JPY 2,200,000 (<euro>15,000) and JPY 3,300,000 (<euro>22,000) per LYG for the patients, respectively, who weighed less than 50 kg, and 60-75 kg. In the sensitivity

analysis, the period of trastuzumab efficacy was the most influential parameter for the result of cost-effectiveness. However, even if the trastuzumab efficacy were to continue for only 2 years, at least, which is a conservative setting judging from the joint analysis (NSABP B-31 and NCCTG N9831 trials), the ICER remains acceptable for any weight class. **CONCLUSION:** These results suggest that the 1-year adjuvant trastuzumab treatment is cost-effective. Both clinical and economic benefits were superior for the 1-year adjuvant trastuzumab treatment group compared with the observation group.

Title: Cost-effectiveness of magnetic resonance angiography versus intra-arterial digital subtraction angiography to follow-up patients with coiled intracranial aneurysms.

Reference: Schaafsma JD, Koffijberg H, Buskens E, Velthuis BK, van der Graaf Y, dekk. Stroke. 2010 Aug;41(8):1736-42. doi: 10.1161/STROKEAHA.110.585083. Epub 2010 Jul 1.

Source: MEDLINE (through PubMed)

Abstract: **BACKGROUND AND PURPOSE:** To follow up patients with coiled intracranial aneurysms, magnetic resonance angiography (MRA) is a promising noninvasive alternative to current standard intra-arterial digital subtraction angiography (IA-DSA). MRA test results do not always concord with those of IA-DSA, and the impact of discrepancies on health benefits and costs is unknown. We evaluated the cost-effectiveness of follow-up with MRA vs IA-DSA to assess whether in this setting MRA may replace IA-DSA. **METHODS:** We studied aneurysm occlusion on MRA in addition to follow-up IA-DSA in 310 patients with 341 coiled intracranial aneurysms. The observed sensitivity (82%) and specificity (89%) of MRA for detection of reopening with IA-DSA as a reference were used as input for a Markov decision-analytic model. Other determinants were derived from the literature. We compared life expectancy, quality-adjusted life-years (QALY), costs, and expected number of events for the two strategies. **RESULTS:** Follow-up with MRA yielded similar life expectancy (MRA, 26.66 years; IA-DSA, 26.63 years; difference, 0.03 years; 95% CI, -0.17-0.23) and QALY (MRA, 10.96; IA-DSA, 10.95; difference, 0.01 QALY; 95% CI, -0.05-0.08) at lower costs (MRA, \$7003; IA-DSA, \$8241 per patient; difference, -\$1238; 95% CI, -2617--36). The expected number of events was comparable except for complications from IA-DSA. **CONCLUSIONS:** MRA provided equivalent health benefits as IA-DSA and was cost-saving. MRA dominates and should replace routine IA-DSA to follow-up patients with coiled aneurysms.

Title: From accuracy to patient outcome and cost-effectiveness evaluations of diagnostic tests and biomarkers: an exemplary modelling study.

Reference: Koffijberg H, van Zaane B, Moons KG. BMC Med Res Methodol. 2013 Jan 31;13:12.

Source: MEDLINE (through PubMed)

Abstract: **BACKGROUND:** Proper evaluation of new diagnostic tests is required to reduce overutilisation and to limit potential negative health effects and costs related to testing. A decision analytic modelling approach may be worthwhile when a diagnostic randomized controlled trial is not feasible. We demonstrate this by assessing the cost-effectiveness of

modified transesophageal echocardiography (TEE) compared with manual palpation for the detection of atherosclerosis in the ascending aorta. **METHODS:** Based on a previous diagnostic accuracy study, actual Dutch reimbursement data, and evidence from literature we developed a Markov decision analytic model. Cost-effectiveness of modified TEE was assessed for a life time horizon and a health care perspective. Prevalence rates of atherosclerosis were age-dependent and low as well as high rates were applied. Probabilistic sensitivity analysis was applied. **RESULTS:** The model synthesized all available evidence on the risk of stroke in cardiac surgery patients. The modified TEE strategy consistently resulted in more adapted surgical procedures and, hence, a lower risk of stroke and a slightly higher number of life-years. With 10% prevalence of atherosclerosis the incremental cost-effectiveness ratio was € 4,651 and € 481 per quality-adjusted life year in 55-year-old men and women, respectively. In all patients aged 65 years or older the modified TEE strategy was cost saving and resulted in additional health benefits. **CONCLUSIONS:** Decision analytic modelling to assess the cost-effectiveness of a new diagnostic test based on characteristics, costs and effects of the test itself and of the subsequent treatment options is both feasible and valuable. Our case study on modified TEE suggests that it may reduce the risk of stroke in cardiac surgery patients older than 55 years at acceptable cost-effectiveness levels.

Title: Effectiveness and costs of screening for aneurysms every 5 years after subarachnoid hemorrhage.

Reference: Wermer MJ, Koffijberg H, van der Schaaf IC; ASTRA Study Group.

Source: MEDLINE (through PubMed)

Abstract: **BACKGROUND:** Patients who survive after subarachnoid hemorrhage (SAH) are at risk for a recurrence despite successful treatment of the ruptured aneurysm and may therefore benefit from screening for new aneurysms.

METHODS: We screened 610 patients with SAH with CT angiography 2-18 years after clipping of the aneurysms. Results of screening were used as input for a Markov decision model. We compared the expected number of recurrent hemorrhages, life expectancy, quality-adjusted life-years (QALYs), and costs associated with the strategies 'screening every 5 years' and 'no screening.' **RESULTS:** Screening individuals with previous SAH prevented almost half of the recurrences, slightly increased life expectancy (from 21.06 to 21.08 years), but reduced QALYs (from 12.18 to 12.04) and increased costs (from \$2,750 to \$4,165 per patient). Screening was cost-saving without increasing QALYs in patients with a more than twofold risk above baseline of both aneurysm formation and rupture and it was cost-saving while increasing QALYs if both risks were at least 4.5 times higher. In patients with reduced quality of life because of fear for a recurrence, screening increased QALYs at a maximum cost of \$17,422 per QALY. **CONCLUSIONS:** In general, screening patients with previous subarachnoid hemorrhage (SAH) cannot be recommended. Screening can save costs and increase quality-adjusted life-years (QALYs) in patients with a relatively high risk of both aneurysm formation and rupture, and increases QALYs at acceptable costs in patients with fear for a recurrence. More data are needed on risk factors

for aneurysm formation and rupture in patients with previous SAH and on management of fear for a recurrence to identify patients who can benefit from screening.

Title: Optimal screening strategy for familial intracranial aneurysms: a cost-effectiveness analysis.

Reference: Bor AS, Koffijberg H, Wermer MJ, Rinkel GJ. *Neurology*. 2010 May 25;74(21):1671-9.

Source: MEDLINE (through PubMed)

Abstract: OBJECTIVE: Individuals with a family history of subarachnoid hemorrhage (SAH), defined as 2 or more affected first-degree relatives, have an increased risk of aneurysm formation and rupture. Screening such individuals for intracranial aneurysms is advocated, but its effectiveness and cost-effectiveness are unknown, as are the optimal age ranges and interval for screening. METHODS: With a Markov model and Monte Carlo simulations we compared screening with no screening in individuals with a family history of SAH. We varied age ranges (starting screening at 20, 30, or 40 years old, ending screening at 60, 70, or 80 years old) and screening intervals (2-, 3-, 5-, 7-, 10-, and 15-year interval), and analyzed the impact in costs and quality-adjusted life years (QALY). RESULTS: Screening individuals with a family history of SAH is cost-effective. The strategy with the lowest costs per QALY was to screen only twice, at 40 and 55 years old. Sequentially lengthening the screening period and decreasing the screening interval yielded additional health benefits at acceptable costs up to screening from age 20 to 80 every 7 years. More frequent screening within this age range still provided extra QALYs, with an incremental cost-effectiveness ratio more favorable than 26,308/QALY (\$38,410/QALY). CONCLUSION: This study provides evidence for recommendations to screen individuals with 2 or more first-degree relatives with subarachnoid hemorrhage. The optimal screening strategy according to our model is screening from age 20 until 80 every 7 years given a cost-effectiveness threshold of 20,000/quality-adjusted life year (QALY) (\$29,200/QALY).

APPENDIX V

STATISTICAL METHODS TO CORRECT FOR CONFOUNDING IN NON-RANDOMISED (DIRECT EVIDENCE) STUDY APPROACHES

In this Appendix we briefly explain different methods that can be used in non-randomised studies to limit or adjust the potential for bias or confounding bias (see also the references in main text).²⁵

A. Methods used at the study design (set-up) stage

A1. Restriction

In this method, the aim is to create two groups that are similar on one or more known and major confounding variables by restricting the inclusion or exclusion criteria. For example, if age, gender or a specific disease variant are known to have a major influence on the mechanism/pathway of the device's benefits, we can restrict the two study groups only to a specific gender, age range, or disease variant. Another method is to exclude subgroups which are known *a priori* to influence the pathway of the device's effect. A disadvantage of restriction is that the results may have limited applicability – reduced generalizability of device use – because they are based on a more homogeneous and less representative study group.

A2. Matching

Matching is the process of searching, for each individual in the index (device) group, a subject in the control group who is similar on the set of most important confounding

25 Cousens S., Hargreaves J., Bonell C., Armstrong B., Thomas J., Kirkwood B.R., Hayes R., Alternatives to randomisation in the evaluation of public-health interventions: statistical analysis and causal inference. *J Epidemiol Community Health* 2011;65:576-581; McNamee R., Regression modelling and other methods to control confounding. *Occup Environ Med* 2005;62:500-506 doi:10.1136/oem.2002.001115.

variables. For example, for an index-group female, 40 years of age, no previous disease, particular disease stage, we would search and include a control-group female with a similar age range (40-45), no previous disease and same disease stage. One limitation to matching is that it is tedious work and therefore an expensive undertaking when there are many known *a priori* confounding variables. It is therefore often performed only for the two or three most important confounders.

B. Methods used during statistical analysis²⁶

B1. Stratification

This method involves stratifying the study participants by the main confounders for which results should be controlled (e.g. by age groups or gender), and estimating the effect of the device per stratum (e.g. per age or gender group). In order to ensure that strata with a larger group of participants receive larger weights when estimating the association between the difference in device use and the observed benefits, we simply estimate a weighted average for the benefits of the device over these strata. We then end up knowing the unbiased and valid effect of the device's use compared to the control group, and adjust for those confounding variables. The most common statistical weighting approach used in this case is the Mantel-Haenszel method. Stratification also allows us to assess whether the device has different benefits in different subgroups, provided we have large enough numbers per subgroup. One disadvantage of this method is that strata with larger numbers of participants will generate more precise estimates of the device's benefits (with a smaller standard error) than smaller strata.

B2. Regression modelling

This statistical approach is used to control for many confounders simultaneously, unlike all the above methods, which control for one or a few confounders. Regression modelling adjustment is the most widely used and perhaps the best method to adjust for other influential factors when studying the benefits of device use versus a control group in a non-randomised study. The use of this method only requires that investigators predefine the known influential or confounding factors for the device and outcome being studied and subsequently measure the presence or absence of each predefined confounder in each study participant in both study groups. One disadvantage of these analyses is that confounders can only be controlled for if they are known *a priori* and properly measured in each study participant.

²⁶ To correct for confounding during the analysis phase, the process of designing and conducting a non-randomised study on the benefits of a device must meet certain requirements.

B3. Propensity score methods

This relatively novel approach – which also uses a regression modelling approach – is used most notable in Non Randomised Studies on the benefits of interventions where the number of study participants is limited and the number of confounders is relatively large. A propensity score is also called a multiple confounder score, with the value of each study participant’s characteristics and confounders being summarised to a single variable. It has two major stages. First, when comparing the two non-randomised groups – participants with the device versus those without the device – we estimate the probability that each individual had received the device according to their characteristics and confounding variables (using a logistic regression model). The probability per participant is known as the propensity score. Second, we then restrict, match or stratify the two study groups according to this propensity score variable, or use a regression model. There has been extensive research comparing the advantages and disadvantages of propensity methods for confounder control to the traditional methods described above.

APPENDIX VI

CONCISE COMPARISON

MEDICAL DEVICES VERSUS

PHARMACEUTICALS

In this appendix we compare medical devices and medicinal products (pharmaceuticals) on all aspects relevant to clinical research. Table 1 summarises the similarities between the two groups. In essence, basic regulatory systems are in place for both pharmaceuticals and medical devices. There are European directives (three for devices and one for pharmaceuticals) defining what a device or pharmaceutical is, and extensive guidelines for market access are available in both cases. These guidelines describe a process that must meet certain requirements rather than request that a specific approach be used to produce predefined results. Decisions are always on a case-by-case basis, depending on the evidence presented. Evidence is based on a dossier provided by the manufacturer. Whereas for pharmaceuticals global harmonisation has been achieved, it is still underway for medical devices. Once the product has been marketed for a specific indication or intended purpose, off-label use is common for medical devices and pharmaceuticals. As in the case of pharmaceuticals, postmarketing surveillance (PMS) is required for medical devices. For medical devices implementation of PMS could improve.

Table 1: Similarities between regulation of medical devices and pharmaceuticals

Medical devices (see Appendix III)	Pharmaceutical products
3 European directives covering all products (MDD, AIMD, IVDD, see Appendix III)	1 European directive (2001/83/EC)
CE marking for (technical) safety	Good Manufacturing Practice (GMP) guidelines according to ICH
NEN-ISO 1455 for GCP guidelines	Good Clinical Practice (GCP) guidelines according to ICH
GHTF is working on harmonisation but is not yet finished > difference USA & EU	International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) global harmonisation of guidelines for access
Clinical evaluation based on the literature, existing studies or new clinical investigations according to risk category and depending on equivalence data	General requirements for phase 1, 2 and 3 studies + judgement by experts at the EMA's CHMP
One device is often useful for many different indications; implicit frequent off-label use	Off-label use is frequent
Postmarketing Surveillance (PMS) is required; in place but improvement is needed	PMS is required, and in place.

Table 2 indicates differences between medical devices and pharmaceuticals that are beyond the remit of this committee. One relevant issue is that the European Medicines Agency (EMA) is an EU agency, whereas the Notified Bodies that judge the application dossiers are independent and thus more difficult to monitor. The main differences are related to differences between the two industrial sectors, medical devices and pharmaceuticals, i.e. the life cycle and innovation process. The pharmaceutical business model is under pressure owing to an empty pipeline, with only a small number of innovative pharmaceutical products entering the market. In contrast, the medical devices industry (often SMEs) is still considered innovative and collaborates closely with health care professionals and hospitals. One major difference is that, once approved, a pharmaceutical product does not change during its lifetime and so the relevant clinical data remain valid. In medical devices, regular incremental changes (and improvements) alter the product over time, making it more difficult to interpret the clinical data in light of the latest version. Another difference is the stronger focus on efficacy at the point of market access for pharmaceuticals. The CE marking procedure tends to look more closely at technical safety and performance than at clinical benefit. Health technology assessment (both theory and practice) is applied regularly for pharmaceuticals, and the reimbursement system in the Netherlands builds on different legislation.

Table 2: Differences between medical devices and pharmaceuticals beyond the remit of the committee

Medical devices	Pharmaceutical products
Notified Bodies and Competent Authorities responsible for market access	EMA procedure for all new medicinal products to obtain market access
R&D trajectory ~ 1-5 years; life cycle of product ~ 2.5 years	R&D trajectory ~ 15 years; life cycle of product > 20 years
Product may change every 18 months; clinical data are no longer valid for the new version	Product remains unchanged for 20 years; clinical data therefore remain valid
Dynamic industry with many SMEs: Buyer's market (prices depend on market, with budget for research and marketing therefore under pressure)	Static, conservative, almost institutional industry. Seller's market (high prices for a quick ROI)
Innovation, often in close collaboration with clinicians/industry	Big pharma has an empty research and development pipeline
Little application of health economics, cost-effectiveness and HTA	Frequently application of health economics, cost-effectiveness, and HTA
Reimbursement based on <i>Stand van Wetenschap en praktijk</i> (CvZ) and DBCs (NZA)	Reimbursement based on pharmacoeconomic dossier for new drug, me-toos in cluster with average pricing; CvZ responsible

Table 3 compares differences between medical devices and pharmaceuticals that are crucial for designing research into device safety, performance and clinical benefit.²⁷ This relates to the risk categories, the mode of action, user interference, direct and

²⁷ Therapeutic and diagnostic devices outcomes research. S.J. Ackerman, M. Dix Smith, J. Ehreth, R. Eldessouki, E. Sullivan (Eds). *ISPOR 2011*: ISBN 978-09743289-2-8. Page 12.

indirect effect, learning curves and all aspects that are described in more detail in Chapters 2 and 4 of this report. These differences are reflected in the way clinical evaluation is designed, and in the burden of proof that is required by the experts judging the data. This is why the committee is not in favour of applying a burden of proof for medical devices similar to that required for pharmaceuticals.

Table 3: Differences between medical devices and pharmaceuticals with consequences for clinical study approaches

Medical devices	Pharmaceutical products
Different risk categories (high to low)	No specific risk categories
Performance (of the product) compared to risk	Efficacy (in subject receiving the drug) is based on benefit/risk ratio
Frequent indirect effect, user interference or intermediate decision	Always direct effect without user interference
Local, supportive, or information-generating effect, based on mechanical engineering, materials or physical, electrophysical or chemical principles	Biologically active, based on pharmacological, immunological, or metabolic action
Clinical evaluation based on literature, existing studies or new clinical investigations according to risk category and depending on equivalence data	General requirements for phase 1, 2 and 3 studies + judgement by experts at the EMA's CHMP

APPENDIX VII

INTERVIEWS AND EXTERNAL CONSULTATIONS

Consulted people (with affiliation)

Hans Arendzen, Leiden University Medical Center (LUMC), Innovative Medical Devices Initiative (IMDI)
Gert Bos, BSI Group and European Association of Notified Bodies for Medical Devices
Bart Blokhuis, Medtech Partners
Robert Geertsma, National Institute for Public Health and the Environment (RIVM)
Michiel Jannink, Demcon
Marijke Janssens, Netherlands Organisation for Health Research and Development (ZonMw)
Eric Klasen, Medtronic
Jos Kraus, Academic Medical Center (AMC)
Frits Lekkerkerker, Nederlandse Vereniging Medisch Ethische Toetsings Commissies (NVMETC)
Gerry Ligtenberg, Health Insurance Board (CvZ)
Susanne Ludgate, Medicines and Healthcare products Regulatory Agency (MHRA), European Commission and Global Harmonization Task Force (GHTF)
Maarten Simoons, Erasmus MC
Veronica van Nederveen, Ministry of Health, Welfare and Sport, The Netherlands
Hans Reiber, Medis, IMDI
Kees Smaling, Siemens Nederland
Henk Viëtor, Skyline Diagnostics,
Fokko Wieringa, TNO Science & Industry
Gert-Jan van der Wilt, UMC St Radboud, Nijmegen

Review Committee

At the request of the Board of the KNAW a draft of this report was reviewed by four reviewers with different backgrounds. They are not responsible for the final report.

Bert Boer, Health Insurance Board (CvZ)
Adam Cohen, Center for Human Drug Research (CHDR)
Hans Hofstraat, Philips Research
Chris Hyde, University of Exeter and National Institute for Health and Care Excellence (NICE)

