OPEN

Beyond the RCT: When are Randomized Trials Unnecessary for New Therapeutic Devices, and What Should We Do Instead?

Arsenio Páez, PhD,†‡ Maroeska Rovers, PhD,‡§ Katrina Hutchison, PhD,‡¶ Wendy Rogers, PhD,‡¶|| Baptiste Vasey, MMed,‡** and Peter McCulloch, MD‡**✉, for the IDEAL Collaboration*

Objective: The aim of this study was to develop an evidence-based framework for evaluation of therapeutic devices, based on ethical principles and clinical evidence considerations

Summary Background Data: Nearly all medical products which do not work solely through chemical action are regulated as medical devices. Their huge range of purposes, mechanisms of action and risks pose challenges for regulation. High-profile implantable device failures have fuelled concerns about the level of clinical evidence needed for market approval. Calls for more rigorous evaluation lack clarity about what kind of evaluation is appropriate, and are commonly interpreted as meaning more randomized controlled trials (RCTs). These are valuable where devices are genuinely new and claim to offer measurable therapeutic benefits. Where this is not the case, RCTs may be inappropriate and wasteful.

Methods: Starting with a set of ethical principles and basic precepts of clinical epidemiology, we developed a sequential decision-making algorithm for identifying when an RCT should be performed to evaluate new therapeutic devices, and when other methods, such as observational study designs and registry-based approaches, are acceptable.

Results: The algorithm clearly defines a group of devices where an RCT is deemed necessary, and the associated framework indicates that an IDEAL 2b study should be the default clinical evaluation method where it is not.

Conclusions: The algorithm and recommendations are based on the principles of the IDEAL-D framework for medical device evaluation and appear eminently practicable. Their use would create a safer system for monitoring innovation, and facilitate more rapid detection of potential hazards to patients and the public.

Keywords: clinical trials, evaluation, evidence, medical devices, RCT, regulation, safety

From the *Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK; †Northeastern University, Bouvé College of Health Sciences, Boston, MA; ‡IDEAL Collaboration, Nuffield Department of Surgical Sciences, University of Oxford, UK; §Departments of Operating Rooms and Health Evidence, Radboud University Medical Center, Nijmegen, The Netherlands; ¶Department of Philosophy, Macquarie University, Sydney, Australia; ||Department of Clinical Medicine, Macquarie University, Sydney, Australia; and **Oxford University Hospitals, IDEAL Collaboration, Nuffield Department of Surgery, John Radcliffe Hospital, Oxford, UK.

✉peter.mcculloch@nds.ox.ac.uk.

Authors Contributions: A.P.: literature search, analysis, interpretation, figures, writing, editing; M.R.: analysis, interpretation, writing; K.H.: analysis, interpretation, figures, writing; W.R.: analysis, interpretation, writing; B.V.: analysis, interpretation, writing; P.M.: analysis, interpretation, figures, writing, oversight.

The authors report no conflicts of interest.

Ethics Committee Approval: No human subjects or participant data was used in this work. No ethics committee approval was required.

ISSN: 0003-4932/21/27502-0324

DOI: 10.1097/SLA.0000000000005053

(*Ann Surg* 2022;275:324–331)

THE PROBLEM

Nearly all medical products which do not work solely through chemical action, from condoms to surgical robots, are regulated as medical devices.[1] Given the huge range of purposes, mechanisms of action and inherent risk of devices, they pose unique challenges for regulation.[2,3] The medical device industry has grown rapidly, with device sales nearly doubling over 10 years to US$172 billion in the United States (2015) and €110 billion in the European Union (EU) in 2015.[4] Alongside industry expansion, concern about the adequacy of the clinical evidence needed for market approval has been growing,[2] fuelled by high-profile implantable device failures, (silicone breast implants,[5] hip devices,[6,7] implantable defibrillators,[8]) and adverse effects (transvaginal mesh[9]) from licensed products. As a result there have been growing calls for more stringent regulation of device safety and clinical effectiveness, especially for implantable devices.[2,3,10–14]

There are actually 3 interlocking problems here. First, there are important defects in current regulatory evidence requirements, as described below. Second, the numbers of new devices to be evaluated are huge, especially now that e-health and software applications are classified as medical devices (Software as Medical Devices: SaMD) by both the Food and Drug Administration (FDA) and the new Medical Device Regulations (MDR) in Europe.[15,16] Third, calls for more rigorous evaluation lack clarity about what kind of evaluation is most appropriate, and are commonly interpreted as meaning more randomized controlled trials (RCTs). This may not be the answer, however, and insufficient attention has been paid to what else we could do. In this article, we integrate ethical and practical considerations to offer a principled framework for deciding when an RCT is appropriate to evaluate new therapeutic devices, and, when it is not, what should be done instead.

CURRENT SYSTEMS AND THEIR WEAKNESSES

The standard of evidence presently required for approval of medical devices is generally lower than for drugs.[1,2,14] Although new drugs must show "substantial evidence of effectiveness and safety" through clinical trials, medical devices need only demonstrate a reasonable assurance of safety and effectiveness in the United States, and safety and "performance" in the EU.[17,18] Both systems classify devices according to the level of risk to the patient, and require more stringent evidence with predicted higher risk. Additionally, although developers of new drugs can receive 20 years of patent protection, and ≥5 years of market exclusivity (US jurisdiction), there are no provisions for medical device exclusivity.[19] Device manufacturers may seek more rapid FDA clearance by claiming substantial equivalence to an existing approved device, yet file a patent application claiming device novelty.[19,20] The US and EU systems differ markedly, but both have weaknesses which allow some devices to gain market approval with little or no clinical evidence (Boxes 1 and 2).

Box 1 The US FDA System

The FDA (US) System

In the United States, the Food and Drug Administration (FDA) requires evidence of safety and efficacy from rigorous ''pivotal'' clinical trials for pre-market approval (PMA) of more invasive and higher risk devices, such as implantable heart valves (Table 1).[22] Class III devices in the FDA system require evidence of effectiveness and safety from ''clinical trials,'' but intermediate-risk (class II) medical devices can enter the market with limited review, and potentially no clinical evidence.[18,22,65] This occurs when ''me too'' devices gain approval through the 510(k) notification process by claiming the same intended use as existing devices and equivalence, but not superiority, in performance while claiming advantages in other areas, such as cost, speed, durability, ease of use, software accessibility, among others.[1,10,12,13,21,22] Rigorous direct evidence of safety and effectiveness from clinical trials is not required, and average total review time is 54 days versus 238 days for PMA applications.[1]

Problem: This creates a loophole allowing devices to be approved if they are substantially equivalent to legally marketed ''predicate'' devices that are not subject to PMA.[1,3,13,22] If a predicate device is withdrawn for safety concerns, devices approved based on ''substantial equivalence" can remain on the market. For example, the ProteGen Sling vaginal mesh was recalled in 1999, but devices approved via equivalence with ProteGen were marketed for another two decades.[66] The practice of ''grandfathering'' greatly increases possibilities for exploiting the 510K loophole. Device sponsors can identify a cascade of predicate devices in a sequence reaching back to devices marketed before 1976, when the current regulations were enacted (and which are therefore exempt from PMA requirements).[67]

Box 2 The EU MDR System

The EU System

Until recently, many devices were authorized in the European Union without evidence from well-designed clinical trials.[25,26] In 2017, however, the, Conformité Européenne (CE) mark requirements were reformed in new medical device regulations (MDR) legislation due to take effect in on 26th of May, 2021. The MDR focuses on clinical performance rather than effectiveness. Nevertheless, the new MDR sets a stricter burden of proof and requires evidence of effectiveness for all class 2b, 3, and implantable devices (Table 1). In addition it requires that evidence must be updated for existing, approved devices to meet the new standards (Table 1).

Problem: The new MDR sets a stricter burden of proof and requires evidence of effectiveness for all class 2b, 3, and implantable devices but effectiveness is not clearly defined. The EU system places responsibility on the manufacturer to design clinical trials. Their design must satisfy a national regulator (Competent Authority) and the results must satisfy a ''Notified Body'' in relation to safety and performance, but neither of these is permitted to advise manufacturers on study design. The quality of evidence and the place of RCTs therefore remain undefined. The MDR mandates enhanced regulatory requirements post-market (ie, once CE marking has been granted), but these are also ill-defined.[2,3,14]

In the United States, ''me too'' devices can be authorized to enter the market through the 510(k) notification process, by claiming equivalence to (but not superiority over) existing devices, allied to advantages in other areas, such as cost, speed, durability, ease of use, or software accessibility.[1,10,12,13,21,22] The practice of ''grandfathering,'' through which device sponsors can identify a cascade of predicate devices, each supposedly substantially equivalent to earlier devices, greatly increases possibilities for exploiting the 510K process. The metal-on-metal hip implant, for example, was cleared citing 95 predicate devices, but without any direct clinical performance data, ultimately leading to considerable patient harm.[23] In a 2014 study, only 10% to 15% of 510(k) submissions contained any clinical data.[12] Reforms to FDA regulations have made it more difficult for high-risk devices to obtain authorization without clinical studies, but have not eliminated the loophole.[24]

Until recently, many devices were authorized in the EU without evidence from well-designed clinical trials.[25,26] In 2017, however, the Conformité Européenne (CE) mark requirements were reformed in new medical device regulations (MDR) legislation due to take effect in 2021. The MDR focuses on clinical performance, (whether the device does what the manufacturer intended it to do), rather than effectiveness, (whether it improves patient outcome). Nevertheless, the new MDR sets a new stricter burden of proof and requires clinical evidence of effectiveness for all class 2b, 3, and implantable devices (Table 1) and fresh, updated evidence to meet the new standards for existing, approved, devices.[27] Neither the types of studies nor the reporting standards required are clearly defined however, creating uncertainty about the evidence required to achieve certification.[28]

Regulation influences device development throughout their life cycle, from early development through manufacturing, clinical evaluation, and post-marketing surveillance. Overly burdensome regulation may hinder device innovation by increasing the cost and complexity of clinical evaluation requirements for device certification, particularly impacting small manufacturers, start-up device companies, and device innovators.[28,29] Conversely, a lack of regulatory clarity or stringency may also hinder device innovation[29,30] and risk patient safety.[31]

The regulation of medical devices is therefore intended to strike a pragmatic balance between bringing innovative technology to the public as rapidly as possible and assuring that it is safe and effective.[1,31] However, the desire for flexibility has created systems in which no precise rules govern what degree of change between devices requires new clinical studies, and loopholes allow approval of higher risk devices without substantial direct clinical evidence. A large number of new devices have therefore been introduced with little or no clinical evidence because they do not make claims of superior efficacy or effectiveness.[13,32]

Paradoxically, incremental changes to devices may have greater potential to cause harm than substantial innovations, because clinical performance data requirements are more likely to be minimal. Unfortunately even minor changes in design can impact safety, as when a change in the location of the etched company symbols or device numbers on the femoral component of hip replacement devices were found to cause premature device fracture.[6,7] The demands for better evidence that devices, particularly implantable devices, are safe and clinically effective therefore seem reasonable.[10]

WHY MORE RCTS MAY NOT BE THE ANSWER

Internationally, drug licensing systems effectively require RCT evidence of benefit or equivalence for nearly all new drugs. There is, therefore, a case for suggesting that all new devices should also require RCT evidence, but there are several reasonable objections to this approach.[33] RCTs provide the best evidence for whether a treatment has greater beneficial effects than another intervention or placebo.[34,35] However, RCTs are increasingly costly and time-consuming,[36] whilst device commercial life-cycles are often short, and many ''me too'' devices make no claim of superior effectiveness. Therapeutic devices also share with surgical operations and other complex interventions several important characteristics which make RCTs challenging to perform and necessitate specific preliminary studies, as highlighted in the IDEAL Framework.[37–39] The primary concern of device regulation is the risk of harm, for which other study designs are more appropriate evaluation tools than RCTs.[40–42]

TABLE 1. Device classifications in the United States[68] and EU[69]

	US Device Regulations		
	Class I	**Class II**	**Class III**
Description	Low risk	Intermediate risk	High risk: "supports or sustains life, is implanted in the body, or has the potential for unreasonable risk of illness or injury"
Requirements	"General controls" Good manufacturing practices Standards and reporting adverse events Registration General recordkeeping requirements	"General controls with special controls" Labeling requirements Device-specific mandatory performance standards Device-specific testing requirements	"General controls and premarket approval"
Examples	Surgical gloves, condoms, oxygen masks	Knee prosthetics, single use scalpels	Pace-makers, breast implants
Requirement for clinical trial	No	Maybe	Yes

	EU Regulations			
Description	**Low, low/medium Class I and IIa**	**Medium Class IIb**	**Medium-high Class III**	**High Class III**
Requirements	—	—	—	—
Example	Reusable surgical instruments, Indwelling urinary catheters	Surgically invasive devices intended for transient use, Peripheral vascular grafts and stents	Blood bags	Surgically invasive devices in contact with the CNS
Requirement for clinical trial	No	Yes	Yes	Yes

Finally, there is a significant mismatch between the number of new devices introduced per year and the available capacity for conducting RCTs. Many therapeutic devices also undergo frequent modifications, with possible unintended consequences. The rapidly growing number of health software applications classified as medical devices (Software as Medical Devices: SaMD), pose unique challenges which worsen this problem. Software applications can control devices, provide both decision-triggering information (eg, blood glucose meters) and decision support (eg, ECG interpretation), and utilize artificial intelligence to continually improve device performance.[43,44] Frequent modifications or updates, including automatic updates, are part of their normal life cycle, but can alter the core functionality of the device, requiring risk and performance reassessment.[45–47] The most appropriate way to evaluate their effectiveness and safety is still debated, but the volume of SaMDs and their plasticity over time makes it infeasible to conduct RCTs for each of them.[47–49]

HOW CAN ETHICS HELP US?

In key respects, the problem is an ethical one, and ethical principles should be integrated into any framework informing device evaluation and regulation. We therefore draw on the well accepted principles of beneficence, non-maleficence, respect for patient autonomy, and justice to help frame the requirements proposed here for regulatory device evaluation.[52] Given the potential for untested devices to cause harm, the ethical principle of non-maleficence is paramount. However, as indicated above, the nature and likelihood of harm are unknown in many cases, so the precautionary principle is also highly relevant. This principle is useful in situations in which there is limited evidence about the potential outcomes associated with various choices.[50,51] It proposes that where decisions may cause harm, the option least likely to cause harm should be chosen, other things being equal.[51] This leaves open the option of choosing a riskier option if other factors strongly favor it. Taken together, non-maleficence and the precautionary principle support rigorous evaluation of safety risks and rapid responses to signals of increased risk of harm. To achieve this requires rigorous prospective studies of obvious potential risks, and large enough high-quality datasets to identify unexpected safety risks and harm signals as early as possible.

Beneficence, or doing good, is next in importance. Evaluation to measure therapeutic benefit is essential, as if there is no therapeutic benefit, all harms become ethically unacceptable. Third, justice requires that we act on the basis of fair adjudication between competing claims and fair distribution of scarce resources.[52] Justice can be served by showing that a device is cost-effective, the benefit it provides outweighing any opportunity costs, and it is less costly than alternative treatments. Justice also requires that evaluation should not unreasonably increase costs or delays, so that patients can receive the benefits of new devices in a timely manner and innovation is not stifled. Finally, the principle of respect for autonomy can be respected by ensuring that patients have full information about both risks and benefits, including an explanation of what remains unknown about these, and free choice in deciding whether or not to use the device.

Taken together, these principles support the idea that rigorous evaluation of safety risks and rapid responses to signals of increased risk of harm should be the most important focus of regulatory evaluation.

Translating these principles into a prescription for action points to the need for:

- Use of the most sensitive methods possible for estimating risks and detecting signals of harm, consistent with sustainable costs and delays.
- Convincing direct clinical evidence that most patients derive measurable benefit is essential for therapeutic devices.
- Valid evidence of cost-effectiveness, consistent with the stage of evolution of the device, is highly desirable.
- Evaluation that does not reduce or delay benefits by imposing disproportionate costs or delays on innovation. Evaluation should therefore be as simple and cost-effective as possible, utilizing agreed minimum datasets[53] and high-quality real-world evidence where appropriate (eg, administrative claims data, and registries.)
- Transparency about risk: All patients should give informed consent including full explanation of potential risks and benefits, the extent of current experience and knowledge, and the potential existence of undiscovered risks.
- Asking patients for consent to the re-use of their data in future studies: this will improve longer-term analysis of safety, maximize the value of information, and benefit others.
- Minimizing restrictions on access to data, consistent with considerations of cost, conflicts of interest, and patient confidentiality.

When Should Devices be Subjected to an RCT? Our Proposal and Decision-Making Tool

How do these principles help us decide when an RCT is required for a new device? A key step in deciding whether to adopt a new treatment is to compare its effectiveness as fairly as possible with current best practice. The main value of RCTs is to provide a valid unbiased estimate of relative efficacy between therapeutic interventions. However, as discussed above, many new devices do not claim to be ''new'' in principle, mode of action or materials, nor do they claim superior relative efficacy.[41] If superiority is not claimed, and harm signals are best evaluated using other methods, the basis for requiring an RCT is seriously undermined, especially since it may entail important costs and delays, and may be impractical due to the volume of devices being produced. We argue, based on the principle of justice, that devices which do not claim superior efficacy and do not claim to be new in terms of mechanism of action should not be subjected to randomized trials. A requirement for noninferiority RCTs to show that efficacy is not significantly worse than current treatment would introduce additional complications as noninferiority trials are usually larger and therefore more expensive than conventional RCTs.[54,55] Given the acceptance of pragmatic arguments against conventional RCTs for devices that do not claim superiority, the case for noninferiority RCTs is difficult to justify. Doing an RCT when it can serve no useful purpose is a form of research waste, and should be avoided. Based on the above considerations, we offer a framework for evaluating new devices and in particular, identifying when an RCT is required (Fig. 1), and if not required, identifying what type of evaluation should be performed instead.

When proposing guidance which will be applied to unforeseeable future situations, the precise use of words becomes important. In this case it is important to define what is meant by ''new'' and what is meant by ''superiority.'' The issue of ''newness'' has previously been considered in precisely this context by some of the authors.[42] Treatments are clearly new in a general sense if their principle or mode of action is different from what has been used before, or in some cases if the materials used are different. Treatments which are new in this sense merit an RCT to evaluate their relative efficacy.[56] In circumstances where there is real doubt as to whether an existing treatment works in a specific subgroup of patients (eg, children), or an existing tool works in a new anatomical location or clinical context (eg, low-income country surgery),[42] a separate RCT might sometimes be justified. Where manufacturers claim a device is innovative (ie, new in principle) an RCT to demonstrate safety and efficacy is required. Conversely, treatments which are new to a hospital, theatre team or individual surgeon may be associated with some risks of harm, but an RCT for each new setting would be inappropriate for determining comparative efficacy.

''Superiority'' in the context of therapeutic devices relates to superior efficacy, effectiveness or cost-effectiveness in achieving specific therapeutic aims compared with current best treatment.[56,57] The question of how much improvement in these measures constitutes meaningful superiority is one which can only be answered in the context of the values of the patients and clinicians involved. For some patient-reported outcomes, this can be estimated from clinimetric studies and minimally important clinical difference data.[58] When a new device claims equal efficacy to other devices but superior cost-effectiveness because it is significantly cheaper, this could be adequately explored in economic modeling studies[59] if the device's efficacy is widely accepted, or if it becomes cheaper to produce without altering its design. Such studies provide the evidence required to assess questions of justice and to limit waste. If the claim of superior cost effectiveness is through a new mechanism of action, however, an RCT would be appropriate.

A pragmatic rule of thumb might be that where there is insufficient belief among stakeholders in the possible superiority of the new device to make a trial feasible, and/or there is no claim of device superiority, RCTs would not be worthwhile. It should be noted that there is no sustainable ethical basis for neglecting to do an RCT in the converse situation, where stakeholder belief in the superiority of the new device is high in the absence of adequate evidence, leading to ''loss of clinical equipoise.''

What to do When an RCT Is Not Required

Arguably the greatest need in present regulatory science is clear guidance on the type and extent of clinical studies to be used when an RCT is not considered appropriate. Our ethical principles indicate that clinical evaluation is needed primarily to evaluate risks (nonmaleficence and the precautionary principle), but also to demonstrate effectiveness (beneficence).[51,52] The rigor of the analysis applied should be determined by the perceived patient risk associated with the treatment, to avoid imposing unnecessary costs and delays where the risk is low. These prescriptions can largely be fulfilled by prospective cohort studies with specific characteristics to ensure that evaluation addresses the most important realities of device development. Where devices or their uses are modified during studies, this needs to be explained, and data collected to detect any relevant change in outcomes. It will often be unclear initially whether the indications for use are optimized, that is, which patient groups might benefit most, and whether subgroups might be at higher risk of harms. There may be ways of implanting, activating or using the device which result in unforeseen risks or additional benefit, and there may be risks associated with operator learning curves.[2,3] Therefore, although device use remains relatively limited, data should be collected on all cases, using a dataset rich enough to capture information on variability of outcome associated with device changes, modes of use and patient subpopulations. Data should be linked to date of use and to operators and institutions, so that operator learning curve issues can be evaluated. Generating high quality, relevant evidence is essential for informed consent from patients (autonomy) and system-level decisions about the allocation of limited health care budgets (justice).[51,52]

Later in the life cycle, once clinical adoption is widespread, evaluation should be more focused on weak or delayed signals of harm, subgroup analysis and comparison, and trends in performance. Successful analysis of these problems is best supported by data from simple, representative and unbiased datasets covering a large proportion of patients treated, such as registries. Confounding by

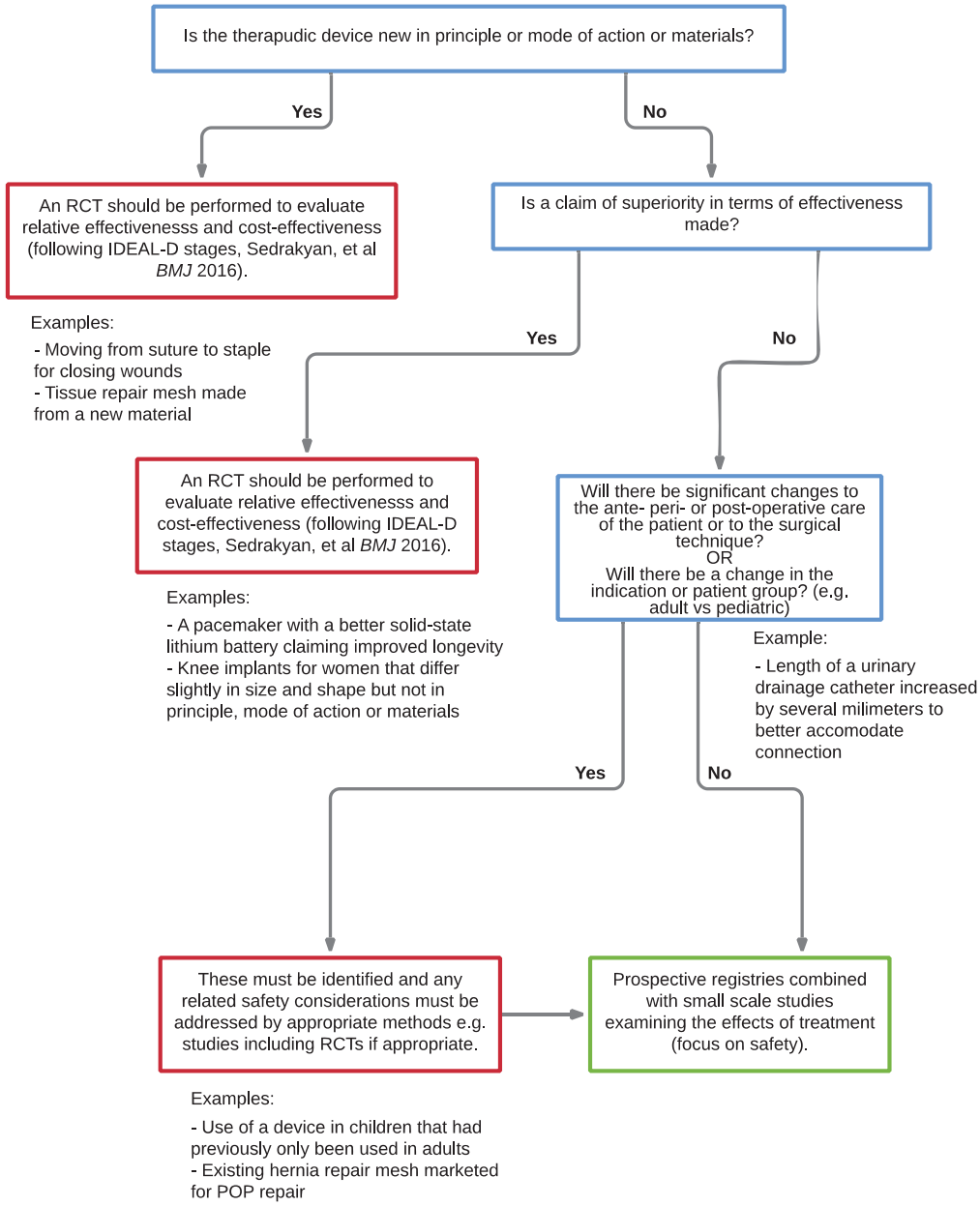

FIGURE 1. Decision-making framework for therapeutic device evaluation.

indication remains an issue for subgroup analysis in real world data, but such confounding can also be an issue in subgroup analysis in RCTs, and methods including propensity scores matching can be used to reduce the risk of confounding in real world and registry data.[60,61] The funding and governance of such registries should minimize conflicts of interest and support continuous review and monitoring for emerging risks, utilizing strategies such as statistical process control.

Guidance From the IDEAL Framework

The IDEAL framework provides actionable guidance for the evaluation of innovation in surgery, medical devices,[2] and complex health care interventions such as physiotherapy,[62] and radiotherapy.[63] The requirements set out above for earlier stage device studies are largely fulfilled by the IDEAL-D recommendations for evaluation of medical devices,[2] whereas those for the later stages in the life cycle are fulfilled by IDEAL-D Stage 4.

IDEAL was initially developed in 2009[1] and updated in 2019,[64] to address the unique problems posed by the evaluation of surgery and complex interventions, including the initial instability of interventional procedures and the effects of practitioner learning curves on the intervention's effectiveness. IDEAL describes 5 sequential stages in the life-cycle of a complex therapy and provides stage-specific recommendations for evaluation at each stage

Stage of Innovation	Study Design Principles & Data Sources
Idea (phase 1)	
Find in human report of new procedure or technology.	Case report focusing on explanation and description. Include details on what was learned from prior failures. Dedicated data collection defined by study protocol, plus some clinical records data.
Development (phase 2a)*	
Early experience (typically 20-30 cases) with technical modifications of the innovation. Stage ends when Technique or device reaches a stable form.	Prospective case series, usually from a single center, with consecutive reporting of results case by case, and explanation of changes to the intervention or indications. Should provide data on changes in outcomes following modifications. Dedicated data collection defined by study protocol, plus some clinical records data.
Exploration (phase 2b)*	
Experience extends to hundreds of cases and is replicated by others. Focus is on establishing proficiency standards for the new procedure.	Prospective multicenter cohort study. Focus is on understanding outcomes in patient subgroups, evaluating quality of delivery and operator learning curves and evaluating the feasibility of a randomized trial. Increased role for electronic health records (EHR) data but bespoke collection required for learning curve analysis
Assessment (phase 3)	
Testing the efficacy of the new intervention compared with current best practice among clinicians who are proficient.	Multicenter randomized clinical "efficacy" trial compared with the next best alternative. Dedicated data collection defined by study protocol, and/or electronic health records data.
Long-term study (phase 4)	
Monitoring late and rare outcomes, understanding effectiveness in the real world, and cataloguing changes in use and quality of delivery over time.	Population-based study clinical registry or EHR/administrative data. Focus on heterogeneity in clinical proficiency and on outcomes too rare or long term than trials.

* For Devices, IDEAL-D recommends a combined 2a and 2b cohort study, evaluating both device modification and quality of delivery

FIGURE 2. Summary of IDEAL Stages with their associated study design and reporting Recommendations, and Data Sources.

(Fig. 2).[37] Common to all stages is an emphasis on transparency, safety, and prospective data collection. Stage 1 (Idea) describes proof of concept (first in human) use. Stage 2a (Development) focuses on safety and short-term outcomes during iterative improvement of the intervention. Modifications and associated outcome changes are prospectively recorded until the procedure reaches stability. Stage 2b (Exploration) prepares for large-scale comparison with present standard of care via multicenter prospective cohort studies. These facilitate consensus on effect estimates, operator learning curves, participant eligibility criteria and feasibility for a large-scale clinical trial. In IDEAL-D, Stages 2a and b are merged. Stage 3 (Assessment) evaluates the comparative effectiveness of the intervention through randomized controlled trials, when possible. Stage 4 (Long-term follow-up) focuses on quality assurance, and the detection of rare and late adverse events.

Once devices' clinical adoption becomes widespread, the IDEAL Stage 4 recommendations are more relevant, especially:

1. Use of clearly defined, widely accepted data definitions.
2. Collection of a limited dataset which captures key outcome measures (positive effect estimates and adverse outcomes) and the most important known confounders in a prospective registry.
3. Widest possible inclusion of data from devices used for the same purpose in a single registry regardless of manufacturer.
4. Minimum barriers to data access and use consistent with confidentiality, cost, and conflicts of interest.
5. Curation of registries by government or professional bodies, with clear separation of funding sources from decisions on data access, use, and definitions.

SUMMARY

We have offered a principles-based framework and decision-making aid for identifying when an RCT should be performed to evaluate new therapeutic devices, and, when RCTs are not appropriate, what type of clinical evaluation should be required. RCTs are valuable where devices are genuinely new and claim to offer a measurable therapeutic benefit to patients compared to existing treatments. Where this is not the case, they may be inappropriate and wasteful. Ethical considerations determine the principles for evaluation where RCTs are not required. We propose that in this situation, the expectation for early evaluation of devices with any significant risk of harm should be observational studies which collect a detailed dataset in a limited population. This allows resolution of early-stage uncertainties about device properties, optimal use, delivery quality, operator learning, and specific indications. With increasing use, these studies could be widened into a registry-based approach, with collection of a more minimal dataset on a high percentage of all cases, to offer maximum sensitivity for signals of harm at minimum cost, potentially reduced further by use of Real World Data.

If this kind of evaluation was required by regulators for higher-risk devices, the result would be a safer system for monitoring innovation. Potential hazards would be detected more rapidly, and

less harm would befall patients. Regulatory and coverage decision-makers, as well as those responsible for governance in hospitals and integrated care systems, could determine if devices fit the proposed criteria for requiring evidence from RCTs, guided by their clinical and methodological experts, and could make this a requirement for their approval. Where the algorithm suggests an RCT is not required, the minimum evaluation standard would be and IDEAL-D Stage 2 study for devices which pose any significant patient risk, ensuring greater uniformity and fairness in the system, and reducing ''loop-holes.'' We have not considered the criteria for requiring long-term registries in this paper, but risk, cost and volume of device use are clearly key issues. Whether these changes would result in higher or lower costs overall is impossible to calculate at present, but costs of re-work and of unnecessary RCTs would be reduced, whereas additional registry costs could be defrayed by a tax on manufacturer profits. Thus, an ethical system for device evaluation could also be efficient and highly cost-effective.

REFERENCES

1. Hines JZ, Lurie P, Yu E, et al. Left to their own devices: breakdowns in United States Medical Device premarket review. *PLoS Med*. 2010;7:e1000280.
2. Sedrakyan A, Campbell B, Merino JG, et al. IDEAL-D: a rational framework for evaluating and regulating the use of medical devices. *BMJ*. 2016;353:i2372.
3. Fraser AG, Butchart EG, Szymański P, et al. The need for transparency of clinical evidence for medical devices in Europe. *Lancet*. 2018;392:521–530.
4. Donahue G, King G. Estimates of Medical Device Spending in the United States. Washington, DC; 2015.
5. Marotta JS, Goldberg EP, Habal MB, et al. Silicone gel breast implant failure: evaluation of properties of shells and gels for explanted prostheses and meta-analysis of literature rupture data. *Ann Plast Surg*. 2002;49:227–242.
6. Louboutin L, Viste A, Desmarchelier R, et al. Long-term survivorship of the CorailTM standard stem. *Orthop Traumatol Surg Res*. 2017;103:987–992.
7. Jalbert JJ, Ritchey ME, Mi X, et al. Methodological considerations in observational comparative effectiveness research for implantable medical devices: AN epidemiologic perspective. *Am J Epidemiol*. 2014;180:949–958.
8. Steinbrook R. The controversy over guidant's implantable defibrillators. *N Engl J Med*. 2005;353:221–224.
9. Keltie K, Elneil S, Monga A, et al. Complications following vaginal mesh procedures for stress urinary incontinence: An 8 year study of 92,246 women. *Sci Rep*. 2017;7:1–9.
10. Medical devices and the public's, health. *J Clin Eng*. 2011;36:175–192.
11. Government Accountability Office. Medical Devices: FDA Should Take Steps to Ensure That High-Risk Device Types Are Approved through the Most Stringent Premarket Review Process.Publication No. JT GAO-09-190 Washington, DC; 2009
12. Zuckerman D, Brown P, Das A. Lack of publicly available scientific evidence on the safety and effectiveness of implanted medical devices. *JAMA Intern Med*. 2014;174:1781–1787.
13. Zannad F, Stough WG, Piña IL, et al. Current challenges for clinical trials of cardiovascular medical devices. *Int J Cardiol*. 2014;175:30–37.
14. Campillo-Artero C. A full-fledged overhaul is needed for a risk and value-based regulation of medical devices in Europe. *Health Policy (New York)*. 2013;113:38–44.
15. U.S. Department of Health and Human Services Food and Drug Administration. Device Software Functions Including Mobile Medical Applications. Available at: https://www.fda.gov/medical-devices/device-software-functions-including-mobile-medical-applications/examples-software-functions-which-fda-will-exercise-enforcement-discretion. Published 2019. Accessed January 4, 2020.
16. Medical I, Regulators D. International medical device regulators forum terms of reference. *IMDRF Website*. 2014;2014:1–9. http://www.imdrf.org/%0Ahttp://www.imdrf.org/index.asp.
17. Kramer DB, Xu S, Kesselheim AS. Regulation of medical devices in the United States and European Union. *N Engl J Med*. 2012;366:848–855.
18. Van Norman GA. Drugs and devices: comparison of European and U.S. approval processes. *JACC Basic to Transl Sci*. 2016;1:399–412.
19. Institute of Medicine of the NA. Public Health Effectiveness of the FDA 510 (k) Clearance Process: Balancing Patient Safety and Innovation: Workshop Report. Vol 510. (Wizemann T, ed.).; 2010. Available at: https://www.ncbi.nlm.nih.gov/books/NBK209791/pdf/Bookshelf_NBK209791.pdf.
20. Mitchell W, Review L, Wallenfelt BP. Hatch-Waxman and medical devices. *William Mitchell Law Rev*. 2014;40:1407–1426. http://open.mitchellhamline.edu/wmlr/vol40/iss4/6.
21. US FDA. US Food and Drug Administration: Classify Your Medical Device. Available at: https://www.fda.gov/medicaldevices/resourcesforyou/consumers/ucm142523.htm. Published 2018.
22. U.S. Department of Health and Human Services, FDA. The 510(k) Program: Evaluating Substantial Equivalence in Premarket Notifications [510(k)] Guidance for Industry and Food and Drug Administration Staff [Internet]. Vol. 510. 2014. Available from: https://www.fda.gov/media/82395/download.
23. Ardaugh BM, Graves SE, Redberg RF. The 510(k) ancestry of a metal-on-metal hip implant. *N Engl J Med*. 2013;368:97–100.
24. CDRH FDA Guidance. The Special 510(k) Program Draft Guidance for Industry and Food and Drug Administration Staff [Internet]. Vol. 510. 2019. Available from: https://www.fda.gov/media/116418/download.
25. Stordeur S, Vinck I, Neyt M, et al. Mise sur le marché européen des dispositifs médicaux innovants à haut risque: l'efficacité clinique et la sécurité sont-elles garanties? *Rev Epidemiol Sante Publique*. 2013;61:105–110.
26. Hulstaert F, Neyt M, Vinck I, et al. Pre-market clinical evaluations of innovative high-risk medical devices in Europe. *Int J Technol Assess Health Care*. 2012;28:278–284.
27. Commission E. Regulations (EU) 2017/745 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC. Off J Eur Union. 2020;L(117):1. doi: http://eur-lex.europa.eu/pri/en/oj/dat/2003/l_285/l_28520031101en00330037.pdf
28. Ben-Menahem SM, Nistor-Gallo R, Macia G, et al. How the new European regulation on medical devices will affect innovation. *Nat Biomed Eng*. 2020;4:585–590.
29. Stern AD. Innovation under regulatory uncertainty: evidence from medical technology. *J Public Econ*. 2017;145:181–200.
30. Stewart L. The impact of regulation on innovation in the United States: a cross-industry literature review. *Inf Technol Innov Found*. 2010;9:1–29. https://www.ncbi.nlm.nih.gov/books/NBK189657/.
31. Curfman GD, Redberg RF. Medical devices—balancing regulation and innovation. *N Engl J Med*. 2011;365:975–977.
32. Rome BN, Kramer DB, Kesselheim AS. FDA Approval of cardiac implantable electronic devices via original and supplement premarket approval pathways, 1979-2012. *JAMA*. 2014;311:385–391.
33. Bernard A, Vaneau M, Fournel I, et al. Methodological choices for the clinical development of medical devices. *Med Devices Evid Res*. 2014;7:325–334.
34. Sackett DL, Rosenberg WMC, Gray JAM, et al. Evidence based medicine: what it is and what it isn't. *Br Med J*. 1996;312:71–72.
35. Beard DJ, Campbell MK, Blazeby JM, et al. Considerations and methods for placebo controls in surgical trials (ASPIRE guidelines). *Lancet*. 2020;395:828–838.
36. Raftery J, Young A, Stanton L, et al. Clinical trial metadata: defining and extracting metadata on the design, conduct, results and costs of 125 randomised clinical trials funded by the National Institute for Health Research Health Technology Assessment programme. *Health Technol Assess (Rockv)*. 2015;19:1–138.
37. McCulloch P, Altman DG, Campbell WB, et al. No surgical innovation without evaluation: the IDEAL recommendations. *Lancet*. 2009;374:1105–1112.
38. Neugebauer EAM, Rath A, Antoine SL, et al. Specific barriers to the conduct of randomised clinical trials on medical devices. *Trials*. 2017;18:1–10.
39. Rothwell PM. External validity of randomised controlled trials: ''To whom do the results of this trial apply?''. *Lancet*. 2005;365:82–93.
40. Johnson J, Rogers W, Lotz M, et al. Ethical challenges of innovative surgery: a response to the IDEAL recommendations. *Lancet*. 2010;376:1113–1115.
41. Hutchison K, Sparrow R. What pacemakers can teach us about the ethics of maintaining artificial organs. *Hastings Cent Rep*. 2016;46:14–24.
42. Rogers WA, Lotz M, Hutchison K, et al. Identifying surgical innovation. *Ann Surg*. 2014;259:273–278.
43. U. S. Food & Drug Administration. *Artificial Intelligence and Machine Learning in Software as a Medical Device*. 2019. Available at: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device. Accessed November 20, 2019.
44. U.S. Department of Health and Human Services Food and Drug Administration. Device Software Functions Including Mobile Medical Applications.

45. U.S. Department of Health and Human Services Food and Drug Administration, Center for Devices and Radiological Health Center for Biologics Evaluation and Research. *Deciding When to Submit a 510 (k) for a Change to an Existing Device Draft Guidance for Industry And*. Vol 510. 2016. Available at: https://www.fda.gov/media/99785/download.
46. Commission European. Guidance document Medical Devices - Scope, field of application, definition—Qualification and Classification of stand alone software - MEDDEV 2.1/6. 2016:19. Available at: https://ec.europa.eu/docsroom/documents/17921.
47. International Medical Device Regulators Forum. International Medical Device Regulators Forum, Terms of Referrence [Internet]. 2011. Available from: http://www.imdrf.org/documents/documents.asp.
48. U.S. Department of Health and Human Services Food and Drug Administration, Center for Devices and Radiological Health Center for Biologics Evaluation and Research. Deciding When to Submit a 510 (k) for a Change to an Existing Device: Guidance for Industry and Food and Drug Administration Staff. [Internet]. Vol. 510. 2017. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/deciding-when-submit-510k-change-existing-device.
49. European Commission. Guidance document: Medical Devices - Scope, field of application, definition. Qualification and classification of stand alone software - MEDDEV 2.1/6 [Internet]. 2016. Available from: https://ec.europa.eu/docsroom/documents/17921.
50. Meyerson D. Innovative surgery and the precautionary principle. *J Med Philos (United Kingdom)*. 2013;38:605–624.
51. Resnik DB. The precautionary principle and medical decision making. *J Med Philos (United Kingdom)*. 2004;29:281–299.
52. Gillon R. Medical ethics: four principles plus attention to scope. *BMJ*. 1994;309:184–1184.
53. Gargon E, Williamson PR, Altman DG, et al. The COMET Initiative database: progress and activities from 2011 to 2010. *Trials*. 2014;15:279.
54. Rehal S, Morris TP, Fielding K, et al. Non-inferiority trials: are they inferior? A systematic review of reporting in major medical journals. *BMJ Open*. 2016;6:e012594.
55. Treadwell JR, Uhl S, Tipton K, et al. Assessing equivalence and noninferiority. *J Clin Epidemiol*. 2012;65:1144–1149.
56. Christensen E. Methodology of superiority vs. equivalence trials and non-inferiority trials. *J Hepatol*. 2007;46:947–954.
57. Golish SR, Groff MW, Araghi A, et al. Superiority claims for spinal devices: a systematic review of randomized controlled trials. *Glob Spine J*. 2019;10:322–345.
58. Revicki D, Hays RD, Cella D, et al. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol*. 2008;61:102–109.
59. Grutters JPC, Govers T, Nijboer JN, et al. Problems and promises of health technologies: the role of early health economic modeling. *Int J Heal Policy Manag*. 2019;8:575–582.
60. Groenwold RHH, Donders ART, van der Heijden GJMG, et al. Confounding of subgroup analyses in randomized data. *Arch Intern Med*. 2009;169:1532–1534.
61. Williamson E, Morley R, Lucas A, et al. Propensity scores: From naïve enthusiasm to intuitive understanding. *Stat Methods Med Res*. 2012;21:273–293.
62. Beard D, Hamilton D, Davies L, et al. Evidence-based evaluation of practice and innovation in physical therapy using the IDEAL-physio framework. *Phys Ther*. 2018;98:108–121.
63. Verkooijen HM, Kerkmeijer LGW, Fuller CD, et al. R-IDEAL: a framework for systematic clinical evaluation of technical innovations in radiation oncology. *Front Oncol*. 2017;7. doi:10.3389/fonc.2017.00059.
64. Hirst A, Philippou Y, Blazeby J, et al. No surgical innovation without evaluation: evolution and further development of the IDEAL framework and recommendations. *Ann Surg*. 2019;269:211–220.
65. Petro J. Drugs and devices. *Am J Cosmet Surg*. 2016;33:109–111.
66. Heneghan CJ, Goldacre B, Onakpoya I, et al. Trials of transvaginal mesh devices for pelvic organ prolapse: a systematic database review of the US FDA approval process. *BMJ Open*. 2017;7:1–8.
67. Institute of Medicine. Medical Devices and the Public's Health: the FDA 510(k) Clearance Process at 35 Years. *Natl Acad Press*. 2011;510.
68. U.S. Food and Drug Administration. Step 1: Device Discovery and Concept.
69. European Commission. Medical Devices: Guidance Document—Classification of Medical Devices; 2010.