

Progress in clinical research in surgery and IDEAL



Peter McCulloch, Joshua Feinberg, Yiannis Philippou, Angelos Kolias, Sean Kehoe, Gillian Lancaster, Jenny Donovan, Tatjana Petrinic, Riaz Agha, Christopher Pennell

The quality of clinical research in surgery has long attracted criticism. High-quality randomised trials have proved difficult to undertake in surgery, and many surgical treatments have therefore been adopted without adequate supporting evidence of efficacy and safety. This evidence deficit can adversely affect research funding and reimbursement decisions, lead to slow adoption of innovations, and permit widespread adoption of procedures that offer no benefit, or cause harm. Improvement in the quality of surgical evidence would therefore be valuable. The Idea, Development, Exploration, Assessment, and Long-term Follow-up (IDEAL) Framework and Recommendations specify desirable qualities for surgical studies, and outline an integrated evaluation pathway for surgery, and similar complex interventions. We used the IDEAL Recommendations to assess methodological progress in surgical research over time, assessed the uptake and influence of IDEAL, and identified the challenges to further methodological progress. Comparing studies from the periods 2000–04 and 2010–14, we noted apparent improvement in the use of standard outcome measures, adoption of Consolidated Standards of Reporting Trials (CONSORT) standards, and assessment of the quality of surgery and of learning curves, but no progress in the use of qualitative research or reporting of modifications during procedure development. Better education about research, integration of rigorous evaluation into routine practice and training, and linkage of such work to awards systems could foster further improvements in surgical evidence. IDEAL has probably contributed only slightly to the improvements described to date, but its uptake is accelerating rapidly. The need for the integrated evaluation template IDEAL offers for surgery and other complex treatments is becoming more widely accepted.

Introduction

Just over 20 years ago, *The Lancet* published one of its most provocative Commentaries since the era of Thomas Wakley. Entitled “Surgical research or comic opera”, the Commentary lampooned clinical research in surgery, contrasting current practice with the principles of the evidence-based medicine (EBM) movement.¹ Surgeons reacted with anger, claiming that they faced special problems that frequently invalidated an approach based solely on randomised controlled trials (RCTs).^{2,3} Their protests were not widely accepted, but the controversy ignited eventually proved constructive. Inquiry began into why RCTs seemed to be so difficult to do in surgery and from these studies a picture emerged that partly justified the original objections from surgeons.^{4,5} Expert consensus conferences involving surgeons, EBM experts, and others developed a credible description of the evolutionary process of innovative treatments in surgery and other disciplines in which complex skilled procedures require adaptation to each individual patient. This construct was termed the IDEAL Framework, referencing the terms used for the sequential stages in the evolutionary process (Idea, Development, Exploration, Assessment, and Long-term study).^{6–8} The Framework provided a theoretical basis for arguments against randomisation from the first patient in these complex interventions. Each stage in the IDEAL Framework frames specific questions to be addressed, which led logically to the development of the IDEAL Recommendations, a set of guidelines for study design and reporting aimed at answering these questions for each stage (figure 1).^{8,9}

The IDEAL Recommendations have been widely recognised as a rational approach to developing an integrated evaluation pathway for surgery and other complex interventions. Since the Recommendations

describe desirable properties for clinical studies of surgery, they have obvious potential as a tool for judging the methodological progress of surgical research. In this Health Policy we look first at how surgical research has developed in the past 20 years, using adherence to the IDEAL Recommendations as a measure of progress. We then consider what effect IDEAL has had to date, how surgical research could be further improved, and what role IDEAL could play in that process in the future.

What is surgical research?

The past 60 years has ushered in cardiac and vascular surgery, organ transplantation, joint replacement, minimally invasive surgery, and most recently robotic surgery, all of which are advances based on surgical research, and are of unquestionable importance to patients. However, many leading university departments of surgery emphasise research on topics such as the molecular genetics of diseases treated with surgery, the immunology of organ transplantation and rejection, or stem cell treatments. Perhaps because of this, the direct study of outcomes from technical innovation such as minimally invasive and robotic surgery, and new surgically implanted devices, has been taken forward just as much by non-academic surgeons as by professional researchers.

In this report we have considered only studies of the outcomes of surgical techniques—ie, publications in which the question addressed is around the effects of an operation, since the focus of methodological criticism of surgical research has always been on this assessment of surgical efficacy and effectiveness.

How has surgical research changed?

Published surgical research has been steadily increasing in volume year on year. Using search terms based on the

Nuffield Department of Surgical Science, University of Oxford, Oxford, UK (Prof P McCulloch MD, Y Philippou MRCS); Department of Surgery, Maimonides Hospital, Brooklyn, NY, USA (J Feinberg MD, C Pennell MD); Division of Neurosurgery, School of Clinical Medicine, Cambridge University, Cambridge, UK (A Kolias PhD); Institute of Cancer and Genomics, University of Birmingham, Birmingham, UK (Prof S Kehoe MD); Institute of Primary Care and Health Sciences, Keele University, Newcastle, UK (Prof G Lancaster PhD); School of Social and Community Medicine, University of Bristol, Bristol, UK (Prof J Donovan PhD); Cairns Library, Oxford University Hospitals Foundation Trust, Oxford, UK (T Petrinic MSc); and Plastic Surgery Department, Guys and St Thomas' NHS Foundation Trust, London, UK (R Agha MSc)

Correspondence to: Prof Peter McCulloch, Nuffield Department of Surgical Science, University of Oxford, John Radcliffe Hospital, Oxford OX 39DU, UK peter.mcculloch@nds.ox.ac.uk

	IDEAL Framework (stage of evolution of intervention)	IDEAL Recommendations (stage-specific study design and reporting)
Stage 1: Idea First in human	Purpose: proof of concept Number and types of patients: single digit, highly selective Number and types of surgeons: very few, innovators* Output: description of intervention or procedure Status of intervention: evolving, at inception stage Reporting methods: structured case reports Outcomes reported: proof of concept, technical achievement, dramatic success, adverse events, surgeon views of the procedure Stage endpoint: once a decision is made to do a series of cases—ie, to proceed to stage 2a	<ul style="list-style-type: none"> • Provide full details of patient selection, technique, and outcomes, and of patients not selected during the timeframe, and why • Use standard well defined measures for reporting outcome and patient characteristics • Use a structured reporting system—eg, SCARE checklist; Make the above information available to peers regardless of whether outcome is favourable or not • Informed consent should clearly explain status of procedure and impossibility of quantifying risks
Stage 2a: Development Iterative modification of the intervention until a stable version is achieved. Design: single centre or single intervention, case series or prospective cohort.	Purpose: development of procedure Number and types of patients: few, selected Number and types of surgeons: few, innovators, and some early adopters Output: technical description of procedure and its development, with explanation of reasons for changes Intervention: evolving, procedure development towards a stable optimised version Methods: prospective development studies (small prospective cohort studies) Outcomes: mainly safety, technical, and procedural success Stage endpoint: when the procedure is considered optimised, and stable enough to allow replication in stage 2b, there should be no intent at this point to make further major modifications	<ul style="list-style-type: none"> • Make protocol for study available • Use standard well defined measures for reporting outcome and patient characteristics • Report and explain all exclusions • Report all cases sequentially with annotation and explanation of changes to indication or procedure, and when and why they took place • Display main outcomes graphically showing cases sequentially to illustrate when changes took place • Informed consent should explain status of intervention and consequent uncertainties around risk†
Stage 2b: Exploration Collaborative prospective data collection and analysis aimed at achieving consensus on key issues, to determine if an RCT is feasible, and to define its design features. Intended as a bridge from rational to comparative evaluation.	Purpose: achieving consensus between surgeons and centres on the parameters for an RCT (if possible) Number and types of patients: many, broadening indication to include all potential beneficiaries Number and types of surgeons: many, innovators, early adopters, early majority* Outputs: effect estimate for the intervention based on a large sample, allowing power calculations, analysis of learning curves, estimate of influence of prespecified technical variants and patient subgroups on outcome, qualitative research to determine operator and patient values, increased mutual confidence amongst operators Intervention: the procedure is stable in individual hands but variation exists between operators, acceptable variants are subsequently defined by analysis of pooled results Method: prospective multicentre exploration cohort study or pilot, or feasibility multicentre RCTs Outcomes: safety, clinical outcomes (specific or graded), short-term outcomes, patient-centred or reported outcomes, feasibility outcomes Stage endpoints: demonstrate that technique can be more widely adopted, and demonstrate that progression to RCT is desirable and feasible	<ul style="list-style-type: none"> • Make protocol for study available • Use standard well defined measures for reporting outcome and patient characteristics • Participate in collaborative multicentre cooperative data collection, incorporating feasibility issues (such as estimating effect size, defining intervention quality standards, assessing learning curves, exploring subgroup differences, eliciting key stakeholder values and preferences, and analysis of adverse events) • Hold a preplanned consensus meeting prior to progressing to an RCT, to identify feasibility and ability to recruit, operator eligibility on basis of learning curve analysis, intervention and comparator definitions, appropriate patient selection criteria, primary endpoint
Stage 3: Assessment Definitive comparison of main efficacy and safety aspects of a new technique against current best treatment.	Purpose: comparative effectiveness testing Number and types of patients: many, expanded indications (well defined) Number and types of surgeons: many, early majority* Output: comparison with current standard therapy Intervention: stable, with acceptable variations clearly defined Method: RCT with or without additions or modifications, alternative designs (cluster, preference RCTs, stepped wedge, adaptive designs) Outcomes: clinical outcomes (specific and graded), potentially patient-reported outcomes, health economic outcomes Stage endpoints: clear valid evidence on relative effectiveness of innovation, and identification of issues requiring long-term monitoring	<ul style="list-style-type: none"> • Make protocol for study available • Register on an appropriate international register (eg, clinicaltrials.gov) • Use standard well defined measures for reporting outcome and patient characteristics • Incorporate information about patient and clinician values and preferences in design of consent information and procedures, and outcome measures • Adhere to reporting guidelines (CONSORT update of 2010 with extension for non-pharmacological treatments, COMET, TIDieR, SPIRIT) for RCT protocol design
Stage 4: Long-term monitoring	Purpose: surveillance Number and types of patients: all eligible Number and types of surgeons: all eligible Output: description, audit, recording of regional and local variations, quality assurance, risk adjustment, detection of indication creep; Intervention: stable Method: registry, routine database, rare-case reports Outcomes: rare events, long-term outcomes, quality assurance Stage endpoints: dependent on lifecycle of device or procedure, registries for devices – IDEAL-D, registries at earlier stages of IDEAL	<ul style="list-style-type: none"> • Make protocol for study available • Registries may begin from the earliest stages of human use • Registry datasets should be defined by the clinical community with patient input • Datasets should be simple, cheap, and easy to collect • Curation of registries by clinical community is desirable • Funding of registries should be agreed between government and commercial interests but kept separate from curation • Patient consent for use of registry data in research should be broad and where possible automatic

narrow definition of surgical research that we have described in this report (appendix), we identified 41 surgical RCTs in PubMed for the year 2000, and 246 for 2014. The number of RCTs recorded yearly on PubMed increased from 11 515 in 2000 to 27 426 in 2014. Non-randomised surgical outcome studies also showed over a four-fold increase during this time. Randomised trials might be increasing somewhat as a proportion of surgical outcomes studies, but this trend is not yet clear enough to be sure. In 2000, RCTs represented 30% (41 of 138) of all outcome studies identified by our search criteria, which rose to just under 50% (208 of 422) by 2011, but the figures of 48% (218 of 458) in 2012 and 38% (217 of 564) in 2014 did not support the impression of a rising trend (figure 2).

Methodological trends in quality: an IDEAL analysis

An analysis of surgical research based solely on the proportion of studies that are RCTs gives an inadequate view of the changes in surgical research over time, because this type of analysis does not acknowledge the importance of pre-RCT studies of innovations still undergoing modification. The IDEAL Recommendations specify several uncontroversial desirable features for clinical studies of surgical interventions, especially for these earlier stages of surgical research. Examining whether these specific features have become more prevalent in published work over the years allows study of the progress of surgical research methods in greater depth and detail. The IDEAL Framework provides clear justification for the fact that RCTs make up only a small proportion of published surgical research. However, for surgical research to make progress the validity and accuracy of the studies during the earlier stages of the evolution of new techniques before an RCT needs to improve in quality, as well as the RCTs themselves. The critiques of the surgical literature in the 1990s were based largely on analysis of contemporary studies of techniques in these earlier stages, for which the retrospective case series was, at that time, the standard (and grossly inadequate) format for publication. The IDEAL Recommendations specify desirable characteristics for studies in the early stages, which case series plainly did not have. We were interested in whether compliance with specific IDEAL Recommendations had improved since the first IDEAL publication, suggesting progress in the direction IDEAL recommended. We therefore did a

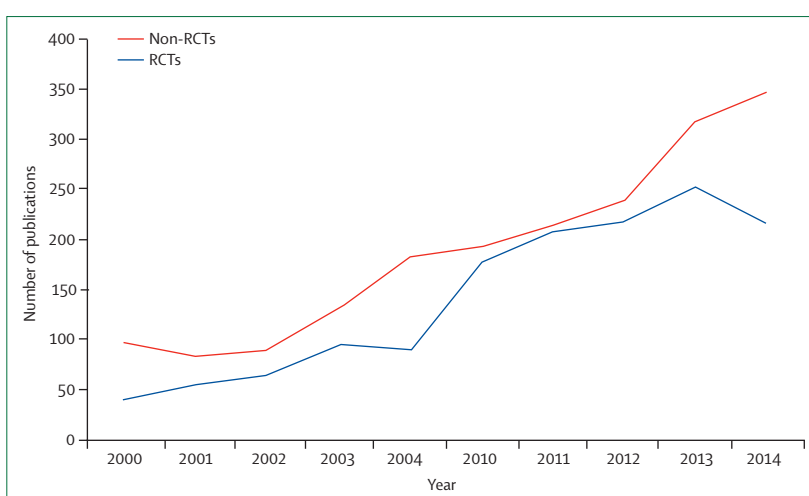


Figure 2: Studies of surgical outcomes identified from 2000–14 showing RCTs and other study designs RCTs=randomised controlled trials.

sampling exercise looking at studies from 2000–04 and from 2010–14 (the first full 5-year period after the publication of the original IDEAL articles; appendix). This exercise was not powered to show statistical significance, so our comments on the trends we found are necessarily tentative (table).

See Online for appendix

The proportion of papers reporting prospective (as opposed to retrospective) studies remained about the same over the decade between the two eras: 60% (15 of 25) in 2000–04 and 64% (16 of 25) in 2010–14 (table). The procedure was reasonably well described in nearly all cases in both cohorts. By contrast information about the quality of surgery was rarely provided, although with some apparent improvement over time.

The IDEAL Recommendations for early development studies in stage 2a stress the need for an account of changes to the procedure or indications during the development process, which remains rare in both eras (four [16%] of 25).

The IDEAL Recommendations suggest prospective collaborative collection of non-randomised data as a preparatory step towards multicentre RCTs. We found that the proportion of randomised trials which referred to such data collection nearly doubled in the later sample (nine [36%] of 25) compared with the earlier one (five [20%] of 25). Although none of the sampled papers referenced the IDEAL Recommendations, stepwise progression from collaborating on prospective data collection to doing a trial together appears to be becoming more common.

Blinding or masking of outcome assessors in RCTs was reported in 24% (six of 25) of trials in 2000–04 and in 40% (ten of 25) in 2010–14. No examples of the use of preparatory qualitative studies were found in any of the RCTs sampled in either epoch. The proportion of studies in which action was taken to address the issue of bias introduced by operator learning curves increased from

Figure 1: The IDEAL Framework and Recommendations, 2017

COMET=Core Outcome Measures in Effectiveness Trials. CONSORT=Consolidated Standards of Reporting Trials. RCT=randomised control trial. IDEAL=Idea, Development, Exploration, Assessment and Long-term follow-up. IDEAL-D=IDEAL for Devices. SCARE=surgical consensus-based guidelines for case reports. SPIRIT=Standard Protocol Items: Recommendations for Interventional Trials. TIDieR=template for intervention description and replication. *Terms used under this heading refer to the classification by Rogers.²⁰ †Patient consent should always be informed by a summary of the outcomes from previous IDEAL stages.

	2000–04 (n=25)	2010–14 (n=25)	Absolute difference (95% CI), percentage points
Use of standardised terminology	11 (44%)	15 (60%)	16 (–13.6 to 2.3)
Definition and description of procedure	20 (80%)	22 (88%)	8 (–15.8 to 31.1)
Prospective data collection	15 (60%)	16 (64%)	4 (–24.1 to 31.3)
Explanation of modifications during early studies (2a)*	4 (16%)	4 (16%)	0 (–23.5 to 23.5)
Previous analysis of learning curves in pre-RCT studies†	5 (20%)	11 (44%)	24 (–4.5 to 48.2)
Use of quality control measures	0	5 (20%)	20 (–1 to 41.3)
Use of qualitative research to define RCT questions (3)†	0	0	0 (–16.6 to 16.6)
Use of prior prospective cohort study to prepare for RCTs (3)†	5 (20%)	9 (36%)	16 (–11.4 to 40.7)
Mention of pilot or feasibility studies to prepare for RCT (3)†	2 (8%)	1 (4%)	–4 (–23.9 to 15.5)
Masking reported in RCTs (3)†	6 (24%)	10 (40%)	16 (–12.2 to 41.2)

Proportions are reported for random samples of surgical outcome studies from the whole search population, except when indicated. RCTs=randomised controlled trials. *Assessed in a random sample taken from the population of all cohort studies reporting less than 50 cases over less than 0 years in less than three centres (a surrogate for development stage studies). †Assessed in a random sample from the population of all RCTs.

Table: Proportion of randomly selected surgical outcome studies meeting IDEAL Recommendations in 2000–04 and 2010–14

five (20%), in 2000–04, to 11 (44%) in 2010–14. Only two (8%) of the randomised trials sampled between 2000 and 2004 explicitly mentioned the use of pilot or feasibility work before the RCT. None made reference to the Consolidated Standards of Reporting Trials (CONSORT) guidelines. By 2010–14 reference to previous pilot work had not increased (one study), but five studies (20%) made reference to the CONSORT guidelines, and 12 (48%) included a study flowchart. In both epochs absence of evidence of difference ($p>0.05$) in underpowered studies was routinely but incorrectly taken to be evidence of no difference between techniques.

Analysis and forward view

Surgical research seems to be changing for the better, although not as fast as hoped. Over 10 years the improvements are clear in both the proportion of studies that use CONSORT and standardised terminology to report key data items, and an increase in the proportion of surgical RCTs that have been developed from previous efforts to collect prospective collaborative data (as recommended by IDEAL for stage 2b). A higher proportion of surgical RCTs now describe masking of outcome assessors, and assessment of the quality with which surgery is delivered has improved (including assessment of learning curves). Surgeons are beginning to appreciate that properly designed preliminary studies are usually necessary before a successful surgical RCT, but the IDEAL proposition of an integrated evaluation pathway with identifiable stages has not yet become widely accepted and understood. Evidence of the challenges still to overcome includes the persistently high proportion of retrospective case series in the

surgical literature and the rarity of qualitative research to inform the development of RCTs. Another indicator of the persistent weaknesses of surgical assessment is the list of surgical procedures introduced during the periods under study in which widespread adoption without an adequate research base has harmed patients or driven up costs, including robotic prostatectomy¹⁰ and the metal-on-metal hip resurfacing techniques.¹¹

Little is known about what influences methodological decision making among surgical researchers, but some factors appear fairly obvious. Career incentives for publishing poor research in surgery are still strong, particularly retrospective case series, and clinical journals that will accept them uncritically are easy to find.

Retrospective case series are generally exempt from many of the regulatory hurdles that challenge prospective research and therefore represent a cheap, rapid route to publication. As long as this form of research retains its value for career advancement, it is unlikely to disappear.

Conversely, the very difficulties that stimulated the development of IDEAL still make surgical trials difficult to organise, and there is nothing comparable to the abundant commercial funding available for pharmaceutical trials. The weakness of basic training in the principles of clinical research for surgeons is a third important obstacle to overcome.

What effect has IDEAL had since its inception in 2009? Citation growth suggests an accelerating upward trajectory (BMC surgery, MS under review; unpublished data) but just as occurred with the EBM movement, understanding lags behind familiarity and practical use is still further behind. Use in the health technology assessment community has, however, made a strong start with programmes in Canada¹² and the Netherlands¹³ that use IDEAL to guide assessment of medical technologies and devices. EXCITE International,¹⁴ an international consortium offering device manufacturers a comprehensive assessment service for innovative new products, has embraced IDEAL as a central part of its methodology. The National Institute of Health and Research, UK, has specified IDEAL 2b-type studies in several funding calls, and discussions continue with the National Health Service (NHS) England and with the National Institute for Health and Care Excellence, UK, about the practical use of IDEAL in the NHS Commissioning through Evaluation project.

What changes could encourage the development of high-quality research in surgery? People who can influence the research environment are those responsible for educating and training surgeons, and those who decide the research that gets funding and publication.⁸ The adoption of higher standards by surgical journals than at present would be a major step forwards. Journal editors could, reasonably easily, challenge the publication of traditional case series and request IDEAL Development or Exploration studies instead, except in rare situations. This change could help to reduce research waste by

lowering the prevalence of small ad-hoc RCTs or single-arm studies trying out a new technique with no future implementation pathway or plan. Clear support from research funders for composite proposals incorporating IDEAL-type pre-RCT pilot or feasibility studies would quickly modify the behaviour of the clinical research community.

The complaint, voiced 20 years ago, that surgeons were surprisingly ill-educated about the principles of scientific methodology for investigating treatments is equally valid nowadays, even in countries where EBM has flourished. Medical schools and surgical training programmes are still failing to produce graduates who understand the methodological basis of clinical research, and who are able to apply this knowledge. Correcting this failure will require more than just courses of study, since this type of complex applied knowledge can only be thoroughly incorporated through experiential learning. Surgical training in most countries separates academic and clinical work in an artificial and unhelpful way, with segments for dedicated research time, and separate career tracks for academic and non-academic surgeons. As seen with the minimally invasive and robotic revolutions, this separation risks creating a situation in which the drive to innovate and the ability to assess are separated, to the detriment of evidence-based progress. Having a clinical workforce who are not afraid to set up simple early-stage data collection efforts that will yield valid and useful results, would be a major contribution to improving both quality and quantity in surgical research. One welcome development has been the successful collaborative movement for surgical trainee research in the UK.¹⁶ These groups give trainees practical participative experience of research, and have shown their ability to recruit patients at impressive rates and minimal cost. The groups' ability to develop RCTs on important questions is limited by their membership situation, but they have enormous potential to undertake IDEAL 2b studies rapidly and effectively, and to use these studies to drive the funding and development of a subsequent RCT.

As well as opportunities to do clinical research as an integral part of training, achieving true integration of research and practice will require appropriate incentives for both trainees and established surgeons in non-academic posts. The distinctions made between research, audit, and quality improvement are often unhelpful, and talk of involvement in scientific evaluation might be more useful. Linking involvement in high-quality assessment to appraisal, tenure, and rewards structures might be helpful, but so might public recognition, and opportunities to develop other initiatives. Making institutional approval for innovative practice conditional on agreement to collect and submit appropriate data to the host institution, or for publication, would be a powerful way of enhancing evidence accumulation. Regulation might also help. The regulatory framework can influence the success and integration of registries into clinical practice, as shown by

the success of Sweden in this area.¹⁵ However regulation is usually most effective when the ideas behind it have become popular opinion, as the examples of restrictions on drink driving and smoking in public places illustrate.

Vision for the future

Surgical research is getting better, although it still has a long way to go. Research was unfairly maligned in the first place because understanding of the real problems faced by surgical research was underdeveloped 20 years ago. The IDEAL Framework and Recommendations have probably only contributed in a minor way to the improvements seen so far, but their influence is growing, they are useful as a tool to measure progress, and they represent a serious attempt to create a new framework for surgical research methodology. The idea of a logical series of study questions and types based on the realities of how surgical operations evolve is clearly of value, and situations in which IDEAL proves imperfect will probably lead to either modification or replacement by a better version, rather than a return to methodological anarchy. Because complex interventions typically require a period of iterative improvement before reaching a stable form, they cannot be subjected to valid comparisons with alternatives until this phase is over. The influences on outcome which can be generated by patient heterogeneity, and by variations in the quality of intervention delivery, especially while operators are learning, are practically infinite. Defining a patient group and a version of the intervention for a trial will therefore require a basis for decision making and considerable negotiation. Substantial prospective empirical data clearly represent a more reliable basis than theory combined with small datasets full of contextual biases. Hence collaborative collection of non-randomised data to assist with decision making is justified, wherever complexity impedes definition of the study population or the intervention.

These principles apply beyond the confines of surgery, and beyond conventional academic research. The concept of integrated stepwise assessment, beginning with a study of development of the innovation, and proceeding through more comprehensive assessment of the innovation's properties and uses before comparing it with alternatives, seems applicable to complex interventions in many fields, in health care and other domains. A version of IDEAL for assessing therapeutic devices was published last year,¹⁷ and plans for the use of IDEAL in radiotherapy,¹⁸ physiotherapy,¹⁹ and acupuncture (Prof Sun, Chinese Cochrane Centre, West China Medical Centre of Sichuan University, China, personal communication) are being implemented. Psychological therapies, and complex social or quality improvement interventions, are other areas in which IDEAL evaluation might prove useful.

The current version of the IDEAL Framework and Recommendations is clearly a work in progress. How IDEAL might need to be modified depends to a large

extent on how clinical research methodology itself evolves. For example, IDEAL's current form is predicated on the assumption that RCTs will remain the gold standard methodology for comparing treatments, which seems likely. However, the increasing expense and regulatory complexity of RCTs, combined with competition from the application of sophisticated risk adjustment techniques to large observational datasets, and reluctance to randomise in cultures that value the individual's right to choose as paramount, might reduce RCT pre-eminence in the future. Even a major change, such as this, would merely require the Framework to be adapted, rather than abolished; in other words, if IDEAL did not already exist, it would need to be invented. Now that a viable alternative is beginning to emerge, the culture of surgical case series and other weak study designs should be consigned to history, and an integrated evaluation pathway for surgery with methods appropriate to the task should be adopted.

Contributors

The Health Policy was conceived by PM following discussion with The Lancet editorial team. PM wrote the first and final drafts, and supervised the work of other authors. JF scanned and catalogued the literature search results, organised the literature sampling analysis programme, and reviewed the results, acting as general coordinator and guardian of data. TP completed and advised on literature searches and bibliometrics. YP, JF, PM, AK, SK, GL, RA, JD, and CP all contributed to analysis of the literature. All authors contributed to and commented on drafts of the manuscript. PM, JF, AK, SK, CP and RA are members of the IDEAL Collaboration.

Declaration of interests

PM, JF, AK, SK, CP, and RA are members of the IDEAL collaboration. This work originated from the IDEAL International Conference at St Katherine's College Oxford, Oxford, UK, in April, 2016, which was funded by Oxford Academic Health Science Network, Medtronic plc, Johnson and Johnson, and the Health Foundation. AK is the specialty lead for neurosurgery for IDEAL. JD reports grants from the Medical Research Council ConDuCT Hub for trial methodology research during the conduct of this study. All other authors declare no competing interests.

Acknowledgments

We thank members of the IDEAL Collaboration who took part in online work to identify early development-stage interventions from the literature, and the IDEAL Coordinator, Allison Hirst, for critically reviewing the report. The IDEAL Collaboration members who assisted with this paper were: Chris Bainbridge, Dimitri Barski, Elena Berti, Egon Cholakian, Pierre-Alain Clavien, Tammy Clifford, Andrew Currie, Thorston Ecke, Liane Friedman, Holger Gerullis, Trish Groves, Judith Hargreaves, Catherine Hatzantonis, Emma Jones, Joey Kwong, Mario Lobao, Aaron Lawson McLean, Jose Merino, Heather McIntosh, Kemp Narayanasetty, Arsenio Paez, Jan Rongen, Markus Schafer, Shafaque Shaikh, Matthew Smith, Roxane Diane Staiger, Casper Tax, Ajay Upadhyay, Bolivar Villacis Bermeo, Rebecca Wong, D R Yadev.

References

- Horton R. Surgical research or comic opera: questions, but few answers. *Lancet* 1996; **347**: 984–85.
- Bell PR. Surgical research. *Lancet* 1996; **347**: 1479.
- Ruckley CV. Surgical research. *Lancet* 1996; **347**: 1479–80.
- Solomon MJ, Laxamana A, Devore L, McLeod RS. Randomized controlled trials in surgery. *Surgery* 1994; **115**: 707–12.
- McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: problems and possible solutions. *BMJ* 2002; **324**: 1448–51.
- Barkun JS, Aronson JK, Feldman LS, al. Evaluation and stages of surgical innovations. *Lancet* 2009; **374**: 1089–96.
- Ergina PL, Barkun JS, McCulloch P, Cook JA, Altman DG. IDEAL framework for surgical innovation 2: observational studies in the exploration and assessment stages. *BMJ* 2013; **346**: f3011.
- McCulloch P, Altman DG, Campbell WB, et al. No surgical innovation without evaluation: the IDEAL recommendations. *Lancet* 2009; **374**: 1105–12.
- Pennell CP, Hirst AD, Campbell WB, et al. Practical guide to the idea, development and exploration stages of the IDEAL Framework and Recommendations. *Br J Surg* 2016; **103**: 607–15.
- Mirheydar HS, Parsons JK. Diffusion of robotics into clinical practice in the United States: process, patient safety, learning curves, and the public health. *World J Urol* 2013; **31**: 455–61.
- Marshall DA, Pykerman K, Werle J, et al. Hip resurfacing versus total hip arthroplasty: a systematic review comparing standardized outcomes. *Clin Orthop Relat Res* 2014; **472**: 2217–30.
- Sampietro-Colom L, Martin J. Hospital-based health technology assessment: the next frontier. In: Sampietro-Colom L, Martin J, eds. Hospital-based health technology assessment. Cham, Switzerland: Springer, 2016: 3–11.
- Heikens JT, Gooszen HG, Rovers MM, van Laarhoven CJ. Stages and evaluation of surgical innovation: a clinical example of the ileo neorectal anastomosis after ulcerative colitis and familial adenomatous polyposis. *Surg Inno* 2013; **20**: 459–65.
- Levin L, Krahn M, Goeree R, et al. Excite—a new collaborative model of pre-market evaluation of health technologies. *Value in Health* 2013; **16**: A264.
- Levy C. Policies to foster quality improvement registries: lessons from the Swedish case. *J Intern Med* 2016; **279**: 160–72.
- Bhangu A, Kolias AG, Pinkney T, Hall NJ, Fitzgerald JE. Surgical research collaboratives in the UK. *Lancet* 2013; **382**: 1091–92.
- Sedrakyan A, Campbell B, Merino JG, Kuntz R, Hirst A, McCulloch P. IDEAL-D: a rational framework for evaluating and regulating the use of medical devices. *BMJ* 2016; **353**: i2372.
- 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**: 59.
- 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* 2017; published online Oct 23. DOI:10.1093/ptj/pzx103.
- Rogers EM. Diffusion of innovations, 4th edn. New York, NY: The Free Press, 1995.