Are we nearly there yet? Progress in clinical research in surgery viewed through an IDEAL lens perspective.

Peter McCulloch,1 Joshua Feinberg,2 Yiannis Philippou3, Angelos Kolas4, Sean Kehoe5, Gillian Lancaster6, Jenny Donovan17, Tatjana Petrinic,1 Riaz Agha2, Christopher Pennell2

1. Professor of Surgical Science & Practice, University of Oxford FULL PROFESSOR, MD
2. Resident, Department of Surgery, Maimonides Hospital, Brooklyn, NY. MD
3. NIHR Academic Clinical Fellow, Nuffield Department of Surgical Science, University of Oxford MRCS
4. Clinical Lecturer, Division of Neurosurgery, School of Clinical Medicine, Cambridge University PhD
5. Professor of Gynaecological Cancer, Institute of Cancer and Genomics, University of Birmingham FULL PROFESSOR, MD
6. Professor, Institute of Primary Care and Health Sciences, Keele University FULL PROFESSOR, PhD
7. Professor of Social Medicine, University of Bristol FULL PROFESSOR, PhD
8. Senior Librarian, Cairns Library, Oxford University Hospitals Foundation Trust MSc
9. Specialist Registrar in Plastic Surgery, Guys and St Thomas’ NHS Foundation Trust, London

SUMMARY

The quality of clinical research in surgery has long attracted criticism over the last 20 years. High quality randomised trials have proved difficult to conduct in surgery and any surgical treatments have therefore been adopted without adequate supporting evidence of efficacy and safety. This evidence deficit can adversely affect research funding and re-imbursement decisions, slowing adoption of innovations, and can also permit widespread adoption of procedures which offer no benefit or cause harm. Improvement in the quality of surgical evidence would therefore be very valuable. Clinical research has become more sophisticated in recent years, and the IDEAL Framework and Recommendations, launched in 2009, have addressed some of the methodological problems in surgical research, 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. We therefore reviewed methodological progress in surgery, and the uptake and influence of IDEAL and identified the challenges to further methodological progress.

We used a small sampling study, using compliance with IDEAL Recommendations to compare quality in surgical research. Comparing studies from the periods 2000 - 2004 and 2010-2014 in two epochs ten years apart, to focus a review of progress, and looked for evidence about the adoption and impact of IDEAL. We noted apparent improvement in some areas, including the use of standard outcome measures, adoption of CONSORT standards and evaluation of the quality of surgery and of learning curves, but no progress in others, such as the use of qualitative research or transparent reporting of modifications changes in procedures during early procedure development. Better education about research, integration of evaluation work into routine practice and training, regulation and linkage of evaluation work to assessment and 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/pathway it 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 Editorials since the era of Thomas Wakley. Entitled “Surgical research or comic opera”, it lampooned clinical research in surgery, contrasting current practice with the principles of the Evidence-Based Medicine movement(1). Surgeons reacted with anger, claiming that they faced special problems which frequently invalidated an approach based solely on randomised controlled trials(2, 3). Their protests were not widely accepted, but the controversy ignited eventually proved constructive. Enquiry into why randomised controlled trials (RCTs) seemed to be so difficult to conduct in surgery, and from these studies a picture emerged which partly justified the surgeons’ original objections(4, 5). Expert consensus conferences involving surgeons, EBM experts and others developed a credible description of the process of evolution of innovative treatments, in surgery and other disciplines where 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; and, this 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 (See Table 1 for a recently updated version of both Framework and Recommendations)(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 they describe desirable properties for clinical studies of surgery, they have obvious potential as a yardstick for judging the methodological progress of surgical research. In this article we look first at how surgical research has developed in the last 20 years, using adherence to the IDEAL Recommendations as a measure of progress. We then consider what impact IDEAL has had to date, how surgical research could be further improved, and what role IDEAL could play in that process in the future, and second at the development of IDEAL itself, assessing its evolution and its progress in terms of acceptance and adoption in the surgical community internationally.

WHAT IS SURGICAL RESEARCH?

Defining surgical research is challenging, and different definitions may be useful depending on the reasons for needing one. During the last 60 years we have seen the development of cardiac and vascular surgery, organ transplantation, joint replacement, minimally invasive surgery and most recently robotic surgery, advances of unquestionable importance to patients which have been based on surgical research. However many leading university departments of surgery emphasise research on topics such as the molecular genetics of diseases treated by 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 article we have considered only studies of the outcomes of surgical techniques, i.e. papers in which the question addressed is around the effects of an operation—either simply reporting the outcomes or comparing them to other treatment options—since the focus of methodological criticism of surgical research has always been on this evaluation of surgical efficacy and effectiveness, and comparisons of different time periods are clearly simpler if the types of study compared are restricted.

HOW HAS IT CHANGED?

Published surgical research has been steadily increasing in volume year on year. Using search terms based on the above narrow definition of surgical research (see Appendix 1) we identified 41 surgical RCTs in PubMed for the year 2000, and 246 for 2014. During the same time, the number of RCTs recorded yearly on PubMed as a whole increased from 11,515 in 2000 to 27,426. Non-randomised surgical outcome studies also showed a several-fold increase during this time. Randomised trials may be increasing somewhat as a proportion of surgical outcomes studies, but this trend is not yet so clear that we can be sure. In 2000, RCTs represented 30% (41 of 138) of all outcome studies identified by our search criteria, and this 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 1).

METHODOLOGICAL TRENDS IN QUALITY: AN IDEAL ANALYSIS

An analysis of surgical research based solely on the proportion of studies which are RCTs gives an inadequate view of the changes in surgical research over time, specifically because it does not acknowledge the importance of pre-RCT studies of innovations still undergoing modification. The IDEAL Recommendations specify a number of 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 us to look at the progress of surgical research methodology in greater depth and detail. The IDEAL Framework provides clear justification for the fact that it remains true that only a minority of published surgical research is made up of RCTs... and indeed the IDEAL Framework provides clear justification for this. However, in order for surgical research to make progress the validity and accuracy of the studies during the earlier stages of the evolution of new techniques prior to an RCT needs to improve in quality, as well as the RCTs themselves. The 1990s critiques of the surgical literature were based largely on analysis of contemporary studies of techniques in these earlier stages IDEAL stages 2a and 2b, 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 early stage studies which the case series plainly lacked. 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 conducted a sampling exercise looking at studies sampled studies from 2000-2004 and from 2010-2014, the latter being the first full 5 year period after the publication of the original IDEAL articles (See Appendices 1 and 2 for details). This sampling exercise was not powered to demonstrate statistical significance, so our comments on the trends we found are necessarily tentative.
conducted a literature search designed to identify clinical studies of the outcomes of surgical
techniques (for details of our search strategy see Appendix 1). We then took random samples of 25
papers from each of the two five-year subpopulations (2000-2004 and 2010-2014), and compared
them for the presence or absence of desirable characteristics specified by the IDEAL
Recommendations (See Table 1). Several different random samples were taken to allow us to study
different IDEAL Recommendations in the appropriate subgroup of studies, since some
Recommendations apply only to specific stages in the IDEAL Framework. Further details of our
methods and reasoning can be found in Appendix 2. Additional details and a list of the papers
reviewed are available from the corresponding author.

- Description of iterative changes in the procedure would be expected only in IDEAL 2a
Development studies, and for this Recommendation we therefore sampled only from amongst
papers which appeared to describe innovations at this stage, using the rules of thumb described by
Pennell et al. (9) to identify these.
- IDEAL recommends that operator learning curves are evaluated before embarking on RCTs;
we therefore searched only amongst reports of RCTs for evidence that learning curves were
considered whilst deciding which centres should enter patients.
- IDEAL recommends the use of pilot and feasibility studies and qualitative studies to inform
protocol development during the planning of RCTs, and we therefore selected only from amongst
RCTs when analysing compliance with these recommendations. The Recommendations we studied
and the classes of study from which the samples were selected are shown in Table 2.

IDEAL RECOMMENDATIONS APPLICABLE TO ALL OUTCOMES STUDIES

The proportion of papers reporting prospective (as opposed to retrospective) studies remained
about the same over the decade between the two eras; it was 60% (15/25) in 2000-2004 and 64%
(16/25) in 2010-14. 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. In the later sample there was some attempt to evaluate the
quality of surgery in 5 papers (20%), but was none (0%) in any of the papers from 2000-2004.

The use of well recognised standardised measures for outcomes, patient characteristics and other
important data showed an apparent modest increase from 11/25 (44%) of 2000-2004 papers to
15/25 (60%) of papers from 2010-2014.

IDEAL RECOMMENDATIONS FOR PRE-RCT STUDIES—STAGES 2A AND 2B

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. This remains
relatively rare (16% (4/25) in both eras). Results at this stage should be presented sequentially so
that trends and their associations with changes made may be more evident. Only four papers (16%)
followed these recommendations during the series in the 2000-2004 cohort, and the same
proportion did so in 2010-2014, demonstrating no trend over time.

IDEAL RECOMMENDATIONS APPLICABLE PREPARATORY STUDIES FOR RCTS (STAGE 2B) AND RCTS
(STAGE 3).

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 (9/25 or 36%) compared to the earlier one (5/25 papers, 20%). Although none of the sampled papers referenced the IDEAL Recommendations, it appears that stepwise progression from collaborating on prospective data collection to doing a trial together is becoming more common. Prospective collaborative cohort studies increased considerably as a proportion of all non-RCTs between the eras sampled, whereas small case series with less than 50 patients have declined. Changes in both medicine and IT which have made collaborative data collection exercises less costly and inconvenient during the period under study, may have contributed to the change. Blinding or masking of outcome assessors in RCTs was reported in 6/25 (24%) of trials in the earlier period and in 10/25 (40%) in 2010-14. There were no examples of the use of preparatory qualitative studies 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 5 (20%) in the earlier epoch to 11 (44%) in 2010-2014. Only two (8%) of the randomised trials sampled between 2000 and 2004 explicitly mentioned the use of pilot/feasibility work prior to the RCT, although 16 studies made reference to previous related work by one or more of the co-authors. None made reference to the CONSORT guidelines. Sample sizes ranged from 20 to 1289, with the largest studies testing 'equivalence'. All studies carried out hypothesis testing but only 7 studies carried out a sample size calculation and were adequately powered for such testing. By 2010-2014 reference to previous pilot work had not increased (one study), but although again 16 studies referenced previous related work, however five studies (20%) made reference to the CONSORT guidelines, and 12 included a study flowchart. Sample sizes ranged from 11 to 597, with the largest study once again being an equivalence trial, and 12 studies reported the sample size calculation. 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. The proportion of studies in which reference to previous related work was made increased from 4 (16%) in the earlier era to 23 (46%) in 2010-2014.

ANALYSIS AND FORWARD VIEW

Surgical research seems to be changing for the better, although not as fast as we might wish. Over ten years there have been clear improvements in the proportion of studies using CONSORT and using standardised terminology to report key data items, and an increase in the percentage of surgical RCTs that have been developed from prior prospective collaborative data collection efforts (as recommended by IDEAL for Stage 2b). The quality of surgical RCTs themselves appears to have improved in some respects too, as indicated by the higher proportion of surgical RCTs trials now describing blinding of outcome assessors, and there is better evaluation of the quality with which surgery was delivered (including evaluation of learning curves). Sit is clear that surgeons are beginning to appreciate that properly designed preliminary studies are usually necessary before a successful surgical RCT if it is going to succeed, but the IDEAL proposition of an integrated evaluation pathway with identifiable stages building towards an RCT has not yet become widely accepted and understood. Evidence of the distance still untravelled includes the persistently high proportion of retrospective case series in the literature and the rarity of qualitative research has been used to inform the development of RCTs. Another indicator of the persistent weaknesses of surgical evaluation is the list of surgical procedures introduced during the periods under study whose widespread adoption without an adequate research base has harmed patients or driven up costs – including robotic prostatectomy(10) and the metal-on-metal hip resurfacing techniques(11). An
IDEAL-type evaluation pathway would probably have mitigated these adverse outcomes and given us much clearer evidence on risks, benefits and costs.

We know little about what influences methodological decision-making amongst surgical researchers, but some factors appear fairly obvious. There are still strong career incentives for publishing poor research in surgery, particularly retrospective case series, and it is easy to find clinical journals that will accept them uncritically. They are generally exempt from many of the regulatory hurdles which 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. On the other hand, the very difficulties which stimulated the development of IDEAL still make surgical trials difficult to organise, and there is no correlate for the abundant commercial funding available for pharma trials. The weakness of basic training in the principles of clinical research for surgeons is a third important obstacle to progress. The issues addressed by IDEAL are very familiar to surgeons, but most are unaware of previous work on the potential solutions.

What impact has IDEAL had since its inception in 2009? It is difficult to evaluate the rate of diffusion of understanding of IDEAL through the surgical community. Citation growth suggests an accelerating upward trajectory (BMC surgery, MS under review) but just as occurred with the EBM movement at its inception, understanding lags behind familiarity, and practical use is still further behind. UPTake and use in the Health Technology Assessment community has however, made a strong start, with programmes in Canada (12) and the Netherlands (13) that use IDEAL to guide evaluation of medical technologies and devices. A recently launched enterprise to offer device manufacturers a comprehensive evaluation service for innovative new products, EXCITE International (14), has embraced IDEAL as a central part of its methodology, has also been held with NHS England and with NICE.

Looking more widely than the adoption of IDEAL, what changes could encourage speed the development of high quality research in surgery? TAs we noted in 2009, those who can influence the research environment are those responsible for educating and training surgeons and those who decide which research gets funding and publication (8). The adoption of higher standards by surgical journals (especially surgical journals) would be a major step forwards— and some, such as the International Journal of Surgery, have begun to specify that IDEAL format studies are welcome (15), and the adoption of these formats, and the integrated pathway they belong to, could help to reduce research waste by lowering the prevalence of small ad-hoc mini RCTs or single arm studies trying out a new technique with no future implementation pathway or plan. In very rare situations where a case series is the only reasonable study design, this would be an important contribution to reducing research waste. Clear support messages from research funders for that they will support composite proposals incorporating IDEAL-type pre-RCT pilot/feasibility studies to prepare the ground for a trial would quickly ill be important in modifying the behaviour of the clinical research community. It is encouraging that bodies such as the UK’s NIHR are beginning to experiment with IDEAL 2b-like studies, but they have as yet made no clear policy statement. Whilst IDEAL has received virtually no direct opposition, there is a tangible reluctance amongst some funders to extend it a welcome. This appears to be based on the (incorrect) perception that it challenges the principles of conventional medical research methodology, because of its championing of...
uncontrolled study designs for some questions where no comparison between treatments is involved. Presenting a case for IDEAL which reconciles true believers and sceptics on the role of RCTs in medical research is an impossible task, but open debate needs to be encouraged. Silent passive resistance which evades discussion is unscientific and can only retard the search for a better framework for evaluating surgery and other complex interventions.

If our aim is to continually enhance surgical research quality, the historical origins of modern science can point us to what catalyses a spirit of systematic enquiry within communities. There is an excellent case that the Enlightenment, which ushered in the scientific revolution, was bought about by the rejection of dogma and the enhancement of basic educational attainment (16).—The complaint, voiced 20 years ago, that surgeons (and doctors in general) were surprisingly ill-educated about the principles of scientific methodology for investigating medical treatments is equally valid today, even in the countries where EBM has flourished. Medical schools and postgraduate-surgical training programmes are still failing to produce graduates who understand the methodological basis of clinical research and are able to apply this knowledge. Correcting this will require more than just courses of study, since this type of complex applied knowledge can only be thoroughly incorporated through experiential learning. Current 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 we saw with the minimally invasive and robotic revolutions, this risks creating a situation where the drive to innovate and the ability to evaluate 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 which will yield valid and useful results would be a major contribution to improving both quality and quantity in surgical research. One very welcome recent development has been the successful movement to establish surgical trainee research collaborative movements in the UK*. These groups give trainees practical participative experience of research, and have demonstrated their ability to recruit patients at impressive rates and minimal cost. Their ability to develop RCTs on important questions is limited by the short tenure and lack of political power of their memberships’ situation, but they have enormous potential to conduct IDEAL 2b studies rapidly and effectively, and to use these 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 we currently make between research, audit and quality improvement are often unhelpful to this integration, and it may be more useful to talk of involvement in scientific evaluation. Linking involvement in high-quality evaluation to appraisal, tenure and rewards structures may be helpful, but so may public recognition, and opportunities to develop other initiatives. Making institutional approval for the development of innovative practice conditional on agreement to collect and submit appropriate data and submit it to the host institution, or for publication, would be a powerful way of enhancing evidence accumulation. Regulation may also help. The current pharmaceutical industry evaluation paradigm, which has hugely improved the evaluation of new drugs, can be largely attributed to regulatory action in the wake of the Thalidomide disaster(17).—The regulatory framework can influence the success and integration into clinical practice of Registries, as the success of Sweden in this area shows(158). However regulation is usually most effective where popular opinion has already been won round to the ideas behind it, as the examples of restrictions on drink driving and smoking in public places illustrate. Regulation may also fail to predict future research trends and developments.
As this review has shown, surgical research is getting better, although it still has a long way to go. It was unfairly maligned in the first place, as understanding of the real problems it faced was under-developed 20 years ago. The IDEAL Framework and Recommendations have probably contributed only in a minor way to the improvements seen so far, but their influence is growing, they are very useful as a yardstick to measure progress, and they represent a serious attempt to create a new paradigm for surgical research methodology in the future. 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 where IDEAL proves imperfect it is likely to be either modified or replaced by a better version, rather than by 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 immediate valid comparisons with alternatives until this phase is over. The variety of influences on outcome which can be generated by subject heterogeneity and by variations in the quality of intervention delivery, especially whilst operators are learning, is practically infinite. Defining a subject group and a version of the intervention for a trial will therefore require a both 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 of "no innovation without evaluation" apply beyond the confines of surgery, surgery and quasi-surgical treatments, and beyond conventional academic research. The concept of integrated stepwise evaluation, beginning with a study of development of the innovation and proceeding through more comprehensive evaluation of its properties and uses before comparing it with alternatives, seems applicable to complex interventions in many fields, both in healthcare and other domains. A version of IDEAL for evaluating therapeutic devices was published last year [160], and plans for using IDEAL in radiotherapy [1720], physiotherapy [18][Paez A et al, Physical Therapy, accepted, in press], and acupuncture [Prof REF Xin Sun, personal communication] are being implemented. Psychological therapies and complex social or quality improvement interventions are other areas where it may prove useful.

The current version of the IDEAL Framework and Recommendations is clearly not the last word on the subject, but a work in progress. How it may 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. This seems likely, given their obvious structural superiority, but there is a danger that their increasing expense and regulatory complexity, combined with reluctance to participate amongst patients in cultures which value the individual’s right to choose as paramount, and competition from the application of sophisticated risk adjustment techniques to very large observational datasets may reduce their pre-eminence in the future. However 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 be necessary to invent it.

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 using methodology appropriate to the task should be adopted.

AUTHOR CONTRIBUTIONS AND COMPETING INTERESTS

The article was conceived by PM following discussion with the Lancet editorial team. He 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 co-ordinator and guardian of data. TP conducted and advised on literature searches and bibliometrics. YP, JF, PM, AK, SK, GL, RA & CP all contributed to analysis of the literature. All authors contributed to and commented on drafts of the MS. 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 in April 2016, which was funded by Oxford AHSN, Medtronic plc, Johnson and Johnson and the Health Foundation.

ACKNOWLEDGEMENTS

We would like to thank members of the IDEAL Collaboration who took part in online work to identify early Development stage interventions from the literature, and the IDEAL Co-ordinator, Ms Allison Hirst for critically reviewing the MS. Thanks to Bill Summerskill of the Lancet for his guidance, advice and support.

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 A. Villacís Bermeo
Rebecca Wong
DR Yadev


<table>
<thead>
<tr>
<th>Stage 1</th>
<th>IDEAL Framework (Description of stage of evolution of intervention)</th>
<th>IDEAL Recommendations (For stage-specific study design and reporting)</th>
</tr>
</thead>
</table>
| Idea   | **Purpose:** Proof of concept  
          **Number & Types of Patients:** Single digit; highly selective.  
          **Number & 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 conduct a series of cases, i.e. to proceed to stage 2a. | • Provide full details of patient selection, technique and outcomes and of patients not selected during the time frame, 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 |
| Development | **Purpose:** Development of procedure  
          **Number & Types of Patients:** Few; Selected  
          **Number & 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. | • 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 the above.  
• Informed consent should explain current status of intervention and consequent uncertainties around risk ** |
<table>
<thead>
<tr>
<th>Stage 2b</th>
<th>Exploration</th>
</tr>
</thead>
</table>
| **Purpose:** Achieving consensus between surgeons and centres on the parameters for an RCT (if possible)  
**Number & Types of Patients:** Many; broadening indication to include all potential beneficiaries  
**Number & 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 pre-specified 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 multi-centre exploration cohort study or pilot/feasibility multicentre RCTs.  
**Outcomes:** Safety; clinical outcomes (specific/graded); short-term outcomes; patient centred/reported outcomes; feasibility outcomes  
**Stage Endpoints:** Demonstrate that technique can be more widely adopted; and, Demonstrate that progression to RCT is desirable and feasible |
| **- Make protocol for study available**  
**- Use standard well-defined measures for reporting outcome and patient characteristics**  
**- Participate in collaborative multi-centre co-operative data collection, incorporating feasibility issues such as:**  
  - estimating effect size,  
  - defining intervention quality standards,  
  - evaluating learning curves,  
  - exploring subgroup differences,  
  - eliciting key stakeholder values and preferences,  
  - analysis of adverse events:  
**- Hold a pre-planned 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.** |

<table>
<thead>
<tr>
<th>Stage 3</th>
<th>Assessment</th>
</tr>
</thead>
</table>
| **Purpose:** Comparative effectiveness testing  
**Number & Types of Patients:** Many; expanded indications (well-defined)  
**Number & 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/modifications; alternative designs (cluster, preference RCTs, stepped wedge, adaptive designs)  
**Outcomes:** Clinical outcomes (specific and graded); potentially Patient Reported outcomes, Health |
| **- Register on an appropriate international register (e.g., 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 |
<table>
<thead>
<tr>
<th>Economic outcomes</th>
<th>SPIRIT (for RCT protocol design)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage Endpoints: Clear valid evidence on relative effectiveness of innovation; and, Identification of issues requiring long term monitoring.</td>
<td></td>
</tr>
</tbody>
</table>

| Stage 4 Long term monitoring | Purpose: Surveillance  
Number & Types of Patients: All eligible  
Number & 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 |  
- 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 |

| Stage Endpoints: dependent on lifecycle of device/procedure.  
Registries for devices – IDEAL-D  
Registries at earlier stages of IDEAL |  
@ Terms used under this heading refer to the classification of Everett Rogers (Diffusion of Innovations, 4th Ed, 1995)  
*Registries should be organised according to the IDEAL recommendations and should be available for enrolment at any Stage  
**Patient consent should always be informed by a summary of the outcomes from previous IDEAL Stages |
<table>
<thead>
<tr>
<th>IDEAL Recommendation – no./total no. (%)</th>
<th>2000-2004</th>
<th>2010-2014</th>
<th>Absolute difference (95% CI) percentage points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use of standardised terminology</td>
<td>11/25 (44)</td>
<td>15/25 (60)</td>
<td>16 (-13.6 to 42.3)</td>
</tr>
<tr>
<td>Definition and description of procedure</td>
<td>20/25 (80)</td>
<td>22/25 (88)</td>
<td>8 (-15.8 to 31.1)</td>
</tr>
<tr>
<td>Prospective data collection</td>
<td>15/25 (60)</td>
<td>16/25 (64)</td>
<td>4 (-24.1 to 31.3)</td>
</tr>
<tr>
<td>Explanation of modifications during early studies (2a)*</td>
<td>4/25 (16)</td>
<td>4/25 (16)</td>
<td>0 (-23.5 to 23.5)</td>
</tr>
<tr>
<td>Prior analysis of learning curves in pre-RCT studies$</td>
<td>5/25 (20)</td>
<td>11/25 (44)</td>
<td>24 (-4.5 to 48.2)</td>
</tr>
<tr>
<td>Use of quality control measures</td>
<td>0/25 (0)</td>
<td>5/25 (20)</td>
<td>20 (-1 to 41.3)</td>
</tr>
<tr>
<td>Use of qualitative research to define RCT questions (3)$</td>
<td>0/25 (0)</td>
<td>0/25 (0)</td>
<td>0 (-16.6 to 16.6)</td>
</tr>
<tr>
<td>Use of prior prospective cohort study to prepare for RCTs (3)$</td>
<td>5/25 (20)</td>
<td>9/25 (36)</td>
<td>16 (-11.4 to 40.7)</td>
</tr>
<tr>
<td>Mention of pilot or feasibility studies to prepare for RCT (3)$</td>
<td>2/25 (8)</td>
<td>1/25 (4)</td>
<td>- 4 (-23.9 to 15.5)</td>
</tr>
<tr>
<td>Blinding reported in RCTs (3)$</td>
<td>6/25 (24)</td>
<td>10/25 (40)</td>
<td>16 (-12.2 to 41.2)</td>
</tr>
</tbody>
</table>

Proportions are reported for random samples of surgical outcome studies from the whole search population, except for:

* A random sample taken from the population of all cohort studies reporting < 50 cases over < 10 years in < 3 centres (A surrogate for Development stage studies)

$ A random sample from the population of all RCTs
Figure 1. Surgical Outcomes studies identified by our search strategy for 2000-2014, showing RCTs and other study designs.