



McCulloch, P., Feinberg, J., Philippou, Y., Kolia, A., Kehoe, S., Lancaster, G., Donovan, J., Petrinic, T., Agha, R., & Pennell, C. (2018). Progress in clinical research in surgery and IDEAL. *Lancet*, 392(10141), 88-94. [https://doi.org/10.1016/S0140-6736\(18\)30102-8](https://doi.org/10.1016/S0140-6736(18)30102-8)

Peer reviewed version

Link to published version (if available):  
[10.1016/S0140-6736\(18\)30102-8](https://doi.org/10.1016/S0140-6736(18)30102-8)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Elsevier at [http://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(18\)30102-8/fulltext](http://www.thelancet.com/journals/lancet/article/PIIS0140-6736(18)30102-8/fulltext). Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: <http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Progress in clinical research in surgery viewed through an IDEAL lens.

Peter McCulloch<sup>1</sup>, Joshua Feinberg<sup>2</sup>, Yiannis Philippou<sup>3</sup>, Angelos Koliass<sup>4</sup>, Sean Kehoe<sup>5</sup>, Gillian Lancaster<sup>6</sup>, Jenny Donovan<sup>7</sup>, Tatjana Petrinic<sup>8</sup>, Riaz Agha<sup>9</sup>, Christopher Pennell<sup>2</sup>

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
9. Specialist Registrar in Plastic Surgery, Guys and St Thomas' NHS Foundation Trust, London MSc

## SUMMARY

The quality of clinical research in surgery has long attracted criticism. High quality randomised trials have proved difficult to conduct 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 re-imbursement decisions, slow adoption of innovations, and permit widespread adoption of procedures which offer no benefit or cause harm. Improvement in the quality of surgical evidence would therefore be very valuable. The 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 - 2004 and 2010-2014, we noted apparent improvement in the use of standard outcome measures, adoption of CONSORT standards, evaluation 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 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 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 began 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.

#### WHAT IS SURGICAL RESEARCH?

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, since the focus of methodological criticism of surgical research has always been on this evaluation of surgical efficacy and effectiveness.

#### 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).

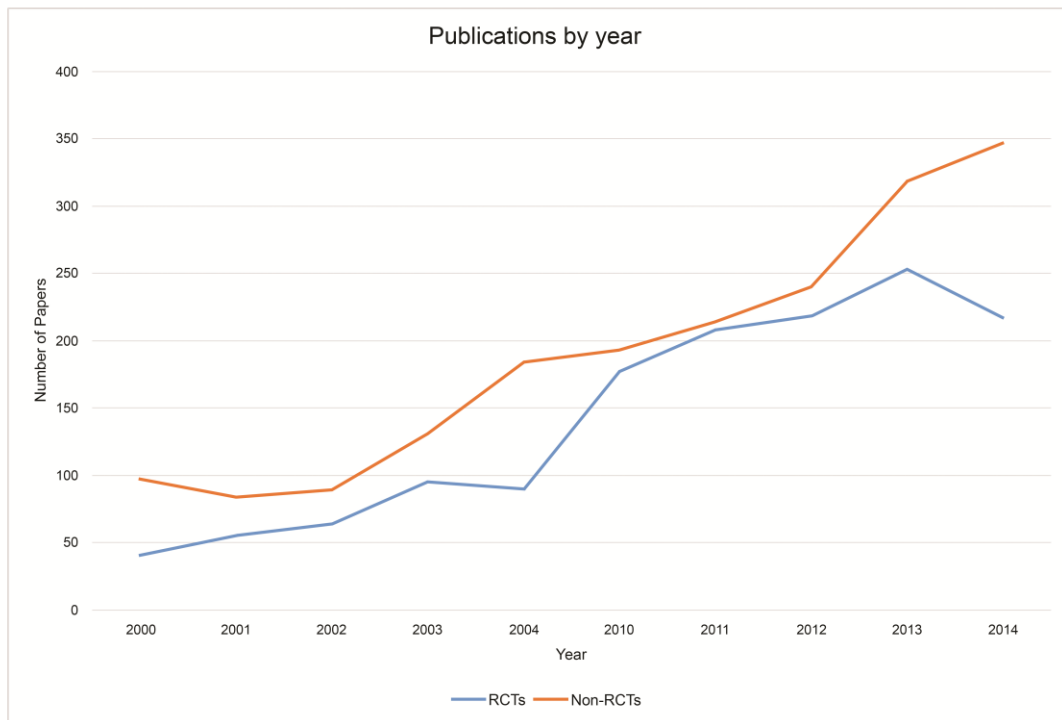


Figure 1. Surgical Outcomes studies identified by our search strategy for 2000-2014, showing RCTs and other study designs

#### 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, 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 only a minority of published surgical research is made up of RCTs. However, 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, 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 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 exercise was not powered to

demonstrate statistical significance, so our comments on the trends we found are necessarily tentative. The Recommendations we studied and the classes of study from which the samples were selected are shown in Table 2.

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.

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).

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.

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. None made reference to the CONSORT guidelines. By 2010-2014 reference to previous pilot work had not increased (one study), but five studies (20%) made reference to the CONSORT guidelines, and 12 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 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). A higher proportion of surgical RCTs now describe blinding of outcome assessors, and there is better evaluation of the quality with which surgery is delivered (including evaluation 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 distance still untravelled includes the persistently high proportion of retrospective case series in the literature and the rarity of qualitative research 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).

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.

What impact has IDEAL had since its inception in 2009? Citation growth suggests an accelerating upward trajectory (BMC surgery, MS under review) 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 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. NIHR has specified IDEAL 2b-type studies in several recent calls, and discussions about practical use of IDEAL in the NHS Commissioning through Evaluation project have been held with NHS England and with NICE.

What changes could encourage the development of high quality research in surgery? 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 would be a major step forwards. Journal editors could, relatively easily, challenge the publication of traditional case series and request IDEAL Development or Exploration studies instead, except in very rare situations. This 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.. Clear support from research funders for composite proposals incorporating IDEAL-type pre-RCT pilot/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 today, 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 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 surgical trainee research collaborative movement 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 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, 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 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 may also help. The regulatory framework can influence the success and integration into clinical practice of Registries, as the success of Sweden in this area shows(15). 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

#### VISION FOR THE FUTURE

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. 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 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 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 apply beyond the confines of surgery, 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(16), and plans for using IDEAL in radiotherapy(17), physiotherapy(18), and acupuncture[Prof 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 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, but there is a danger that their increasing expense and regulatory complexity, combined with reluctance to participate 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.

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  
Bolívar A. Villacís Bermeo  
Rebecca Wong  
DR Yadev

## REFERENCES

1. Horton R. Surgical research or comic opera: questions, but few answers. *Lancet* (London, England). 1996;347(9007):984-5.
2. Bell PR. Surgical research. *Lancet* (London, England). 1996;347(9013):1479; author reply 82.
3. Ruckley CV. Surgical research. *Lancet* (London, England). 1996;347(9013):1479-80; author reply 82.
4. Solomon MJ, Laxamana A, Devore L, McLeod RS. Randomized controlled trials in surgery. *Surgery*. 1994;115(6):707-12.
5. McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: problems and possible solutions. *BMJ*. 2002;324(7351):1448-51.
6. Barkun JS, Aronson JK, Feldman LS, Maddern GJ, Strasberg SM, Altman DG, et al. Evaluation and stages of surgical innovations. *Lancet* (London, England). 2009;374(9695):1089-96.
7. Ergina PL, Barkun JS, McCulloch P, Cook JA, Altman DG, group I. IDEAL framework for surgical innovation 2: observational studies in the exploration and assessment stages. *BMJ* (Clinical research ed ). 2013;346:f3011.
8. McCulloch P, Altman DG, Campbell WB, Flum DR, Glasziou P, Marshall JC, et al. No surgical innovation without evaluation: the IDEAL recommendations. *Lancet*. 2009;374(9695):1105-12.
9. Pennell CP, Hirst AD, Campbell WB, Sood A, Agha RA, Barkun JS, et al. Practical guide to the Idea, Development and Exploration stages of the IDEAL Framework and Recommendations. *The British journal of surgery*. 2016;103(5):607-15.
10. Mirheydar HS, Parsons JK. Diffusion of robotics into clinical practice in the United States: process, patient safety, learning curves, and the public health. *World journal of urology*. 2013;31(3):455-61.
11. Marshall DA, Pykerman K, Werle J, Lorenzetti D, Wasylak T, Noseworthy T, et al. Hip resurfacing versus total hip arthroplasty: a systematic review comparing standardized outcomes. *Clinical orthopaedics and related research*. 2014;472(7):2217-30.
12. Sampietro-Colom L, Martin J. Hospital-Based Health Technology Assessment: The Next Frontier. *Hospital-Based Health Technology Assessment*: Springer; 2016. p. 3-11.
13. 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. *Surgical innovation*. 2013;20(5):459-65.
14. Levin L, Krahn M, Goeree R, Mamdani M, Easty T, Fergussoon D, et al. Excite—A New Collaborative Model Of Pre-Market Evaluation Of Health Technologies. *Value in Health*. 2013;16(3):A264.
15. Levay C. Policies to foster quality improvement registries: lessons from the Swedish case. *Journal of internal medicine*. 2016;279(2):160-72.
16. 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.
17. Verkooijen HM, Kerkmeijer LGW, Fuller CD, Huddart R, Faivre-Finn C, Verheij M, et al. R-IDEAL: A Framework for Systematic Clinical Evaluation of Technical Innovations in Radiation Oncology. *Frontiers in oncology*. 2017;7:59.
18. Beard D, Hamilton D, Davies L, Cook J, Hirst A, McCulloch P, Paez A. Evidence-based evaluation of practice and innovation in Physical Therapy using the IDEAL-Physio framework.