

1 ~~Are we nearly there yet?~~ Progress in clinical research in surgery ~~viewed through from~~ an IDEAL
2 ~~lensperspective.~~

3 Peter McCulloch¹, Joshua Feinberg², Yiannis Philippou³, Angelos Koliass⁴, Sean Kehoe⁵, Gillian
4 Lancaster⁶, Jenny Donovan⁷, Tatjana Petrinic⁸, Riaz Agha⁹, Christopher Pennell²

- 5 1. Professor of Surgical Science & Practice, University of Oxford FULL PROFESSOR, MD
- 6 2. Resident, Department of Surgery, Maimonides Hospital, Brooklyn, NY. MD
- 7 3. NIHR Academic Clinical Fellow, Nuffield Department of Surgical Science, University of Oxford
- 8 MRCs
- 9 4. Clinical Lecturer, Division of Neurosurgery, School of Clinical Medicine, Cambridge University
- 10 PhD
- 11 5. Professor of Gynaecological Cancer, Institute of Cancer and Genomics, University of
- 12 Birmingham FULL PROFESSOR, MD
- 13 6. Professor, Institute of Primary Care and Health Sciences, Keele University FULL PROFESSOR,
- 14 PhD
- 15 7. Professor of Social Medicine, University of Bristol FULL PROFESSOR, PhD
- 16 ~~—~~Senior Librarian, Cairns Library, Oxford University Hospitals Foundation Trust
- 17 8.
- 18 ~~8.~~Specialist Registrar in Plastic Surgery, Guys and St Thomas' NHS Foundation Trust, London
- 19 MSc
- 20 9.

21 SUMMARY

22 The quality of clinical research in surgery has ~~long attracted come under sustained~~ criticism ~~over the~~
23 ~~last 20 years.~~ High quality randomised trials ~~have proved are more~~ difficult to conduct in surgery
24 ~~than for drug treatments, and .~~ Many surgical treatments have therefore been adopted without
25 adequate supporting evidence of efficacy and safety. This evidence deficit can adversely affect
26 research funding and re-imburement decisions, slowing adoption of innovations, and ~~can also~~
27 permit widespread adoption of procedures which ~~offer no benefit or cause harm turn out to be~~
28 ~~harmful or of no benefit.~~ Improvement in the quality of surgical evidence would therefore be very
29 valuable. ~~T~~Clinical research has become more sophisticated in recent years, and the IDEAL
30 Framework and Recommendations, ~~launched in 2009, have addressed some of the methodological~~
31 ~~problems in surgical research specify desirable qualities for surgical studies, and outline an~~
32 integrated evaluation pathway for surgery and similar complex interventions. We used the IDEAL
33 Recommendations to assess methodological progress in surgical research over time, assessed We
34 ~~therefore reviewed methodological progress in surgery, and~~ the uptake and influence of IDEAL and
35 identified the challenges to further methodological progress.

36 ~~We used a small sampling study, using compliance with IDEAL Recommendations to compare quality~~
37 ~~in surgical research Comparing studies from the periods 2000 - 2004 and 2010-2014 in two epochs~~
38 ~~ten years apart, to focus a review of progress, and looked for evidence about the adoption and~~
39 ~~impact of IDEAL.~~ ~~We~~ noted apparent improvement in ~~some areas, including~~ the use of standard
40 outcome measures, adoption of CONSORT standards, ~~and~~ evaluation of the quality of surgery and of
41 learning curves, but no progress in ~~others, such as~~ the use of qualitative research or ~~transparent~~
42 reporting of ~~modifications changes in procedures~~ during ~~early procedure~~ development. Better
43 education about research, integration of evaluation work into routine practice and training,
44 regulation and linkage of evaluation work to assessment and awards systems could foster further
45 improvements in surgical evidence. IDEAL has probably contributed only slightly to the

Formatted: List Paragraph,
Numbered + Level: 1 + Numbering
Style: 1, 2, 3, ... + Start at: 1 +
Alignment: Left + Aligned at: 0.63
cm + Indent at: 1.27 cm

Formatted: List Paragraph,
Numbered + Level: 1 + Numbering
Style: 1, 2, 3, ... + Start at: 1 +
Alignment: Left + Aligned at: 0.63
cm + Indent at: 1.27 cm

46 improvements described ~~to date~~, but its uptake is accelerating rapidly, ~~and~~ the need for the
47 integrated evaluation ~~template pathway~~ it offers for surgery and other complex treatments is
48 becoming more widely accepted.

49

50 INTRODUCTION

51 Just over 20 years ago, the Lancet published one of its most provocative Editorials since the era of
52 Thomas Wakley. Entitled “Surgical research or comic opera”, it lampooned clinical research in
53 surgery, contrasting current practice with the principles of the Evidence-Based Medicine
54 movement(1). Surgeons reacted with anger, claiming that they faced special problems which
55 frequently invalidated an approach based solely on randomised controlled trials(2, 3). Their protests
56 were not widely accepted, but the controversy ignited eventually proved constructive. Enquiry
57 began into why randomised controlled trials (RCTs) seemed to be so difficult to conduct in surgery,
58 and from these studies ~~a~~ picture emerged which partly justified the surgeons’ original objections(4,
59 5). Expert consensus conferences involving surgeons, EBM experts and others developed a credible
60 description of the process of evolution of innovative treatments, in surgery and other disciplines
61 where complex skilled procedures require adaptation to each individual patient. This construct was
62 termed the IDEAL Framework, referencing the terms used for the sequential stages in the
63 evolutionary process (Idea, Development, Exploration, Assessment and Long-term study)(6-8). The
64 Framework provided a theoretical basis for arguments against randomisation from the first patient
65 in these complex interventions. Each stage in the IDEAL Framework frames specific questions to be
66 addressed, ~~and~~ this led logically to the development of the IDEAL Recommendations, a set of
67 guidelines for study design and reporting aimed at answering these questions for each stage (See
68 Table 1 for a recently updated version of both Framework and Recommendations)(8, 9).

69 The IDEAL Recommendations have been widely recognised as a rational approach to developing an
70 integrated evaluation pathway for surgery and other complex interventions. Since they describe
71 desirable properties for clinical studies of surgery, they have obvious potential as a yardstick for
72 judging the methodological progress of surgical research. In this article we look first at how surgical
73 research has developed in the last 20 years, using adherence to the IDEAL Recommendations as a
74 measure of progress, ~~We then consider what impact IDEAL has had to date, how surgical research
75 could be further improved, and what role IDEAL could play in that process in the future. and second
76 at the development of IDEAL itself, assessing its evolution and its progress in terms of acceptance
77 and adoption in the surgical community internationally.~~

78 WHAT IS SURGICAL RESEARCH?

79 ~~Defining surgical research is challenging, and different definitions may be useful depending on the
80 reasons for needing one.~~ During the last 60 years we have seen the development of cardiac and
81 vascular surgery, organ transplantation, joint replacement, minimally invasive surgery and most
82 recently robotic surgery, advances of unquestionable importance to patients which have been based
83 on surgical research. However many leading university departments of surgery emphasise research
84 on topics such as the molecular genetics of diseases treated by surgery, the immunology of organ
85 transplantation and rejection, or stem cell treatments. Perhaps because of this, the direct study of
86 outcomes from technical innovation such as minimally invasive and robotic surgery and new
87 surgically implanted devices has been taken forward just as much by “non-academic” surgeons as by
88 professional researchers.

89 In this article we have considered only studies of the outcomes of surgical techniques, i.e. papers in
90 which the question addressed is around the effects of an operation, ~~either simply reporting the~~
91 ~~outcomes or comparing them to other treatment options, since t~~The focus of methodological
92 criticism of surgical research has always been on this evaluation of surgical efficacy and
93 effectiveness, ~~and comparisons of different time periods are clearly simpler if the types of study~~
94 ~~compared are restricted.~~

95 HOW HAS IT CHANGED?

96 Published surgical research has been steadily increasing in volume year on year. Using search terms
97 based on the above narrow definition of surgical research (see Appendix 1) we identified 41 surgical
98 RCTs in PubMed for the year 2000, and 246 for 2014. During the same time, the number of RCTs
99 recorded yearly on PubMed as a whole increased from 11,515 in 2000 to 27,426. Non-randomised
100 surgical outcome studies also showed a several-fold increase during this time. Randomised trials
101 may be increasing somewhat as a proportion of surgical outcomes studies, but this trend is not yet
102 so clear that we can be sure. In 2000, RCTs represented 30% ~~(41 of 138)~~ of all outcome studies
103 identified by our search criteria, and this rose to just under 50% ~~(208 of 422)~~ by 2011, but the figures
104 of 48% ~~(218 of 458)~~ in 2012 and 38% ~~(217 of 564)~~ in 2014 did not support the impression of a rising
105 trend (Figure 1).

106 METHODOLOGICAL TRENDS IN QUALITY: AN IDEAL ANALYSIS

107 An analysis of surgical research based solely on the proportion of studies which are RCTs gives an
108 inadequate view of the changes in surgical research over time, ~~specifically~~ because it does not
109 acknowledge the importance of pre-RCT studies of innovations still undergoing modification. The
110 IDEAL Recommendations specify a number of uncontroversial desirable features for clinical studies
111 of surgical interventions, especially for these earlier stages of surgical research. Examining whether
112 these ~~specific~~ features have become more prevalent in published work over the years allows us to
113 look at the progress of surgical research methodology in greater depth and detail. ~~The IDEAL~~
114 ~~Framework provides clear justification for the fact that it remains true that~~ only a minority of
115 published surgical research is made up of RCTs, ~~and indeed the IDEAL Framework provides clear~~
116 ~~justification for this.~~ However, ~~in order~~ for surgical research to make progress the validity and
117 accuracy of the studies during the earlier stages of the evolution of new techniques prior to an RCT
118 ~~needs to~~ improve in quality, as well as the RCTs themselves. The 1990s critiques of the surgical
119 literature were based largely on analysis of contemporary studies of techniques in ~~these earlier~~
120 ~~stages IDEAL stages 2a and 2b~~, for which the retrospective case series was, at that time, the
121 standard (and grossly inadequate) format for publication. The IDEAL Recommendations specify
122 desirable characteristics for early stage studies which the case series plainly lacked. We were
123 interested in whether compliance with specific IDEAL Recommendations had improved since the
124 first IDEAL publication, suggesting progress in the direction IDEAL recommended. We therefore
125 ~~conducted a sampling exercise looking at studies sampled studies~~ from 2000-2004 and ~~from~~ 2010-
126 2014, the latter being the first full 5 year period after the publication of the original IDEAL articles
127 ~~(See Appendices 1 and 2 for details)~~. This ~~sampling~~ exercise was not powered to demonstrate
128 statistical significance, so our comments on the trends we found are necessarily tentative.

129 SAMPLING SURVEY

130 ~~This article is principally intended as discussion and debate, and this sampling study of the changes~~
131 ~~in surgical research over time supports this by providing some objective evidence on trends: its'~~
132 ~~methodology does not however allow us to draw firm quantitative conclusions from it. We~~

133 conducted a literature search designed to identify clinical studies of the outcomes of surgical
134 techniques (for details of our search strategy see Appendix 1). We then took random samples of 25
135 papers from each of the two five-year subpopulations (200-2004 and 2010-2014), and compared
136 them for the presence or absence of desirable characteristics specified by the IDEAL
137 Recommendations (See Table 1). Several different random samples were taken to allow us to study
138 different IDEAL Recommendations in the appropriate subgroup of studies, since some
139 Recommendations apply only to specific stages in the IDEAL Framework. Further details of our
140 methods and reasoning can be found in Appendix 2. Additional details and a list of the papers
141 reviewed are available from the corresponding author.

142 •—— Description of iterative changes in the procedure would be expected only in IDEAL 2a
143 Development studies, and for this Recommendation we therefore sampled only from amongst
144 papers which appeared to describe innovations at this stage, using the rules of thumb described by
145 Pennell et al(9) to identify these.

146 •—— IDEAL recommends that operator learning curves are evaluated before embarking on RCTs;
147 we therefore searched only amongst reports of RCTs for evidence that learning curves were
148 considered whilst deciding which centres should enter patients.

149 •—— IDEAL recommends the use of pilot and feasibility studies and qualitative studies to inform
150 protocol development during the planning of RCTs, and we therefore selected only from amongst
151 RCTs when analysing compliance with these recommendations. The Recommendations we studied
152 and the classes of study from which the samples were selected are shown in Table 2.

153

154 IDEAL RECOMMENDATIONS APPLICABLE TO ALL OUTCOMES STUDIES

155 The proportion of papers reporting prospective (as opposed to retrospective) studies remained
156 about the same over the decade between the two eras; it was 60% (15/25) in 2000-2004 and 64%
157 (16/25) in 2010-14. The procedure was reasonably well described in nearly all cases in both cohorts.
158 By contrast information about the quality of surgery was rarely provided, although with some
159 apparent improvement over time. In the later sample there was some attempt to evaluation of the
160 quality of surgery in 5 papers (20%), but there was none (0%) in any of the papers from 2000-2004.
161 The use of well-recognised standardised measures for outcomes, patient characteristics and other
162 important data showed an apparent modest increase from 11/25 (44%) of 2000-2004 papers to
163 15/25 (60%) of papers from 2010-2014.

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

165 The IDEAL recommendations for early development studies in Stage 2a stress the need for an
166 account of changes to the procedure or indications during the development process. This remains
167 relatively rare (16% (4/25) in both eras). Results at this stage should be presented sequentially so
168 that trends and their associations with changes made may be more evident. Only four papers (16%)
169 followed these recommendations during the series in the 2000-2004 cohort, and the same
170 proportion did so in 2010-2014, demonstrating no trend over time.

171 IDEAL RECOMMENDATIONS APPLICABLE PREPARATORY STUDIES FOR RCTS (STAGE 2B) AND RCTS 172 (STAGE 3);

173 The IDEAL Recommendations suggest prospective collaborative collection of non-randomised data
174 as a preparatory step towards multicentre RCTs. We found that the proportion of randomised trials

Formatted: Normal, No bullets or numbering

175 which referred to such data collection nearly doubled in the later sample (9/25 or 36%) compared to
176 the earlier one (5/25 papers, 20%). Although none of the sampled papers referenced the IDEAL
177 Recommendations, it appears that stepwise progression from collaborating on prospective data
178 collection to doing a trial together is becoming more common. ~~Prospective collaborative cohort
179 studies increased considerably as a proportion of all non-RCTs between the eras sampled, whereas
180 small case series with less than 50 patients have declined. Changes in both medicine and IT which
181 have made collaborative data collection exercises less costly and inconvenient during the period
182 under study, may have contributed to the change.~~

183 Blinding or masking of outcome assessors in RCTs was reported in 6/25 (24%) of trials in the earlier
184 period and in 10/25 (40%) in 2010-14. There were no examples of the use of preparatory qualitative
185 studies in any of the RCTs sampled in either epoch. The proportion of studies in which action was
186 taken to address the issue of bias introduced by operator learning curves increased from 5 (20%) in
187 the earlier epoch to 11 (44%) in 2010-2014.

188 Only two (8%) of the randomised trials sampled between 2000 and 2004 explicitly mentioned the
189 use of pilot/feasibility work prior to the RCT, ~~although 16 studies made reference to previous related
190 work by one or more of the co-authors.~~ None made reference to the CONSORT guidelines. ~~Sample
191 sizes ranged from 20 to 1289, with the largest studies testing 'equivalence'. All studies carried out
192 hypothesis testing but only 7 studies carried out a sample size calculation and were adequately
193 powered for such testing.~~ By 2010-2014 reference to previous pilot work had not increased (one
194 study), ~~but although again 16 studies referenced previous related work. However~~ five studies (20%)
195 made reference to the CONSORT guidelines, and 12 included a study flowchart. ~~Sample sizes ranged
196 from 11 to 597, with the largest study once again being an equivalence trial, and 12 studies reported
197 the sample size calculation.~~ In both epochs absence of evidence of difference ($p > 0.05$) in
198 underpowered studies was routinely but incorrectly taken to be evidence of no difference between
199 techniques. ~~The proportion of studies in which action was taken to address the issue of bias
200 introduced by operator learning curves increased from 5 (20%) in the earlier epoch to 11 (44%) in
201 2010-2014.~~

202 ANALYSIS AND FORWARD VIEW

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

220 ~~IDEAL-type evaluation pathway would probably have mitigated these adverse outcomes and given~~
221 ~~us much clearer evidence on risks, benefits and costs.~~

222 We know little about what influences methodological decision-making amongst surgical
223 researchers, but some factors appear fairly obvious. There are still strong career incentives for
224 publishing poor research in surgery, particularly retrospective case series, and it is easy to find
225 clinical journals that will accept them uncritically. They are generally exempt from many of the
226 regulatory hurdles which challenge prospective research and therefore represent a cheap, rapid
227 route to publication. As long as this form of research retains its value for career advancement it is
228 unlikely to disappear. On the other hand, the very difficulties which stimulated the development of
229 IDEAL still make surgical trials difficult to organise, and there is no correlate for the abundant
230 commercial funding available for pharma trials. The weakness of basic training in the principles of
231 clinical research for surgeons is a third important obstacle to progress. ~~The issues addressed by~~
232 ~~IDEAL are very familiar to surgeons, but most are unaware of previous work on the potential~~
233 ~~solutions.~~

234 ~~What impact has IDEAL had since its inception in 2009? It is difficult to evaluate the rate of diffusion~~
235 ~~of understanding of IDEAL through the surgical community.~~ Citation growth suggests an accelerating
236 ~~upward~~ trajectory (BMC surgery, MS under review) but just as occurred with the EBM movement ~~at~~
237 ~~its inception~~, understanding lags behind familiarity, and practical use is still further behind.
238 ~~U~~Practical uptake and use in the Health Technology Assessment community has however, made a
239 strong start, with programmes in Canada(12) and the Netherlands(13) that use IDEAL to guide
240 evaluation of medical technologies and devices. A recently launched enterprise to offer device
241 manufacturers a comprehensive evaluation service for innovative new products, EXCITE
242 International(14), has embraced IDEAL as a central part of its methodology. ~~ical strategy.~~ ~~NIHR has~~
243 ~~specified IDEAL 2b-type studies in several recent calls, and d~~Discussions about practical use of IDEAL
244 in the NHS Commissioning through Evaluation project have ~~also~~ been held with NHS England and
245 with NICE.

246 ~~W~~Looking more widely than the adoption of IDEAL, what changes could ~~encouragespeed~~ the
247 development of high quality research in surgery? ~~T~~As we noted in 2009, those who can influence
248 the research environment are those responsible for educating and training surgeons and those who
249 decide which research gets funding and publication(8). The adoption of higher standards by ~~surgical~~
250 ~~journals (especially surgical journals)~~ would be a major step forwards. ~~and some, such as the~~
251 ~~International Journal of Surgery, have begun to specify that IDEAL-format studies are welcome(15).~~
252 Journal editors could, relatively easily, challenge the publication of traditional case series and
253 request IDEAL Development or Exploration studies instead, except in ~~very the rare situations where~~
254 ~~a case series is the only reasonable study design.~~ ~~This~~ Adoption of these formats, and the
255 ~~integrated pathway they belong to,~~ could help to reduce ~~research waste by lowering~~ the prevalence
256 of small ad-hoc mini RCTs or single arm studies trying out a new technique with no future
257 implementation pathway or plan. ~~this would be an important contribution to reducing research~~
258 ~~waste.~~ Clear ~~support messages~~ from research funders ~~for that they will support~~ composite
259 proposals incorporating IDEAL-type pre-RCT pilot/feasibility studies ~~to prepare the ground for a trial~~
260 ~~would quickly ill be important in modify transforming~~ the behaviour of the clinical research
261 community. ~~It is encouraging that bodies such as the UK's NIHR are beginning to experiment with~~
262 ~~IDEAL 2b-like studies, but they have as yet made no clear policy statement.~~ ~~Whilst IDEAL has~~
263 ~~received virtually no direct opposition, there is a tangible reluctance amongst some funders to~~
264 ~~extend it a welcome.~~ ~~This appears to be based on the (incorrect) perception that it challenges the~~
265 ~~principles of conventional medical research methodology, because of its championing of~~

Formatted: Highlight

266 uncontrolled study designs for some questions where no comparison between treatments is
267 involved. Presenting a case for IDEAL which reconciles true believers and sceptics on the role of
268 RCTs in medical research is an impossible task, but open debate needs to be encouraged. Silent
269 passive resistance which evades discussion is unscientific and can only retard the search for a better
270 framework for evaluating surgery and other complex interventions.

271 If our aim is to continually enhance surgical research quality, the historical origins of modern science
272 can point us to what catalyses a spirit of systematic enquiry within communities. There is an
273 excellent case that the Enlightenment, which ushered in the scientific revolution, was bought about
274 by the rejection of dogma and the enhancement of basic educational attainment (16). The
275 complaint, voiced 20 years ago, that surgeons (and doctors in general) were surprisingly ill-educated
276 about the principles of scientific methodology for investigating medical treatments is equally valid
277 today, even in the countries where EBM has flourished. Medical schools and postgraduate surgical
278 training programmes are still failing to produce graduates who understand the methodological basis
279 of clinical research and are able to apply this knowledge. Correcting this will require more than just
280 courses of study, since this type of complex applied knowledge can only be thoroughly incorporated
281 through experiential learning. Current surgical training in most countries separates academic and
282 clinical work in an artificial and unhelpful way, with segments for dedicated “research time” and
283 separate career tracks for academic and non-academic surgeons. As we saw with the minimally
284 invasive and robotic revolutions, this risks creating a situation where the drive to innovate and the
285 ability to evaluate are separated, to the detriment of evidence-based progress. Having a clinical
286 workforce who are not afraid to set up simple early-stage data collection efforts which will yield
287 valid and useful results would be a major contribution to improving both quality and quantity in
288 surgical research. One very welcome recent development has been the successful movement to
289 establish surgical trainee research collaborative movements in the UK*. These groups give trainees
290 practical participative experience of research, and have demonstrated their ability to recruit patients
291 at impressive rates and minimal cost. Their ability to develop RCTs on important questions is limited
292 by the short tenure and lack of political power of their memberships’ situation, but they have
293 enormous potential to conduct IDEAL 2b studies rapidly and effectively, and to use these to drive the
294 funding and development of a subsequent RCT.

295 As well as opportunities to do clinical research as an integral part of training, achieving true
296 integration of research and practice will require appropriate incentives for both trainees and
297 established surgeons in “non-academic” posts. The distinctions we currently make between
298 research, audit and quality improvement are often unhelpful to this integration, and it may be more
299 useful to talk of involvement in scientific evaluation. Linking involvement in high-quality evaluation
300 to appraisal, tenure and rewards structures may be helpful, but so may public recognition, and
301 opportunities to develop other initiatives. Making institutional approval for the development of
302 innovative practice conditional on agreement to collect and submit appropriate data and submit it to
303 the host institution, or for publication, would be a powerful way of enhancing evidence
304 accumulation. Regulation may also help. The current pharmaceutical industry evaluation paradigm,
305 which has hugely improved the evaluation of new drugs, can be largely attributed to regulatory
306 action in the wake of the Thalidomide disaster (17). The regulatory framework can influence the
307 success and integration into clinical practice of Registries, as the success of Sweden in this area
308 shows (158). However regulation is usually most effective where popular opinion has already been
309 won round to the ideas behind it, as the examples of restrictions on drink driving and smoking in
310 public places illustrate. Regulation may also fail to predict future research trends and developments.

311 VISION FOR THE FUTURE

312 | ~~As this review has shown,~~ surgical research is getting better, although it still has a long way to go. It
313 | was unfairly maligned in the first place, as understanding of the real problems it faced was under-
314 | developed 20 years ago. The IDEAL Framework and Recommendations have probably contributed
315 | only in a minor way to the improvements seen so far, but their influence is growing, they are very
316 | useful as a yardstick to measure progress, and they represent a serious attempt to create a new
317 | paradigm for surgical research ~~methodology in the future~~. The idea of a logical series of study
318 | questions and types based on the realities of how surgical operations evolve is clearly of value, and
319 | where IDEAL proves imperfect it is likely to be either modified or replaced by a better version, rather
320 | than by a return to methodological anarchy. Because complex interventions typically require a
321 | period of iterative improvement before reaching a stable form, they cannot be subjected to
322 | ~~immediate~~ valid comparisons with alternatives until this phase is over. The variety of influences on
323 | outcome which can be generated by subject heterogeneity and by variations in the quality of
324 | intervention delivery, especially whilst operators are learning, is practically infinite. Defining a
325 | subject group and a version of the intervention for a trial will therefore require ~~a~~ both a basis for
326 | decision making and considerable negotiation. Substantial prospective empirical data clearly
327 | represent a more reliable basis than theory combined with small datasets full of contextual biases.
328 | Hence collaborative collection of non-randomised data to assist with decision making ~~is~~ justified
329 | wherever complexity impedes definition of the study population or the intervention.

330 | ~~These principles - vision of "no innovation without evaluation"~~ applies beyond the confines of
331 | ~~surgery surgery and quasi-surgical treatments~~, and beyond conventional academic research. The
332 | ~~concept-principle~~ of integrated stepwise evaluation, beginning with a study of development of the
333 | innovation and proceeding through more comprehensive evaluation of its properties and uses
334 | before comparing it with alternatives, seems applicable to complex interventions in many fields,
335 | both in healthcare and other domains. A version of IDEAL for evaluating therapeutic devices was
336 | published last year(169), and plans for using IDEAL in radiotherapy(1720), physiotherapy(18) [Paez A
337 | et al, Physical Therapy, accepted, in press], and acupuncture [Prof REF Xin Sun, personal
338 | communication] are being implemented. Psychological therapies and complex social or quality
339 | improvement interventions are other areas where it may prove useful.

340 | The current version of the IDEAL Framework and Recommendations is clearly ~~not the last word on~~
341 | ~~the subject, but~~ a work in progress. How it may need to be modified depends to a large extent on
342 | how clinical research methodology itself evolves. For example, IDEAL's current form is predicated on
343 | the assumption that RCTs will remain the "gold standard" methodology for comparing treatments.
344 | This seems likely, ~~given their obvious structural superiority~~, but there is a danger that their
345 | increasing expense and regulatory complexity, combined with reluctance to participate ~~amongst~~
346 | ~~patients~~ in cultures which value the individual's right to choose as paramount, and competition from
347 | the application of sophisticated risk adjustment techniques to very large observational datasets may
348 | reduce their pre-eminence in the future. However even a major change such as this would merely
349 | require the Framework to be adapted, rather than abolished: in other words, if IDEAL did not already
350 | exist, it would be necessary to invent it.

351 | Now that a viable alternative is beginning to emerge, the culture of surgical case series and other
352 | weak study designs should be consigned to history, and an integrated evaluation pathway for
353 | surgery using methodology appropriate to the task should be adopted.

354 | AUTHOR CONTRIBUTIONS AND COMPETING INTERESTS

355 | The article was conceived by PM following discussion with the Lancet editorial team. He wrote the
356 | first and final drafts and supervised the work of other authors. JF scanned and catalogued the

357 literature search results, organised the literature sampling analysis programme and reviewed the
358 results, acting as general co-ordinator and guardian of data. TP conducted and advised on literature
359 searches and bibliometrics. YP, JF, PM, AK, SK, GL, RA & CP all contributed to analysis of the
360 literature. All authors contributed to and commented on drafts of the MS. PM, JF, AK, SK, CP and RA
361 are members of the IDEAL Collaboration. This work originated from the IDEAL International
362 Conference at St Katherine's College Oxford in April 2016, which was funded by Oxford AHSN,
363 Medtronic plc, Johnson and Johnson and the Health Foundation.

364 ACKNOWLEDGEMENTS

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

369 The IDEAL Collaboration members who assisted with this paper were:

370 Chris Bainbridge
371 Dimitri Barski
372 Elena Berti
373 Egon Cholakian
374 Pierre-Alain Clavien
375 Tammy Clifford
376 Andrew Currie
377 Thorston Ecke
378 Liane Friedman
379 Holger Gerullis
380 Trish Groves
381 Judith Hargreaves
382 Catherine Hatzantonis
383 Emma Jones
384 Joey Kwong
385 Mario Lobao
386 Aaron Lawson McLean
387 Jose Merino
388 Heather McIntosh
389 Kemp Narayanasetty
390 Arsenio Paez
391 Jan Rongen
392 Markus Schafer
393 Shafaque Shaikh
394 Matthew Smith
395 Roxane Diane Staiger
396 Casper Tax
397 Ajay Upadhyay
398 Bolívar A. Villacís Bermeo
399 Rebecca Wong
400 DR Yadev

Formatted: English (U.K.)

Formatted: Spanish (International Sort)

Formatted: English (U.K.)

401
402
403

404 REFERENCES

405

- 406 1. Horton R. Surgical research or comic opera: questions, but few answers. *Lancet* (London,
407 England). 1996;347(9007):984-5.
- 408 2. Bell PR. Surgical research. *Lancet* (London, England). 1996;347(9013):1479; author reply 82.
- 409 3. Ruckley CV. Surgical research. *Lancet* (London, England). 1996;347(9013):1479-80; author
410 reply 82.
- 411 4. Solomon MJ, Laxamana A, Devore L, McLeod RS. Randomized controlled trials in surgery.
412 *Surgery*. 1994;115(6):707-12.
- 413 5. McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: problems
414 and possible solutions. *BMJ*. 2002;324(7351):1448-51.
- 415 6. Barkun JS, Aronson JK, Feldman LS, Maddern GJ, Strasberg SM, Altman DG, et al. Evaluation
416 and stages of surgical innovations. *Lancet* (London, England). 2009;374(9695):1089-96.
- 417 7. Ergina PL, Barkun JS, McCulloch P, Cook JA, Altman DG, group I. IDEAL framework for surgical
418 innovation 2: observational studies in the exploration and assessment stages. *BMJ* (Clinical research
419 ed). 2013;346:f3011.
- 420 8. McCulloch P, Altman DG, Campbell WB, Flum DR, Glasziou P, Marshall JC, et al. No surgical
421 innovation without evaluation: the IDEAL recommendations. *Lancet*. 2009;374(9695):1105-12.
- 422 9. Pennell CP, Hirst AD, Campbell WB, Sood A, Agha RA, Barkun JS, et al. Practical guide to the
423 Idea, Development and Exploration stages of the IDEAL Framework and Recommendations. *The*
424 *British journal of surgery*. 2016;103(5):607-15.
- 425 10. Mirheydar HS, Parsons JK. Diffusion of robotics into clinical practice in the United States:
426 process, patient safety, learning curves, and the public health. *World journal of urology*.
427 2013;31(3):455-61.
- 428 11. Marshall DA, Pykerman K, Werle J, Lorenzetti D, Wasylak T, Noseworthy T, et al. Hip
429 resurfacing versus total hip arthroplasty: a systematic review comparing standardized outcomes.
430 *Clinical orthopaedics and related research*. 2014;472(7):2217-30.
- 431 12. Sampietro-Colom L, Martin J. Hospital-Based Health Technology Assessment: The Next
432 Frontier. *Hospital-Based Health Technology Assessment*: Springer; 2016. p. 3-11.
- 433 13. Heikens JT, Gooszen HG, Rovers MM, van Laarhoven CJ. Stages and evaluation of surgical
434 innovation: a clinical example of the ileo neorectal anastomosis after ulcerative colitis and familial
435 adenomatous polyposis. *Surgical innovation*. 2013;20(5):459-65.
- 436 14. Levin L, Krahn M, Goeree R, Mamdani M, Easty T, Fergusson D, et al. Excite—A New
437 Collaborative Model Of Pre-Market Evaluation Of Health Technologies. *Value in Health*.
438 2013;16(3):A264.
- 439 ~~15. — International Journal of Surgery. Guide for Authors: Elsevier; 2017 [Guide for authors].~~
440 ~~Available from: [https://www.elsevier.com/journals/international-journal-of-surgery/1743-](https://www.elsevier.com/journals/international-journal-of-surgery/1743-9159/guide-for-authors)~~
441 ~~[9159/guide-for-authors](https://www.elsevier.com/journals/international-journal-of-surgery/1743-9159/guide-for-authors).~~
- 442 ~~16. — Herman A, Bishop J. The Scottish Enlightenment: The Scots' invention of the modern world:~~
443 ~~Fourth Estate London; 2003.~~
- 444 ~~17. — Greene JA, Podolsky SH. Reform, Regulation, and Pharmaceuticals — The Kefauver-Harris~~
445 ~~Amendments at 50. *New England Journal of Medicine*. 2012;367(16):1481-3.~~
- 446 ~~158.~~ Levay C. Policies to foster quality improvement registries: lessons from the Swedish case.
447 *Journal of internal medicine*. 2016;279(2):160-72.
- 448 ~~169.~~ Sedrakyan A, Campbell B, Merino JG, Kuntz R, Hirst A, McCulloch P. IDEAL-D: a rational
449 framework for evaluating and regulating the use of medical devices. *BMJ*. 2016;353:i2372.
- 450 ~~1720.~~ Verkooijen HM, Kerkmeijer LGW, Fuller CD, Huddart R, Faivre-Finn C, Verheij M, et al. R-
451 IDEAL: A Framework for Systematic Clinical Evaluation of Technical Innovations in Radiation
452 Oncology. *Frontiers in oncology*. 2017;7:59.

453 | [18. Beard D, Hamilton D, Davies L, Cook J, Hirst A, McCulloch P, Paez A. Evidence-based](#)
454 | [evaluation of practice and innovation in Physical Therapy using the IDEAL-Physio framework.](#)

455

Table 1. Summary of the IDEAL Framework and Recommendations 2017

	IDEAL Framework (Description of stage of evolution of intervention)	IDEAL Recommendations (For stage-specific study design and reporting)
<p>Stage 1 Idea <i>First in human</i></p>	<p>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</p> <p>Stage Endpoint: Once a decision is made to conduct a series of cases, i.e. to proceed to stage 2a.</p>	<ul style="list-style-type: none"> • 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
<p>Stage 2a Development</p> <p>Iterative modification of the intervention until a stable version is achieved. Design: Single centre/single intervention; case series/prospective cohort</p>	<p>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.</p> <p>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.</p>	<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 the above. • Informed consent should explain current status of intervention and consequent uncertainties around risk **

<p>Stage 2b Exploration</p> <p>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 observational to comparative evaluation.</p>	<p>Purpose: Achieving consensus between surgeons and centres on the parameters for an RCT (if possible)</p> <p>Number & Types of Patients: Many; broadening indication to include all potential beneficiaries</p> <p>Number & Types of Surgeons: Many; innovators, early adopters, early majority[@]</p> <p>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.</p> <p>Intervention: The procedure is stable in individual hands but variation exists between operators; acceptable variants are subsequently defined by analysis of pooled results</p> <p>Method: Prospective multi-centre exploration cohort study or pilot/feasibility multicentre RCTs.</p> <p>Outcomes: Safety; clinical outcomes (specific/graded); short-term outcomes; patient centred/reported outcomes; feasibility outcomes</p> <p>Stage Endpoints: Demonstrate that technique can be more widely adopted; and, Demonstrate that progression to RCT is desirable and feasible</p>	<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 multi-centre co-operative data collection, incorporating feasibility issues such as: <ul style="list-style-type: none"> ○ 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.
<p>Stage 3 Assessment</p> <p>Definitive comparative evaluation of main efficacy and safety aspects of new technique against current best treatment.</p>	<p>Purpose: Comparative effectiveness testing</p> <p>Number & Types of Patients: Many; expanded indications (well-defined)</p> <p>Number & Types of Surgeons: Many; early majority[@]</p> <p>Output: Comparison with current standard therapy</p> <p>Intervention: Stable, with acceptable variations clearly defined</p> <p>Method: RCT with or without additions/modifications; alternative designs (cluster, preference RCTs, stepped wedge, adaptive designs)</p> <p>Outcomes: Clinical outcomes (specific and graded); potentially Patient Reported outcomes , Health</p>	<ul style="list-style-type: none"> • 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

	<p>Economic outcomes</p> <p>Stage Endpoints: Clear valid evidence on relative effectiveness of innovation; and, Identification of issues requiring long term monitoring.</p>	<p>SPIRIT (for RCT protocol design)</p>
<p>Stage 4 Long term monitoring</p>	<p>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</p> <p>Stage Endpoints: dependent on lifecycle of device/procedure. Registries for devices – IDEAL-D Registries at earlier stages of IDEAL</p>	<ul style="list-style-type: none"> • 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

@ 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 2

	2000-2004	2010-2014	Absolute difference (95% CI)
IDEAL Recommendation – no./total no. (%)			<i>percentage points</i>
Use of standardised terminology	11/25 (44)	15/25 (60)	16 (-13.6 to 42.3)
Definition and description of procedure	20/25 (80)	22/25 (88)	8 (-15.8 to 31.1)
Prospective data collection	15/25 (60)	16/25 (64)	4 (-24.1 to 31.3)
Explanation of modifications during early studies (2a) [*]	4/25 (16)	4/25 (16)	0 (-23.5 to 23.5)
Prior analysis of learning curves in pre-RCT studies [§]	5/25 (20)	11/25 (44)	24 (-4.5 to 48.2)
Use of quality control measures	0/25 (0)	5/25 (20)	20 (-1 to 41.3)
Use of qualitative research to define RCT questions (3) [§]	0/25 (0)	0/25 (0)	0 (-16.6 to 16.6)
Use of prior prospective cohort study to prepare for RCTs (3) [§]	5/25 (20)	9/25 (36)	16 (-11.4 to 40.7)
Mention of pilot or feasibility studies to prepare for RCT (3) [§]	2/25 (8)	1/25 (4)	- 4 (-23.9 to 15.5)
Blinding reported in RCTs (3) [§]	6/25 (24)	10/25 (40)	16 (-12.2 to 41.2)

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

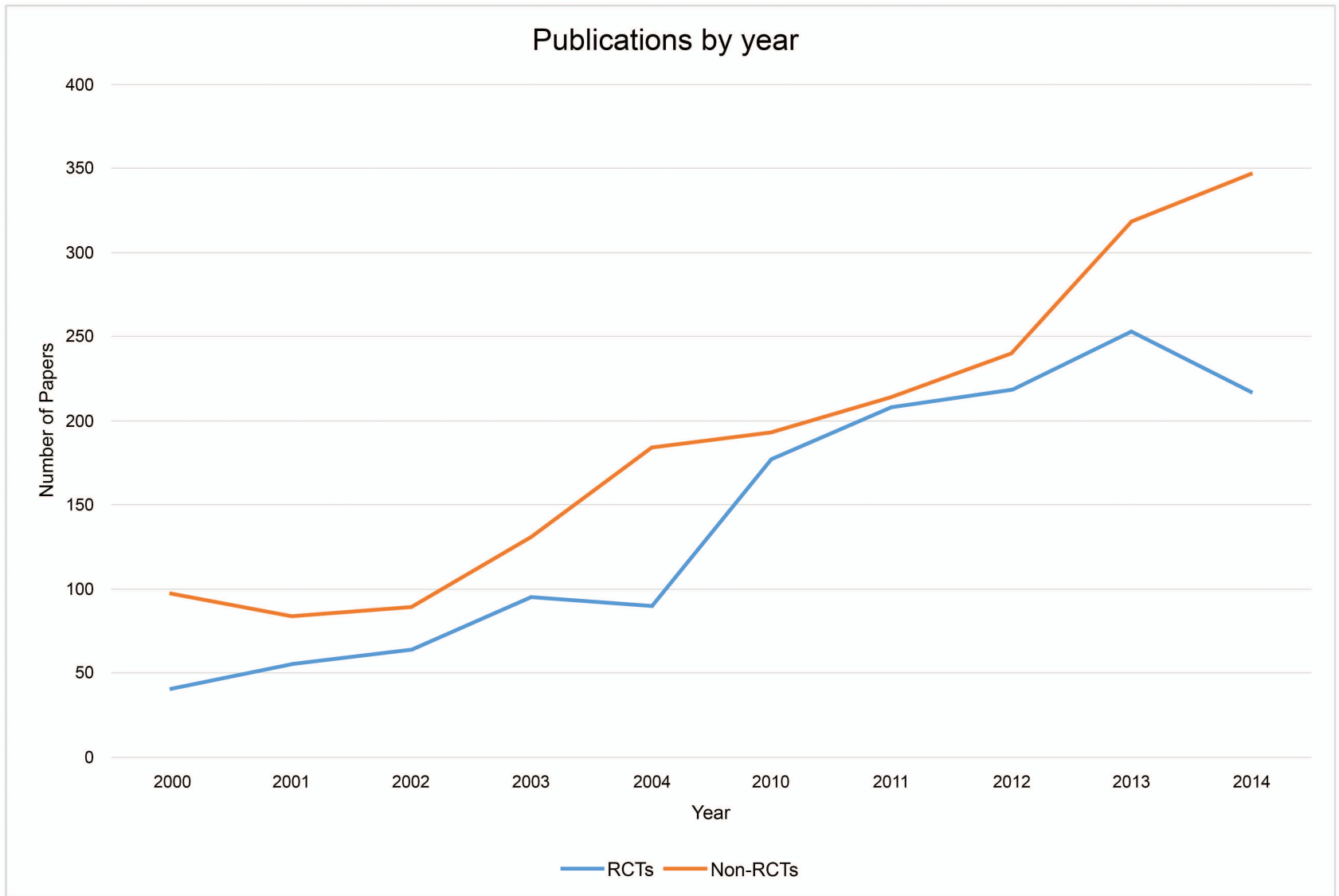


Figure 1. Surgical Outcomes studies identified by our search strategy for 2000-2014, showing RCTs and other study designs