No surgical innovation without evaluation: Evolution and further development of the IDEAL Framework and Recommendations.

This paper updates the IDEAL Framework and Recommendations following review at international expert meetings held during IDEAL Conferences in Oxford, UK in 2016 and New York, USA in 2017.

Allison Hirst, Project Manager/Researcher, IDEAL Collaboration [1]*

Yiannis Philippou, NIHR Academic Clinical Fellow in Urology [1]*

Jane Blazeby, Prof of Surgery and Honorary Consultant Surgeon [2]

Bruce Campbell, Honorary Professor [3]

Marion Campbell, Prof of Health Services Research [4]

Joshua Feinberg, General Surgery Resident [5]

Maroeska Rovers, Professor in Evidence-based Surgery [6]

Natalie Blencowe, NIHR Clinical Lecturer, Surgery [2]

Christopher Pennell, Resident Physician [5]

Tom Quinn, Prof of Nursing & Associate Dean for Research & Innovation [7]

Wendy Rogers, Professor of Clinical Ethics [8]

Jonathan Cook, Assoc Professor, Centre for Statistics in Medicine [9]

Angelos G. Kolias, Clinical Lecturer in Neurosurgery [10]

Riaz Agha, Speciality Registrar [11]

Philipp Dahm, Prof of Urology [12]

Art Sedrakyan Vice-Chair IDEAL Collaboration, Professor of Health Services Research and Policy [13 #]

Peter McCulloch, Chair, IDEAL Collaboration and Professor of Surgical Science & Practice [1 #]

** AH and YP Contributed equally as joint first authors*

[#] *PMcC and AS Contributed equally as joint senior authors*

on behalf of IDEAL Collaboration (attendees at 2016 Oxford consensus meeting listed in full in Acknowledgements with permission)

Authors affiliations:

1. Nuffield Dept of Surgical Sciences, University of Oxford, UK
2. Bristol Centre for Surgical Research, University of Bristol, UK
3. University of Exeter Medical School, UK
4. Health Services Research Unit, University of Aberdeen, UK
5. Maimonides Medical Center, Brooklyn, NY, USA
6. Radboud University Medical Center, Nijmegen, The Netherlands
7. Faculty of Health, Social Care and Education, Kingston University and St George's, University of London, UK
8. Department of Philosophy and Department of Clinical Medicine, Macquarie University, Sydney, Australia
9. Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, UK
10. Division of Neurosurgery, Dept of Clinical Neurosciences, Addenbrooke's Hospital and University of Cambridge, UK and Surgery Theme, Cambridge Clinical Trials Unit, UK
11. Department of Plastic Surgery, Guy's and St. Thomas' NHS Foundation Trust, London, UK
12. Minneapolis VAMC and University of Minnesota, Dept of Urology, USA
13. Weill Cornell Medical College, New York, USA

Corresponding authors (and reprints):

Prof Peter McCulloch, Professor of Surgical Science & Practice, University of Oxford,

Nuffield Dept of Surgical Sciences (Level 6), John Radcliffe Hospital, Oxford OX3 9DU

Tel. +44 (0)1865 612276

peter.mcculloch@nds.ox.ac.uk

Allison Hirst, IDEAL Collaboration Project Manager/Researcher, University of Oxford,

Nuffield Dept of Surgical Sciences (Level 6), John Radcliffe Hospital, Oxford OX3 9DU

allison.hirst@nds.ox.ac.uk

Funding: No specific funding was obtained for this update of IDEAL. The discussion meeting was funded as part of the IDEAL International Conference in Oxford on 7th April 2016 by several sponsors as follows: the Oxford Academic Health Science network (Oxford AHSN), The Health Foundation, Medtronic and Johnson and Johnson (J&J). It was also supported by BMJ and the Lancet. IDEAL is funded by the Oxford NIHR Biomedical Research Centre, educational grants and meetings income (PMcC and AH). JMB is an NIHR Senior Investigator supported in part by the Medical Research Council (MRC) ConDuCT-II (Collaboration and innovation for Difficult and Complex randomized controlled Trials In Invasive procedures) Hub (MR/K025643/1) and the NIHR Bristol Biomedical Research Centre at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol. The views expressed are those of the author and not necessarily those of the UK National Health Service, National Institute for Health Research, or Department of Health.

Running Head: IDEAL Framework & Recommendations Update

Mini-Abstract: The IDEAL Recommendations provide a logical stepwise pathway for evaluating surgical operations, therapeutic devices and other complex interventions through five definable life-cycle stages (Idea, Development, Exploration, Assessment and Long term study). This new paradigm for research has matured since its launch in 2009, and this comprehensive update clarifies its use.

Structured Abstract:

Objective: To update, clarify and extend IDEAL concepts and recommendations.

Background: New surgical procedures, devices and other complex interventions need robust evaluation for safety, efficacy and effectiveness. Unlike new medicines, there is no internationally agreed evaluation pathway for generating and analysing data throughout the life cycle of surgical innovations. The IDEAL Framework and Recommendations were designed to provide this pathway and they have been used increasingly since their introduction in 2009. Based on a Delphi survey, expert workshop and major discussions during IDEAL conferences held in Oxford (2016) and New York (2017), this article updates and extends the IDEAL Recommendations, identifies areas for future research, and discusses the ethical problems faced by investigators at each IDEAL stage.

Methods: The IDEAL Framework describes five stages of evolution for new surgical therapeutic interventions – Idea, Development, Exploration, Assessment and Long Term Study. This comprehensive update proposes several modifications. First, a "Pre-IDEAL" stage describing pre-clinical studies has been added. Second we discuss potential adaptations to expand the scope of IDEAL (originally designed for surgical procedures) to accommodate therapeutic devices, through an IDEAL-D variant. Third, we explicitly recognise the value of comprehensive data collection through registries at all stages in the Framework and fourth, we examine the ethical issues that arise at each stage of IDEAL and underpin the recommendations. The Recommendations for each stage are reviewed, clarified and additional detail added.

Conclusions: The intention of this article is to widen the practical use of IDEAL by clarifying the rationale for and practical details of the Recommendations. Additional research based on the experience of implementing these Recommendations is needed to further improve them.

INTRODUCTION

Surgery is a complex intervention with properties which make it more difficult to evaluate rigorously than drug treatments. Evaluation methods that fail to address this complexity have led to much controversy and wasted effort through poor study design, inadequate reporting and failure to reach agreement on standards for high quality trials. The resulting adverse consequences have included widespread adoption of new techniques or devices which later proved to be harmful and of refusal by healthcare funders to reimburse for innovations with an inadequate evidence base, as well as large scale failures of surgical research to compete successfully for public funding.

The IDEAL Framework and Recommendations represent a new paradigm for the evaluation of surgical operations, invasive medical devices and other complex therapeutic interventions. IDEAL began with a series of meetings at Balliol College, Oxford during 2007-9 to discuss the specific challenges of evaluating surgical innovation, recognising, analysing and proposing solutions for the challenges which arise as new procedures move from proof of concept towards a randomised controlled trial. These discussions resulted in the publication of a five-stage Framework describing the natural stages of surgical innovation (Idea, Development, Exploration, Assessment and Long Term Study), together with recommendations for a rigorous stepwise surgical research pathway, and suggestions for appropriate study methodology for the questions which characterise each stage. (1, 2, 3) This was subsequently followed up by publications offering methodological guidance. (4, 5, 6, 7)

Each stage is defined by a key research question:

- IDEA - STAGE 1: What is the new treatment concept and why is it needed?
- DEVELOPMENT - STAGE 2a: Has the new intervention reached a state of stability sufficient to allow replication by others?
- EXPLORATION - STAGE 2b: Have the questions that might compromise the chance of conducting a successful RCT been addressed?
- ASSESSMENT - STAGE 3: How does the new intervention compare with current practice?
- LONG-TERM STUDY- STAGE 4: Are there any long-term or rare adverse effects or changes in indications or delivery quality over time?

Various users and funders of research have acknowledged the utility of IDEAL. (Table 1) Surgical researchers are also increasingly citing and using the study designs and reporting formats recommended by IDEAL (552 papers cited key IDEAL papers [Web of Science searched 19 October 2017]). Despite these signs of interest, international use of IDEAL remains limited. It is clear that researchers need more detailed guidance about how to implement the recommendations (8, 9, 10) which were initially outlined in a generalised way.

We revised and updated the Framework and Recommendations using a three-step modified Delphi process comprising a two-round online questionnaire survey between December 2015 – April 2016 followed by an expert consensus meeting at the IDEAL Conference at St Catherine's College, Oxford on 8th April 2016 (Appendix 1- link 1). The first round required participants to address 22 questions aimed at clarifying the IDEAL Recommendations, and the second round was focused on 11 areas where consensus was not clear in round 1, with remaining areas of controversy dealt with at the Oxford meeting. The international group of

experts comprised 56 participants (47 attended the final meeting) including surgeons, methodologists, clinical trialists, ethicists, journal editors, HTA professionals, purchasers of healthcare and device industry representatives. Further details of the process and findings are provided (Appendix 2).

SCOPE

Discussions about updating IDEAL covered four main new areas. First, the need for a pre-IDEAL, pre-clinical Stage. Second, the application of IDEAL to other complex health interventions. Third, a reconsideration of the place of registries in the IDEAL pathway, and finally an explicit examination of the ethical issues that arise at each stage of IDEAL and underpin the recommendations

Publications reporting pre-clinical studies prior to 'first in human' studies as "IDEAL Stage 0" have already appeared (11). However, due to the challenge of drafting recommendations for conduct and reporting of such a potentially broad and varied set of different study types, we recommend the use of the term 'Pre-IDEAL' for these studies rather than adding a formal "Stage 0" to the Framework.

IDEAL's potential for application to therapeutic medical device evaluation was quickly recognised (Appendix 1-link 2), and the necessary modifications have already been summarized in a description of an IDEAL-D (Devices) variant (12) supported by an international Delphi process (13). Other specific variants have also been proposed, such as the R-IDEAL tool from the MRI-Linear Accelerator Consortium (14) and IDEAL-Physio (15). However, the consensus group agreed not to broaden the scope of the original IDEAL Framework, but to welcome and monitor the work of groups investigating potential wider uses.

With the help of academic ethicists, the scope of the IDEAL Framework and Recommendations has been expanded to include explicit ethical guidance at each stage. Ethical issues arising in each of the IDEAL stages are examined in detail in an accompanying paper. (16)

UPDATE OF IDEAL: STAGE BY STAGE

The 2009 Framework comprised five stages: Idea (1), Development (2a), Exploration (2b), Assessment (3) and Long-term study (4). (3) The main purpose of each stage is summarised above.

We updated the original Framework to improve its usability and transparency. This update includes description of each IDEAL stage using the PICO framework (patients/operators, interventions, comparators, outcomes), identifying appropriate reporting guidelines and identifying when to progress to the next stage. We also underline the key research question for each IDEAL stage and highlight areas for future research. The revised and original versions are shown side by side to clarify the changes made in this update. (Table 2)

Pre-IDEAL studies

Purpose and description

Pre-IDEAL research is essential prior to first in human trials of an innovation.

Patients and operators: Although patients do not receive treatments in Pre-IDEAL studies, consultation is desirable to evaluate the societal need and perceived value of the proposed intervention. (17) Where the investigators are engineers or scientists with no clinical background, clinicians should also be consulted.

Intervention and comparator: Pre-clinical studies usually do not involve a comparator. However, testing (e.g. surface wear, battery drain), simulation or modelling studies may allow comparison with the current standard.

Outcomes: Depending on the type of pre-clinical study, the aims and methods may vary widely. Typically, key studies focus on demonstrating that the intervention brings about the intended physical changes. Following this, studies to estimate reliability and safety, qualitative studies with stakeholders to determine potential demand and acceptability, and modelling studies to predict overall impact on health care costs/efficiency may be desirable. This is especially relevant for new devices for purpose of coverage and reimbursement.

Appropriate study designs

Pre-clinical studies include material testing, simulator, cadaver, animal, modelling and cost-effectiveness studies. (18, 19, 20)

Controlled experiments will often be appropriate and feasible for laboratory or bench studies. Because of the wide range of study objectives and methods, comprehensive recommendations about study design are impractical. We therefore recommend following the best available authoritative guidance on particular study types. (Appendix 1- link 3)

Recommended reporting guidelines for protocols and full reports

Explicit reporting of Pre-IDEAL findings is both practically and ethically necessary to support optimal development of Stage 1 studies. However protection of professional or commercial advantage may inhibit full reporting. Transparency around study design, subjects and outcomes is essential, but recommendations for reporting of sensitive details of mechanism or technique at this stage may prove impractical. We recommend using the numerous high

quality, subject-specific guidelines that exist for reporting different kinds of preclinical studies and protocols which can be found on the EQUATOR and Meridian Network databases (Appendix 1- link 4), or at the National Institutes of Health (Appendix 1- link 5)

Ethical aspects

Animal studies should abide by recognised ethical guidelines, including the imperative to reduce, refine, and replace animal use whenever possible (Appendix 1- link 6). Animal models must be valid and the results applicable to humans. Pre-IDEAL simulation and bench testing can raise ethical considerations of justice where assumptions built into testing (e.g. about typical body shapes) may limit the applicability of the results, or expose subgroups to greater risk of harm (16).

Identifying stage endpoints

In principle, the pre-IDEAL stage should be complete before the first in human procedure is done. Any feasible studies which can be expected to identify avoidable and/or predictable risks of failure or harm to the first patient should be performed.

"Idea" Stage 1 - 'First in human' use

Purpose and description

IDEAL Stage 1 describes the first use of a new procedure or device in a patient, either as part of a planned approach or in an unplanned emergency situation. (21) Where planned, the patient or patients are usually highly selected.

Patients and operators: Patient selection criteria should be explained in detail: if any proposed patients refuse the procedure, or are excluded, this should be explained with reasons. The operators are usually an individual or small team in a single centre.

Intervention*:* A full technical description of the new procedure or technology, in sufficient detail to allow an equivalently skilled operator to reproduce it, should be provided. There is normally no explicit comparator.

***Outcomes**:* Commonly include technical success, safety and short-term physiological and clinical measures. Transparency demands that all adverse events must be fully disclosed. Whenever feasible, in all stages of IDEAL, outcomes should be described using widely accepted standardised definitions and terms, preferably selected from a core outcomes set (Appendix 1- link 7).

Appropriate study designs

Stage 1 involves a single case or a few cases. If enough suitable patients are available progression may occur within a short time. It is recommended that reports explain the need for the new treatment concept and why it might be better than currently available treatment. Video recording and sharing is highly recommended and can be part of on-line publication.

Recommended reporting guidelines for protocols and full reports

We recommend Stage 1 studies be reported as a single case report. The SCARE Guidelines provide a useful standard reporting structure. (22) Menon et al (11) demonstrate a well reported IDEAL model of surgical innovation in the development of robotic kidney transplantation (RKT) with regional hypothermia.

Ethical aspects

Significant ethical issues arising in Stage 1 include: minimising patient harm; ensuring adequately informed patient consent; optimising communication about the innovation in the

surgical team; identifying and managing conflicts of interest; and obligatory full and accurate reporting of outcomes to prevent avoidable harms to future patients. Reflection is important to confirm the potential for the new innovation to solve a real clinical problem. Planned innovations should be conducted in compliance with local hospital and research ethics frameworks such as IRBs. Local organisations should be responsible for ensuring that review supports rather than discourages innovation (16).

Identifying stage endpoints

Once Stage 1 is completed, deciding whether to progress to Stage 2a depends on proof of concept, technical achievement, apparent safety and potential efficacy.

Areas for future work & research

The IDEAL Collaboration strongly recommends registration of all first-in-human procedures. There is an ethical obligation on investigators to make their research available to others contemplating similar interventions. Registration of unsuccessful interventions is critical to prevent repetition of harmful errors. However, investigators may be discouraged if transparency exposes them to legal and other challenges. Therefore regulatory and governance frameworks need to evolve to accommodate these tensions. Public interest protection rules similar to those shielding accident investigations in the transport industries could achieve this. The recently enacted UK “Access to Medical Treatments (Innovations) Bill” allows the creation of a “medical innovations register” which, with appropriate protection could represent an IDEAL Stage 1 register (Appendix 1- link 8)

Incentives may be necessary to ensure that innovations with adverse outcomes are registered. Potential mechanisms to achieve this include making reporting a professional or legal obligation or requiring registration as a pre-requisite to publishing first-in-human case reports.

"Development" (Stage 2a) – Towards stabilisation of the technique

Purpose and description

In IDEAL Stage 2a, procedures are typically undergoing iterative modification towards a final, stable version.

Patients and operators*:* Usually only a few tens of patients and a few operators are involved, within a single institution for technique-based innovations. For new devices the number of operators and centres may be larger.

Intervention*:* A technical description of the initial version of the intervention is needed and an account of when, why and how modifications to either technique or selection criteria were made (23). No comparator is involved this stage. For example, Diez del Val et al (23) report the development of robotically assisted oesophagectomy by a two-surgeon team from the first robotic case onwards, presenting the cases sequentially and showing key outcomes (blood loss, robotic operating time, lymph node yield, length of stay and complications) for each patient. The reasons for rejecting cases for robotic surgery are explained, and all changes to technique or indication are highlighted, showing when they occurred and explaining why they were instituted.

Outcomes*:* Relevant outcomes include short-term clinical, technical and safety outcomes.

Appropriate study designs

Normally a small single centre prospective cohort. A typology which deconstructs interventions into their component parts may help with precise definition of procedures, and clarify description of which parts of the procedure change as it is modified and updated. (24)

Recommended reporting guidelines for protocols and full reports

IDEAL advocates that Stage 2a studies should report:

- Patient inclusion and exclusion criteria, how many patients were assessed for treatment, which candidates were excluded, and the reasons for their exclusion. Patients considered for, but ultimately not offered the intervention may also be described, together with their outcomes.
- Consecutive presentation of case-specific outcomes for all cases treated. May involve use of statistical monitoring techniques eg, CUSUM (25)
- When and why modifications to the technique or indications occurred, to reduce the risk of avoidable harm by preventing repetition of unsuccessful modifications. A graphic representation showing when technical modifications occurred during the series is useful.

Ethical aspects

Stage 2a studies are planned, and are therefore subject to appropriate institutional research ethics review. Issues specific to Stage 2a are minimising harm during development of the new procedure, informed consent and transparent publication of outcomes. Consent should include information about Stage 1 outcomes, acknowledging that risks cannot be reliably predicted or quantified at this stage. Collecting and reporting outcomes in accessible ways minimises avoidable harm to future patients and guides the fair and equitable uptake of innovation. In this and subsequent stages an independent oversight group to monitor outcomes during the study is highly desirable particularly for high risk procedures. (16)

Identifying stage endpoints

By the end of this stage, the procedure and indications should be stable enough to permit multi-centre replication during Stage 2b. The phase of intentional iterative change should be complete, although limited further modifications may still occur in Stage 2b.

Areas for future work & research

Further work is needed to guide development of protocols to plan Stage 2a studies, and to develop methods to evaluate when Stage 2a endpoints have been reached. (Appendix 1- link 9)

"Exploration" (Stage 2b) – Bridge to a pivotal trial

Purpose and description

In the 'Exploration' phase a greater experience of the new intervention is gained in a wider group of surgeons and patients to collect information which will determine whether and how to progress to a definitive comparison against current best treatment. The data should be used to promote consensus on the design and conduct of a future RCT, and to improve its feasibility by minimising potential barriers to delivery. (26, 27) Chen at al (27) conducted a 20-centre non-randomised prospective cohort study of treatments for uterine fibroids, in which patients chose to receive hysterectomy, myomectomy or HIFU therapy. The very large differences reported in complications and short term recovery ruled these out as primary outcomes for a future RCT, as equipoise appeared impossible. Using QoL measures at 6 months was also infeasible, as the very similar results found implied an enormous trial population requirement. There remained, however, an answerable question about recurrence after treatment.

Patients and operators*:* The patient group is less selective than for 2a studies, involving more patients in more centres; case numbers will commonly be over 100. More surgeons are learning and undertaking the new technique, and may disagree on inclusion criteria.

Intervention and comparator: The intervention will now be relatively well defined, but minor differences in technique are still common and can be explored through pre-planned subgroup analyses. A comparator intervention may or may not exist.

Outcomes*:* These include safety, a precise effect estimate of short term clinical outcome useful for trial size calculations, identification of possible subgroup outcome differences, assessment of surgical quality and learning curves, qualitative evaluation of trial feasibility and definition of core outcomes measures for the future trial. Validated core outcome measures should be used if available (Appendix 1- link 7).

Appropriate study designs

IDEAL Exploration studies are typically collaborative multi-centre prospective cohort studies and feasibility RCTs designed to enhance investigator consensus on key issues. The goal is to resolve the problems which most commonly prevent surgical RCTs: in doing so, however, Exploration studies may sometimes make it clear that an RCT is inappropriate or infeasible. In such cases Exploration studies may facilitate "next best" approaches such as propensity scoring of the observational data collected, or point to the need for a registry approach. We therefore recommend an early pre-planned consensus meeting to evaluate short term results and agree on whether an RCT can be done, and where feasible, to develop its design. Statistical analysis of learning curves may be useful. (28)

Recommended reporting guidelines for protocols and full reports

Guidance to assist design and reporting of Exploration Stage studies includes the recent CONSORT extension for pilot and feasibility RCTs (29) and the updated STROBE guideline for

observational studies (30). IDEAL's development of a reporting checklist for this stage is underway.

Ethical aspects

Stage 2b studies require institutional research ethics review. Potential harms from the learning curve should be minimised by training and mentoring. Informed consent should be based upon information from Stage 2a. For reasons of justice and equity patient selection should reflect the population in which the innovation can be expected to be effective; and the data set should include outcomes of importance to patients.(16) An independent study oversight group to monitor outcomes is recommended.

Identifying Stage Endpoints

By the end of Stage 2b, all the requirements needed to progress to a pivotal RCT should be complete. These fall into two groups;

a) Endpoints that demonstrate that the technique can be widely adopted by surgeons:

1. Agreement on the definition of the intervention (and acceptable variants) for the purposes of an RCT;
2. Agreement on quality standards for delivery of the intervention; and
3. Assessment of learning curves to allow decisions on admission of clinicians into an RCT and how to evaluate their outcomes to avoid bias in a comparison versus standard treatment.

b) Endpoints that demonstrate that progression to a formal randomised controlled trial is feasible:

1. Confirmation of the appropriate target patient population. Disagreements on details of patient selection are a common cause of failure to achieve consensus for an RCT;
2. Confirmation of the appropriate comparator treatment;

3. Confirmation of the appropriate primary endpoint for outcome assessment in the RCT. The DELTA guidance on specifying the target difference for a RCT is recommended (Table 2- link 10).
4. Evidence of consensus amongst surgeons and patients that they are willing to accept randomisation between the proposed treatment options.

We recommend the use of nested qualitative studies to explore the attitudes and values of participants.

Areas for future research

Successful Stage 2b Exploration studies require the collaboration of multiple researchers and institutions. These type of studies may therefore be especially suitable to cooperative groups such as the UK Surgical Trainee collaborative. (31) Methods for establishing that the intervention is sufficiently defined and stable for RCT evaluation need further work. Empirical evaluation of the impact of 2b studies on the probability of developing a successful RCT is needed. If large scale procedure registry systems or permanent audits exist, methods for nesting 2b studies within the system are needed.

"Assessment" (Stage 3) – Pivotal study/RCT

Purpose and description

In the 'Assessment' Stage a pivotal comparative evaluation occurs, usually against current standard treatment. This should take place soon after the new procedure is stable, and prior to its extensive use, to avoid loss of equipoise among clinicians and patients.

Patients and operators*:* This will typically involve a substantial number of patients and operators at multiple centres. Clear patient and operator selection criteria should be based on data gathered in previous Stages. (26, 27)

Intervention and comparator*:* The new intervention is compared with a clinically relevant comparator. Typically this will be "best usual care" but on occasions, a placebo or sham surgery control might be appropriate. (32)

Outcomes: Outcomes reflecting the values of patients and operators should be identified during the Stage 2b collaborative study and consideration given to developing a new core outcome set (COS) if one does not already exist (33). IDEAL 2b studies provide opportunities to identify and pilot test the primary outcome for a definitive trial.

Appropriate study designs

Where feasible a multi-surgeon, multi-centre randomised trial (6, 34) should be performed. Variants, including cluster-randomised or expertise-based RCTs (35, 36) or stepped wedge designs may be appropriate. Where circumstances preclude randomisation, acceptable alternatives include controlled interrupted time series or observational designs using efficient post-hoc techniques (e.g. propensity scoring) to minimise known sources of bias. Nesting medical device based trials within national population/disease based registries is recommended to facilitate larger scale pragmatic trials (12)

Recommended reporting guidelines for protocols and full reports

Several robust guidelines for reporting surgical RCTs are recommended including the updated 2010 CONSORT Statement (37), its extension for NPT (non-pharmacological treatments) (38) and a template for intervention description and replication (TIDieR Statement). (39)

The SPIRIT statement provides excellent guidance for writing protocols for RCTs. (40) These and other more specific reporting guidance for surgery can be found at the EQUATOR website (Appendix 1- link 4).

Ethical aspects

Stage 3 trials require institutional research ethics review and independent oversight. Relevant ethical issues concern the generation of valid data; fair inclusion and exclusion criteria; access and equity in research participation; use of outcomes relevant to patients; measures to minimise surgeon bias; efforts to minimise patient harm due to the learning curve; fair treatment of the data and prompt publication of all relevant results. Placebo or sham surgery controls may be justifiable where they offer the best chance to resolve uncertainty, do not involve unacceptable risks to patients and are acceptable to key stakeholders (16). All RCTs should be registered on an appropriate international register, for example https://clinicaltrials.gov/

Identifying endpoints of stage

The two main endpoints of Stage 3 are a) valid evidence on the intervention's relative effectiveness; and b) identification of aspects which require long-term monitoring (typically late and rare outcomes).

Areas for future work and research

Many current initiatives will contribute to improvements in surgical trial design, conduct and reporting. These include Trial Forge (Appendix 1- link 11) to improve trial efficiency (41), and PRECIS-2, a tool to improve trial design (42). Research to establish better methods for early phase IDEAL studies is also being undertaken at the Bristol Centre for Surgical Research in the UK (Appendix 1- link 12).

"Long-term study" (Stage 4) – Identifying rare and later outcomes

Purpose and description

The IDEAL Framework proposes registries for data collection in Stage 4 (Long-term study). Their strength lies in recognising late or uncommon safety outcomes and identifying changes in the use of procedures - so called "indication creep", and trends in outcomes which may reveal variations in the quality of surgery. Registries allow evaluation of "real world" outcomes and, very importantly, can allow ongoing feedback to clinicians and manufacturer (43). This update of IDEAL introduces the recommendation for the use of registries at a much earlier stage in the framework.

- **Registries for device surveillance and life cycle assessments.**

Registries or 'registry like' systems can detect long term or rare safety problems with devices such as the failures of metal-on-metal hip implants (44, 45), problems with vascular closure devices (46), and surgical meshes. (47, 48) Comprehensive coverage requires commitment from users and sustainable partnership with stakeholders. Registries should have independent supervision, use standard universally applicable definitions of outcomes and relevant confounders, and cover equivalent devices from all relevant manufacturers. However, manufacturers are often involved in registry funding, and are naturally wary of exposure of their data to competitors. IDEAL promotes comprehensive, high quality registry/big data development whilst recognising the conflicts and the need to address stakeholder concerns (49).

- **Introducing registries at an early stage.**

Whereas IDEAL originally envisaged registries as confined to Stage 4, registry-type data collection can usefully begin from IDEAL Stage 1. Device manufacturers typically develop competitor products more or less simultaneously, and evaluate them through exclusive arrangements with a relatively small pool of clinicians. The evident scientific and public interest in pooling of these datasets is in tension with competition between manufacturers. Starting a registry at an early stage can allow data harmonization and pooling of resources and data if agreements can be reached (50 & Appendix 1- link 13).

- **Developing trials within registries**

IDEAL Stage 3 RCTs can be 'nested' within registries or cohort studies (51, 52 & Appendix 1 - link 14), potentially enhancing trial recruitment, but informed consent processes require careful consideration (see below).

Patients & operators: The patient population is dictated by the stage at which the registry is introduced. Registries introduced at an early stage are mostly procedure or device driven with clinicians entering all patients treated with the innovation. Disease-driven registries, consisting of all known patients with a specific medical condition, are preferable scientifically but are not easy to implement and rarely sustainable. A feasible compromise approach which facilitates nested trials and other secondary data usage is the "all comers" model which collects data on available treatments in a specific condition or area of practice (53). Advances in electronic records systems' capacity to code accurately for both diagnosis and treatment will rapidly improve our ability to develop such registries in the future.

Intervention & Comparator: Registries may be procedure/device-driven or disease-driven. Which design is optimal depends on the objectives of the registry and should be clinician- and

patient driven. Registries can be utilised within specific trials where appropriate to increase their efficiency, for example the UK REBOA (Resuscitative Endovascular Balloon Occlusion of the Aorta) RCT will be using data collected via the UK trauma registry (Appendix 1 –link 15).

Outcomes: Registries allow for analysis of long term outcomes that may not be captured within the lifetime of an RCT. Early stage registries may be particularly useful in allowing co-operating innovators to pool learning on procedural modifications in order to arrive at the optimal technique more rapidly.

Appropriate study designs

Key design issues for registries centre on the dataset and on fostering engagement. Datasets should be as small and cheap to collect as possible, whilst reliably capturing patient and device/procedure identity, diagnosis, and the key influences on outcome. Standardised terminology should be used for all data items. Careful design of contributor recognition, data entry systems and feedback methods should maximise incentives for and minimise barriers to full data submission.

Recommended reporting guidelines for protocols and full reports

Clear plans for analysis and reporting of data should be developed at the outset, specifying time intervals, numbers of patients registered, or pre-specified safety signals which will trigger analysis. The form, authorship, and means of distribution of reports needs to be agreed in advance, giving all contributors appropriate recognition, notice of reports and access to the data.

Ethical aspects

In Stage 4, interventions are no longer within a research ethics framework, but part of routine clinical practice. However, as registry data is collected this raises issues about patient consent

for future use of data including the development of pragmatic nested RCTs. Integrated verbal consent can be an alternative to written consent in some contexts. (54) Issues of access and equity may arise in Stage 4 if the innovation is more expensive than alternatives, or concentrated in specialist centres. Conflicts of interest created by financial or reputational rewards and/or aggressive marketing may bias practice towards the innovation even for indications where there is little evidence. (16)

Areas for further research

Research is needed on the specific value of different features of registries, and on the factors which facilitate or prevent their introduction. Further analysis of real world application of trials within registries is required to inform practical guidance and future IDEAL recommendations.

Significant variation exists in the outcome measures used in international patient registries within the same clinical area, reducing their utility by making comparison, linkage, and aggregation of data more difficult. International collaborations are emerging to advance these issues (Appendix 1 links 16 and 17).

LIMITATIONS

The recommendations in this article arose from a relatively small selected group of experts. Different perspectives may have emerged with a wider range of participants. The use and uptake of IDEAL is continuously evolving, and this paper presents an update at a specific point in time. It will undoubtedly be refined further in the future.

CONCLUSIONS

The idea of an integrated stepwise evaluation pathway for complex interventions is generally accepted as desirable. The IDEAL approach meets this need in a logical and widely accepted

manner. However, IDEAL requires on-going review and updates. Experience in the planned use of the IDEAL Framework and Recommendations is currently limited. Empirical analysis of the outcomes and impact of using IDEAL will be an important driver for future incremental and evidence-based modifications.

References

1. Barkun JS, Aronson JK, Feldman LS et al.; Balliol Collaboration. Evaluation and stages of surgical innovations. The Lancet. 2009 Sep 26; 374(9695):1089-1096.

2. Ergina PL, Cook JA, Blazeby JM et al.; Balliol Collaboration. Challenges in evaluating surgical innovation. The Lancet. 2009 Sep 26; 374(9695):1097-1104.

3. McCulloch, P, Altman, DG, Campbell, W et al.; Balliol Collaboration. No surgical innovation without evaluation: the IDEAL recommendations. The Lancet. 2009 Sep 26; 374(9695):1105-1112.

4. McCulloch P, Cook JA, Altman DG et al.; IDEAL Group. IDEAL Framework for surgical innovation 1: the idea and development stages. BMJ. 2013 Jun 18; 346:f3012.

5. Ergina PL, Barkun JS, McCulloch P,et al.; IDEAL Group. IDEAL Framework for surgical innovation 2: observational studies in the exploration and assessment stages. BMJ. 2013 Jun 18; 346:f3011

6. Cook JA, McCulloch P, Blazeby JM et al.; IDEAL Group. IDEAL Framework for surgical innovation 3: randomised controlled trials in the assessment stage and evaluations in the long term study stage. BMJ. 2013 Jun 18; 346:f2820

7. Pennell CP, Hirst AD, Campbell WB et al. Practical guide to the Idea, Development and Exploration stages of the IDEAL Framework and Recommendations. Br J Surg. 2016 Apr; 103(5):607-615.

8. Gudeloglu A, Fleischman SR, Dahm, P. Awareness of the IDEAL recommendations among the members of the Robotic-Assisted Microsurgery and Endoscopic Society (RAMSES)

Abstracts from the IDEAL Collaboration Conference 2016, 7 April 2016, Oxford, UK. Int J Surg. 2016 Dec; 36(Suppl 2):S143

9. Blazeby JM, Blencowe NS, Titcomb DR et al. Demonstration of the IDEAL recommendations for evaluating and reporting surgical innovation in minimally invasive oesophagectomy. Br J Surg. 2011 Apr; 98(4):544-551.

10. Currie A, Brigic A, Blencowe NS, et al. Systematic review of surgical innovation reporting in laparoendoscopic colonic polyp resection. Br J Surg. 2015 Jan; 102(2):e108-116.

11. Menon M, Abaza R, Sood A et al. Robotic Kidney Transplantation with Regional Hypothermia: Evolution of a Novel Procedure Utilizing the IDEAL Guidelines (IDEAL Phase 0 and 1). European Urology, 2014 65(5), pp.1001-1009.

12. Sedrakyan A, Campbell B, Merino JG, et al. IDEAL-D: a rational framework for evaluating and regulating the use of medical devices. BMJ. 2016 Jun 9; 353:i2372.

13. Pennell CP, Hirst A, Sedrakyan A, et al. Adapting the IDEAL Framework and Recommendations for medical device evaluation: A modified Delphi survey. Int J Surg. 2016 Apr; 28:141-148.

14. Verkooijen HM, Kerkmeijer LGW, Fuller CD, et al.; R-IDEAL: A Framework for Systematic Clinical Evaluation of Technical Innovations in Radiation Oncology. Front Oncol. 2017 Apr; 7:59

15. 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. Phys Ther. 2018 Feb 1;98(2):108-121

16. Rogers et al – accompanying ethics and IDEAL paper

17. Pound P, Blaug R. Transparency and public involvement in animal research. Altern Lab Anim. 2016 May; 44(2):167-173

18. Zendejas B, Brydges R, Hamstra SJ, et al. State of the evidence on simulation-based training for laparoscopic surgery: a systematic review. Ann Surg. 2013 Apr; 257(4):586-593

19. van der Worp HB, Howells DW, Sena ES, et al. Can animal models of disease reliably inform human studies? PLoS Med. 2010 Mar 30; 7(3):e1000245.

20. McAteer H, Cosh E, Freeman G, et al. Cost-effectiveness analysis at the development phase of a potential health technology: examples based on tissue engineering of bladder and urethra. J Tissue Eng Regen Med. 2007 Sep-Oct; 1(5):343-349.

21. Stannard, A, Eliason J, Rasmussen, T. Resuscitative Endovascular Balloon Occlusion of the Aorta (REBOA) as an Adjunct for Hemorrhagic Shock. J Trauma. 2011 Dec; 71(6):1869-1872

22. Agha RA, Fowler AJ, Saeta A,et al.; SCARE Group. The SCARE Statement: Consensus-based surgical case report guidelines. Int J Surg. 2016 Oct; 34:180-186.

23. Diez del Val I, Loureiro C, McCulloch P. The IDEAL prospective development study format for reporting surgical innovations. An illustrative case study of robotic oesophagectomy. Int J Surg. 2015 Jul;19: 104-111.

24. Blencowe NS, Mills N, Cook JA, et al. Standardizing and monitoring the delivery of surgical interventions in randomized clinical trials. Br J Surg. 2016 Sep; 103(10):1377-1384.

25. Biau DJ, Porcher R. A method for monitoring a process from an out of control to an in control state: application to the learning curve. Statistics in Medicine 2010; 29: 1900–9.

26. Penna M, Hompes R, Arnold S, et al.; TaTME Registry Collaborative. Transanal Total Mesorectal Excision: International Registry Results of the First 720 Cases. Ann Surg. 2016 Oct 4. [Epub ahead of print]

27. Chen J, Li Y, Wang Z, McCulloch P, et al. Committee of the Clinical Trial of HIFU versus Surgical Treatment for Fibroids. Evaluation of HIFU Ablation for Uterine Fibroids: an IDEAL Prospective Exploration Study. BJOG. 2017 Apr 19. doi: 10.1111/1471-0528.14689. [Epub ahead of print]

28. Cook JA, Ramsay CR, Fayers P. Statistical evaluation of learning curve effects in surgical trials. Clin Trials. 2004;1(5):421-427

29. Eldridge SM, Chan CL, Campbell MJ, et al. ; PAFS consensus group. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. BMJ. 2016 Oct 24; 355:i523930.

30. The PLOS Medicine Editors. Observational Studies: Getting Clear about Transparency. PLoS Med. 2014; 11(8): e1001711.

31. Jamjoom AA, Phan PN, Hutchinson PJ, et al. Surgical trainee research collaboratives in the UK: an observational study of research activity and publication productivity. BMJ Open. 2016 Feb 4; 6(2):e010374.

32. Wartolowska K, Judge A, Hopewell S, et al. Use of placebo controls in the evaluation of surgery: systematic review. BMJ. 2014; 348: g3253. doi:10.1136/bmj.g3253

33. Williamson PR, Altman DG, Bagley H et al. The COMET Handbook: version 1.0. Trials. 2017 Jun 20;18(Suppl 3):280. doi: 10.1186/s13063-017-1978-4

34. Blencowe NS, Cook JA, Pinkney T, et al. Delivering successful randomized controlled trials in surgery: Methods to optimize collaboration and study design. Clin Trials. 2017 Jan 1:1740774516687272. doi: 10.1177/1740774516687272. [Epub ahead of print]

35. Devereaux PJ, Bhandari M, Clarke M, et al. Need for expertise based randomised controlled trials. BMJ 2005; 330: 88

36. Cook JA, Elders A, Boachie C, et al . A systematic review of the use of an expertise-based randomised controlled trial design. Trials 2015, 16:241.)

37. Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ. 2010 Mar23; 340:c332.

38. Boutron I, Moher D, Altman DG, et al. for the CONSORT group. Methods and Processes of the CONSORT Group: Example of an Extension for Trials Assessing Nonpharmacologic Treatments. Ann Intern Med. 2008:W60-W67.

39. Hoffmann TC, Glasziou PP, Boutron I, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ. 2014 Mar 7; 348:g1687. doi: 10. 1136/bmj. g1687.

40. Chan A-W, Tetzlaff JM, Altman DG, et al. SPIRIT 2013 Statement: Defining standard protocol items for clinical trials. Ann Intern Med 2013; 158:200-207.

41. Treweek S, Altman DG, Bower P, et al. Making randomised trials more efficient: report of the first meeting to discuss the Trial Forge platform. Trials. 2015 Jun 5; 16:261

42. Loudon K, Treweek S, Sullivan F, et al. The PRECIS-2 tool: designing trials that are fit for purpose. BMJ. 2015 May 8; 350:h2147

43. Sedrakyan A, Campbell B, Graves S, et al. Surgical registries for advancing quality and device surveillance. Lancet. 2016 Oct 1; 388(10052):1358-1360.

44. Smith AJ, Dieppe P, Vernon K, et al.; National Joint Registry of England and Wales. Failure rates of stemmed metal-on-metal hip replacements: analysis of data from the National Joint Registry of England and Wales. Lancet. 2012 Mar 31; 379(9822):1199-1204.

45. Smith AJ, Dieppe P, Howard PW, et al.; National Joint Registry for England and Wales.. Failure rates of metal-on-metal hip resurfacings: analysis of data from the National Joint Registry for England and Wales. Lancet. 2012 Nov 17; 380(9855):1759-1766.

46. Resnic FS, Majithia A, Marinac-Dabic D, et al. Registry-Based Prospective, Active Surveillance of Medical-Device Safety. N Engl J Med. 2017 Feb 9; 376(6):526-535.

47. Chughtai B, Mao J, Buck J, et al. Use and risks of surgical mesh for pelvic organ prolapse surgery in women in New York state: population based cohort study. BMJ. 2015 Jun 2; 350:h2685. doi: 10.1136/bmj.h2685. Erratum in: BMJ. 2015; 350:h3060.

48. Chughtai B, Barber MD, Mao J, et al. Association Between the Amount of Vaginal Mesh Used With Mesh Erosions and Repeated Surgery After Repairing Pelvic Organ Prolapse and Stress Urinary Incontinence. JAMA Surg. 2016 Nov 30. doi: 10.1001/jamasurg.2016.4200. [Epub ahead of print]

49. Krucoff MW, Sedrakyan A, Normand SL. Bridging Unmet Medical Device Ecosystem Needs With Strategically Coordinated Registries Networks. JAMA. 2015 Oct 27; 314(16):1691-1692.

50. Campbell WB, Poole RL, Ray AF, et al. Collecting data on treatments for varicose veins: Getting the focus right. Phlebology. 2016 Jun 20. Pii: 0268355516655247. [Epub ahead of print] PubMed PMID: 27329668.

51. van der Velden JM, Verkooijen HM, Young-Afat DA, et al. The cohort multiple randomized controlled trial design: a valid and efficient alternative to pragmatic trials? Int J Epidemiol. 2016 Apr 26. pii: dyw050. [Epub ahead of print]

52. Li G, Sajobi TT, Menon BK, Korngut L, et al; 2016 Symposium on Registry-Based Randomized Controlled Trials in Calgary. Registry-based randomized controlled trials- what are the advantages, challenges, and areas for future research? J Clin Epidemiol. 2016 Dec; 80:16-24. doi:10.1016/j.jclinepi.2016.08.003.

53. Institute of Medicine. 2001. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: The National Academies Press. https://doi.org/10.17226/10027.

54. Kim SY, Miller FG. Informed consent for pragmatic trials--the integrated consent model. N Engl J Med. 2014 Feb 20; 370(8):769-772.