Bryan Luce

Affiliate Professor, School of Pharmacy, University of Washington and former CSO at PCORI

Critical Review of Explanatory vs. Pragmatic Trials from the Perspective of Regulators, Payers, Patients and Health Systems: Synopsis of June 2016 Summit Discussion

---

Draft Guidance for EXCITE International (Excellence in Clinical Innovation and Technology Evaluation) Scientific Collaboration

I. Introduction

EXCITE International (EI) convened high-level health care leaders at an inaugural multijurisdictional Summit on June 2-3 at the MaRS Discovery District, Toronto's science, technology and innovation hub. The Summit is the first comprehensive attempt to answer one of health technology's most pressing challenges: How do we accelerate the development and adoption of promising, disruptive, new technologies so they reach the patient faster and more quickly impact the effectiveness of health care delivery?

EI is attracting international interest to a new approach. Bringing together key experts from Canada, the UK, USA, and the Netherlands, the Summit examined how new health technologies could be evaluated through multinational studies using a single protocol into which the needs of each stakeholder party are included upfront in the pre-market space.

During the Summit a multijurisdictional panel discussion was held to explore the issue of trial design in the development of new innovative medical technologies. Partners represented regulators, payers, patients, health systems, industry, scientists and expert end users. Discussion centered on the fundamental question of "what elements of trial design work for whom and under what circumstances?" The attributes of Pragmatic Trials (pRCTs) and Explanatory Randomized Controlled Trials (RCTs) and the pragmatic-to-explanatory continuum were examined.

II. Background

The first RCT was published in 1948 and during the subsequent twenty years RCTs were increasingly implemented to evaluate medications. In 1967, Daniel Schwartz and Joseph Lellouch proposed that there were, in effect, two kinds of trials: "explanatory" trials (those aiming to "verify a biological hypothesis") and "pragmatic" trials (those aiming to "choose between two treatments").[1] Thorpe et al. expand on the statisticians' findings, describing explanatory trials (RCTs) as those designed to test causal hypotheses and pragmatic trials (pRCTs) as those designed to help users choose between options for care.[2] Schwartz and Lellouch proposed that most RCTs are, in fact, a mix of both elements.

Panelists affirmed that all randomized trials are situated on a pragmatic-to-explanatory continuum. PRECIS is a tool developed to help trialists make design decisions in accordance with the intended purpose of their trial, and the updated PRECIS-2 tool is comprised of nine domains, each of which can be designed on a range from very pragmatic to very explanatory using a Likert 5-point scale (Figure 1). An in-depth description of the PRECIS-2 tool and guidance on how to use it are presented in a 2015 article by Loudon et al.[3] As PRECIS-2 is focused solely on the issue of applicability, the authors advise trialists to use other tools to assess internal validity and other aspects of their trial design.

---

[1] Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. J Chronic Dis 1967;20:637-48. [Reprinted in J Clin Epidemiol. 2009;62:499-505.]

[2] Thorpe K, Zwarenstein M, Oxman A, Treweek S, Furberg C, Altman D, Tunis S, Bergel E, Harvey I, Magid D, Chalkidou K. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. J Clin Epidemiol. 2009 May;62(5):464-75.

[3] Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe K, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. BMJ 2015;350:h2147.

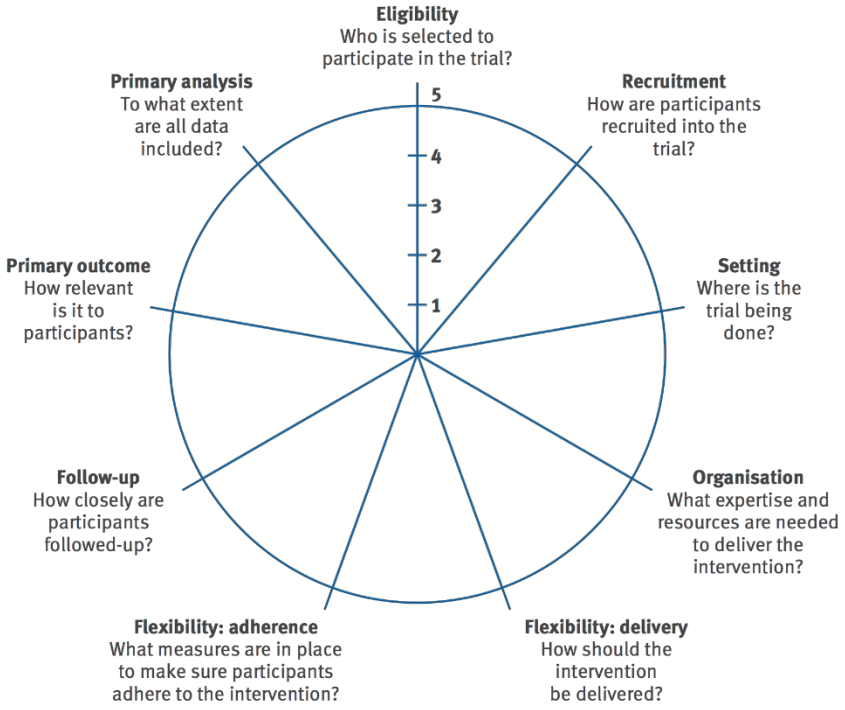

Figure 1: The PRagmatic-Explanatory Continuum Indicator Summary 2 (PRECIS-2) wheel.
(Adapted from BMJ 2015;350:h2147)

While specific trial attributes may be more pragmatic or explanatory in nature, they should not be regarded as dichotomous. Several trial attributes are presented in Table 1.

Table 1: Attributes of Pragmatic Trials (pRCTs) and Explanatory Randomized Controlled Trials (RCTs)

Pragmatic Trial (pRCT) ⟷ Explanatory Randomized Controlled Trial (RCT)	
The movement toward open data and methods of individual patient-level meta-analysis create the potential for probing heterogeneity and also for repurposing trials to new questions	
Tends toward measuring effectiveness	Tends toward measuring efficacy
Goal: choosing among alternatives in a real world situation; usually designed with this goal in mind	Goal: measuring the causal relationship between an intervention and a specified outcome
Usually designed to maximize generalizability	Usually designed to optimize internal validity; less concerned with external validity
The intervention is generally: ✓ flexibly applied in "usual care"	The intervention is generally: ✓ optimized through study design and monitoring
Inclusion criteria tend to: ✓ emulate the real world - participants, practitioners and institutions are typical of the range reflected in usual conditions of health care delivery ✓ not monitor or correct participant adherence	Inclusion criteria tend to: ✓ eliminate poorly adherent participants and/or comorbidity, both of which act to dilute the effect of the intervention
The trial is typically conducted in "usual care"	The trial is typically conducted in optimal conditions to "control" the behavior of the patient and maximize the effect of the intervention
Generally focused on evaluating outcomes of importance to patients, funders and health care practitioners (e.g., quality of life, long term survival, severe morbidity); typically not focused specifically on regulatory requirements	To the extent they are focused on regulatory requirements, outcomes of interest may not be fully oriented to outcomes of primary interest to patients, funders and health care practitioners

Major Benefit: ✓ tend to be widely generalizable (based in usual care and real world situations)	Major Benefit: ✓ lack of bias
Limitations: ✓ there is a widespread belief (but little evidence) that there is modest reduction of effect size in comparison with more explanatory trials ✓ there is a belief (but no evidence) that pragmatic trials are more biased as a tradeoff for wider real applicability	Limitations: ✓ lack of generalizability (results cannot necessarily be extrapolated to the wider population)
Trialists and policy makers alike should recognize that what constitutes "real life" and "usual care" in one jurisdiction may be different in another	By design, there is typically no attempt to match conditions found in the real world

III. Key Results and Recommendations of Summit Panel Discussion (June 2 & 3, 2016)

A. **Harmonized multijurisdictional clinical trial.** Stakeholders including regulators, payers, patients, health systems, industry, scientists and expert end users recommend the development of a common clinical trial framework encompassing pre- through postmarket medical technology (device) development as well as across national boundaries.
ACTION: The EI framework shifts medical technology assessment from a postmarket cycle to a harmonized pre-market model, using evidence to drive innovation.

B. **Sufficiency of evidence.** A key challenge in the regulatory space is the shortfall of clinical evidence generated to support the regulatory requirements of safety and efficacy. This shortfall exists throughout international jurisdictions.[4] As a result, the ability to conduct a health technology assessment is hampered, causing delays in medical technology funding and patient access.
ACTION: Increase the prevalence of clinical trials, especially Pragmatic Trials, to support safety, efficacy and/or comparative effectiveness propositions.

C. **Raising the evidence bar.** In addition to the goal of speeding patient access to effective new treatments, the EI model represents an unprecedented opportunity to improve the quality, rigor and patient-centredness of evidence. "*Collaboration should improve the quality of information available while also reducing barriers to positive change"* – Panelist.
ACTION: EI's framework and execution streamline and facilitate the evaluation process *and* deliver high-quality data, establishing the clinical trial as the gold standard for medical technology evaluation.

D. **The medical technology device as a "complex intervention."** Complex interventions are commonly described as interventions with multiple interacting parts and as such research protocols of them require special consideration. Programmatic, organizational or system interventions which, by nature, are made up of multiple components and are employed across multiple contexts are classic examples of complex interventions. Certain medical devices can also be considered as complex interventions, for instance when a device is coupled with complex surgery. In other cases, devices may simplify interventions, such as allowing minimally-invasive or robotic surgery. The UK's Medical Research Council (MRC) suggests that complex interventions present specific challenges to evaluation, requiring the following measures: a multi-phased approach which is not necessarily linear; a preference for experimental designs over observational designs even though the former are not always practicable; a focus on outcome evaluation vs. an understanding of process; allowing adaptation to local settings vs. adhering to a strict protocol; and study reports which detail the intervention for ease of replication. In addition, the MRC advises using a range of measures vs. a single primary outcome, to make best use of the data and that large sample sizes may be necessary to counteract variability in individual-level outcomes. [5]

[4] Sorenson C, Drummond M. Improving medical device regulation: the United States and Europe in perspective. Milbank Q. 2014 Mar;92(1):114-50.
[5] Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008;337:a1655.

ACTION: When a medical device is considered to be a complex intervention, regulators and payers should take notice of that fact and foster appropriate designs of clinical trials to evaluate them.

E. **Surgical safety and quality as part of protocol development.** Panelists emphasized the need to address the issues of surgical safety and quality as part of protocol development, using the following two components: A valid review for surgical quality (e.g., IDEAL (Idea, Development, Exploration, Assessment and Long-term study))***** and a human factors and usability review by experts (e.g., ECRI and Global eHealth).
ACTION: Medical device innovations should include a valid review for surgical quality and engage experts in human factors and usability to undertake an analysis as appropriate.

F. **Consideration of the device-operator interaction as regards the "learning curve" during the HTA process**. A panelist who participated in the multijurisdictional study "Methods for Health Technology Assessment of Medical Devices: a European Perspective" (MedtecHTA) noted the study found that the device-operator learning curve was often not explicitly considered in the methods guidelines of organizations conducting HTAs; the study specifies that the device-operator interaction can produce a learning curve effect which, if not taken into account, may lead to bias in estimating the size of the [device's] benefits.[6]
ACTION: When conducting a medical device HTA, include the device-operator interaction and possible learning-curve effect in the methods guidelines as appropriate. In addition, use post-marketing surveillance not only to monitor safety, but also to assess device effectiveness in usual care and provide data on the device-user learning curve and the organizational impact of the medical device.

G. **Robust outcome measures.** Currently, outcome measures are often poorly specified.
ACTION: A validated inventory of outcomes should be developed so that payers (and others) can understand how a medical technology impacts study outcomes. Clarity on statistical differences in outcomes should be specified. This is critical for functional outcomes such as depression and for studies assessing patient-reported outcomes (PROs) such as in diseases of chronic pain and neurodegeneration. Minimally-important differences should be defined and levels of magnitude of effect should be quantified.

H. **Patient as partner in outcome development.** Patients report that many issues of concern to them are not on the research agenda. Furthermore, patient expertise is believed to be greatly undervalued in the current clinical research system.
ACTION: Researchers should involve patients in the development of primary and secondary outcomes and ensure that trial outcomes are relevant and meaningful to patients.

I. **Greater patient involvement throughout the life cycle of technology evaluation**. Patients are playing an increasingly greater role in their own care, have increasingly higher expectations of the health care system and are demanding greater involvement in decisions about their care. Likewise, as users of medical technology in "real life," patients can provide researchers with valuable information and expertise. In joining the EI process, patients offer:
    - a deep experience of the qualitative aspects under study, especially for the adoption stage and in value assessment
    - the capacity to significantly and positively affect patient recruitment & retention in RCTs
    - knowledge across a product life cycle in technology adoption, uptake among diverse populations and off label uses
    - education to research teams on the experience of living with a disease
    - education to research teams on the use of a product in "real life" and over time

ACTION: Include the patient in clinical trial design throughout the life cycle of medical technology evaluation.
Table 2 describes stakeholder suggestions regarding how to gather patient-centric evidence.

***** A detailed description of the IDEAL method is presented in **APPENDIX I**.

[6] European Commission. CORDIS. Final Report Summary – MEDTECHTA (Methods for Technology Assessment of Medical Devices: a European Perspective). Project reference: 305694. From 2013-01-01 to 2015-12-31, closed project. Available at: http://cordis.europa.eu/project/rcn/105665_en.html. Accessed July 26, 2016.

Table 2: Panel suggestions for gathering patient-focused evidence	
In generating patient-focused evidence, how can we?	Solution
❖ produce patient focused outcomes	✓ overlay large trials with smaller patient studies
❖ more rapidly answer patient focused questions	✓ incorporate simple examples of electronic case report forms in trials (eCRFs)
❖ investigate why some patients do well in a trial and others do not	✓ nest small qualitative research studies in the pRCT design
❖ disseminate patient experience	✓ add a patient-focused section to published papers
❖ satisfy the market's demand for fewer outcomes, not more	✓ conduct a patient survey for desired outcomes & incorporate results in study design
❖ engage physicians, who play a key role in user adoption of new technologies	✓ engage physician associations (e.g., the UK's NICE has incorporated the Royal College under their umbrella & focuses on a full set of indicators including economic value) ✓ include patients in clinical guideline development

J. **Adverse events (AEs) in the evaluation and integration of medical technology.** Adverse events are commonly described as any untoward medical occurrence associated with the use of the intervention [medical technology device] in humans, and broadly speaking can be classified as: serious or non-serious (according to a graduated system); expected or unexpected; and study-related, possibly study-related, or not study-related. Safety is a key element of any study protocol when evaluating a new technology, so adverse event data should be collected, reviewed and reported to regulatory authorities. Additionally, adverse events may occur during the integration phase when staff in a health care system interacts with the new technology. The technology itself may not be at fault, but mistakes may occur due to a lack of testing in the care-delivery process and/or lack of user understanding and may be related, for example, to software issues, system compatibility issues, user interfaces and clinical-decision supports.

ACTION: Adverse events should be identified and reported, in the health care delivery process as well as during clinical trials. More thorough testing and user education should occur during the implementation process. National/international data repositories of AEs should be encouraged so that providers can learn from others' mistakes. Action should also be proactive; learnings from the human factors laboratory could anticipate adverse events so that mitigation occurs hand-in-hand with dissemination.

K. **Innovative models of collaboration.** There is a need for innovative models of collaboration between regulators, HTA, and reimbursement agencies. One panelist noted that efforts at regulatory reform fall short of stakeholder needs and recommended they also include:

- requiring pre-approval evidence to match the potential risk of a new device
- requiring companies to provide clinical trials for the efficacy & safety of a high-risk device
- greater centralization of regulatory approval across jurisdictions, with common risk classification rules
- better links between device identifiers (UDIs) and electronic data collection systems (EHRs)
- greater use of registries to ensure safe use of devices in the real world
- enhanced surveillance at post-market, providing real world data (e.g., effectiveness, user learning curve, organizational impact, etc.)

ACTION: In implementing its collaborative model, EI can effect positive changes in the evaluation structure and pathway.

L. **Use of the Pragmatic Clinical Trial design (pRCT).** There was general consensus that pRCTs, while not replacing RCTs, are an important element of medical technology evaluation and may be increasingly acceptable to end users such as providers and payers. One stakeholder stated the pRCT design is also increasingly able to accommodate the needs of regulators. Table 3 describes multijurisdictional stakeholder input regarding some of the methodological challenges of pRCTs and possible ways to address them.

ACTION: Outreach to regulators is necessary in order to promote implementation of pRCTs. All trials have an end date, and in the case of pRCTs, observational studies may be needed in many cases to establish the durability of the

treatment effect and monitor for adverse events. It is also important to acknowledge that real world safety and/or effectiveness issues will always remain. It takes large numbers to discern that an intervention increases the risk of an adverse event that is already common in the treated population, and due to size and duration, trials are not well suited to do so.

Table 3: Methodological challenges presented by pRCTs and stakeholder input on solutions

pRCT methodological consideration	Solution
❖ how does industry address diluting effect of heterogeneity in pRCTs?	✓ Current evidence that heterogeneity of trials (i.e., more pragmatism) dilutes effect size is mixed. If lack of association between pragmatism and effect size is confirmed, this fear may be unfounded. Repeats of this work on medical technologies need to be funded using standardized methods ✓ Create a common framework for payer expectations of trials and for communicating those expectations to industry. (For example, BCBS is piloting a new product called Evidence Street: "Where the Market Meets Evidence." Industry will have access to the evidence reviews used by the 36 BCBS Plans, representing 106 million members. Industry will also have access to a framework of evidence expectations and gaps analysis.)
❖ what happens when a pRCT ends but real world safety and/or effectiveness issues remain?	✓ Continue follow-up of all patients in RCTs beyond end of trial date by seeking patient consent at start of trials to use administrative data for very long term follow-up (5 to 15 years post RCT), in order to identify long term outcomes – safety and durability of effect and how they differ between patients (initially) randomized to different arms ✓ Conduct an observational study to determine durability of treatment effect and monitor adverse events
❖ how to address the heterogeneity and/or discrepancy of results when there are several robust RCTs with differing results?	✓ Conduct a systematic assessment of the potential sources of heterogeneity ✓ Quantify the degree of discrepancy - i.e., how far apart are the results considering confidence intervals ✓ Consider weighting evidence from more pragmatic trials more strongly ✓ Prefer evidence from trials whose design matches real world setting of patients and practitioners in which the technology will be used once it has been released to the market ✓ Examine whether treatment effect is consistent across jurisdictions ✓ Consider mini-trials in each jurisdiction and stratified analysis based on jurisdiction, then determine outcomes by jurisdiction
❖ what is the threshold for conditional coverage?	✓ Dependent on strength of statistical consistency – determine effect at the alpha level (e.g., if trial has power of 80% and alpha 0.005, then provides sufficient security)
❖ how does industry address methodological challenges such as?: • blinding • measurement of subjective endpoints	✓ If effect size is large enough – make case that blinding may not be necessary, even with "humanistic" outcomes ✓ Obtain clarity on payer specifications for

• placebo & Hawthorne effect • demand for large quantity of data – culture of "needing to keep to status quo," even though FDA is increasingly working on pRCTs and "lean" studies	"minimally important differences" ✓ Use of PROs vs. "hard" outcomes when appropriate: *"if bias is introduced, it is usually on the part of physicians, not patients"* - Panelist ✓ Consider the use of an adaptive clinical trial design

M. **Randomization.** Several panelists felt that randomization is necessary in many instances.
ACTION: A repository is needed for pRCT randomization methods and applications.

N. **Use of an adaptive design.** In medical technology evaluation, sometimes an RCT is not practicable in the long term and a pRCT does not encompass the needs of all stakeholder parties in getting the product to market *and* providing outcomes of significance to patients. Thought should be given for a continuous adaptive design that morphs from an explanatory design platform (e.g., to address regulatory internal validity concerns) to an increasingly pragmatic design platform to address real world "pragmatic" decision-maker concerns. The FDA defines an adaptive design for a medical device clinical study as "a clinical trial design that allows for prospectively planned modifications based on accumulating study data without undermining the trial's integrity and validity," and describes adaptive design styles, uses, methodologies, advantages and limitations in its document *Adaptive Designs for Medical Device Clinical Studies: DRAFT GUIDANCE.*[7] The recommendation here goes beyond the FDA draft guidance (which focuses on the design of a specific trial) by suggesting adapting from one trial design to another over time to meet different stakeholder evidence needs.
ACTION: EI pilots alternative adaptive clinical trial design schemes for medical technology evaluation.
A repository of RCT designs, statistical methodology and examples of their implementation is needed.
"EI could pilot an adaptive process whereby one starts with a traditional explanatory approach and over time adapts to a more pragmatic design at the postmarket stage. This model would include the input of stakeholder parties and describe a continuous process from concept through to post adoption." Bryan Luce

IV. Conclusion

One of the key challenges faced by emerging new medical technologies is mounting a clinical trial which is robust, provides high-quality evidence with clear and appropriate outcomes and meets the needs of regulators, payers and health care systems. EXCITE International, in its collaborative multijurisdictional approach is poised to assist in making this happen. EI is intent on profiling and incorporating collaborative and imaginative efforts by international experts (e.g., NICE, IDEAL, NOCRI, Mayo Clinic, MedValue, ECRI, UHN Health Human Factors, etc.) and shifting medical technology assessment from a postmarket cycle to a harmonized pre/post-market model, using evidence to drive innovation. Trial results would not only satisfy licensing and regulatory requirements but also showcase the value of the technology to health systems and payers and more readily lead to successful adoption.

The author thanks panel members Naomi Aronson, Peter Juni, Ian Roberts, Rod Taylor and Merrick Zwarenstein.
The author gratefully acknowledges Mary Spayne, MPH for her assistance in writing this document.

[7] Draft Guidance for Industry and Food and Drug Administration Staff - Adaptive Designs for Medical Device Clinical Studies: May 18, 2015. Available at: http://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm446729.pdf. Accessed June 23, 2016.

ACRONYMS:

AE - Adverse Event
BCBS - Blue Cross Blue Shield Association (US)
eCRF - Electronic Case Report Form
ECRI - ECRI Institute (US)
EHR - Electronic Health Record
EI - EXCITE International (Excellence in Clinical Innovation and Technology Evaluation)
FDA - U.S. Food and Drug Administration (US)
HTA - Health Technology Assessment
IDEAL - The IDEAL Collaboration (Idea, Development, Exploration, Assessment, Long-term study) (UK)
MRC – Medical Research Council (UK)
NICE - National Institute for Health and Care Excellence (UK)
NOCRI - NIHR Office for Clinical Research Infrastructure (UK)
pRCT - Pragmatic Trial
PRO – Patient Reported Outcome
RCT - Explanatory Randomized Controlled Trial
UDI - Unique Device Identifier
UHN – University Health Network (Canada)

APPENDIX I:

Peter McCulloch
Professor of Surgical Science & Practice, University of Oxford
Chair, IDEAL Collaboration

SURGICAL EVIDENCE AND SAFETY
The IDEAL method - A rational approach to the evaluation of complex treatments

At the EXCITE International Summit on June 2-3 in Toronto, Peter McCulloch, surgeon and chair of UK's IDEAL Collaboration, presented the IDEAL Framework and Recommendations and discussed how they provide the basis for an integrated evaluation pathway from first-in-human use through to maturity for new therapeutic devices.

I. Background

The Balliol Colloquium at Balliol College, Oxford University, was comprised of a group of clinicians and methodologists who met in three conferences between 2007 and 2009 to examine the topic of Surgical Innovation and Evaluation. Initiated by Dr. Jonathan Meakins, professor of surgery at Oxford and McGill Universities, the colloquium arose in response to a general perception, shared by major international journals, that surgery lacked a satisfactory basis in evidence, with surgeons lagging behind other clinicians in their employment of randomized controlled trials (RCTs).([8] [9])

In a climate pessimistic that treatments were based in science, an April 1995 study at a district hospital in Oxford set out to evaluate the degree to which treatments in general medicine were evidence based.([10]) Researchers examined treatments delivered to 109 patients for whom diagnoses had been reached over the month and found that 82% of treatments were in fact evidence based. Similarly, in 1997 Howes et al. sought to evaluate the degree to which surgical treatments were evidence based in an audit of 100 surgical inpatients in a general surgical/vascular unit at an urban teaching hospital.([11]) Howes et al. found that while 95% of surgical treatments were evidence based, the proportion supported by RCT evidence was much smaller than that found in general medicine and recommended improvements in the ways in which evidence was classified and described in therapeutic studies.

The Balliol Colloquium resulted in the publishing in 2009 of three core articles, proposing a practical, evidence-based approach for improving clinical research in surgery.([12] [13] [14]) The first article presented a model to describe the development of surgical innovation, the second described the challenges inherent in evaluating surgical intervention and the third introduced the IDEAL Framework and Recommendations, identifying five stages of surgical innovation (Idea, Development, Exploration, Assessment and Long-term study) and linking them systematically with evaluation.

Collaborators posited that while the challenges and practical considerations associated with research on surgical interventions apply also to other medical settings, research on surgery is complicated by the fact that these factors often coincide, making it all the more difficult to conduct RCTs. Some of these challenges include surgeon-related factors (knowledge, training, experience, skills, between-surgeon variability, learning curve of a surgical intervention), outcome evaluation factors (non-standardised terminology to define surgical outcomes, patient-related outcomes vs. surgeon-selected outcomes) and a hierarchical master-student model which can hinder the emergence of new models. Ergina et al. concluded that surgery did not lack evaluation; rather, it was in need of a standardised and systematic set of guidelines for generating and reporting high-quality evidence on which surgical practice could be based.(13)

The IDEAL Framework describes the stages through which interventional therapy innovation normally passes and the characteristics of each stage. The IDEAL Recommendations endorse a number of suggestions for specific study designs and reporting standards which are recommended at different stages of the Framework. These suggestions are rooted on a

[8] Pollock A. The rise and fall of the random controlled trial in surgery. Theoretical Surgery. 1989;4:163–170.
[9] Pollock A. Surgical evaluation at the crossroads. Brit. J. Surg. 1993;80:964-966.
[10] Ellis J, Mulligan I, Rowe J. Sackett D. Inpatient general medicine is evidence based. Lancet 1995;364:407-410.
[11] Howes N, Chagla L, Thorpe M, McCulloch P. Surgical practice is evidence based. Brit. J. Surg. 1997;84:1220-1223.
[12] Barkun J, Aronson J, Feldman L, et al. Evaluation and stages of surgical innovations. Lancet 2009;374:1089-96.
[13] Ergina P, Cook J, Blazeby J, et al. Challenges in evaluating surgical innovation. Lancet 2009; 374:1097-104.
[14] McCulloch P, Altman D, Campbell W, et al. No surgical innovation without evaluation: the IDEAL recommendations. Lancet 2009;374:1105-12.

series of general principles for design and reporting associated with the different questions to be addressed and the challenges faced at each stage of the process.

II. The IDEAL model and the EXCITE International process

The IDEAL Framework & Recommendations provide the basis for an integrated evaluation pathway from first-in-human use through to maturity for new therapeutic devices. In this way, the Framework and Recommendations also provide an excellent template for the design of EXCITE International studies through the Scientific Collaboration.

Using the IDEAL Framework, it is possible to assess the stage of evolution of a new device and classify it on the basis of available literature; a practical guide to help investigators in their assessment has been developed.([15]) Depending on the stage of the new device, an appropriate IDEAL-format study design can then be recommended for use by the Collaboration.

For devices where the human experience is currently minimal, it may be necessary to recommend that the evidence is developed through the IDEAL stages as far as an RCT - that is to say, using a Development study first, then an Exploration study and finally an RCT if appropriate.

IDEAL Recommendations provide the flexibility to customise EXCITE International's approach depending on the situation. For example, in the Exploration stage, a study on clinicians' learning curves for a new device may or may not be appropriate; also in the Exploration stage, studies can be designed to compare a new device with a comparator if necessary, etc.

IDEAL, however, does not address health economic questions or public utility prioritization. It is recommended that, wherever relevant, these issues should be addressed using the approach developed by Professor Rovers et al. at Radboud University prior to commencing an IDEAL-format study.

III. The IDEAL Framework and Recommendations

In considering the IDEAL model, McCulloch et al. in 2009 proposed a set of guiding principles (Table 1) as follows:(14)

Table 1: IDEAL guiding principles

- ✓ Surgery and other invasive therapies are complex interventions, the assessment of which is challenged by factors that depend on operator, team, and setting, such as learning curves, quality variations, and perception of equipoise.
- ✓ We propose recommendations for the assessment of surgery based on a five-stage description of the surgical development process.
- ✓ We also encourage the widespread use of prospective databases and registries. Reports of new techniques should be registered as a professional duty, anonymously if necessary when outcomes are adverse.
- ✓ Case series studies should be replaced by prospective development studies for early technical modifications and by prospective exploration studies (collaborative cohort studies) for later pre-trial evaluation.
- ✓ Protocols for these studies should be registered publicly.
- ✓ Statistical process control techniques can be useful in both early and late assessment. Randomised trials should be used whenever possible to investigate efficacy, but adequate pre-trial data are essential to allow power calculations, clarify the definition and indications of the intervention, and develop quality measures.
- ✓ Difficulties in conducting randomised clinical trials should be addressed by measures to evaluate learning curves and alleviate equipoise problems. Alternative prospective designs, such as interrupted time series studies, should be used when randomised trials are not feasible.
- ✓ Established procedures should be monitored with prospective registries to analyse outcome variations and to identify late and rare events.
- ✓ Achievement of improved design, conduct, and reporting of surgical research will need concerted action by editors, funders of health care and research, regulatory bodies, and professional societies.

[15] Pennell C, Hirst A, Campbell W, Sood A, Agha R, Barkun J, McCulloch P. Practical guide to the Idea, Development and Exploration stages of the IDEAL Framework and Recommendations. Br J Surg. 2016 Apr;103(5):607-15.

A summary of key features of the IDEAL Framework, Recommendations and Proposals are presented in Table 2 and Table 3.* The information is drawn from the Lancet articles arising from the Balliol Colloquium (12 13 14), from three subsequent articles published in the BMJ in 2013([16] [17] [18]) and from a McCulloch PowerPoint presentation.

The IDEAL Framework:

Table 2: Defining characteristics of IDEAL framework phases

Stage 1 IDEA	Stage 2a DEVELOPMENT	Stage 2b EXPLORATION	Stage 3 ASSESSMENT	Stage 4 LONG-TERM MONITORING
Initial report Innovation may be planned, accidental or forced Focus on explanation and description	"Tinkering"(rapid iterative modification) Small experience from one centre Focus on technical details and feasibility	Technique now more stable Replication by others Focus on adverse effects and potential benefits Learning curves important Definition and quality parameters developed	Gaining wide acceptance Considered as possible replacement for current treatment Comparison against current best practice (RCT if possible)	Monitoring late and rare problems, changes in use and quality of surgical performance

Each of the five stages of surgical innovation is defined by one key question and studies at each stage should be designed to answer this question. IDEAL Recommendations describe study formats which are designed to do so. The key question at each stage is noted below:

- STAGE 1: What is the new treatment concept?
- STAGE 2a: Have we perfected it?
- STAGE 2b: Can we reach agreement on the questions which need to be answered before an RCT can succeed?
- STAGE 3: Is it better than current practice?
- STAGE 4: Are there any surprises?

* The IDEAL Collaboration: http://www.ideal-collaboration.net/about-ideal/ideal-summary-tables/ (Accessed July 19, 2016).

[16] McCulloch P, Cook J, Altman D, Heneghan C, Diener M; IDEAL group. IDEAL framework for surgical innovation 1: the idea and development stages. BMJ. 2013 Jun 18;346:f3012.
[17] Ergina P, Barkun J, McCulloch P, Cook J, Altman D; IDEAL group. IDEAL framework for surgical innovation 2: observational studies in the exploration and assessment stages. BMJ. 2013 Jun 18;346:f3011.
[18] Cook J, McCulloch P, Blazeby J, Beard D, Marinac-Dabic D, Sedrakyan A; 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.

The IDEAL Recommendations:

Table 3: Key recommendations for research design at each IDEAL phase

IDEAL – An integrated evolution pathway

1 IDEA (First-in-human) ***Professional Innovation Database***	2a DEVELOPMENT ***Prospective Development Studies***	2b EXPLORATION ***Prospective Exploration Study(collaborative uncontrolled pros. study)***	3 ASSESSMENT ***Surgical RCT***	4 LONG TERM MONITORING ***Prospective Registries***
✓ Complete technical description ✓ Explanation of patient selection ✓ Registration of report	✓ Prospective Cohort Study (PDS) ✓ Transparent consecutive reporting of cases ✓ Explanation of changes in technique or indication	✓ Prospective collaborative cohort study (PES) ✓ Evaluation of learning curves ✓ Definition of QC parameters ✓ Estimation of power calculations ✓ Early joint analysis leading to RCT ✓ Feasibility/Pilot RCT	✓ Definitive RCT ✓ Removal of investigator bias from recruitment	✓ Registry to detect late/rare events ✓ Monitoring of indication and performance creep
Compulsory reporting of all new innovations on a *Professional Innovation Database* Confidential entry allowed to encourage reporting of failed innovations (similar to CHRP system) Hospital or institution to be informed separately as a professional duty	Detailed description of selection criteria Detailed technical description Prospective account of ALL cases consecutively, including those NOT treated with new technique/device Clear STANDARDISED definitions of outcomes reported Description of ALL modifications, and when they were made during the series Registration of PROTOCOL before study starts Use of Statistical Process Control (SPC) methods to evaluate progress	To evaluate technique prospectively and cooperatively To develop a consensus over *Definition* of the procedure, *quality standards* and *patient selection* To gather *data* for *power calculations* To evaluate and monitor *learning curves* To evaluate *preferences* and *values* amongst patients and clinicians To achieve consensus on future *trial question* To develop a *multi-centre randomised trial (RCT)*	Consider blinding Standardise Terms Use PES data to decide: •Power calculations •Definition of intervention •Quality measures •Learning curve eligibility Deal with preferences •Expertise-based trials •Qualitative research •Third party randomisation •Decision support aids •Cohort/RCT mixtures Use modified RCTs or recognised alternative if RCT not feasible: •Pilot & Feasibility RCTs •Stepped-wedge design •Controlled-interrupted time series •TWICS (Trials within Cohort Studies)	To detect late and rare problems Quality Control Monitoring "indication creep" SPC used for quality control (Shewart charts, CUSUM, VLAD)

McCulloch notes that **Prospective Development Studies (PDS)** are important for the following reasons:

- Techniques in the DEVELOPMENT stage are not yet stable
- Reporting changes and their reasons allows others to learn faster and not repeat mistakes
- They are thereby ethically superior and should increase the speed of the development process

EXAMPLES of PDS in Robotic Surgery:
(Stage 2a)

Menon M, Sood A, Bhandari M, Kher V, Ghosh P, Abaza R, Jeong W, Ghani KR, Kumar RK, Modi P, Ahlawat R. Robotic kidney transplantation with regional hypothermia: a step-by-step description of the Vattikuti Urology Institute-Medanta technique (IDEAL phase 2a). Eur Urol. 2014 May;65(5):991-1000.

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;19:104-11.

McCulloch notes that **Prospective Exploration Studies (PES)** are important for the following reasons:

- Organising a surgical RCT is difficult but important: it requires TRUST and UNDERSTANDING between surgeons
- PES studies improve trust and understanding between surgeons
- PES studies improve ownership and belief in data
- PES studies allow questions which worry surgeons to be answered, e.g.:
  - Which variations of the procedure are acceptable?
  - Are some colleagues still learning?
  - Are we all agreed upon which patients are suitable?
  - How many patients will we need?
- As a result of the above factors and questions, PES improve the feasibility of RCTs

EXAMPLE of PES:
(Stage 2b study)

Degiuli M, Sasako M, Ponzetto A, et al. Extended lymph node dissection for gastric cancer: results of a prospective, multi-centre analysis of morbidity and mortality in 118 consecutive cases. Eur J Surg Oncol. 1997 Aug;23(4):310-4.

Degiuli M, Sasako M, Ponti A. Morbidity and mortality in the Italian Gastric Cancer Study Group randomized clinical trial of D1 versus D2 resection for gastric cancer. Br J Surg. 2010 May;97(5):643-9.

8 Italian centres performed a cohort study together which identified that 3 centres had not overcome their learning curves. The remaining 5 centres subsequently performed the RCT together.

IV. The IDEAL Framework – modification for the development of medical devices:

Stating that current evidence requirements for regulation and large scale purchasing of invasive therapeutic devices fall short of safety and efficacy needs, Sedrakyan et al.([19]) support a "total product life cycle" (TPLC) approach to data collection as evidence is generated and promote the IDEAL model. The authors further propose that medical device development and evaluation may necessitate specific modifications to the IDEAL framework including:

- The need for a Stage 0
  While in surgery most iterative changes occur in Stage 2a (development), most changes in the development of devices occur before in-human use. Sedrakyan et al. therefore suggest that the IDEAL framework for devices should begin in pre-clinical development during product design and testing before first-in-human use (Stage 1).

[19] Sedrakyan A, Campbell B, Merino J, Kuntz R, Hirst A, McCulloch P. IDEAL-D: a rational framework for evaluating and regulating the use of medical devices. BMJ 2016;353:i2372.

Citing a need to balance intellectual property against proof of safety and reliability, the authors recommend consensus between scientists, engineers, regulators, and industry for this stage.

- An integrated Stage 2 (i.e., a merging of 2a and 2b)
  While innovation in surgery requires for iterative "tinkering" in Stage 2a and consensus among surgical teams (Prospective Exploration Studies) in Stage 2b, most technical iterations in devices occur between Stages 0-1 and usually involve a single manufacturer. The authors suggest, therefore, that a single integrated Stage 2 may be sufficient for devices, in readiness for a definitive randomized controlled trial in Stage 3. In addition, if a Stage 2 study were a regulatory requirement for device innovation, the authors propose that the overall quality of clinical evidence would be enhanced.
- Stage 3 - Recognised alternative trial designs and/or RCTs for lower-risk devices?
  Given the wide range of risk represented by medical devices, Sedrakyan et al. state that requiring lower-risk devices to undergo an RCT in order to provide evidence of "substantial equivalence" to a predicate device may be impractical. Commercial, societal and patient interests need to be balanced by policymakers in deciding which categories of devices should routinely undergo RCT evaluation, or equivalent. IDEAL allows for valid experimental trial designs other than RCTs in Stage 3, including tracker trials, adaptive designs and studies based on economic modelling whereby incremental innovations and device comparators could also be included.
- Earlier and continuous use of registries, beginning with first-in-human (Stage 1)
  IDEAL recommends the implementation of registries in Stage 4 (long-term follow-up). The authors suggest that using registries in device evaluation is possible throughout all four stages, particularly with the increasing availability of unique device identifiers in registries and other data systems. Prospective registries beginning with Stage 1 could be useful for safety monitoring of higher-risk devices, also allowing for "nested" RCTs and for using risk adjustment techniques in large registry datasets to study small or long-term effects in situations with multiple confounders, where an RCT might not be possible.

V. Conclusion

A significant obstacle in the development and evaluation of innovative therapeutic devices is the generation of high-quality evidence which meets the needs of regulators, payers and health care systems. The IDEAL Framework and Recommendations provide the basis for an integrated evaluation pathway from first-in-human use through to maturity for new therapeutic devices and a practical template for EXCITE International studies. Specifically, the IDEAL model facilitates an assessment of where devices are on an evolutionary path and allows for critical study questions to be identified and appropriate studies to be designed.

The author gratefully acknowledges Mary Spayne, MPH for her assistance in writing this document.

ACRONYMS:

CUSUM – Cumulative Sum Control Chart
EI - EXCITE International (Excellence in Clinical Innovation and Technology Evaluation)
IDEAL - The IDEAL Collaboration (Idea, Development, Exploration, Assessment, Long-term study) (UK)
RCT - Explanatory Randomized Controlled Trial
SPC – Statistical Process Control
TPLC – Total Product Life Cycle
TWICS – Trial Within Cohort Studies
VLAD – Variable Life Adjusted Display Chart