Surgical Innovation and Evaluation 3

No surgical innovation without evaluation: the IDEAL recommendations

*Peter McCulloch, Douglas G Altman, W Bruce Campbell, David R Flum, Paul Glasziou, John C Marshall, Jon Nicholl, for the Balliol Collaboration**

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 research databases 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 doing 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 databases 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.

Lancet 2009; 374: 1105–12

See **Editorial** page 1037

See **Comment** page 1039

This is the third in a **Series** of three papers on surgical innovation and evaluation

*For members see *Lancet* 2009; **374:** 1089–96

Nuffield Department of Surgery, University of Oxford, Oxford, UK (P McCulloch MD); **Centre for Statistics in Medicine, University of Oxford, Oxford, UK** (Prof D G Altman DSc); **National Institute for Health and Clinical Excellence, London, UK** (Prof W B Campbell FRCS FRCP); **Departments of Surgery, Health Services, and the Surgical Outcomes Research Center, University of Washington, Seattle, WA, USA** (Prof D R Flum MD); **Department of Primary Health Care, University of Oxford, Oxford, UK** (Prof P Glasziou PhD); **Department of Surgery, University of Toronto, St Michael's Hospital, Toronto, ON, Canada** (Prof J C Marshall MD); **and School of Health and Related Research, University of Sheffield, Sheffield, UK** (Prof J Nicholl DSc)

Correspondence to:
Mr Peter McCulloch, Nuffield Department of Surgery, John Radcliffe Hospital, Headley Way, University of Oxford, Oxford OX3 9DU, UK
peter.mcculloch@nds.ox.ac.uk

Introduction

Development and evaluation of surgical and interventional techniques proceeds through stages similar to those for drug development, but with important differences.[1] In general, however, the appropriate model for surgery has probably more in common with complex interventions in areas such as psychological and physical therapies.

The UK Medical Research Council (MRC) set out guidance for the assessment of complex interventions in 2000,[2] which was updated in 2008.[3] Complex interventions are defined as methods consisting of several interacting components or involving the use of difficult or complex techniques, which may be applied in various ways. These are also defining characteristics of surgical procedures, but surgery has a specific combination of attributes that causes additional problems. We outlined these issues in the first two papers of the Series,[1,4] and they have guided the development of these recommendations.

The MRC recommendations include: development and evaluation through iterative phases; use of experimental rather than observational designs whenever possible; measurement of outcomes as well as process; reporting detailed descriptions of interventions to improve reproducibility, evidence synthesis, and wider implementation. We have tried to tailor these recommendations to the surgical setting.

We know that, if our proposals are going to improve the quantity and quality of surgical research, they should be practical and not create issues for the continuing development of procedures. Unrealistically demanding standards could hinder surgical innovation. In the short term, we cannot change how surgical innovation happens and so we need to adapt our methods to the process rather than doing the opposite.

Recommendations are futile without a viable mechanism to enable their adoption. We therefore address the roles that funding bodies, regulators, and journal editors could have in encouraging improvements in the conduct and reporting of surgical evaluation. We recognise that these roles, together with the attitudes of the surgical community and the public, vary worldwide depending on culture and legislation, resulting in differences in the process.

Surgical innovation and the factors affecting its evaluation have been described in the first two papers of this Series.[1,4] Here, we address the difficulties in the assessment of surgical innovations and propose some possible solutions. We focus on the evaluation of new invasive techniques and procedures rather than changes in associated aspects of care, such as diagnostic imaging or accelerated recovery programmes.

Stages in the development and assessment of surgical innovations

In the first paper of this Series,[1] we describe how surgical innovation happens and how innovations are adopted. By contrast with the formalised approach for drug development, the process in surgery has been unregulated, unstructured, and variable. Nevertheless, it seems to proceed in phases;[5] we have developed this idea in a descriptive model delineating stages of innovation, development, exploration, assessment, and long-term study (the IDEAL model, see table).

Stage 1 of innovation happens when a surgeon or small group of surgeons try out a procedure for the first time. If early reports suggest benefits, some early adopters may take up the innovation (stage 2a). In this phase (development stage), the focus is on the technical development of the procedure. Subsequently, attention is on the investigation of indications for use of the procedure, understanding its potential benefits and harms, and increasing effectiveness to an optimum (stage 2b, exploration phase). Early adopters refine their skills, moving up their learning curve. During stage 3, the key question is posed: is this technique better than established methods in terms of clinical efficacy and cost-effectiveness? (assessment phase). Definitive studies are needed, but a tipping point may occur once there is an optimised procedure and a sufficiently large group of surgeons skilled in its use. If the opportunity for robust evaluation is not seized, widespread adoption may happen without adequate evidence. Finally (stage 4), when the procedure has become widely adopted, its effectiveness in routine use should come under scrutiny (long-term study phase). Rare and long-term outcomes might become clear at this stage, and outcome variability can lead to clarification of indications or of important technical details. Study results may be generalised to routine practice, and indications may be widened.

Simulator or animal studies before stage 1, if they exist, could be regarded as stage 0. The stages represent a model of development, but in practice innovation will not always proceed in an orderly, linear fashion; for example, if development work in animal models and simulators has been extensive, it might be appropriate for the first use in people to be in stage 2. Stages may also overlap, with evaluations occurring in parallel. Innovation processes are naturally iterative, reverting to earlier stages when substantial difficulties arise. During each phase, however, planning, evaluation, and reporting are needed.

Stage 1: innovation

The stage of innovation describes the first use of a new procedure in a patient, prompted by the need for a new solution to a clinical problem. This situation might occur in an emergency (eg, the development of damage control surgery for polytrauma[11]) or in a patient whose condition allows time for planning. If time allows, we suggest that the surgeon informs the hospital of the intention to undertake a new procedure. At this stage, research ethics approval is not appropriate, although full and clear informed consent is an ethical obligation for competent patients.

All new procedures should be reported automatically, whether successful or not. It is perhaps even more important to report adverse events and failures than successes, to avoid their repetition in the future. Hospitals need to be informed; however, surgeons should also report the new procedures in an online register available to all surgeons. This approach would need some infrastructure and a cultural change among surgeons. The option of anonymous reporting of adverse outcomes, as in the aviation confidential human factors incident reporting programme (CHIRP)

	1 Idea	2a Development	2b Exploration	3 Assessment	4 Long-term study
Purpose	Proof of concept	Development	Learning	Assessment	Surveillance
Number and types of patients	Single digit; highly selected	Few; selected	Many; may expand to mixed; broadening indication	Many; expanded indications (well defined)	All eligible
Number and types of surgeons	Very few; innovators	Few; innovators and some early adopters	Many; innovators, early adopters, early majority	Many; early majority	All eligible
Output	Description	Description	Measurement; comparison	Comparison; complete information for non-RCT participants	Description; audit, regional variation; quality assurance; risk adjustment
Intervention	Evolving; procedure inception	Evolving; procedure development	Evolving; procedure refinement; community learning	Stable	Stable
Method	Structured case reports	Prospective development studies	Research database; explanatory or feasibility RCT (efficacy trial); diseased based (diagnostic)	RCT with or without additions/ modifications; alternative designs	Registry; routine database (eg, SCOAP, STS, NSQIP); rare-case reports
Outcomes	Proof of concept; technical achievement; disasters; dramatic successes	Mainly safety; technical and procedural success	Safety; clinical outcomes (specific and graded); short-term outcomes; patient-centred (reported) outcomes; feasibility outcomes	Clinical outcomes (specific and graded); middle-term and long-term outcomes; patient-centred (reported) outcomes; cost-effectiveness	Rare events; long-term outcomes; quality assurance
Ethical approval	Sometimes	Yes	Yes	Yes	No
Examples	NOTES video[6]	Tissue engineered vessels[7]	Italian D2 gastrectomy study[8]	Swedish obese patients study[9]	UK national adult cardiac surgical database[10]

RCT=randomised controlled trial. SCOAP=Surgical Clinical Outcomes Assessment Programme. STS=Society of Thoracic Surgeons. NSQIP=National Surgical Quality Improvement Program. NOTES=natural orifice translumenal endoscopic surgery.

Table: **Stages of surgical innovation**

system,[12] might aid acceptance. These reports should contain clear anonymous details of the patient, their condition, the rationale and background for use of the procedure, exactly what was done, and adequate details of relevant outcomes.

Stage 2a: development

Development involves the planned use of a procedure in an initial small group of patients (rarely more than 30 and sometimes less than ten) to support experience with its first use and often to refine or modify the precise technique. A good example is the report of tissue engineered autologous grafts for haemodialysis.[7] The traditional method of reporting this experience in retrospective case series studies has been justifiably criticised.[13] Instead, we recommend that protocols for prospective development studies are registered before patient recruitment begins, describing patient selection principles, operative methods, and outcomes to be measured. Protocols should be registered beyond the surgeon's institution and should undergo some form of ethical approval. A responsive and fast system, and a presumption in favour of innovation, will however be needed to avoid stifling progress.

Technical modifications may be common during stage 2a: their nature and timing should be meticulously recorded to allow understanding of their possible effect on outcomes. Learning curves are also an important issue in this phase, and clear sequential outcome reporting of all cases should be done, without omissions. Ethical considerations require that all reasonable precautions are taken to avoid harm to patients during the learning curve, including, when possible, mentoring. Agreement should exist about who is responsible for ensuring risk minimisation between the surgeon, the institution, and their ethics committee.

Reporting during this stage needs to include: selection criteria and proportion of eligible patients selected; a clear description of the procedure and each modification, with timing; and relevant outcomes, with recognised standard definitions of important categories, such as specific complications. Retrospective case series studies may have a minor role in hypothesis generation, but should, as a minimum requirement, describe consecutive patients without exclusions, and use a clear, unambiguous standard reporting protocol, perhaps developed in a way similar to CONSORT,[14] STROBE,[15] and other templates.

Stage 2b: exploration

Exploration occurs once the procedure has been described and the main technical aspects worked out. Experience with the procedure may still be scarce, however, and outcomes with larger numbers of patients are usually needed (up to a few hundred) before a randomised clinical trial that compares the new procedure with traditional management is feasible. At this stage, the procedure is likely to be adopted by surgeons in more than one unit, making the issues of mentoring and learning-curve evaluation especially important. Data should be captured systematically for every patient having the procedure, especially to ensure that adverse outcomes are documented.

To achieve these goals, prospective research databases are valuable. These carefully planned, prospective but uncontrolled clinical studies could run as parallel additions to smaller feasibility or explanatory randomised clinical trials that might be appropriate at this stage (table). These uncontrolled studies could also be an integral preparatory stage for a major randomised trial, as in the case of the Italian study of radical gastrectomy[8] (the stage 2S idea[5]), but may often be the main study method in this phase. Controversy persists around the appropriate timing of randomised trials, with theoretical arguments for early randomisation balanced against practical ones for delay. The learning curve is likely to affect which surgeons participate in randomised trials and when they become involved. Statistical methods for continuous performance monitoring during prospective database accrual can be helpful in making these decisions.[16,17]

Well characterised and relevant outcome measures are important for both research databases and randomised trials. These should include technical, clinical, and patient-reported outcomes. Research databases also need to provide information about the population presenting for the new treatment. How many were treated by the new procedure, by alterative procedures, or managed conservatively? This method of reporting represents a shift from traditional procedure-based to disease-based. Finally, research databases should report quality control measures to enable clear understanding of the accuracy with which the procedure has been reproduced. As with stage 2a, a restricted role for retrospective case series studies might remain in stage 2b, but the same strictures about reporting standards apply.

Stage 3: assessment

Previous stages focused on the development of a new technique and the description of its outcomes; this stage aims to assess effectiveness against current standards. The new method should now be sufficiently evolved to warrant full evaluation, which does not mean that it will not evolve further. This stage should be seen as a milestone on a learning path; the key issue is to decide which is the best feasible comparator for the new procedure.

Randomised trials should be the default option in this stage, but trials of surgical techniques are sometimes unnecessary,[18] sometimes not feasible, and sometimes might need adaptations or additional features. Trials might be unnecessary when an advance that cannot be explained by either chance or bias is clear and substantial. For example, tracheostomy for tracheal obstruction, suturing for repairing large wounds, and ether for anaesthesia were such clear advances that trials were

Panel 1: **Study design and reporting ideas for improving evidence on surgical innovation**

Design

- Prospective development studies
- Prospective research databases
- Alternatives to randomised clinical trials
  - Case-matching studies
  - Controlled interrupted-time series designs
  - Step-wedge designs
- Modified randomised clinical trials
  - Randomisation variants: expertise-based, third party
  - Tracker trials
  - Phase 2S transition from database to randomised clinical trial
  - Feasibility randomised clinical trial (where study size and endpoints are aimed at determining the feasibility of a definitive study)
  - Explanatory randomised clinical trial (where contextual confounders are minimised to ensure the best possible comparison between experimental and control treatments)
- Additions to randomised clinical trials
  - Learning curve evaluation
  - Quality control and compliance measures

Reporting

- Mandatory registry for procedures thought to be first in man with anonymous reporting option
- Protocol and study registries for prospective development studies in surgery
- Registries for surveillance of specific established techniques
- Development of agreed reporting standards and definitions for key outcomes
- Reporting of continuous quality control measures (eg, CUSUM in stages 2b to 4)

not needed. Most operations, however, are smaller advances prone to overoptimistic assessment by their developers and, therefore, need controlled randomised studies, when possible.

Although randomised trials are generally more difficult in surgery and interventional procedures than with pharmaceuticals, several successful trials have been done in surgery. Many of these were comparisons of different techniques, but others assessed whether surgery was better than alternative management (eg, coronary bypass grafting, carotid endarterectomy, and knee arthroscopy[19–21]). However, randomised trials may not be feasible for ethical or pragmatic reasons, such as recruitment difficulties. In these cases, alternative designs are necessary and some options are listed below.

Parallel group non-randomised studies

Reasonable controls can be obtained without randomisation by matching procedures, perhaps with propensity scores. For example, in the Swedish obese patients study,[9] when a patient underwent bariatric surgery, a matched control patient was selected from a database of obese individuals. Of course, this can only match known confounders, but is better than unmatched or historical controls.

Controlled interrupted-time series studies

These studies allow rapid and simple comparison with a parallel control group and with the pre-interruption results of the previous technique in the study group, but cannot eliminate selection bias.[22]

Step-wedge designs (randomised roll-outs)

These designs are used when study centres commence the new method in a random order. These have been used in public health interventions,[23] but not in surgery. A limitation is that interpretation of the eligibility criteria may change as a centre switches to the new technique, and hence non-comparable groups can be created despite randomisation.

Tracker trials

These studies have been proposed[24] to make the randomisation process more palatable by allowing each clinician to randomly assign trial groups they think are reasonable alternatives, including those added during the course of the study, thereby easing recruitment difficulties and providing an improved dataset.

Expertise-based randomised trials

These studies are done when patients are randomly assigned by a third party to surgeons or interventionists (ie, doctors who are not surgeons but do complex invasive procedures), who then treat all their patients with their preferred intervention. Although surgeons may be more willing to participate in expertise-based trials,[25] these are not without analytic issues: randomised allocation to a surgeon often introduces important confounding factors, such as postoperative care regimens.

Lack of equipoise might be an issue for surgeons[26] or for patients in trials comparing very different options. For these situations, consent and randomisation by trained third parties can provide a solution, as for the ProTect trial.[27]

Whatever evaluation is used, the quality of the intervention needs to be monitored, and this may mean taking into account not only the surgery but also the surgical environment, including preoperative and postoperative care.

Stage 4: long-term study

In this stage, established procedures are assessed for rare and long-term outcomes, and for variations in outcome. Learning curves may be less important now than in previous stages, but differences in selection criteria or in the quality of surgery or aftercare may become apparent

through unexpected outcome variation between study centres. The typical study design is a registry: making registries disease-based in this setting may be impractical, in which case careful monitoring of indications for use of the procedure is needed to make sure that changes in outcome are not due to changes in case selection. The value of this type of study depends on its representativeness; therefore, only key outcomes and relevant information should be obtained to encourage complete data entry.

Depending on the frequency of the procedure studied, large numbers of cases may be available for analysis, allowing observation and investigation of outcome variations among subgroups. Risk adjustment for patient comorbidity is a major and a very complex issue. Most surgeons are sensitive about potentially unfair comparisons of their results with those of colleagues who might deal with a less challenging patient group. Recent work[28] suggests that the safest use of adjustment is to allow longitudinal comparison of a unit or surgeon against themselves, because comparisons between units are confounded by inconstancy of the risk associated with any adjustment factor, such as the American Society of Anesthesiologists (ASA) score. The use of CUSUM (ie, cumulative sum) and other statistical process control methods may enable this type of study to become a quality control method.[5,17]

Discussion

The IDEAL model is based on the suggestion that surgical innovation and evaluation can and should evolve together in an ordered manner from concept, through exploration, to validation by randomised trials (panel 1). This order does not always reflect real situations, in which the timing and nature of evaluation might depend on the type of development. Gradual evolution of a technique may culminate in a systematically different approach, as in delayed surgical intervention for infected necrotising pancreatitis[29] (panel 2), or non-operative management of splenic injuries. The feasibility of a novel technique may be assessed in animal models or simulation before full evaluation in patients with a disease. The development of natural orifice translumenal endoscopic surgery (NOTES) through animal studies is an example of this approach.[6] Another common variation is when comparison of an established surgical technique with a non-operative strategy becomes necessary. An example would be the evaluation of carotid endarterectomy to prevent stroke.[20]

The ethical principles of human experimentation are expressed in the amended Declaration of Helsinki. These recognise the central role of research in advancing medical care, but emphasise that well-being of the individual is the main concern. These principles also underline the importance of transparent reporting of research protocol and context, and of ethical approvals to register research initiatives before patient enrolment, and to report outcomes in peer-reviewed publications.

Although application of these principles to the conceptualisation and comparison pathways seems straightforward, this is not so for the earlier stages of the evolutionary pathway in which innovation may only become apparent in retrospect. A surgeon may first use a new approach out of necessity for an otherwise insoluble problem. Repeating the approach, he or she becomes aware of having discovered something potentially useful. Thus, the first two stages of our model may be mainly completed before the innovative nature of the process has been appreciated. No precise timepoint exists to indicate when an approach becomes an innovation needing a formal scientific and ethical framework, because crucial change occurs within the consciousness of the investigator, specifically the belief that the innovation represents a novel approach and a desire to communicate the innovation and to assess its usefulness.

Ethically, research consent is required at this point because the investigator believes that the intervention is novel, and therefore not part of accepted practice. Evaluation may also entail additional investigations focused on the procedure rather than the patient. Moreover, communication of the innovation requires dissemination of information about the patient. But other obligations also arise from the recognition that innovation is now being pursued, and these ethical and scientific considerations dictate the adoption of four measures that are familiar from the stage-based description of the process.

First, the investigator should develop a protocol describing the nature of the innovation, its rationale, its

Panel 2: Acute pancreatitis—evolution of a surgical strategy

Acute pancreatitis is a disorder the clinical expression of which ranges from mild, self-limited abdominal pain to a complex and life-threatening illness associated with substantial morbidity and long intensive care unit stay. In the 1970s, when a disease severity staging system was first developed (Ranson's criteria), the mortality rate for patients with the most severe form of the disease approached 100%.[30] Nowadays, 70–80% of these patients survive. The improved prognosis reflects advances in resuscitation and intensive care, but also a fundamental shift in surgical approach.

In the 1980s, early aggressive surgical debridement was a popular management strategy for pancreatitis, because of widespread acceptance of an analogy with early definitive intervention to trauma and major burns. However, surgical intervention was commonly complicated by uncontrollable retroperitoneal bleeding, because tissue planes between viable and non-viable tissue were poorly demarcated. Case series studies showed that delayed surgery was associated with an improved outcome,[31,32] a conclusion that was supported by a small randomised trial.[33] Recent case series studies support the notion that surgery can be deferred, or even avoided,[34] in patients with documented infected necrosis.

The evolution of surgical therapy for infected pancreatic necrosis shows an incremental shift in practice rather than the deliberate introduction of a fundamentally different treatment approach.

Panel 3: Actions to facilitate improvements in surgical evidence

Editors

- Promotion of IDEAL reporting and design standards
- Assistance by editors with development of registries of surgical protocols and reports
- Calls for specific prospective study designs

Funders (both service and research)

- Provide specific funding for well-designed early-stage surgical innovation
- Demand evidence of benefit for new techniques
- Link funding to adequate scientific evaluation
- Support well-designed surgical databases, registries, and reporting systems

Regulators

- Provide rapid, flexible, and expert ethical oversight for early-stage innovation
- Link provisional approval to evaluation or registration of all cases
- Accept IDEAL approved study designs as evidence of appropriate evaluation
- Raise burden of proof for full licensing of new devices to demonstration of efficacy level

Professional societies

- Ensure guidelines explicitly support IDEAL model of technical development and evaluation
- Require members to use appropriate registers for the various stages of innovation as a condition of specialist recognition

potential risks and benefits, and the approach to be used in its evaluation. Second, the protocol and more detailed investigative plans should be submitted for some type of ethical review before further study. Third, the research programme should be registered with an appropriate registry, which would help to address the potential bias arising from the selective reporting that occurs when studies yielding disappointing or adverse results are suppressed. Previous registration shows whether the final analytic plan or the outcomes reported differ from the initial intention, recognises the primacy of the innovator, and could accelerate sharing of evolving ideas. Dedicated registries for early surgical innovation are needed to record these early steps adequately. Finally, there is an obligation to report outcomes in a public or professional forum. Reports should describe the innovation and the population in which it has been evaluated in sufficient detail to enable replication of the work. Endpoints should reflect three broad areas—technical aspects of the intervention, its potential for harm, and its potential for benefit. The main focus of outcome evaluation will change with the stages, from the possibility of the intervention to the refinement and standardisation of technical details, to the potential for benefit and harm, to comparison with current practice, and finally to quality control, with outcomes focusing on long-term benefit and harms.

The achievement of improved standards for the development, testing, and reporting of surgical innovation will require widespread change, with distinct roles for publishers, payers, regulators, and surgical societies (panel 3).

Journal editors have been influential in the wide uptake of registration for randomised trials by specifying required standards for study acceptance. They could and should have the same role in improving surgical research. Adherence to reporting standards, such as those of CONSORT and STROBE,[14,15] will decrease the inadequate reporting of flawed studies and encourage better methodological approaches and more transparent reporting. Some standards that are most urgently needed include mandatory clarification of reporting issues, such as outcome assessment by observers masked to the treatment given, and adequate description of the intervention and of study conduct issues, such as prospective study design and the inclusion of consecutive patients. The reporting of complications is also in urgent need of standardisation with established schemata.[35] Calls from journals to submit specific study types relevant to surgery, such as those described above, could be helpful. Editors have assisted in promoting registries of randomised trial protocols,[14] and could help to create the surgical study protocol registration we suggest.

Groups that pay for health care and for clinical research—whether private or government funded—also have a key role in improving surgical innovation. Research funders need to recognise the nature of surgical innovation and begin to provide funding for well-designed surgical studies preceding randomised clinical trials to encourage high-quality innovation. Ring-fenced allocations or special calls may be needed initially to stimulate the specialty. Similarly, health-care funders should demand high-quality evidence of benefit and cost-effectiveness for surgical procedures and technology, as for drugs, but should understand and accept appropriate study designs as part of this process.

Major gains in surgical innovation could be achieved through the financing and development of reporting systems (stage 1), protocol registries (stage 2a), comprehensive, disease-specific research databases (stages 2b and 3) and population registries (stage 4). Health-care funders would directly benefit from such systems through accurate information about costs and benefits. A lot of data for innovation (especially unsuccessful innovation) are simply not recorded at present, condemning failed innovations to be repeated by others. Linking health-care financing for innovative procedures to mandatory data gathering would create a learning health-care system that could identify and

advance effective new technology. Coverage with evidence development (CED) has been used to finance innovation while ensuring maximum scrutiny.[36] Broad use of CED could encourage innovation while allowing evaluation of effectiveness and safety with appropriate study designs.

National regulatory and advisory bodies, such as the National Institute for Health and Clinical Excellence (NICE) in the UK, could drive up research quality by insisting on better evidence before approving new techniques and by linking provisional approvals to continuing research and data registration. However, such demands need to be practical within the health-care system. Where new equipment is involved, regulatory pressure on industry to deliver data for efficacy (and comparative effectiveness) could be a very effective lever.

The pharmaceutical industry is closely regulated so that drugs cannot come to market without clear evidence of safety and efficacy, and consequently numerous high-quality randomised clinical trials of new drugs exist. Surgical innovation currently lacks major commercial funding sources for research, partly because device, implant, and technology developers do not have similar barriers to market entry. A device or implant must simply be shown to be safe and do what it says it does, and this weak requirement does not encourage randomised trials. For example, to enter the market, radiofrequency ablation device manufacturers needed to show that they did not hurt patients and ablated tissue, but not that they were an effective treatment for specific indications (eg, liver cancer). Had this been a regulatory requirement, an industry-financed randomised trial of radiofrequency ablation device for liver cancer would by now have been done. Requirements for early registration, however, need to take into account the importance of allowing intellectual property rights to be properly developed: ultimately, patient safety and ethics should trump profit, but excessively detailed regulation should not be allowed to smother creativity.

Finally, surgical societies could also help to change surgical innovation. As leaders and representatives of surgeons, societies have an obligation to encourage safe, effective innovation. All societies can and should communicate to their members the importance of an ethical framework for innovation and expected reporting standards. Guidelines from specialist bodies carry substantial weight with government, industry, and hospital managements, and specific recommendations about study design and reporting could therefore be very helpful. A recent example was the effective control of innovation in laparoscopic colon resection for cancer in the USA for more than 5 years by the American Society of Colon and Rectal Surgeons while a large-scale randomised clinical trial was done.[37]

We believe surgical science can be greatly improved, and progress in surgical care and interventions will become safer, more efficient, and better.

Contributors

The paper was written by the writing group, which includes PM, DGA, WBC, DRF, PG, JCM, JN, and the Balliol Collaboration. The group agreed the form and content of the paper at a meeting on April 3, 2009. PM wrote the initial outline and the final draft, and coordinated and edited contributions from the other authors. Members of the Balliol Collaboration were circulated with drafts, and made comments and contributions.

Conflicts of interest

We declare that we have no conflicts of interest.

Acknowledgments

The Balliol Colloquium has been supported by Ethicon UK with unrestricted educational grants and by the National Institute of Health Research Health Technology Assessment Programme. The Balliol Colloquium was administratively and financially supported by the Nuffield Department of Surgery at the University of Oxford and the Department of Surgery at McGill University.

References

1 Barkun JS, Aronson JK, Feldman LS, Maddern GJ, Strasberg SM, for the Balliol Collaboration. Evaluation and stages of surgical innovations. *Lancet* 2009; **374:** 1089–96.

2 Medical Research Council. A framework for the development and evaluation of RCTs /or complex interventions to improve health. London: Medical Research Council, 2000.

3 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:** 1655.

4 Ergina PL, Cook JA, Blazeby JM, et al. Challenges in evaluating surgical innovation. *Lancet* 2009; **374:** 1097–104.

5 McCulloch P. Developing appropriate methodology for the study of surgical techniques. *J R Soc Med* 2009; **102:** 51–55.

6 Flora ED, Wilson TG, Martin IJ, O'Rourke NA, Maddern GJ. A review of natural orifice transluminal endoscopic surgery (NOTES) for intra-abdominal surgery: experimental models, techniques, and applicability to the clinical setting. *Ann Surg* 2008; **247:** 583–602.

7 McAllister TN, Maruszewski M, Garrido SA, et al. Effectiveness of haemodialysis access with an autologous tissue-engineered vascular graft: a multicentre cohort study. *Lancet* 2009; **373:** 1440–46.

8 Degiuli M, Sasako M, Ponti A, Calvo F. Survival results of a multicentre phase II study to evaluate D2 gastrectomy for gastric cancer. *Br J Cancer* 2004; **90:** 1727–32.

9 Sjöström L, Narbro K, Sjöström CD, et al; Swedish Obese Subjects Study. Effects of bariatric surgery on mortality in Swedish obese subjects. *N Engl J Med* 2007; **357:** 741–52.

10 Keogh BE, Kinsman R, on behalf of the Society of Cardiothoracic Surgeons of Great Britain and Ireland. Fifth national adult cardiac surgical database report 2003. Henley on Thames, UK: Dendrite Clinical Systems, 2004.

11 Burch JM, Ortiz VB, Richardson RJ, Martin RR, Mattox KL, Jordan GL Jr. Abbreviated laparotomy and planned re-operation for critically injured patients. *Ann Surg* 1992; **215:** 476–84.

12 The UK Confidential Reporting Programme for Aviation and Maritime. http://www.chirp.co.uk/information-about.asp (accessed April 29, 2009).

13 Horton R. Surgical research or comic opera: questions, but few answers. *Lancet* 1996; **347:** 984–85.

14 Altman DG. Better reporting of randomised controlled trials: the CONSORT statement. *BMJ* 1996; **313:** 570–71.

15 von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP for STROBE Initiative. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007; **147:** 573–77.

16 Novick RJ, Fox SA, Stitt LW, et al. Assessing the learning curve in off-pump coronary artery surgery via CUSUM failure analysis. *Ann Thorac Surg* 2002; **73:** S358–62.

17 Beiles CB, Morton AP. Cumulative sum control charts for assessing performance in arterial surgery. *ANZ J Surg* 2004; **74:** 146–51.

18 Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. *BMJ* 2007; **334:** 349–51.

19 Detre KM, Takaro T, Hultgren H, Peduzzi P. Long-term mortality and morbidity results of the Veteran Administration randomized trial of coronary artery bypass surgery. *Circulation* 1985; **72:** V84–89.

20 Mayo Asymptomatic Carotid Endarterectomy Study Group. Results of a randomized controlled trial of carotid endarterectomy for asymptomatic carotid stenosis. *Mayo Clin Proc* 1992; **67:** 513–18.

21 Kirkley A, Birmingham TB, Litchfield RB, et al. A randomized trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med* 2008; **359:** 1097–107.

22 Shadish WR, Cook WD, Campbell DT. Quasi-experiments: interrupted time-series designs. In: Experimental and quasi-experimental designs for generalized causal inference. Boston, MA, USA: Houghton Mifflin, 2002: 171–206.

23 De Allegri M, Pokhrel S, Becher H, et al. Step-wedge cluster-randomised community-based trials: an application to the study of the impact of community health insurance. *Health Res Policy Syst* 2008; **6:** 10.

24 Lilford RJ, Braunholtz DA, Greenhalgh R, Edwards SJ. Trials and fast changing technologies: the case for tracker studies. *BMJ* 2000; **320:** 43–46.

25 Bednarska E, Bryant D, Devereaux PJ; Expertise-Based Working Group. Orthopaedic surgeons prefer to participate in expertise-based randomized trials. *Clin Orthop Relat Res* 2008; **466:** 1734–44.

26 McCulloch P, Kaul A, Wagstaff GF, Wheatcroft J. Tolerance of uncertainty, extroversion, neuroticism and attitudes to randomized controlled trials among surgeons and physicians. *Br J Surg* 2005; **92:** 1293–97.

27 Mills N, Donovan JL, Smith M, Jacoby A, Neal DE, Hamdy FC. Perceptions of equipoise are crucial to trial participation: a qualitative study of men in the ProtecT study. *Control Clin Trials* 2003; **24:** 272–82.

28 Mohammed MA, Deeks JJ, Girling A, et al. Evidence of methodological bias in hospital standardized mortality ratios: retrospective database study of English hospitals. *BMJ* 2009; **338:** 817–20.

29 Adler DG, Chari ST, Dahl TJ, Farnell MB, Pearson RK. Conservative management of infected necrosis complicating severe acute pancreatitis. *Am J Gastroenterol* 2003; **98:** 98–103.

30 Ranson JHC, Rifkind KM, Turner JW. Prognostic signs and nonoperative peritoneal lavage in acute pancreatitis. *Surg Gynecol Obstet* 1976; **143:** 209.

31 Aultman DF, Bilton DB, Zibari GB, Mcmillan RW, McDonald JC. Nonoperative therapy for acute necrotizing pancreatitis. *Am Surg* 1997; **63:** 1114–17.

32 Ashley SW, Perez A, Pierce EA, et al. Necrotizing pancreatitis: contemporary analysis of 99 consecutive cases. *Ann Surg* 2001; **234:** 572–80.

33 Mier J, Leon EL, Castillo A, Robledo F, Blanco R. Early versus late necrosectomy in severe necrotizing pancreatitis. *Am J Surg* 1997; **173:** 71–75.

34 Adler DG, Chari ST, Dahl TJ, Farnell MB, Pearson RK. Conservative management of infected necrosis complicating severe acute pancreatitis. *Am J Gastroenterol* 2003; **98:** 98–103.

35 Dindo D, Demartines N, Clavien PA. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg* 2004; **240:** 205–13.

36 Straube BM. How changes in the Medicare coverage process have facilitated the spread of new technologies. *Health Aff* 2005. http://content.healthaffairs.org/cgi/content/full/hlthaff.w5.314/DC1 (accessed April 29, 2009).

37 American Society of Colon and Rectal Surgeons. Approved statement on laparoscopic colectomy. *Dis Colon Rectum* 1994; **37:** 8–12.