COMMISSION ON THE FUTURE OF SURGERY

Submission by Professor Peter McCulloch

"Never make predictions: especially about the future."

Darryl F Zanuck

Thanks you for asking me to contribute to the evidence for this Commission. I consider it an honour and I hope I can make a useful contribution. As the quote suggests I begin from a rather sceptical point of view. Most of the major historical events of my lifetime – the 1970s oil crisis, the fall of the Berlin Wall, 9/11 and the 2008 financial collapse – took society unawares. The chances that predictions based on perceptible trends and influences turn out to be accurate are therefore not particularly high. In making this contribution I am assuming (hoping) that the UK continues to be a stable democratic country with functioning institutions, and avoids wars, economic or environmental disasters severe enough to dismantle current social infrastructure. In this stable situation, some major influences on the direction of future changes in surgery seem clear. Technological and scientific advances in surgical equipment, including medical devices and robotics, will continue, but much more important will be advances in artificial intelligence and linked computer systems. Non-surgical therapies (genomics related and other) may replace surgery or modify its use in some areas, whilst advances in transplantation and stem cell based therapies may increase it in others. In terms of challenges, the unaffordability of optimal care during the current demographic transition ranks high. Major difficulties in training and developing the surgical workforce are already evident but will worsen. I also foresee a series of ethical challenges related to changes in our relationship with patients, and equity in access to treatment depending on genetics, wealth and behaviour. Most relevant to my own interests, we will need both radical solutions to the safety crisis brought about by increasingly complex care, and better methods for scientifically evaluating surgery and quasi-surgical treatments and devices. I will confine the rest of my comments to these two areas.

EVALUATING INNOVATION IN SURGERY

There is now general recognition that there are major problems to be resolved in the evaluation of new operations, invasive devices and other complex therapeutic interventions. Since the widespread adoption of the RCT in medical research, surgery has consistently lagged behind in developing valid evidence for the innovations it has produced. This has hurt the reputation and standing of surgery as a science, and has had major consequences for funding support – as the College's campaigns on this subject eloquently testify. For the same reasons, it has proved difficult to produce adequate evidence for safe regulation of invasive therapeutic devices, resulting in repeated episodes of widespread harm and unnecessary costs. Unlike medicines, neither surgery nor devices need Level 1 (RCT) evidence of relative efficacy before approval for general marketing. The result is a flood of inadequately evaluated new devices, often with only small changes from previous designs, but some of which have unexpectedly large adverse impacts on patients.

Why has this occurred? To subject a surgical technique to a randomised trial involves solving difficult problems arising from the complex nature of the treatment and the way it is developed. Unlike drugs, new procedures are almost invariably modified following the first-in-human experience, normally several times, before a stable version emerges. During this phase an RCT is not feasible, as the new procedure is in flux, and cannot reasonably be compared with anything else. As use

spreads, controversies arise around points of technique and patient selection, and concerns surface that less experienced operators might cause performance bias in any proposed trial during their learning curves. This in turn leads to a need to define adequate quality of delivery of the operation. Lack of consensus in these three areas – procedure definition, patient selection and quality of delivery – frequently present obstacles to developing an RCT.

Devices, especially invasive therapeutic devices such as hip prostheses, vascular stents and artificial valves, suffer from many of the same evaluation problems. They require skilled implantation and activation procedures and usually have a learning curve. There are variants in the way they are used and disagreements over indications. For both devices and surgery, long term surveillance is needed to evaluate rare or late adverse outcomes, and to identify changes in use and in the quality of delivery. These questions are best studied using comprehensive registry data, but the development of registries has been beset by problems of coverage, ownership, cost and consent.

Legislators recognised early on that operations and devices were much more difficult to study than drugs, and therefore avoided including them in the requirements for more thorough evaluation that followed the Thalidomide disaster, but without providing a clear alternative evaluation pathway. Society requires evaluation of new treatments to be effective but not onerous – wanting to avoid slowing the evaluation process or making it unduly expensive, but at the same time ensuring that all new procedures are safe. No regulatory jurisdiction has, as yet, developed a comprehensive evaluation pathway which provides a coherent response to these complex problems. However an attempt to do so has been made by the IDEAL Collaboration, an international network of surgeons, research methodologists, regulators and industry experts.

The IDEAL Framework provides a 5-step description of the evolution of devices and operations from first human use to obsolescence, and the accompanying IDEAL Recommendations propose specific questions to be answered at each stage, and appropriate study formats appropriate for doing so. The early iterative stage of procedure or device modification is termed Development by IDEAL, and the subsequent consensus-building stage Exploration. The Framework and Recommendations have recently been updated and expanded, (see Table, permission required from Lancet).

In the near future it is likely that IDEAL or a similar paradigm will be adopted more widely to help regulators and purchasers make better decisions. To make this effective, regulators will need to modify the current all-or-none system of approval for marketing, which provides no incentives for manufacturers or innovators to invest in further evaluation after approval is granted. If marketing approval is granted step by step, only in the context of specified study types, and linked to the production of clearly specified evidence, it becomes possible to control the rate of market entry for new devices, and allow high quality scientific evaluation at the same time. This “Total Product Life Cycle” approach will ensure much higher quality evidence for new devices and operations. It does however imply the development of a formal regulatory system for evaluating new operations, which have never been officially regulated to date. Since the costs of the Development stage are small, and those of the Exploration stage far smaller than an RCT, and since recruitment to these non-randomised studies is likely to be rapid, this approach should neither slow innovation nor increase its costs dramatically. The development of more effective registries, an important part of IDEAL, poses more challenges, but here advances in IT are likely to be helpful. The capacity to link massive online datasets will make registries much easier and cheaper to create and much more flexible to use. It will become simple (if data protection concerns can be dealt with) to create temporary bespoke virtual registries for specific purposes. One danger arising from these advances is the threat that this massive expansion of “real world evidence” might lead to the abandonment of RCTs. This must be resisted, since even limitless linked data on every known source of bias would not

render RCTs redundant, as the most important property of randomisation is that it protects against biases we did not know existed. There is however a completely justifiable argument for limiting RCTs to a subset of innovations. RCTs are expensive and time consuming, and it is logistically impossible to subject all changes to devices to an RCT. But deciding when it may be acceptable to eschew an RCT is very difficult, and devising a workable set of principles to guide us is one of the major methodological challenges for the future. Changes to regulatory rules of evidence for devices would have profound effects. If these were suddenly changed to mimic those of pharma, most small device companies would quickly be gobbled up by the few larger units able to finance the full evaluation pathway including the RCT. The ambition of small device start-ups would become (as it is in Pharma) to sell the company to one of the major players on the basis of one promising product. Political reluctance to initiate such turmoil will probably lead to small cautious regulatory steps in the right direction, as we have seen in the FDA and EU in recent years. Meanwhile, the rate at which academic surgical research adopts an integrated evaluation pathway depends largely on the incentives provided by funders, publishers and institutions such as the Royal Colleges.

Whether IDEAL adapts and updates to deal with upcoming challenges or is replaced, a rational framework for producing valid evidence about surgery and complex interventions will remain essential. Without it, surgery (and the device industry) will continue to suffer from the current problems, with regular "boom and bust" cycles of reaction to innovations, based on a critical lack of valid evidence. The legal costs (in the billions) from the vaginal mesh scandal illustrate why this change is likely to happen – the status quo is becoming too expensive.

PATIENT SAFETY

Patient safety has become a subject of acute interest to surgery only in the last 20 years. Prior to this it was largely assumed that both failure to help and unintentional harm to patients were an inevitable and largely stochastic consequence of attempts to intervene to prevent suffering and death. Unfortunately the term "patient safety" has rapidly become an all-purpose epithet which has lost any real meaning, which can be mobilised to justify everything from cost-saving concentration of specialist services to attacks on the pay and conditions of trainees. The real meaning of patient safety is the development of highly reliable healthcare systems which deliver what is intended with a very high degree of fidelity. The improvement in health outcomes if care across the NHS was delivered at this level of reliability would be enormous. The big insight of the "discovery" of patient safety as an issue was that medicine had transformed into industrialised process by a rather haphazard evolutionary process over a century or more, and the numerous resulting system flaws had resulted in a very significant burden of real harm to patients, despite the best efforts of dedicated professionals. In a future where surgical treatment, like all of medicine, will become an ever more distributed process, how can we build systems which provide highly reliable treatment in which errors and omissions do not occur, and outcomes are consistently optimised?

The obstacles to safer, more reliable surgical systems are formidable. The combination of a rapidly ageing society and the steadily rising cost of new healthcare technology place intolerable demands on publicly funded health systems, leading to delays, cost-cutting measures and pressure on staff. The person-centred, judgemental paradigm which underpins traditional professional attitudes and cultures in healthcare sits uneasily with the "learning culture" concept, where objective systems analysis of error identifies ways of improving system reliability, rather than focusing on individual accountability. Don Berwick in his analysis of NHS healthcare safety correctly described a "culture of fear" as a major barrier to safer healthcare. Unfortunately the origins of this culture can be traced back to some of our best established institutions, such as the General Medical Council and Nursing and Midwifery Council.

Research on safer healthcare systems has proved difficult, and many questions remain unresolved. Theory has been led largely by analogy with disciplines which have apparently addressed safety and reliability problems successfully. Civil aviation has been over-used as a model, much to the annoyance of many surgeons, whereas in terms of stress, risk and a predictably unpredictable environment, surgery may have at least as much in common with military operations. All such analogies have important flaws: reviewing the available research, three main approaches to improving reliability can be identified across a range of high-risk environments: Interventions to improve co-operation and communication between team members; systems rationalisation methods to drive out error by designing low-risk standardised processes; and technological solutions where machines and computers are integrated into work processes to eliminate failure due to predictable human error. A fourth strand of research centres on leadership, but the nature of this subject makes valid scientific conclusions especially difficult.

How surgery addresses these issues in the future will have important implications for the future role and ethical outlook of the surgeon. Training for surgeons already pays considerable attention to communication and teamwork, but we are still, in the main, not required to use the formal systems of low-risk reliable communication which are insisted on in both military and aviation settings – consider how theatre teams currently communicate, compared to pilots and soldiers. This will gradually change, and protocol-driven surgery will become the norm. This is already the case in some high volume specialities such as cardiac surgery and orthopaedics, but the use of explicit Standard Operating Procedures and checklists will become more extensive in all types of surgery to cut down the risk of error. Adherence to these will become as much a professional standard as scrubbing up properly. Smart robots and imaging systems linked to AI will increasingly be capable of intervening to further decrease risk. In a few years your laparoscope may flash warning queries on the screen if its interpretation of the anatomy does not align with yours. Experts in AI consider that a truly autonomous surgical robot, capable of recognition of anatomy, dissection and adaptive decision making in real time is probably impossible, unless and until quantum computing becomes a reality. However subroutines such as abdominal or skin closure may become automated in next-generation robotic surgery, just as many modern cars can now park themselves.

Reliability will be enhanced by the vastly increased analytic power of linked electronic records systems, which will make performance audit simple and instantaneous. Surgeons will become accustomed to regular updates and alerts based on sophisticated risk-adjusted analysis of outcomes. Simulation will become the preferred method for achieving technical proficiency in new procedures, and trainees will not be permitted to operate live until they have demonstrated a learning curve plateau in the simulator. Simulators may also be used for refresher training if results suggest performance is slipping in established teams, although a systems-based analysis is likely to prescribe a whole-team simulation rather than focusing on the surgeon alone.

Healthcare has yet to embrace continuous quality improvement as an integral part of hospital management, but economic pressures will force it to do so. Most quality improvement in the NHS is currently done either by external consultants or enthusiastic amateurs, but Trusts will increasingly invest in professional-level skills in Human Factors and Quality Improvement. Surgeons will have to learn some of these skills if they are to retain any control over their work, and they will become a standard part of the surgical training curriculum. Many highly reliable organisations allow a significant percentage of time for improvement activity in all employee job plans. This will be difficult in the pressured world of NHS surgery, but incentives to participate may be introduced, such as linking activity to training requirements or appraisal.

A critical issue for surgery will be whether the surgeon remains the person in overall charge of patient care for surgical patients. The principles of high reliability performance do not necessarily support this. Responsibility for care is already shared with a range of professionals in areas like pain relief, nutrition, infection control, rehabilitation etc. In trauma, orthopaedics and surgical emergency departments generalist physicians already play a vital role in pre and postoperative care. Given the economic incentive to maximise the use of the surgeon's unique skills in the operating theatre, there is a risk that the surgeon might become merely another advisor, with special expertise in drains, wounds and complications. Someone needs to orchestrate the team: but why should it be the surgeon? Perhaps in future surgeons will pursue a portfolio career, progressively stepping back from intensive operating earlier in their careers to become co-ordinators, decision makers and counsellors for patients facing difficult decisions. However this may well require some retraining to ensure they have an adequate grasp of the facets of care which are now led by other professionals.

Appropriate investigation of adverse outcomes and incidents of patient harm will be vital to the smooth functioning of high-reliability surgical systems, and for that reason it is likely to become more prevalent. Current systems for investigation in healthcare compare poorly with those in other high-risk environments such as the energy and transport industries. Political pressures and medico-legal concerns have led to considerable unnecessary over-investigation, whilst most public healthcare systems are under too much pressure to facilitate beneficial learning and systems change from investigations. Investigators lack important skill sets, especially Human Factors and Systems Analysis expertise, and the independent status of NHS Trusts encourages internal investigation which is very often affected by significant conflicts of interest. In the near future, protocols and networks for Human Factors based impartial external investigation will be introduced, associated with investment in training for the personnel involved with the process. These investigators will make recommendations as to whether any staff should be referred to their professional bodies or disciplined, and the frequency with which this occurs will likely decline, hopefully beginning to eradicate the climate of fear. NHS investigations may require protection from legal discovery, similar to that enjoyed by maritime and aviation investigations. The role of the regulatory Councils will change significantly, if they survive, and they will be required to take into account a systems-based analysis in deciding on the culpability of individuals. Criminal prosecutions against clinical staff will be confined to cases of extreme recklessness or deliberate malfeasance, but prosecutions against Trusts and institutions may well rise. Such cases may need to be held before specialist courts such as those involved in complex fraud prosecutions, to ensure that juries, if they are involved at all, are able to properly weigh the evidence presented to them.

Table 1. Summary of the IDEAL Framework and Recommendations 2017

	IDEAL Framework (Description of stage of evolution of intervention)	**IDEAL Recommendations (For stage-specific study design and reporting)**
Stage 1 **Idea** *First in human*	**Purpose:** Proof of concept **Number & Types of Patients:** Single digit; highly selective. **Number & Types of Surgeons:** Very few; innovators[@] **Output:** Description of intervention or procedure **Status of Intervention:** Evolving; at inception stage **Reporting Methods:** Structured case reports **Outcomes Reported:** Proof of concept; technical achievement; dramatic success; adverse events, surgeon views of the procedure **Stage Endpoint: Once a decision is made to conduct a series of cases, i.e. to proceed to stage 2a.**	• Provide full details of patient selection, technique and outcomes and of patients not selected during the time frame, and why. • Use standard well-defined measures for reporting outcome and patient characteristics • Use a structured reporting system eg, SCARE checklist. • Make the above information available to peers regardless of whether outcome is favourable or not. • Informed consent should clearly explain status of procedure and impossibility of quantifying risks
Stage 2a **Development** Iterative modification of the intervention until a stable version is achieved. Design: Single centre/single intervention; case series/prospective cohort	**Purpose**: Development of procedure **Number & Types of Patients:** Few; Selected **Number & Types of Surgeons:** Few; innovators and some early adopters **Output:** Technical description of procedure and its development, with explanation of reasons for changes **Intervention:** Evolving; procedure development towards a stable optimised version. **Methods:** Prospective development studies (small prospective cohort studies) **Outcomes:** Mainly safety, technical and procedural success. **Stage Endpoint: When the procedure is considered optimised, and stable enough to allow replication in Stage**	• Make protocol for study available • Use standard well-defined measures for reporting outcome and patient characteristics • Report and explain all exclusions • Report all cases sequentially with annotation and explanation of changes to indication or procedure, and when and why they took place. • Display main outcomes graphically showing cases sequentially to illustrate the above. • Informed consent should explain current status of intervention and consequent uncertainties around risk[**]

	2b. There should be no intent at this point to make further major modifications.	

Stage 2b Exploration
 	
Collaborative prospective data collection and analysis aimed at achieving consensus on key issues, to determine if an RCT is feasible, and to define its design features. Intended as a bridge from observational to comparative evaluation.
	Purpose: Achieving consensus between surgeons and centres on the parameters for an RCT (if possible)
	Number & Types of Patients: Many; broadening indication to include all potential beneficiaries
	Number & Types of Surgeons: Many; innovators, early adopters, early majority[@]
	Outputs: Effect estimate for the intervention based on a large sample, allowing power calculations; Analysis of learning curves; estimate of influence of pre-specified technical variants and patient subgroups on outcome. Qualitative research to determine operator and patient values; Increased mutual confidence amongst operators.
	Intervention: The procedure is stable in individual hands but variation exists between operators; acceptable variants are subsequently defined by analysis of pooled results
	Method: Prospective multi-centre exploration cohort study or pilot/feasibility multicentre RCTs.
	Outcomes: Safety; clinical outcomes (specific/graded); short-term outcomes; patient
		• Make protocol for study available
		• Use standard well-defined measures for reporting outcome and patient characteristics
		• Participate in collaborative multi-centre co-operative data collection, incorporating feasibility issues such as:
		  ○ estimating effect size,
		  ○ defining intervention quality standards,
		  ○ evaluating learning curves,
		  ○ exploring subgroup differences,
		  ○ eliciting key stakeholder values and preferences,
		  ○ analysis of adverse events:
		• Hold a pre-planned consensus meeting prior to progressing to an RCT, to identify feasibility and ability to recruit, operator eligibility on basis of learning curve analysis, intervention and comparator definitions, appropriate patient selection criteria, primary endpoint.

	centred/reported outcomes; feasibility outcomes **Stage Endpoints: Demonstrate that technique can be more widely adopted; and, Demonstrate that progression to RCT is desirable and feasible**	
Stage 3 **Assessment** Definitive comparative evaluation of main efficacy and safety aspects of new technique against current best treatment.	**Purpose:** Comparative effectiveness testing **Number & Types of Patients**: Many; expanded indications (well-defined) **Number & Types of Surgeons:** Many; early majority[@] **Output:** Comparison with current standard therapy **Intervention:** Stable, with acceptable variations clearly defined **Method:** RCT with or without additions/modifications; alternative designs (cluster, preference RCTs, stepped wedge, adaptive designs) **Outcomes**: Clinical outcomes (specific and graded); potentially Patient Reported outcomes , Health Economic outcomes **Stage Endpoints: Clear valid evidence on relative effectiveness of innovation; and, Identification of issues requiring long term monitoring.**	• Register on an appropriate international register (e.g., clinicaltrials.gov) • Use standard well-defined measures for reporting outcome and patient characteristics • Incorporate information about patient and clinician values and preferences in design of consent information and procedures, and outcome measures. • Adhere to Reporting guidelines: CONSORT update of 2010 with extension for non-pharmacological treatments COMET TIDieR SPIRIT (for RCT protocol design)
Stage 4 **Long term monitoring**	**Purpose:** Surveillance **Number & Types of Patients**: All eligible **Number & Types of Surgeons:** All eligible **Output:** Description; audit; recording of regional and local variations; quality assurance; risk adjustment; detection of indication creep **Intervention:** Stable **Method:** Registry; routine database; rare-case reports	• Registries may begin from the earliest stages of human use • Registry datasets should be defined by the clinical community with patient input • Datasets should be simple, cheap and easy to collect • Curation of registries by clinical community is desirable • Funding of registries should be agreed between government and commercial

	Outcomes: Rare events; long-term outcomes; quality assurance **Stage Endpoints: dependent on lifecycle of device/procedure.** **Registries for devices – IDEAL-D** **Registries at earlier stages of IDEAL**	interests but kept separate from curation • Patient Consent for use of registry data in research should be broad and where possible automatic

@ Terms used under this heading refer to the classification of Everett Rogers (Diffusion of Innovations, 4th Ed, 1995)
*Registries should be organised according to the IDEAL recommendations and should be available for enrolment at *any Stage*
**Patient consent should always be informed by a summary of the outcomes from previous IDEAL Stages