The IDEAL framework for Surgical Robotics: Development, Comparative Evaluation, and Long-term Monitoring

Suggested Authorship

Hani J Marcus[1,2,†], Pedro T Ramirez[3], Danyal Z Khan[1,2], Hugo Layard Horsfall[1,2], John G Hanrahan[1,2], Simon C Williams[1,2], David J Beard[4], Rani Bhat[5], Ken Catchpole[6], Andrew Cook[7], Katrina Hutchison[8], Janet Martin[9], Tom Melvin[10], Danail Stoyanov[2], Maroeska Rovers[11], Nicholas Raison[12], Prokar Dasgupta[13], David Noonan[14], Deborah Stocken[15], Georgia Sturt[16], Anne Vanhoestenberghe[17], Baptiste Vasey[18], Peter McCulloch[19,‡], on behalf of the IDEAL Robotics Colloquium.

Affiliations:

[†]h.marcus@ucl.ac.uk

[‡]peter.mcculloch@nds.ox.ac.uk

1. Department of Neurosurgery, National Hospital of Neurology and Neurosurgery, London, UK
2. Wellcome/Engineering and Physical Sciences Research Council (EPSRC) Centre for Interventional and Surgical Sciences (WEISS), London, UK
3. Department of Obstetrics and Gynaecology, Houston Methodist Hospital Neal Cancer Center, USA
4. RCS Surgical Interventional Trials Unit (SITU) & Robotic and Digital Surgery Initiative (RADAR), Nuffield Dept Orthopaedics, Rheumatology and Musculo-skeletal Sciences, University of Oxford, Oxford, UK
5. Department of Gynaecological Oncology, Apollo Hospital, Bengaluru, India
6. Department of Anaesthesia and Perioperative Medicine, Medical University of South Carolina, Charleston, SC, USA
7. NIHR Coordinating Centre and Clinical Trials Unit, University of Southampton, Southampton, UK
8. Department of Philosophy, Macquarie University, Sydney, Australia
9. University of Western Ontario, Ontario, Canada
10. Department of Medical Gerontology, School of Medicine, Trinity College Dublin, Republic of Ireland
11. Department of Medical Imaging, Radboudumc, Nijmegen, The Netherlands
12. Department of Biomedical Engineering and Imaging Sciences, King's College London, UK
13. King's Health Partners Academic Surgery, King's College London, London, UK
14. Moon Surgical, Paris, France
15. RCSEng Surgical Trials Centre, Leeds Institute of Clinical Trials Research, University of Leeds, Leeds, UK
16. Bowel Research UK, London, UK
17. School of Biomedical Engineering & Imaging Sciences, King's College London
18. Department of Surgery, Geneva University Hospital, Geneva, Switzerland
19. Nuffield Department of Surgical Sciences, University of Oxford, John Radcliffe Hospital, Oxford, UK

Counts:

Abstract: 149 words
Manuscript: 8133 words
Figures: 3
Boxes: 3
References: 112

Running head:

IDEAL Framework for Surgical Robotics

Keywords:

Robot-assisted surgery; surgical robots; robotic; surgery; IDEAL; evaluation; artificial intelligence; framework; innovation

Overview of contents

Abstract

The next generation of surgical robotics is poised to disrupt healthcare systems worldwide, requiring new frameworks for evaluation. However, evaluation during a surgical robot's development is challenging due to their complex evolving nature, potential for wider system disruption, and integration with complementary technologies like artificial intelligence. Comparative clinical studies require attention to intervention context, learning curves, and standardized outcomes. Long-term monitoring needs to transition towards collaborative, transparent, and inclusive consortiums for real-world data collection. Here, the Idea, Development, Exploration, Assessment, and Long-term monitoring (IDEAL) Robotics colloquium proposes recommendations for evaluation during development, comparative study, and clinical monitoring of surgical robots — providing practical recommendations for developers, clinicians, patients, and healthcare systems. Multiple perspectives are considered, including economics, surgical training, human factors, ethics, patient perspectives and sustainability. Further work is needed on standardized metrics, health economic assessment models, and global applicability of recommendations.

1. Introduction

Surgical robots may be on the brink of achieving their fundamentally disruptive potential.[1] Since the first surgical robot was introduced in 1985 (the PUMA560, tasked with performing a CT-guided brain biopsy), the field of robotic surgery has expanded in size and scope, offering the potential for enhanced surgical precision, telesurgery and increasingly complex autonomous function. Technological advances in robotic control systems and artificial intelligence (AI) make it likely that the next generation of surgical robots will transform the surgical technology landscape, previously monopolised by a limited number of approved devices such as Intuitive's Da Vinci.[1]

This proliferation of robotic platforms poses important problems for their safe and ethical clinical translation.[1–3] extending beyond the operating room and encompassing wider considerations within healthcare and society.[4,5] The scope of the evaluation challenge is too broad for existing methodological templates[2,6], but current circumstances create a brief window of opportunity to develop a structured framework capable of guiding the evaluation of surgical robots across their development and translation.[7]

Conducting high-quality surgical research is difficult owing to the nature of surgical innovation[REF] leading historically to a methodologically weak approach. Specific problems have included a lack of robust early stage studies addressing the need for transparent reporting of iterative development during its' progress, and subsequent comparative studies which failed to address variations in surgical technique and indications, operator learning curves and lack of equipoise.[1,2] Evaluating surgical robots is subject to all of these challenges, but adds the need to consider unique ethical considerations, profound questions about economic value and sustainability, major impacts on the host healthcare system, and the increasing integration of AI into robotic systems.[8]

The Idea, Development, Exploration, Assessment, and Long-term monitoring (IDEAL) Framework provides a structured evaluation pathway for surgical innovation and devices, from needs analysis and preclinical testing, to long-term studies of widespread use.[9–11] However, the breadth of the evaluation problem of surgical robotics goes beyond both IDEAL and the boundaries of classical evidence-based medicine, with solutions requiring a diverse array of stakeholders to tackle all the aspects which need consideration. The IDEAL Robotics Colloquium was established to make proposals for a comprehensive practical guide for evaluation of surgical robots, using the existing IDEAL study stages as a template (Figure 1 & 2).

In this paper, we present a systematic analysis of the evaluation life cycle of surgical robots, in three parts. First, we dissect the pre-clinical and early clinical study of the safety and feasibility of new robotic concepts (IDEAL Stages 0, 1, & 2a). Next, we review the pivotal phase when the effectiveness of robotic interventions is studied on a larger scale, and compared against current best practice (IDEAL Stages 2b & 3). Finally, we consider IDEAL Stage 4, when the robot has been widely adopted, shifting focus to long-term monitoring of performance in real-world settings. This analysis results in a list of stage specific recommendations for systematic evaluation of robots in surgery.

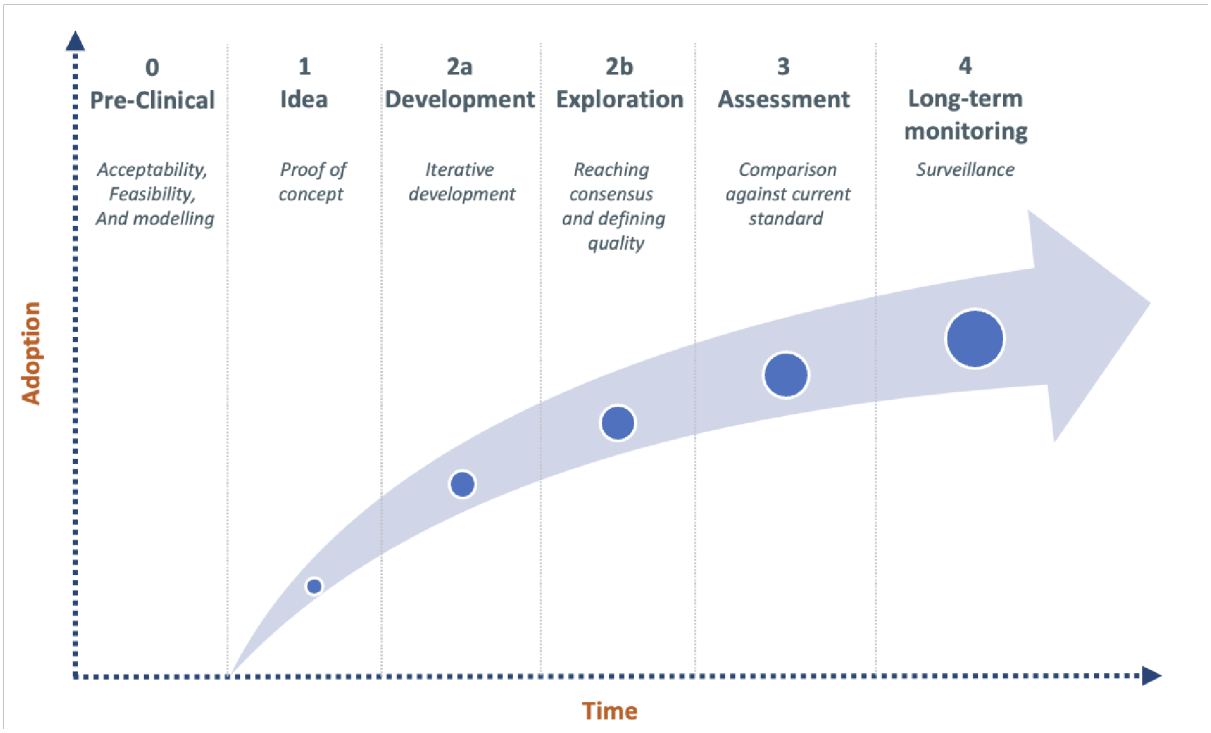

Figure 1: Current IDEAL framework with example study types. The IDEAL framework provides an evaluative pathway for complex innovations and devices, spanning the entire life-cycle, from early adoption to widespread use. IDEAL Stages 0, 1, & 2a explore the safety and feasibility of an intervention; IDEAL Stages 2b & 3 compare the intervention against the current standard to determine effectiveness; and IDEAL Stage 4 involves the long-term monitoring of interventions following widespread uptake and adoption. Adapted from Roodbeen et al., 2019.[12]

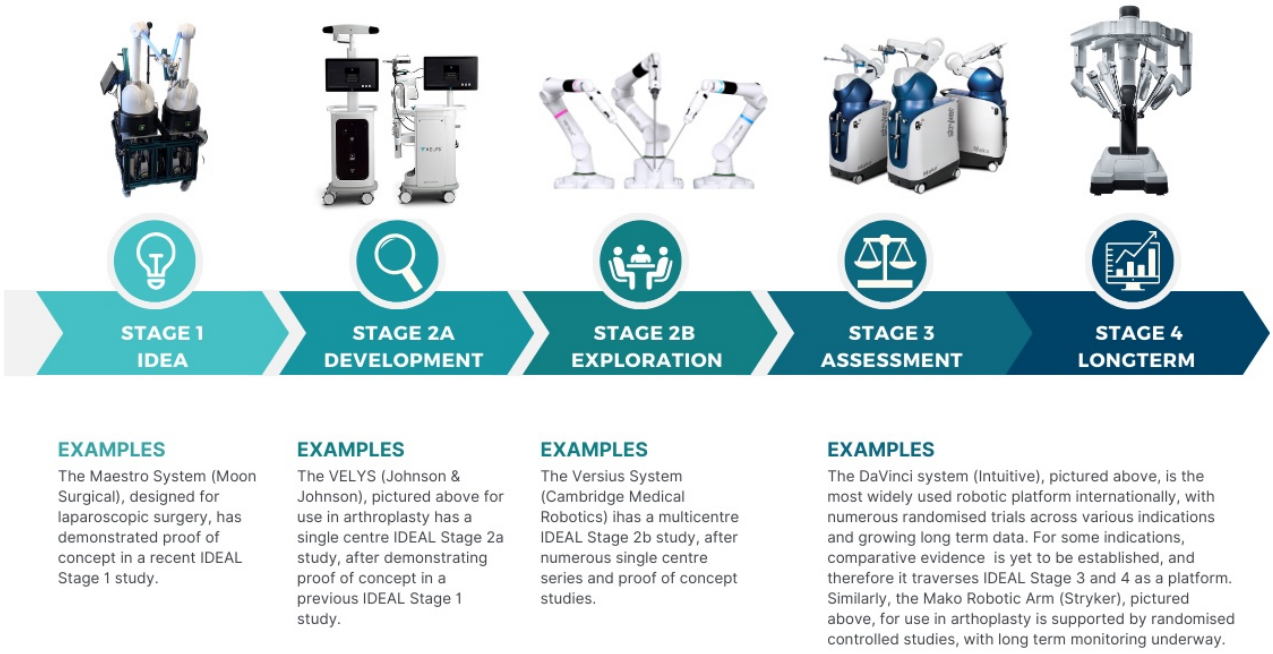

Figure 2: Examples of Current Robotic Systems Across IDEAL Stages of Evaluation. Examples chosen are purely illustrative, representing an array of systems across a variety of specialities. Stage 3 and 4 were combined to reflect how many robots may have comparative evidence (via a randomised control trial) with long term monitoring data for particular indications, but still require further comparative evidence for use in other indications (see Appendix A for a summary of the literature used to determine the IDEAL stage of the selected examples).

2. Methodology

An international interdisciplinary consensus process was completed in several stages. Firstly, eight distinct virtual panels with expertise relevant to important aspects of the challenges to robotic surgery evaluation were devised by the three lead authors (HJM, PTR, PMC). These panels considered: technical evaluation, clinical evaluation, artificial intelligence (AI), human factors, ethics, health economics, surgical training, and patient perspectives.

Panel leaders with relevant expertise were selected from the IDEAL council, and asked to invite 8-12 experts from multiple disciplines (including surgeons, engineers, economists, statisticians, device regulators, patient representatives, ethicists, digital health experts, patient safety experts, system engineers, social scientists, philosophers, and education experts) to join their respective panels. Experts from diverse professional and geographical backgrounds were invited, and were chosen based on leadership roles in relevant organisations (university, hospital, societal and industrial) and/or accomplishments relevant to robotic surgery development and evaluation. Patient representatives were included in each panel. The recruitment and facilitation of these panels and the general strategy for their function were developed in partnership with the Royal College of Surgeons of England and the National Institute for Health and Care Research, and considered the views of an industry roundtable. To avoid bias by association, one co-author conducted MEDLINE searches for publications relevant to each panel, and identified additional potential members, who were invited to join the panel, ensuring that each panel had at least one such member..

Each panel participated in a series of semi-structured virtual meetings chaired by respective panel leaders discussing the key challenges for each panel domain, whether the current IDEAL framework addressed these challenges, and recommendations to address these challenges. Each panel therefore produced a report across each stage of the IDEAL framework to summarise these meetings. Panel reports were then synthesised by a second international diverse team that included experts in sustainability, global health, device regulation and medical statistics - independent from the colloquium panels. A full list of authors, panel members and industry collaborators is found in the footnote. To improve usability, the final recommendations were considered from the perspective of key surgical robotics stakeholders: the device developer, clinician, patient, and wider healthcare ecosystems (Figure 3).[13]

Recommendations according to each of these perspectives were grouped together for IDEAL Stages 0, 1 and 2a, which cover preclinical development and early clinical evaluation; for IDEAL Stages 2b & 3, which cover comparative assessment; and for IDEAL Stage 4, which covers long-term monitoring and technological evolution.

The IDEAL Recommendations are based on three principles: the use of the most rigorous and appropriate methodology to address the key questions at each stage in the intervention's life-cycle; adherence to the fundamental principles of medical ethics (beneficence, non-maleficence, autonomy and justice); and maximum feasible transparency in reporting evaluation outcomes. These principles have allowed the development of coherent proposals for evaluation across a very broad range of complex therapeutic interventions, but they inevitably lead to some recommendations which may not be feasible in many current contexts. In reporting our

recommendations, we have indicated our recognition of this by prefacing certain recommendations with “in principle”, or by qualifying them by explicitly mentioning their conditional feasibility.

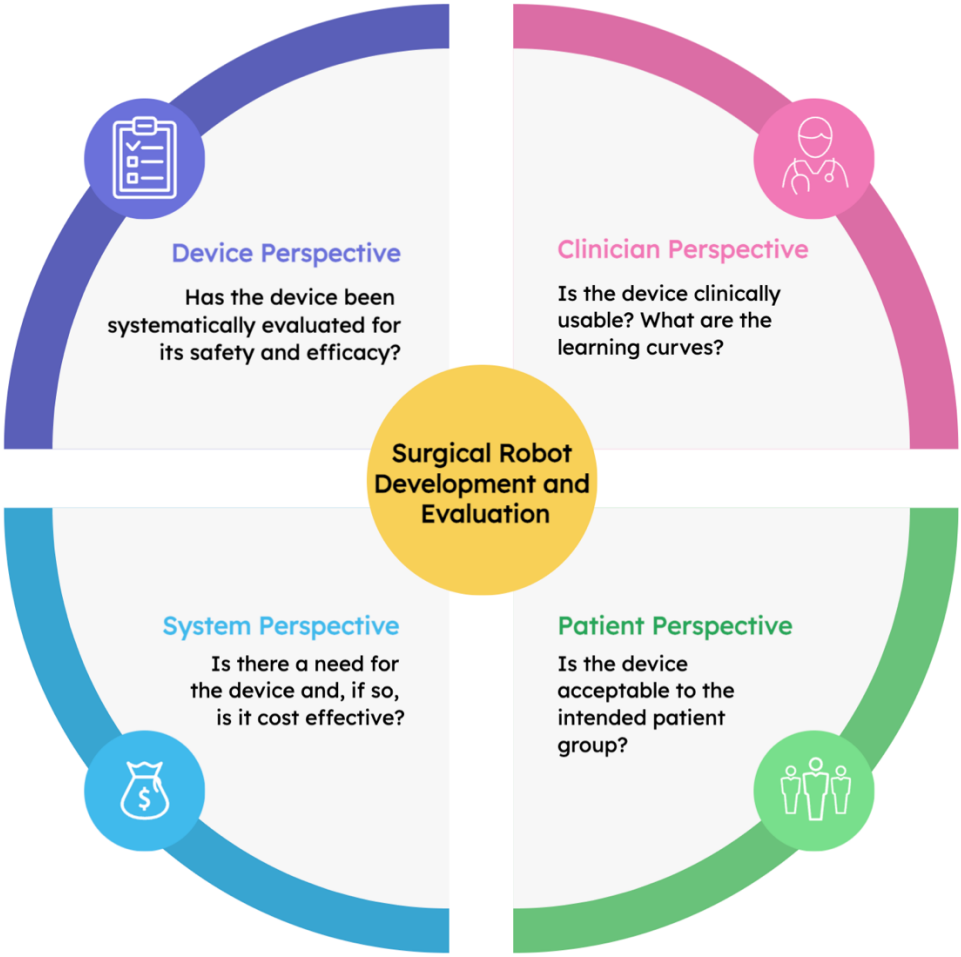

Figure 3: Key stakeholders in the development and evaluation of surgical devices such as surgical robots. Consideration of the key stakeholders is essential to the successful introduction of innovative devices.

3. Preclinical Development and Early Clinical Evaluation (IDEAL Stages 0, 1 & 2a)

An innovative device must first be deemed safe, feasible and acceptable for its successful translation, which is encompassed in the preclinical evaluation (IDEAL Stage 0), first-in-human study (IDEAL Stage 1) and prospective development (IDEAL Stage 2a). In these stages preclinical studies to assess safety and feasibility are followed by a first-in-human use and then iterative development ahead of further collaborative evaluation and comparative assessment. Studies in this phase currently suffer from design flaws, severe reporting bias and methodological heterogeneity, which IDEAL aims to reduce.[13] This stage also commonly encompasses critical progression points such as regulatory approval and financing. The key challenges of this early developmental stage are considered below.

BOX 1

Summary of key recommendations across IDEAL Stages 0, 1, and 2a

Here, we present a summary of the key recommendations for the evaluation of surgical robots in the pre-clinical and early clinical evaluation stages, from the perspectives of four key stakeholders: Device, Clinician, Patient and System.

Device Perspective

- Standardise the publication of technical and clinical data.
- Transparently document changes to devices, indications, patients, and AI models.
- AI-integrated robot evaluation should initially examine AI facets separately, followed by in-silico & simulator based assessment of the integrated robot (IDEAL Stage 0). First in human studies (IDEAL Stage 1) should assess the integrated robot in a clinical context, using clinical outcomes, guided by reporting guidelines (e.g., DECIDE-AI).
- Evaluate robotic autonomy based on level and risk.

Clinician

- Define, analyse, iterate clinician-device integration accounting for stakeholder perspectives, clinician behaviour, and cognitive workload.
- For autonomous systems, evaluate the reliability of handover mechanisms and reasons for human takeover.

Patient

- Ensure transparent consent processes regarding theoretical risks, evidence, system failure mitigation, autonomy level, surgical team experience, and potential conflicts of interest.

System

- Perform early and iterative economic modelling, using exploratory analyses to guide cost-effective development and prevent future research wastage.
- Consider the impact of surgical robots on different healthcare ecosystems, using life cycle assessments, reverse engineering, and frugal design concepts where possible to improve accessibility and sustainability.

3.1 Device Perspective: in IDEAL Stages 0 – 2a

3.1.1 Key Challenges

The complex and rapidly evolving nature of surgical robots poses unique challenges to assessing their safety and effectiveness.[3] Current assessment domains are usually driven by regulatory requirements. In the United Kingdom and European Union, this requires a demonstration of overall safety and performance; in the United States, a reasonable assurance of safety and effectiveness is required.[14,15] However, the implementation of these requirements varies amongst national regulators, and is subject to complex procedural rules and variable decision-making both within and between bodies. Requirements are also influenced by wider geopolitical, economic and legal factors.[16–20] Although international harmonised standards exist, they focus on technical aspects of device assessment, such as software or electrical safety assessments, rather than clinical metrics.[16–19] The nature and quality of scientific evidence developed for device safety, performance and effectiveness may therefore be vastly different for similar systems, being largely defined, verified, and validated internally by each company. Without recording of iterative systematic modification and assessment, key domains may be overlooked, particularly during prototyping and when changes are made during early clinical studies.

As complementary technologies develop they will increasingly be integrated into surgical robotic systems.[21] The most impactful of these technologies will likely be AI – significantly boosting function and increasing the autonomy of robotic systems through integration of sensory inputs, learned computational reasoning and

adaptive behaviour.[7,22–25] However, autonomous systems have no "common sense" and so would not necessarily stop an obviously unsafe action if a specific scenario had not been learned. The integration of AI also adds a further layer of complexity to device development, calibration and evaluation.[7,22–25] AI integrated functions have the potential for rapid self-updating, requiring monitoring and understanding of risk and failure modes. Isolating the "moving" AI components of the robotic system for assessment may be difficult, and assessment frameworks need to address this problem. A recent review on intraoperative AI applications for robotic surgery found that all identified publications reported on pre-clinical development only, and were heterogenous in their evaluation approach, highlighting the need for a robust evaluation framework for early integration of AI into clinical practice.[7]

3.1.2 Recommendations

With these challenges in mind, this Colloquium proposes the following recommendations for early stage evaluation of surgical robotics from a device perspective. When assessing the performance of a robotic system, technical metrics alone are acceptable in earlier studies (Stage 0), however a clinical outcomes-based approach should be used as the primary focus of assessment as early as is feasible.[39]

For early technical assessment of robotic systems, a standardised checklist should be used to summarise performance, safety, and usability for each released version. Assessment should be systematic and transparent, including details of system latency, motion accuracy, instrument safety, operation under load, reliability, internal fault recognition, and online security. Metrics and measurement instruments for each of these domains require further definition. For each domain, performance benchmarks and areas of concern should be clearly stated and shared.

Building on the IDEAL-D pre-clinical device assessment approach of relative risk assessment, the proportionate evaluation of autonomous surgical robots should be guided by its classification along two main axes (autonomy level and risk), before proceeding to clinical studies.[13,22] Autonomy levels are as described by Yang et al.: no autonomy; robot assistance; task autonomy; conditional autonomy; high autonomy; and full autonomy. The pre-clinical evidence requirements should be guided by a Failure Modes and Effects Analysis (FMEA) approach to risk stratification, based on the likelihood and severity of device failure in each cell of the risk/autonomy matrix. Therefore, prior to clinical study, a high-risk, full-autonomy device would require more extensive pre-clinical evidence than a low-risk device with no or low (i.e. task only) autonomy. [13,22]

For the evaluation of AI-integrated robots, the pre-clinical (Stage 0) testing should begin with standalone evaluation of the autonomous component and hardware separately, followed by in-silico and simulator-based assessment of the two integrated into a functional unit in realistic tasks. Later stages (Stage 1 and beyond) should study the performance of the AI algorithms within systems (with the hardware components of that system version) in a clinical context – using clinical outcomes where feasible. Reporting guidelines, such as DECIDE-AI, should be used to guide early clinical evaluation.[40]

The maturation of the system from *in silico* to *in vitro* and *in vivo* versions should revolve around addressing identified clinical unmet needs and should be described with clear identification of the prototype version. This should include documentation of iterative changes to the procedure, device, and patient selection, and describe simulation studies in detail. This information should be recorded prospectively, and a log should be accessible to regulators. In systems with AI integration, the AI component is particularly susceptible to rapid iterations, and therefore changes to input data, algorithm code and model testing should be reported.

3.2 Clinician Perspective in IDEAL Stages 0 – 2a

3.2.1 Key Challenges

From the perspective of clinicians, the introduction of a robotic device within a clinical team is a multifaceted challenge. Investigation of robot interaction with humans (i.e. the surgical team) is crucial, particularly in the domains of usability, trust and failure analysis.[21,26–28] This is particularly pertinent with respect to the integration of AI, which could alter responsibility and liability paradigms.[29] Understanding systems modelling is important, as a device is never integrated into a 'static' system – the act of introducing it will change both the behaviours of the people and the way they think about their work within the operating room. Surgical team trust must also be considered in the evaluation of these systems - especially in systems with an autonomous component which current assessment frameworks do not recognise or evaluate.[22,30] Human factors approaches will be important in developing solutions to these largely unexplored problems. Recent projects such as the Trustworthy Autonomous Systems Hub and Responsible AI UK will aim to bring standardisation and regulation to this rapidly evolving field.

3.2.2 Recommendations

The human-device interface and team-device integration in the operating room should be included in the intervention development and description.

In principle this process starts with robotic development, which should employ user-centred design and involve input from surgical team members.

Robot assessment should include a human factors-based evaluation of team communication (including communication with the robot); intuitiveness of visual displays, control interface usability, feedback mechanisms (e.g., haptic and auditory) and ease of integration with existing workflows.

Human factors assessment of system integration should ideally include directly observed user situational awareness, user workload (mental & physical), task analysis in device use, operational challenges, and potential safety-related issues.[41] Formal qualitative research to study robot user opinion and perceptions may be helpful. The ongoing REINFORCE initiative may provide a framework for this.[42–45]

Human reliability assessments should be used to stratify potential risks and hazards across a wide variety of surgical expertise (i.e., consultants and trainees, those with previous robotic or minimally invasive surgery expertise).[46]

Surgeons' trust in any AI autonomous function and its evolution should be evaluated initially in simulated situations. This should include monitoring for frequency and reasons for taking over control of the robotic system, alongside independent observation, and qualitative assessment. Surgical robotic incorporation of AI poses ethical challenges, including fair distribution of risk and benefits for patients and clinicians. Integration of ethical considerations should occur across key domains for every study stage by addressing the key issues of minimising harm, ensuring autonomy and consent and optimising justice (e.g., in terms of differential access to treatment). Conflicts of interest should be openly addressed.[30]

In principle, a standard process for determining responsibility for errors when AI is integrated should be adopted with suitable expert advice and should be publicly accessible.

3.3 Patient Perspective in IDEAL Stages 0 – 2a

3.3.1 Key Challenges

From a patient perspective, as robotic systems grow in complexity they become increasingly difficult to understand and trust.[31,32] Patients invited to participate in early clinical studies will rarely be able to understand the risks and limitations of the technology, compare these against other treatment options (including other robots), or be aware of potential vested interests of the investigators and healthcare system. The nature of the early IDEAL stages mean patient numbers and operating team experience are limited, resulting in the evaluation of interventions at early stages of the learning curve, with resultant implications for both clinicians and patients.[33] Surgical teams may not know all of the risks, or how learning curves may increase the overall risk relative to that which later patients experience.[33] Provisions to minimise harm and ensure truly informed consent are ethical requirements in this phase of surgical robot evaluation.[30,32,34,35]

3.3.2 Recommendations

Active Patient and Public Involvement is desirable to ensure a patient-centred research design from the outset, and formal qualitative research assessing patient perceptions, understanding of the robotic system, and trust in the intervention may be very informative. Patient information sheets for both research and surgical consent purposes should be developed with patient groups. Crucially, informed consent in early clinical studies (i.e., Stages 1 and 2a) should acknowledge a potentially increased uncertainty of benefit and risk of harm in early cases, as with all new device introductions. Information should include details of previous studies; known risks and the possibility of unknown risks; dependence level on the surgical robot and mitigation plans for system failure; level of AI system autonomy and protocols for the takeover of control; transparency regarding surgical team experience with the system; and any potential conflicts of interest.

3.4 System Perspective in IDEAL Stages 0 – 2a

3.4.1 Key Challenges

When considering the impact of surgical robots in health systems, societal cost must be considered. Currently, health-economic assessments are not standard components of early evaluation frameworks for devices, as illustrated by the lack of ISPOR guidelines (The Professional Society for Health Economics and Outcomes Research) for this stage. Early health economic evaluations are heterogeneous and often unsatisfactory. Economic evaluations at this stage act as exploratory tools to assist decision-making about pursuing further development, and to provide insights into future cost-effectiveness, particularly for complex interventions.[36] The current deficit in early economic evaluation extends to encompass related gaps in the evaluation of the environmental sustainability and global applicability of surgical robots.[37] Early and systematic use of unmet needs analyses, health-economic analyses, and sustainability analyses can and should serve a vital role in guiding the efficient onward development of devices and avoiding waste.[38]

3.4.2 Recommendations

With these complex system challenges in mind, we present the following recommendations: Unmet needs analyses and early economic models should routinely be considered before moving into definitive studies,[47] such as headroom analyses to provide early estimates of cost-effectiveness or economic burden studies to advise on high-priority disease targets. These could pilot metrics for expenditure (including time, money, human resource, and technical resource) and costs of altered downstream care. Iterative exploratory decision-analytic modelling could inform robotic development as part of the early health technology assessment (HTA) process.[36,48]

Value of research studies should identify surgical robots that are unlikely to be successfully implemented into the health system, permitting decision-making on halting research and investment into technologies unlikely to be adopted. Reverse engineering and frugal or alternative surgical robot design (such as handheld platforms) could be explored to reduce cost, improve eventual accessibility across healthcare systems and boost the potential for global health impact.[49,50]

Sustainability metrics should be recorded during pre-clinical (device-only) and clinical (device within system) evaluations.[37] Assessment should integrate a complete life cycle assessment (LCA) model.[51] This includes recording resources required to build, run and maintain each device version, along with device design (e.g., careful material selection, modular system design, reusable parts) from preclinical stages onwards. Interoperability, parts replacement and maintenance by local teams, especially in low resource settings, should be considered at the earliest design stages.

4. Comparative Evaluation (IDEAL Stages 2b & 3)

Once a stable version of an effective and safe robot has been developed, a comparative evaluation with the current surgical standard should follow. Expert consensus is needed on the nature of the patients and procedures to be studied in trials, and on markers of adequate procedure quality, to avoid bias due to learning curves or wide variations in performance. Evidence from collaborative prospective cohort studies (IDEAL Stage 2b) in a range of potentially appropriate settings and indications can provide this, and thereby facilitate definitive randomised comparative studies against an appropriate control group (IDEAL Stage 3).

BOX 2

Summary of key recommendations across IDEAL Stages 2b and 3

Here, we present a summary of the key recommendations for the evaluation of surgical robots in the comparative stage, from the perspectives of four key stakeholders: Device, Clinician, Patient and System.

Device Perspective

- Risks and benefits of surgical robots must be evaluated through prospective data collection using a suitable study design, mutually agreed dataset, appropriate analysis techniques, and assessment of study-specific confounders.
- Robot re-evaluation for alternative indications should be based on risk, autonomy level, and available evidence.

Clinician

- Validated tools and qualitative research should be used to explore human factors.
- The real-world learning curve for surgical robots must be investigated. Metrics should be collected from direct supervision of both real-world and simulator use cases.
- Establish institutional clinical governance policies with consistent specifications on surgeon training, audit, and ethics.

Patient

- Explore robotic surgery acceptability through assessing patient perspectives, understanding, and consent.
- Maintain transparency with participants regarding existing evidence, development stage, conflicts of interest, surgical experience, complications, and alternative treatment.

System

- Economic impact analysis of healthcare costs associated with robotic intervention should be measured in comparative studies, including clinically and system-relevant outcomes over a sufficient length of follow-up.
- Include stakeholders from low-resource settings in modelling capacity, benefit, and risks of robot use, compared against available alternatives.
- Life cycle assessments of surgical robots should be compared to the current gold standard treatment.

Robotic surgery, like most innovative surgical technology, is often introduced without the stepwise testing process routinely used in medical therapeutics.[52] Evaluation of surgical innovation is traditionally through initial small case series documenting feasibility, followed by adoption (which may be fast or slow) based largely on non-comparative retrospective evidence of potential benefits to the patient. Robotics manufacturers engage in active campaigns to promote their products with physicians and directly or indirectly with patients. Uncertainty, desire to improve, and personal biases can lead to innovation without rigorous evaluation, with consequent risks to patient safety. Therefore, frameworks to ensure proper evaluation of patient safety are essential.[21] The importance of adequate comparative evaluation prior to adoption was recently illustrated by the US Food and Drug Administration (FDA) warning against the use of robotic surgery for the treatment of breast and cervical cancers.[53] The recommendation on cervical cancer was based on the results of a prospective randomised trial, and a population-based study comparing open vs minimally invasive surgery (including robotic surgery) showing worse disease-free survival and overall survival in patients who underwent minimally invasive surgery.[54,55] A breakdown of adverse events in robotic surgery recorded by the FDA includes 2,000 events that involved injury to the patient, 17,000 events due to malfunction of specific robots and 294 fatalities.[56] It is not clear how many of these events could have been avoided by more rigorous evaluation at an earlier stage, but it is undeniable that omitting such evaluation reduces our capacity to limit harms.

4.1 Device Perspective in IDEAL Stages 2b & 3

4.1.1 Key challenges

Surgical robots offer great potential technical advantages including improved precision, dexterity, improved ergonomics, and teleoperation, but they demand new or different resources from healthcare systems (e.g., surgical team training, audit, maintenance).[1,57] Few definitive high-quality comparative trials have been published, and from these the evidence of benefits of robot-assisted surgery over comparable minimally invasive surgical approaches has been inconclusive.[54,55,58–60] The literature reveals methodological limitations, such as poor reporting of outcome measures, a lack of agreed core outcomes sets, incomplete efficacy or effectiveness assessment, and variable reporting of safety.[57,61] The rapid evolution of robots poses major problems for evaluation, with newer AI-enabled systems threatening to render current studies outdated before their completion, demanding innovative, iterative evaluation strategies, such as implementation trials. This uncertainty complicates decision-making about when, how and if a definitive randomised clinical trial should be performed within the evaluation cycle of the surgical robot. Some of the technology incorporated in newer surgical robots could itself provide next generation evaluation measures, such as computer vision, a domain of AI applied to operative videos with procedural analytics.[62–65] However, such outcome measures must themselves be robustly validated prior to clinical implementation, and their relation to clinical outcomes fully understood.

4.1.2 Recommendations

The comparative stage poses numerous device-related challenges – this colloquium proposes the following recommendations for surgical robot evaluation in the comparative stage. The benefits and risks of a surgical robot should be documented through well-designed prospective evaluations capturing clearly defined safety and effectiveness outcomes, including patient-reported outcomes, relevant to a given procedure, surgical speciality or patient population. These studies should proceed in a stepwise fashion according to the IDEAL recommendations, considering and adopting seamless designs for efficiency, where plausible.

Measured outcomes must include well-defined clinical outcomes (ideally from existing consensus core outcome sets), technical outcomes (including those derived from robotic kinematic and haptic sensors), patient-reported outcomes (such as quality of life indicators) and wider outcomes which reflect potential robotic disruption (ergonomic benefits, impacts on accessibility to surgery) where relevant. Next-generation outcomes and measures, such as those derived from robotic kinematic, haptic sensors and video data, should be reported where relevant, but should be robustly validated and their associations with clinical outcomes determined.

Randomised controlled trials will serve as the default choice for thorough comparative studies of robotic surgery where preliminary studies suggest a potentially important clinical or economic benefit. Planned prospective implementation trials should be considered only where randomised trials are considered impossible. However, for procedures where a robot system has previously established its superiority over non-robotic surgery in technically similar contexts, and no significant change in the level of risk is expected, further randomised trials for every new procedure may be unnecessary. In this situation, an IDEAL 2b study, a prospective implementation trial or a

prospective registry is ethically necessary to ensure that a meaningful evaluation of effectiveness and safety is performed as indicated by existing decision-support algorithms.[88]

In principle, public pre-registration or registration of protocols and analytic intent is recommended for all studies, with any post hoc changes recorded. Protocols should specify defined data dictionaries, data recording by independent observers, with independent validation, and calculations of interobserver reliability. Data collection and analysis should be sensitive to, and protected against, conflicts of interest and related biases. The privacy and security implications of capturing, storing, and using data from robotic devices should also be considered.

During evaluation, changes to the technology or procedure may result in unexpected outcomes which could warrant re-evaluation at the current or an earlier IDEAL stage. Thresholds for this kind of action should be established in advance, considering trends in outcome data suggesting changes in risk levels, indications for use, or device performance. As a guiding principle, major changes in risk should warrant a return to earlier IDEAL stages. An independent expert panel should be involved and work with regulators in making these decisions, including which IDEAL Stage study is required.

Where a robotic system can perform a procedure which achieves a physiological, clinical, or functional effect which was not previously possible, there may be no reasonable comparator. Independent ethical advice should be sought to determine whether control groups for a randomised trial are acceptable, depending on the nature of the presumed benefits and anticipated risks of the procedure and the outcome data available. Where the clinical outcomes of the novel robotic approach are clearly unachievable by other means, randomisation of participants may be unethical. Alternate designs to study effectiveness and safety, where possible, should be sought.

4.2 Clinician Perspective in IDEAL Stages 2b & 3

4.2.1 Key challenges

Human factors and ergonomics analysis is crucial during the clinical translation of surgical robots to ensure they are usable, and can efficiently integrate into complex teams and workflows.[28,66,67] Concerns about the occupational consequences of surgery has led to an interest in ergonomic innovation in surgery and is a purported benefit of surgical robotics, but the evidence base is conflicted and of limited quality.[67] As surgeons gain experience with the robot, their operative skills are expected to improve, described by a learning curve.[68] Surgeon experience and learning curves are an important source of potential variation and bias in comparative surgical robot trials, with high-quality trials incorporating their effects into analysis.[69–73] A reliable measure of the learning curve can only be achieved by analysis of meaningful measures of operation quality (Kirkpatrick Level 3) and patient outcomes (Kirkpatrick Level 4).[74] Learning curve evaluation is important for fair comparative analysis, and for planning and implementing training programs for the surgical team.[27,75,76] Effective, standardised team training is essential for comparative evaluation and clinical translation, but there is no consensus on developing mandated training programme requirements.[68,75,77,78]

4.2.2 Recommendations

In light of these challenges, we propose the following recommendations. Human factors and behaviour change scientists should be included in the evaluation of surgical robots to examine hypotheses generated in earlier IDEAL stages, evaluating features such as workflow, variations in system use, ergonomic risk assessment, data collection capabilities of the device, teamwork, non-technical skills and workspace analysis.

Analysing learning curves is essential in evaluating new technologies, including surgical robots. Large prospective cohorts (IDEAL 2b studies) offer the first opportunity to capture real-world learning curves for surgical robots, and should be used to study their complexity and improve our tools for evaluating them. Metrics gathered from direct supervision, objectively defined criteria, cadaver laboratories, and simulator training should be standardised and used for assessment of real-world learning curves. The performance plateau should be continuously monitored to detect changes over time, studying the effects of factors that may influence surgical performance such as casemix, team changes and changes in the surgical environment. Criteria for the minimal acceptable level of plateau performance should be agreed for surgeons to practice independently or take part in definitive pivotal comparative studies with the robot, using objective measures of procedural quality. Statistical exploration of learning effects (such as sensitivity analysis or extensions of the primary analysis) should be included in trial protocols to identify and adjust for learning curve bias. Training mechanisms should be audited for impact, and iteratively improved to meet user needs.[42] Programmes should directly attempt to track the learning curves seen in surgical robotics training and investigate techniques such as mentoring approaches to shorten them or minimise any effect on patients. Training courses should be validated by evidence of correlation between course evaluations and clinical performance. Institutional clinical governance policies should require the development and use of consistent criteria pertaining to surgeon training and outcomes to monitor continued learning.

In the case of autonomous systems, learning curves will likely be linked to the evolution of trust in the AI application. Proxies for clinicians' trust in the autonomous components (such as instances of use or if manual override is required) should be studied and presented with learning curve analysis.

4.3 Patient Perspective in IDEAL Stages 2b & 3

4.3.1 Key challenges

Patient acceptability is increasingly important when implementing healthcare interventions, but this is difficult to define or assess in relation to surgical robots.[79,80] Acceptability is important for IDEAL Stage 2b/3 studies as patients must provide fully informed consent to studies prior to enrolment. Very few patients have a comprehensive understanding of what a surgical robot is or does, of the current evidence about the potential and proven risks and benefits of a surgical robot, or of the degree of autonomy of robots during surgery.[81] Patient perceptions of likely benefit or harm may be affected by media "spin" or by industry and marketing psychology.[33] This may affect patient preference for one treatment over another.[31] and this could contribute to the challenges of randomisation or trial recruitment. Therefore, it is important that patients are provided with a clear, accurate non-technical explanation of the evidence on the established benefits, known risks, and gaps in knowledge about robotic surgery in their specific context, and protected from potential bias from developers and robotic enthusiasts.

4.3.2 Recommendations

Although no universal definition of patient acceptability exists for surgical robots, it should be considered as including (1) patient perception (personal and societal views, the degree of trust within a patient-doctor relationship and wider system); (2) patient understanding (procedure, risk, equipoise, device); (3) patient consent (informed consent, full disclosure of conflicts of interest).

The consent process should not be contaminated by surgeon bias or patient misinformation. Potential alternatives to the traditional consent process include using research nurses or computer decision support programs. Surgeons involved in the process of consent for robotic surgery trials should undergo training to minimise unconscious bias.[89]

Patients involved in IDEAL Stage 2b or 3 studies should be informed about their surgeon's current level of experience with the proposed robotic platform and procedure, encompassing information on both local outcomes and complications for robotic and alternative (i.e. standard of care) procedures. If an accurate assessment of the learning curve is available, this should be disclosed.

4.4 System Perspective in IDEAL Stages 2b & 3

4.4.1 Key challenges

A broad systems perspective is needed during comparative surgical robotic evaluation.[43] Surgical robots must be economically viable, and the cost of purchase, maintenance and repairs fully evaluated.[8,21] Increasing attention is being paid to the environmental impact of surgery, thus the impact of robotics should be measured and justified in terms of global Net Zero initiatives.[51,82,83] Adoption of single use robotic tools presents a concern in this regard.

Existing efforts reporting on the resource use, greenhouse gas emissions, and material footprints associated with robots require extension to provide impact comparisons with existing technologies.[84] The Lancet Commission on Global Surgery highlighted the huge global unmet need, , for timely and effective surgical services, particularly in low-income settings.[85] An in-depth understanding of each surgical ecosystem will be needed, prior to a decision on integration of robotics.[38,85,86] This includes understanding challenges such as inconsistent access to electricity, clean water, operating rooms and certified surgeons, equipment sterilisation procedures, maintenance of equipment, and inconsistent funding, which may render robotic surgery infeasible. In resource-poor settings, there is a clear opportunity cost of introducing surgical robots, which may squander scarce resources, and be impossible to maintain, resulting in net harm and perpetuating healthcare inequality.[85]From an ethical viewpoint it is important to consider the impact of robotics on access to care for relatively disadvantaged populations in all healthcare systems.[87]

4.4.2 Recommendations

Analysis of healthcare costs associated with robotic intervention and control treatments should be routinely included in comparative surgical robotic studies. Economic studies should include clinically and system-relevant outcomes over a sufficient length of follow-up to compare a surgical robot to current surgical practice. Established international frameworks such as those published by The Professional Society for Health Economics and Outcomes Research (ISPOR) should be used to evaluate health economics and outcomes research.[90,91] Decision-analytic modelling should

be used in IDEAL Stage 2b studies. IDEAL Stage 3 studies should incorporate formal economic evaluations providing trial-based cost-effectiveness analyses which follow established reporting guidelines, such as the Consolidated Health Economic Evaluation Reporting Standards (CHEERS).

Although the Colloquium acknowledges the significant barriers to implementing robotic surgery programs in low-resource settings, it is possible that future advances may reduce these. Therefore, stakeholders from low-income countries with an interest in robotic surgery should be encouraged to join discussions and provide insights into how robotic surgery might become more feasible and beneficial in such settings once its value in higher income settings is established.

To delineate whether a surgical robot would result in net health benefits whilst remaining cost-effective in low-income settings, a rigorous modelling approach can be applied. This should include metrics on robot effectiveness and safety; health economic and sustainability analysis; and specific capacity metrics for the target healthcare environment – such as basic infrastructure (including energy and information technology services), healthcare infrastructure, necessary human resources (entire surgical team), medical supplies, critical care capacity, and healthcare funding. The goal of this process is to estimate the robot's impact within lower-resourced ecosystems, determining an environment's readiness for downstream robot integration.

A modelling approach can also be applied to identify significant risks to fair distribution of benefits within higher-income contexts. If modelling reveals concerns regarding equity of access, safety, cost-effectiveness, or readiness, then a plan for local capacity building should be developed and its implementation monitored before robots are introduced.

Efforts to uphold fairness by increasing access to successful innovation internationally should be supported by non-governmental organisations, governments and the robotic industry, and by existing infrastructure, such as the SAFROS project to address current inequities in access to safe surgery.

The sustainability and economic evaluation of a surgical robot should include a complete life cycle assessment considering how the surgical robot changes practice in relation to the surgical procedure, manufacture and maintenance, type and amount of waste generated, reusable and single-use items. Any projected increase in carbon footprint compared with continuing with non-robotic surgery should be assessed, minimised and offset where possible (e.g., switching from consumable to reusable components), and should be justified in terms of other quantifiable benefits (e.g., improved patient outcomes, and downstream economic and environmental benefits).

5. Long-term Monitoring and Technological Evolution (IDEAL Stage 4)

Following comparative evaluation and widespread adoption, the focus shifts to long-term monitoring of performance, in real-world settings. Registries are the predominant methodology in this stage of evaluation,[11] but ownership and curation of robotic registries by commercial groups can introduce risks of bias and lack of transparency. Other prospective methods of long-term study, such as observational cohort studies, have limitations including fragmentation, maintenance costs, and lack of comparability. In an increasingly digitalised healthcare landscape, real-world datasets (RWD) leveraging data collected for clinical care or administrative purposes have become important potential data sources for the evaluation of health interventions.[92] However, valid studies based on RWD need standards to guide their design and reporting, and safeguards for privacy and data security. Expanding on the IDEAL framework, targeted recommendations specific to IDEAL Stage 4 study designs are needed to inform their methodologies and analytics.

BOX 3

Summary of key recommendations across IDEAL Stage 4

Here, we present a summary of the key recommendations for the evaluation of surgical robots in the final IDEAL stage – long term monitoring. We consider the perspectives of four key stakeholders: Device, Clinician, Patient and System.

Device Perspective

- Long-term monitoring should be led by real-world datasets tailored to provide high-quality, transparent, and valid data.
- Evaluation of surgical robots must be customised to accommodate for their dynamic nature, specifically with regards to AI-enabled system and to detect device creep.

Clinician

- Standardised training programmes, informed by comparative stage findings, should be employed and recognised by accrediting bodies.
- Surgeon revalidation and credentialling should be performed to ensure robotic surgery skills are maintained to a high standard.
- All adverse events should undergo human and systems factors analysis with dedicated experts.

Patient

- Registries and long-term monitoring studies should be independently procured, readily available, and understandable to patients.
- Patient reported outcome measures should predominate in long-term monitoring studies to ensure outcomes remain patient centred.

System

- Cost-effectiveness analysis of surgical robots should be performed, informed by real-world data driven decision-analytic modelling.
- International forums should assess and mitigate global health inequities introduced by surgical robotics.
- Sustainability and environmental impact assessment are imperative in the long-term stage, guided by regular consultation with expert stakeholders.

5.1 Device Perspective in IDEAL Stage 4

5.1.1 Key Challenges

Long-term monitoring of a surgical robot's real-world performance is critical for the safety, evolution and longevity of a device. This could best be achieved by device developers working with regulators, providers, insurers and other stakeholders to create international surveillance systems.[6] The developers of surgical robots have a duty to ensure that patients and scientific evaluators have the best possible evidence to fulfil the ethical requirements for autonomy and non-maleficence respectively, and this needs comprehensive, unbiased outcome data from real-world settings. Many existing device monitoring systems are criticised as passive, and inconsistent, under-reporting incidents, and therefore lagging behind analogous systems dealing with drug monitoring.[93–95] Given the current lack of incentives to evaluate, it is unsurprising that existing evidence on devices is weak, and efforts to curate data are fragmented, reducing comparability and scope for analysis.[96,97]

Manufacturers, hospitals and insurers curate and maintain datasets, but have few incentives to make them widely accessible, while commercial, and sometimes regulatory, issues also inhibit full disclosure of clinical and

technical data.[93] Registries are currently the predominant methodology for long term monitoring of robotic surgical interventions but currently these are generally in-house datasets focused on a single robotic system, and usually lack independent validation and/or have limited access to real-world data.[11] Efforts to link datasets to facilitate better analysis of larger groups are currently limited in their impact and capacity, partly by regulatory issues around data sharing. Stakeholder collaboration at all levels (individual, organisational, system, international) is required to generate high-quality data, as seen with the US national device and evaluation system MDEPinet, which acts as a registry network for specific surgical devices.[98] To give a full picture, evaluation systems for surgical robots need to go further, supplementing standard outcome measures (i.e. effectiveness, safety, economical) with complementary datasets, including machine-generated activity data, data from human factors analyses, and data to monitor the dynamic nature of AI incorporated into surgical robotics.[99]

5.1.2 Recommendations

To address these key challenges, this colloquium proposes the following recommendations. In principle, best practice should be followed using established design and reporting guidelines, and prospectively collected high quality data.[114] Integration of real-world datasets should be encouraged if quality can be assured. Data should be collected and analysed by groups independent from those producing it. The roles and conflicts of interest of those producing and curating data should be transparent and available.

Datasets should include, but not be limited to, patient population demographics, disease characteristics, device characteristics, device indications, type of setting, clinical outcomes, economic outcomes, low-level technical outcomes, technical failures, adverse events, changes in device capabilities and dedicated metrics monitoring AI-system evolution. Reporting of technical failures (including software failures) and patient safety incidents should be mandatory, supported by national regulators and independent of device manufacturers. Rapidly generated, scalable datasets should be developed for widely adopted innovations. Collection and analysis should be fully automated, with harmonised coding language and core reporting and outcome measures.

Regulatory, political and commercial barriers may limit the feasibility of optimal sharing of real world data. In principle, international collaborative approaches are recommended to produce homogeneous and comparable datasets, with data-sharing agreements giving data access to all stakeholders. Objective and clearly presented summary information for patients concerning the pros and cons of robots should be produced by these systems. Governance of linked datasets should ensure open access to facilitate observational research. Governments, insurers, hospitals, and professional associations all have potential roles in this.

Statistical analyses of real-world data should be transparent in their methods, and show how they account for confounding factors, sources of bias and missing data... Analyses should be made accessible according to the FAIR (Findability, Accessibility, Interoperability and Reuse) principles.[115]

AI-enabled and autonomous systems require particular attention. The initial use and indication of use should be clearly stated, and metrics for monitoring performance and safety established from the outset of clinical use. Performance should be evaluated at regular intervals, with more frequent evaluations of rapidly changing systems. Changes in indication of use, the level of autonomy of the system or performance drift which might increase the level of risk will require detailed evaluation. Changes in machine behaviour during the period should be described, with analysis of how the algorithm has changed where this is possible.

5.2 Clinician Perspective IDEAL Stage 4

5.2.1 Key Challenges

The long-term integration of surgical robots into health systems relies on their adoption by clinicians. The principal challenges from this perspective arise from training, credentialing, and determining accountability for adverse outcomes (particularly in the context of robot autonomy and AI). Technologies that demonstrate safety and efficacy experimentally pose risk to patients in untrained hands, and inadequate training prolongs learning curves, particularly during the long-term study stage, as devices are adopted by new surgical teams.[100–103] Research attempting to elucidate learning curves associated with surgical robots remains sparse but appears to be developing. Whilst standardised robotic training programmes exist for well-established surgical robots, such as the Da Vinci, most robotic surgery training remains inconsistent and non-standardised, particularly for novel robots.[104,105]

There are efforts to address these challenges, such as the multi-institutional validation and assessment of training modalities in robotic surgery (the MARS project), but the optimal strategies for training robotic surgeons are unclear.[105]Ongoing certification and credentialing based on a regular re-examination of skills is not currently required for robotic surgery, which contrasts with practice in comparable high-risk industries involving complex technologies (e.g., aviation).[68] Determining accountability for, and analysing the causation of, adverse events during surgery will be more complicated in a robotic future.[29,106,107] Communication difficulties due to altered spatial relationships in the operating room, telesurgery, input from company technical experts and, in future, increasing machine AI autonomy all have the potential to diffuse responsibility for decisions.[28,108,109] Effective monitoring will require routine recording, storing and analysis of granular data including technical, video, audio, and IT data streams, which may be needed in the analysis of adverse events, aligning surgery with other high-risk, high-technology processes.[110]

5.2.2 Recommendations

Novel training methods should undergo evaluation using appropriate frameworks for determining validity (e.g., Messick's framework[116]). They should specify the aims of the training, and employ an appropriate educational paradigm. These studies should inform standardised training programmes, which receive oversight from recognised accrediting bodies and are independent from industry partners. Where validated methods exist, surgeons using a robotic system should undergo regular revalidation with holistic assessments of performance

through assessment of technical and non-technical skills. Novel methodologies including automated performance metrics, AI-driven credentialing, and operative video assessment should be adopted if validated.

Ongoing credentialing and revalidation should include assessments of skills necessary to operate the device, but also the availability of skills in techniques needed to safely manage emergencies using alternative approaches, whether by the same surgeon or another.

A human factors expert should be included in the analysis of all serious adverse events involving a surgical robot. Adverse events/errors should be analysed using data including technical, usability, interface and system integration failures.

Governance for robotic surgery, particularly where AI systems with autonomy are involved, needs to evolve so that it is capable of determining appropriate responsibility for monitoring, accountability for adverse events, and responsibility for implementing improvements. This will require collaboration between legislators, healthcare organisations, professional bodies, and industry.

Processes for monitoring the unplanned evolution of aspects of machine-learning enabled AI should be iteratively reviewed, as human experience of this activity is in its infancy. Re-evaluation of processes and algorithms should take place at regular intervals, and whenever evolving aspects (e.g., level of autonomy or drift in target population) cause significant changes in performance.

5.3 Patient Perspective IDEAL Stage 4

5.3.1 Key Challenges

As with all IDEAL stages, patients are the most important stakeholder when evaluating surgical robotics, as the recipients of both benefits and harms. Patient perceptions are influenced by exposure to the views and agendas of other stakeholders, for example, manufacturer marketing and clinician enthusiasm. However, patients have limited access to scientific evidence, which may be further restricted due to regulatory/approval processes. They may be falsely reassured that a robotic system is well established and safe, without specific evidence for the indication it is being offered for (procedure creep). They are unlikely to be cognisant of iterative changes to a surgical robot, rendering it different to the device upon which initial evidence was generated (device creep), making it important that this type of information is explicitly mentioned during the consent process.

5.3.2 Recommendations

Taking these issues into account, the Colloquium recommend the following. Comprehensive robotic surgery registries and/or systems for extracting reliable information from existing real-world data sources should be made accessible and understandable to patients by providing lay language explanations of their outputs.

Current data should inform the consenting process: evidence referred to in informed consent must relate to the indication and not simply to the device, since robotic systems may be used for many different procedures (procedure creep).

Informed Consent by patients should routinely seek general consent for future use of anonymised data for research and safety surveillance to maximise the value of health data.

Finally, where mechanisms to facilitate this exist, public and patient involvement should inform the design of IDEAL Stage 4 studies and outcome measures to ensure they remain patient-centred.

5.4 System Perspective IDEAL Stage 4

5.4.1 Key Challenges

The evaluation of the wider systems impact of robotic systems needs to continue in the long term, to track the cost-effectiveness and sustainability of their integration into healthcare systems with varying resources and capacity. Health economic analyses need to be iteratively updated with real-world data, and should remain free from restriction or private interests to maintain transparency.[111] Costs will be impacted by learning curves, technical errors, system failures, dynamic pricing, and other factors. This means that real-world data, including health data, administrative claims data, and prospective observational studies are essential in modelling the true value of robotic systems in IDEAL Stage 4. Potential access and equity issues accompanying these high-cost investments must be considered,[112] meaning resource allocation requires justification in terms of their place amongst competing choices. Ethically, providers must consider the benefits of robots against wider health system needs, and rationally allocate limited resources to high-priority issues.

Similarly, strong arguments are needed in favour of robotic surgery to counterbalance environmental impacts seen through life cycle assessments. This issue makes an argument for innovators to adopt sustainable practices in the development, implementation, and maintenance of robots. Innovators should measure and minimise environmental harms of robots, ideally through open, transparent datasets, such as the HealthcareLCA repository,[113] fostering collaborative investigation of their impact in real-world settings. Outside of experimental evaluation settings, complex interventions enter complex adaptive systems, with potentially unforeseen "emergent" consequences. True performance will only be revealed in real-world settings and must be monitored to avoid unrecognised gradual decline in safety or effectiveness. This demands the development of monitoring infrastructure, processes, and governance.

5.4.2 Recommendations

For the long-term evaluation of surgical robots, we propose the following recommendations. Cost-effectiveness analyses of real-world data employing decision-analytic modelling should evaluate robotic systems by indication, and provide comparable analyses openly available to all stakeholders. These should use validated outcome

metrics and comply with Professional Society for Health Economics and Outcomes Research (ISPOR) guidance. In principle, regular reviews of robotic surgery cost-effectiveness should include an assessment of changes in organisational configurations and their influence on process/outcomes, where the necessary resources are available.

National and international discussion forums involving clinicians, patient advocacy groups, industry, policymakers, ethicists, and economists are needed to consider the potential effects of robotics on equity of healthcare access, and to explore models which might justify use in low-income settings. Advice from public health experts, policymakers, ethicists, and climate scientists should be considered in discussions of how robotic surgery platform design, development, and use could be made more sustainable.

In principle, complete life cycle assessments of surgical robotics should incorporate a broad range of parameters, be guided by environmental experts, produce data without restriction and contribute towards living open-access data repositories. Moreover, complete life cycle assessments of surgical robotics should be iteratively updated against existing care standards in real-world settings to monitor and minimise their environmental impacts through quality improvement. These recommendations will require significant further development of collaborations, datasets and resources.

6. Conclusion

The next generation of surgical robotics is poised to transform healthcare systems around the world. Whether this will result in substantial patient and societal benefit depends critically on whether innovation is guided by appropriate evaluation. This Colloquium has provided key recommendations for the integration of the development, evaluation and monitoring of surgical robots, across the developmental life cycle, mapped to the IDEAL evaluation framework.

Our analysis has bridged the pre-clinical and early evaluation stages (IDEAL Stages 0, 1, & 2a), the comparative stages (IDEAL Stages 2b & 3), and the long-term monitoring stage (IDEAL Stage 4), and presents practical recommendations to guide robotics developers, clinicians, patients, and wider systems as we enter the next era surgical robotics. For all stages of evaluation, all stakeholders should be considered at the outset, including the surgical team (human factors analysis and training), patients (acceptability and rigorous ethical assessment) and the wider system (economic and sustainability evaluation). Further work is needed to establish standardised metrics for technical and clinical outcomes, refine health economic assessment models and assess the global representativeness of these recommendations.

No framework which deals with such a broad range of evaluation challenges can hope to avoid conflicts between recommendations, or situations where recommendations may appear disproportionate to the problem addressed. Such dilemmas with IDEAL recommendations are usually easily resolved by referral to the underlying principles mentioned in Methods. The breadth of the subject also raises the question of which evaluation recommendations are relevant in the context of which particular studies. Clearly, incorporating all possible aspects in any single study would be infeasible and unnecessary, but sensible judgement, involving discussion where necessary with relevant subject experts, should allow this guidance to be of practical use to clinicians, robotic engineers, patients and other stakeholders in the development of robotic surgery.

7. References

1. Lee, N. Robotic surgery: where are we now? *The Lancet* **384**, 1417 (2014).
2. Lancet, T. Robotic surgery evaluation: 10 years too late. *The Lancet* **388**, 1026 (2016).
3. Peters, B. S., Armijo, P. R., Krause, C., Choudhury, S. A. & Oleynikov, D. Review of emerging surgical robotic technology. *Surg Endosc* **32**, 1636–1655 (2018).
4. Christensen, C. M., Baumann, H., Ruggles, R. & Sadtler, T. M. Disruptive innovation for social change. *Harv Bus Rev* **84**, 94–101, 163 (2006).
5. Maynou, L., Pearson, G., McGuire, A. & Serra-Sastre, V. The diffusion of robotic surgery:

   Examining technology use in the English NHS. *Health Policy* **126**, 325–336 (2022).
6. Tan, W. S., Ta, A. & Kelly, J. D. Robotic surgery: getting the evidence right. *Med J Aust*

   217, 391–393 (2022).
7. Vasey, B. *et al.* Intraoperative Applications of Artificial Intelligence in Robotic Surgery: A

   Scoping Review of Current Development Stages and Levels of Autonomy. *Annals of Surgery* 10.1097/SLA.0000000000005700 (2023)

   doi:10.1097/SLA.0000000000005700.

8. REINFORCE: A real-world, in-situ, evaluation of the introduction and scale-up of robotassisted surgical services in the NHS. *ARC* https://arcnenc.nihr.ac.uk/projects/reinforce-a-real-world-in-situ-evaluation-of-the-introductionand-scale-up-of-robot-assisted-surgical-services-in-the-nhs/ (2023).
9. McCulloch, P., Cook, J. A., Altman, D. G., Heneghan, C. & Diener, M. K. IDEAL framework for surgical innovation 1: the idea and development stages. *BMJ* **346**, f3012 (2013).
10. Ergina, P. L., Barkun, J. S., McCulloch, P., Cook, J. A. & Altman, D. G. IDEAL framework for surgical innovation 2: observational studies in the exploration and assessment stages. *BMJ* **346**, f3011 (2013).
11. Cook, J. A. *et al.* IDEAL framework for surgical innovation 3: randomised controlled trials in the assessment stage and evaluations in the long term study stage. *BMJ* **346**, f2820 (2013).
12. Roodbeen, S. X. *et al.* Evolution of transanal total mesorectal excision according to the IDEAL framework. *BMJ Surgery, Interventions, & Health Technologies* **1**, e000004 (2019).
13. Marcus, H. J. *et al.* IDEAL-D Framework for Device Innovation: A Consensus Statement on the Preclinical Stage. *Ann Surg* **275**, 73–79 (2022).
14. UK Statutory Instruments, UK Government. *The Medical Devices Regulations 2002*. *2002 No. 618* (King's Printer of Acts of Parliament, 2023).
15. *Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC (Text with EEA relevance.)*. *OJ L* vol. 117 (2017).

16. IEC 60601-1-2:2014/AMD1:2020 | IEC Webstore | electromagnetic compatibility, EMC, smart city. https://webstore.iec.ch/publication/59644 (2023).
17. 14:00-17:00. IEC 62304:2006. *ISO* https://www.iso.org/standard/38421.html (2023).
18. 14:00-17:00. IEC 80601-2-77:2019. *ISO* https://www.iso.org/standard/68473.html (2023).
19. 14:00-17:00. ISO 13485:2016. *ISO* https://www.iso.org/standard/59752.html (2021).
20. Foote, S. B. *Managing the Medical Arms Race: Innovation and Public Policy in the Medical Device Industry*. (University of California Press, 1992).
21. Future of Surgery Commission Group. *The Commission on the Future of Surgery*. https://www.rcseng.ac.uk/standards-and-research/future-of-surgery/ (2018).
22. Yang, G.-Z. *et al.* Medical robotics-Regulatory, ethical, and legal considerations for increasing levels of autonomy. *Sci Robot* **2**, eaam8638 (2017).
23. Andras, I. *et al.* Artificial intelligence and robotics: a combination that is changing the operating room. *World J Urol* **38**, 2359–2366 (2020).
24. Bhandari, M., Zeffiro, T. & Reddiboina, M. Artificial intelligence and robotic surgery: current perspective and future directions. *Current Opinion in Urology* **30**, 48 (2020).
25. Panesar, S. *et al.* Artificial Intelligence and the Future of Surgical Robotics. *Ann Surg* **270**, 223–226 (2019).
26. Schreyer, J. *et al.* RAS-NOTECHS: validity and reliability of a tool for measuring nontechnical skills in robotic-assisted surgery settings. *Surg Endosc* **36**, 1916–1926 (2022).
27. Raison, N. *et al.* Development and validation of a tool for non-technical skills evaluation in robotic surgery-the ICARS system. *Surg Endosc* **31**, 5403–5410 (2017).
28. Catchpole, K. *et al.* Human Factors Integration in Robotic Surgery. *Hum Factors*

187208211068946 (2022) doi:10.1177/00187208211068946.

29. O'Sullivan, S. *et al.* Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (AI) and autonomous robotic surgery. *Int J Med Robot* **15**, e1968 (2019).
30. Rogers, W. A., Hutchison, K. & McNair, A. Ethical Issues Across the IDEAL Stages of Surgical Innovation. *Ann Surg* **269**, 229–233 (2019).
31. Boys, J. A. *et al.* Public perceptions on robotic surgery, hospitals with robots, and surgeons that use them. *Surg Endosc* **30**, 1310–1316 (2016).
32. J, J. & W, R. Innovative surgery: the ethical challenges. *Journal of medical ethics* **38**, (2012).
33. Angelos, P. Ethics and surgical innovation: challenges to the professionalism of surgeons. *Int J Surg* **11 Suppl 1**, S2-5 (2013).
34. Hutchison, K., Rogers, W., Eyers, A. & Lotz, M. Getting Clearer About Surgical Innovation: A New Definition and a New Tool to Support Responsible Practice. *Ann Surg* **262**, 949–954 (2015).
35. B, H., S, D., W, O., I, C. & D, S. Harmonization of ethics in health technology assessment: a revision of the Socratic approach. *International journal of technology assessment in health care* **30**, (2014).
36. Partington, A. & Karnon, J. It's Not the Model, It's the Way You Use It: Exploratory Early Health Economics Amid Complexity Comment on 'Problems and Promises of Health Technologies: The Role of Early Health Economic Modelling'. *Int J Health Policy Manag* **10**, 36–38 (2020).
37. Rizan, C. *et al.* The Carbon Footprint of Surgical Operations: A Systematic Review. *Ann Surg* **272**, 986–995 (2020).

38. Sullivan, R. *et al.* Global cancer surgery: delivering safe, affordable, and timely cancer surgery. *The Lancet Oncology* **16**, 1193–1224 (2015).

39. Hung, A. J. *et al.* Development and Validation of Objective Performance Metrics for Robot-Assisted Radical Prostatectomy: A Pilot Study. *J Urol* **199**, 296–304 (2018).

40. Vasey, B. *et al.* Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *BMJ* **377**, e070904 (2022).

41. Møller, L. *et al.* Identifying curriculum content for operating room nurses involved in robotic-assisted surgery: a Delphi study. *Surg Endosc* (2022) doi:10.1007/s00464-02209751-4.

42. Lawrie, L. *et al.* Current issues and future considerations for the wider implementation of robotic-assisted surgery: a qualitative study. *BMJ Open* **12**, e067427 (2022).

43. Lawrie, L. *et al.* Barriers and enablers to the effective implementation of robotic assisted surgery. *PLOS ONE* **17**, e0273696 (2022).

44. Woudstra, K., Reuzel, R., Rovers, M. & Tummers, M. An Overview of Stakeholders, Methods, Topics, and Challenges in Participatory Approaches Used in the Development of Medical Devices: A Scoping Review. *International Journal of Health Policy and Management* **12**, (2023).

45. van der Wilt, G. J., Gerhardus, A. & Oortwijn, W. TOWARD INTEGRATION IN THE CONTEXT OF HEALTH TECHNOLOGY ASSESSMENT: THE NEED FOR EVALUATIVE FRAMEWORKS. *Int J Technol Assess Health Care* **33**, 586–590 (2017).

46. Health and Safety Laboratory. Review of human reliability assessment methods. (2009).

47. Scholte, M., Rovers, M. M. & Grutters, J. P. C. The Use of Decision Analytic Modeling in the Evaluation of Surgical Innovations: A Scoping Review. *Value Health* **24**, 884–900 (2021).

48. Grutters, J. P. C. *et al.* Problems and Promises of Health Technologies: The Role of Early Health Economic Modeling. *Int J Health Policy Manag* **8**, 575–582 (2019).

49. Bolton, W. S. *et al.* Disseminating technology in global surgery. *British Journal of Surgery* **106**, e34–e43 (2019).

50. Payne, C. J. & Yang, G.-Z. Hand-held medical robots. *Ann Biomed Eng* **42**, 1594–1605 (2014).

51. Papadopoulou, A., Kumar, N. S., Vanhoestenberghe, A. & Francis, N. K. Environmental sustainability in robotic and laparoscopic surgery: systematic review. *British Journal of Surgery* **109**, 921–932 (2022).

52. Sheetz, K. H., Claflin, J. & Dimick, J. B. Trends in the Adoption of Robotic Surgery for Common Surgical Procedures. *JAMA Network Open* **3**, e1918911 (2020).

53. Micha, J. P., Rettenmaier, M. A., Bohart, R. D. & Goldstein, B. H. Robotic-Assisted Surgery for the Treatment of Breast and Cervical Cancers. *JSLS* **26**, e2022.00014 (2022).

54. Ramirez, P. T. *et al.* Minimally Invasive versus Abdominal Radical Hysterectomy for Cervical Cancer. *New England Journal of Medicine* **379**, 1895–1904 (2018).

55. Melamed, A. *et al.* Survival after Minimally Invasive Radical Hysterectomy for Early-Stage Cervical Cancer. *New England Journal of Medicine* **379**, 1905–1914 (2018).

56. US Food and Drug Administration. *MAUDE - Manufacturer and User Facility Device Experience*. (2023).

57. Hu, Y. & Strong, V. E. Robotic Surgery and Oncologic Outcomes. *JAMA Oncology* **6**, 1537–1539 (2020).

58. Parekh, D. J. *et al.* Robot-assisted radical cystectomy versus open radical cystectomy in patients with bladder cancer (RAZOR): an open-label, randomised, phase 3, noninferiority trial. *The Lancet* **391**, 2525–2536 (2018).

59. Feng, Q. *et al.* Robotic versus laparoscopic surgery for middle and low rectal cancer (REAL): short-term outcomes of a multicentre randomised controlled trial. *The Lancet Gastroenterology & Hepatology* **7**, 991–1004 (2022).

60. Yaxley, J. W. *et al.* Robot-assisted laparoscopic prostatectomy versus open radical retropubic prostatectomy: early outcomes from a randomised controlled phase 3 study. *Lancet* **388**, 1057–1066 (2016).

61. Garfjeld Roberts, P. *et al.* Research quality and transparency, outcome measurement and evidence for safety and effectiveness in robot-assisted surgery: systematic review. *BJS Open* **4**, 1084–1099 (2020).

62. Khan, D. Z. *et al.* Automated operative workflow analysis of endoscopic pituitary surgery using machine learning: development and preclinical evaluation (IDEAL stage 0). *Journal of Neurosurgery* **137**, 51–58 (2021).

63. van Amsterdam, B., Clarkson, M. J. & Stoyanov, D. Gesture Recognition in Robotic Surgery: A Review. *IEEE Trans Biomed Eng* **68**, 2021–2035 (2021).

64. Kiyasseh, D. *et al.* A vision transformer for decoding surgeon activity from surgical videos. *Nat Biomed Eng* **7**, 780–796 (2023).

65. Chen, J. *et al.* Use of Automated Performance Metrics to Measure Surgeon Performance during Robotic Vesicourethral Anastomosis and Methodical Development of a Training Tutorial. *J Urol* **200**, 895–902 (2018).

66. Shouhed, D., Gewertz, B., Wiegmann, D. & Catchpole, K. Integrating Human Factors Research and Surgery: A Review. *Archives of Surgery* **147**, 1141–1146 (2012).

67. Ijy, W., Lj, K. & Jc, N. A systematic review of the true benefit of robotic surgery: Ergonomics. *The international journal of medical robotics + computer assisted surgery : MRCAS* **16**, (2020).

68. Collins, J. W. & Wisz, P. Training in robotic surgery, replicating the airline industry. How far have we come? *World Journal of Urology* **38**, 1645 (2020).

69. Jayne, D. *et al.* Effect of Robotic-Assisted vs Conventional Laparoscopic Surgery on Risk of Conversion to Open Laparotomy Among Patients Undergoing Resection for Rectal Cancer: The ROLARR Randomized Clinical Trial. *JAMA* **318**, 1569–1580 (2017).

70. Johnson, B., Sorokin, I., Singla, N., Roehrborn, C. & Gahan, J. C. Determining the Learning Curve for Robot-Assisted Simple Prostatectomy in Surgeons Familiar with Robotic Surgery. *Journal of Endourology* **32**, 865–870 (2018).

71. Pernar, L. I. M. *et al.* An appraisal of the learning curve in robotic general surgery. *Surg Endosc* **31**, 4583–4596 (2017).

72. Vilallonga, R. *et al.* The Initial Learning Curve for Robot-Assisted Sleeve Gastrectomy: A Surgeon's Experience While Introducing the Robotic Technology in a Bariatric Surgery Department. *Minimally Invasive Surgery* **2012**, e347131 (2012).

73. Wijburg, C. J. *et al.* Learning Curve Analysis for Intracorporeal Robot-assisted Radical Cystectomy: Results from the EAU Robotic Urology Section Scientific Working Group. *Eur Urol Open Sci* **39**, 55–61 (2022).

74. Kirkpatrick, D. L. Techniques for evaluating training programs. *Training and development journal* (1979).

75. Sridhar, A. N., Briggs, T. P., Kelly, J. D. & Nathan, S. Training in Robotic Surgery-an Overview. *Curr Urol Rep* **18**, 58 (2017).

76. Skjold-Ødegaard, B. & Søreide, K. Competency-based Surgical Training and Entrusted Professional Activities – Perfect Match or a Procrustean Bed? *Annals of Surgery* **273**, e173 (2021).

77. Carpenter, B. T. & Sundaram, C. P. Training the next generation of surgeons in robotic surgery. *Robot Surg* **4**, 39–44 (2017).

78. Mark Knab, L. *et al.* Evolution of a Novel Robotic Training Curriculum in a Complex General Surgical Oncology Fellowship. *Ann Surg Oncol* **25**, 3445–3452 (2018).

79. Marcus, H. J. *et al.* IDEAL-D Framework for Device Innovation: A Consensus Statement on the Preclinical Stage. *Annals of Surgery* (2021) doi:10.1097/SLA.0000000000004907.

80. Torrent-Sellens, J., Jiménez-Zarco, A. I. & Saigí-Rubió, F. Do People Trust in Robot-Assisted Surgery? Evidence from Europe. *Int J Environ Res Public Health* **18**, 12519 (2021).

81. Buabbas, A. J., Aldousari, S. & Shehab, A. A. An exploratory study of public' awareness about robotics-assisted surgery in Kuwait. *BMC Med Inform Decis Mak* **20**, 140 (2020).

82. Salas, R. N., Maibach, E., Pencheon, D., Watts, N. & Frumkin, H. A pathway to net zero emissions for healthcare. *BMJ* **371**, m3785 (2020).

83. Rasheed, F. N. *et al.* Decarbonising healthcare in low and middle income countries: potential pathways to net zero emissions. *BMJ* **375**, n1284 (2021).

84. Papa, L. *et al.* Systematic review of clinical research on biomarkers for pediatric traumatic brain injury. *Journal of neurotrauma* **30**, 324–338 (2013).

85. Meara, J. G. *et al.* Global Surgery 2030: evidence and solutions for achieving health, welfare, and economic development. *The Lancet* **386**, 569–624 (2015).

86. Garas, G. *et al.* Surgical Innovation in the Era of Global Surgery: A Network Analysis.

Annals of Surgery **271**, 868 (2020).

87. K, H., J, J. & D, C. Justice and Surgical Innovation: The Case of Robotic Prostatectomy. *Bioethics* **30**, (2016).
88. Páez, A. *et al.* Beyond the RCT: When are Randomized Trials Unnecessary for New Therapeutic Devices, and What Should We Do Instead? *Ann Surg* **275**, 324–331 (2022).
89. Rooshenas, L. *et al.* The QuinteT Recruitment Intervention supported five randomized trials to recruit to target: a mixed-methods evaluation. *J Clin Epidemiol* **106**, 108–120 (2019).
90. Caro, J. J., Briggs, A. H., Siebert, U. & Kuntz, K. M. Modeling Good Research Practices—Overview: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. *Value in Health* **15**, 796–803 (2012).
91. Ramsey, S. D. *et al.* Cost-Effectiveness Analysis Alongside Clinical Trials II—An ISPOR Good Research Practices Task Force Report. *Value in Health* **18**, 161–172 (2015).
92. Dreyer, N. A. Strengthening evidence-based medicine with real-world evidence. *The Lancet Healthy Longevity* **3**, e641–e642 (2022).
93. Kramer, D. B., Xu, S. & Kesselheim, A. S. How Does Medical Device Regulation Perform in the United States and the European Union? A Systematic Review. *PLOS Medicine* **9**, e1001276 (2012).
94. Cooper, M. A., Ibrahim, A., Lyu, H. & Makary, M. A. Underreporting of Robotic Surgery Complications. *The Journal for Healthcare Quality (JHQ)* **37**, 133 (2015).
95. Rajan, P. V., Kramer, D. B. & Kesselheim, A. S. Medical device postapproval safety monitoring: where does the United States stand? *Circ Cardiovasc Qual Outcomes* **8**, 124–131 (2015).

96. Cipriani, A. *et al.* Generating comparative evidence on new drugs and devices after approval. *Lancet* **395**, 998–1010 (2020).

97. Huot, L., Decullier, E., Maes-Beny, K. & Chapuis, F. R. Medical device assessment: scientific evidence examined by the French national agency for health – a descriptive study. *BMC Public Health* **12**, 585 (2012).

98. Sedrakyan, A. *et al.* Advancing the Real-World Evidence for Medical Devices through Coordinated Registry Networks. *BMJ Surgery, Interventions, & Health Technologies* **4**, e000123 (2022).

99. Ficuciello, F., Tamburrini, G., Arezzo, A., Villani, L. & Siciliano, B. Autonomy in surgical robots and its meaningful human control: *Paladyn, Journal of Behavioral Robotics* **10**, 30–43 (2019).

100. van Workum, F. *et al.* Learning Curve and Associated Morbidity of Minimally Invasive Esophagectomy: A Retrospective Multicenter Study. *Ann Surg* **269**, 88–94 (2019).

101. Short-term outcomes and one surgeon's learning curve for thoracoscopic esophagectomy performed with the patient in the prone position | SpringerLink. https://link.springer.com/article/10.1007/s00595-016-1378-5 (2023).

102. Zeuschner, P. *et al.* Three Different Learning Curves Have an Independent Impact on Perioperative Outcomes After Robotic Partial Nephrectomy: A Comparative Analysis. *Ann Surg Oncol* **28**, 1254–1261 (2021).

103. Le Morvan, P. & Stock, B. Medical learning curves and the Kantian ideal. *J Med Ethics* **31**, 513–518 (2005).

104. Dixon, F. & Keeler, B. Robotic surgery: training, competence assessment and credentialing. *Bulletin* **102**, 302–306 (2020).

105. Chen, R. *et al.* A comprehensive review of robotic surgery curriculum and training for residents, fellows, and postgraduate surgical education. *Surg Endosc* **34**, 361–367 (2020).

106. Jamjoom, A. A. B. *et al.* Autonomous surgical robotic systems and the liability dilemma. *Front Surg* **9**, 1015367 (2022).

107. van Wynsberghe, A. Responsible Robotics and Responsibility Attribution. in *Robotics, AI, and Humanity: Science, Ethics, and Policy* (eds. von Braun, J., S. Archer, M., Reichberg, G. M. & Sánchez Sorondo, M.) 239–249 (Springer International Publishing, 2021). doi:10.1007/978-3-030-54173-6_20.

108. Catchpole, K. *et al.* Safety, efficiency and learning curves in robotic surgery: a human factors analysis. *Surg Endosc* **30**, 3749–3761 (2016).

109. Poulsen, J. L., Bruun, B., Oestergaard, D. & Spanager, L. Factors affecting workflow in robot-assisted surgery: a scoping review. *Surg Endosc* **36**, 8713–8725 (2022).

110. van Dalen, A. S. H. M. *et al.* Analyzing and Discussing Human Factors Affecting Surgical Patient Safety Using Innovative Technology: Creating a Safer Operating Culture. *J Patient Saf* **18**, 617–623 (2022).

111. Bai, F. *et al.* More work is needed on cost-utility analyses of robotic-assisted surgery. *J Evid Based Med* **15**, 77–96 (2022).

112. Schneider, M. A. *et al.* Inequalities in access to minimally invasive general surgery: a comprehensive nationwide analysis across 20 years. *Surg Endosc* **35**, 6227–6243 (2021).

113. Drew, J., Christie, S. D., Rainham, D. & Rizan, C. HealthcareLCA: an open-access living database of health-care environmental impact assessments. *The Lancet Planetary Health* **6**, e1000–e1012 (2022).

114. Bilbro, N. A. *et al.* The IDEAL Reporting Guidelines: A Delphi Consensus Statement Stage Specific Recommendations for Reporting the Evaluation of Surgical Innovation. *Annals of Surgery* **273**, 82 (2021).
115. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016).
116. *Messick, S. (1989). Validity. In R. L. Linn (Ed.), Educational measurement (3rd ed., pp. 13-104). New York, NY: American Council on education and Macmillan.*

8. Colloquium Panel Members

Panel 1. **Artificial intelligence & autonomy**

Hani J Marcus[1,2], Baptiste Vasey[18], Aswin Chari[20], Fanny Ficuciello[21], Effy Vayena[22], Christopher Baber[23], Marco Zenati[24], Alan Kuntz[25], Karen Kerr[26], Nigel Horwood[§]

Affiliations:

1. Department of Neurosurgery, National Hospital of Neurology and Neurosurgery, London, UK
2. Wellcome/Engineering and Physical Sciences Research Council (EPSRC) Centre for Interventional and Surgical Sciences (WEISS), London, UK
18. Department of Surgery, Geneva University Hospital, Geneva, Switzerland
20. Department of Neurosurgery, National Hospital of Neurology and Neurosurgery, London, UK
21. PRISMA Lab, University of Naples Federico II, Naples, Italy
22. Health Ethics and Policy Lab, Department of Health Sciences and Technology, Zürich, Switzerland
23. School of Computer Science, University of Birmingham, Birmingham, UK
24. Division of Cardiac Surgery, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA
25. School of Computing and the Robotics Centre, University of Utah, Salt Lake City, USA
26. Medtronic, UK
§ Patient Advocate, No Affiliation

Panel 2. Device evaluation

David Noonan[14], Danail Stoyanov[2], Katherine Anderon[27], Ka-Wai Kwok[28], Rich Mahoney[29], Bill Peine[30], Ferdinando Rodriquez Y Baena[31], Pietro Valdastri[32], Richard Leparmentier[33], Len Evans[§], Rebecca Langley[§]

Affiliations:

14. Moon Surgical, Paris, France
2. Wellcome/Engineering and Physical Sciences Research Council (EPSRC) Centre for Interventional and Surgical Sciences (WEISS), London, UK
27. K2A2 consulting, USA
28. Department of Mechanical Engineering, The University of Hong Kong, Hong Kong
29. Intuitive, California, USA
30. Medtronic, USA
31. Department of Mechanical Engineering, Imperial College London, London, UK.
32. Storm Lab UK, Institute for Robotics, University of Leeds, UK
33. Bionaut Labs, California, USA
§ Patient Advocate, No Affiliation

Panel 3. Clinical evaluation

Pedro Ramirez[3], Janet Martin[9], Garnette Sutherland[34], Sanju Lama[34], Naeem Soomro[35], Justin Collins[2], Mario Leitao[36], James Kinross[37], Prokar Dasgupta[13], Alvin Goh[38], Bernard Park[39]

Affiliations:

1. Department of Obstetrics and Gynaecology, Houston Methodist Hospital Neal Cancer Center, USA
9. University of Western Ontario, Ontario, Canada
34. Project neuroArm Medical Robotics Program, Department of Clinical Neurosciences, University of Calgary, Calgary, Canada
35. Department of Urology, Newcastle Upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK

36. Department of Surgery, Division of Gynecology, Memorial Sloan Kettering Cancer Center, New York, USA
37. Department of General Surgery, Imperial College Healthcare NHS Trust, London, UK
38. Urology Service, Department of Surgery, Memorial Sloan Kettering Cancer Centre, New York, USA
39. Thoracic Surgery Service, Department of Surgery, Memorial Sloan Kettering Cancer Centre, New York, USA

Panel 4. Human factors

Ken Catchpole[6], Matthias Weigl[40], Rebecca Randell[41], Steven Yule[42], Duncan McPherson[43], Laura Pickup[44], Richard Skipworth[42], Jennifer Anger[45], Denny Yu[46], Lora Cavuoto[47], Ann Bisantz[47], Tara Cohen[48]

Affiliations:

6. Department of Anaesthesia and Perioperative Medicine, Medical University of South Carolina, Charleston, SC, USA
40. Institute for Patient Safety, University Hospital, Bonn University, Bonn, Germany
41. Faculty of Health Studies, University of Bradford, Bradford, UK
42. Clinical Surgery, Royal Infirmary of Edinburgh, University of Edinburgh, Edinburgh, UK
43. Medicines and Healthcare products Regulatory Agency (MHRA), London, UK
44. LP Human Factors Ltd, Taunton, UK
45. Division of Urology, Cedars-Sinai Medical Centre, Los Angeles, USA
46. School of Industrial Engineering at Purdue University, Lafayette, USA
47. School of Engineering and Applied Sciences, University at Buffalo, Buffalo, USA
48. Cedars-Sinai Medical Centre, Department of Surgery, Hollywood, USA

Panel 5. Economics

Mirre Scholte[49,50], Guy Maddern[51], Laura Sampietro-Colom[52], Alane Clark, Tammy Clifford[§], Belen Corbacho[53], Cynthia Iglesias[54], Janneke Grutters[55], Maroeska Rovers[11]

Affiliations:

49. Department of Operating Rooms, Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, The Netherlands.
50. Department of Clinical Epidemiology & Medical Technology Assessment (KEMTA), Maastricht University Medical Centre, Maastricht, The Netherlands
51. Discipline of Surgery, The University of Adelaide, The Queen Elizabeth Hospital, Adelaide, Australia
52. Clinic Barcelona University Hospital
53. York Trials Unit, Department of Health Sciences, University of York, York, UK
54. Department of Health Sciences, University of York, York, UK
55. Department of Health Evidence, Radboudumc, Nijmegen, The Netherlands
11. Department of Medical Imaging, Radboudumc, Nijmegen, The Netherlands
[§] Patient Advocate, No Affiliation

Panel 6. Ethics

Katrina Hutchinson[8], Lesley Booth[§], Heather Draper[56], Len Evans[§], Sarah Goering[§], Alex Kon[57], Rebecca Langley[§], Guy Maddern[51], Rob Sparrow[58]

Affiliations:

8. Department of Philosophy, Macquarie University, Sydney, Australia

56 Health Sciences, Warwick Medical School, Warwick, UK
57 Department of Bioethics and Humanities, University of Washington, Seattle, USA
51 Discipline of Surgery, The University of Adelaide, The Queen Elizabeth Hospital, Adelaide, Australia
58 Monash University, Melbourne, Australia
§ Patient Advocate, No Affiliation

Panel 7. Training and education

Nick Raison[12], Prokar Dasgupta[13], Kamran Ahmed[59], Justin Collins[2], Deena Harji[60], Teodor Grantcharov[61], Lars Konge[62], Naeem Soomro[35]

Affiliations:

12. Department of Biomedical Engineering and Imaging Sciences, King's College London, UK
13. King's Health Partners Academic Surgery, King's College London, London, UK
59. Department of Urology, Sheikh Khalifa Medical City, Abu Dhabi, UAE
2. Wellcome/Engineering and Physical Sciences Research Council (EPSRC) Centre for Interventional and Surgical Sciences (WEISS), London, UK
60. Department of Colorectal Surgery, Manchester University NHS Foundation Trust, Manchester, UK
61. Department of Surgery and Clinical Excellence Research Centre, Stanford Medicine & Stanford University, USA
62. Copenhagen Academy for Medical Education and Simulation, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
35. Department of Urology, Newcastle Upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK

9. Acknowledgements

The authors would like to thank all colloquium panel members and experts who took part in the semistructured meetings to devise this framework. This work was supported by the IDEAL Collaboration.

10. Disclosures

H.J.M is supported by grants from the Wellcome (203145Z/16/Z) EPSRC (NS/A000050/1) Centre for Interventional and Surgical Sciences, National Institute for Health Research (NIHR) Biomedical Research Centre at University College London and the National Brain Appeal. He also declares stocks in Panda Surgical, and has previously received consulting fees from Intuitive Ventures.

D.Z.K is supported by an NIHR Academic Clinical Fellowship and a Cancer Research UK Predoctoral Fellowship.

J.G.H is supported by an NIHR Academic Clinical Fellowship.

D.J.B holds a Royal College of Surgeons chair supported by the Rosetrees trust and declares institutional research grant funding from the NIHR.

K.C is financially supported by Agency for Healthcare Research and Quality (Grant Number: R01 HS26491-01).

A.C is supported by NIHR funding from NIHR HTA, PGfAR, PHR, RfPB, and i4i programmes.

K.H is supported by a Discovery Project grant from the Australian Research Council (Grant Number DP200100883) and Australian Research Council Discovery Early Career Researcher Award (Grant Number DE200101301).

T.M has acted as unpaid advisory board member of Pumpinheart Ltd., and previously as senior medical officer in medical devices at the Health Products Regulatory Authority, Ireland and previous co-chair of the Clinical Investigation and Evaluation Working Group of the European Commission.

D.Stoyanov is supported by grants from the Wellcome Trust, UKRI, European Commission and RAEng and employment from Digital Surgery Ltd (Medtronic) and declares stocks in Odin Vision Ltd (Olympus) and Panda Surgical.

M.R is supported by grants from Siemens Healthineers and the Healthcare Institute in the Netherlands.

P.D is supported by grants from the UKRI, The Urology Foundation and the Recordati Foundation and declares consulting fees from Proximie, MysteryVibe and Jiva.AI.

D.N is the Chief Technology Officer of Moon Surgical and holds stocks in Panda Surgical.

D.Stocken is supported by grants from NIHR, CRUK, YCR and BHF. She declares travelling funding from RCSEng and AntiCancer Fund Belgium for conference attendance.

G.S is supported by a grant from the European Union Framework Programme (Horizon 2020).

B.V is supported by the Lord Florey scholarship (PhD scholarship) from the Berrow Foundation.

P.M.C receives grants from Medtronic and CMR (unrestricted educational grants to Oxford University for the IDEAL Collaboration) and the Oxford Biomedical Research Centre.