







Outcome selection, measurement and reporting for new surgical procedures and devices: a systematic review of IDEAL/IDEAL-D studies to inform development of a core outcome set

R. C. Macefield¹ , N. Wilson¹, C. Hoffmann¹ , J. M. Blazeby¹ , A. G. K. McNair^{1,2} , K. N. L. Avery¹  and S. Potter^{1,3} 

¹National Institute for Health Research Bristol Biomedical Research Centre, Bristol Centre for Surgical Research, Bristol Medical School, University of Bristol, and ²Department of Gastrointestinal Surgery and ³Bristol Breast Care Centre, North Bristol NHS Trust, Bristol, UK

Correspondence to: Ms R. C. Macefield, Population Health Sciences, Bristol Medical School, University of Bristol, Canynge Hall, 39 Whatley Road, Clifton, Bristol BS8 2PS, UK (e-mail: r.macefield@bristol.ac.uk)

Background: Outcome selection, measurement and reporting for the evaluation of new surgical procedures and devices is inconsistent and lacks standardization. A core outcome set may promote the safe and transparent evaluation of surgical innovations. This systematic review examined outcome selection, measurement and reporting in studies conducted within the IDEAL (Idea, Development, Exploration, Assessment and Long-term monitoring) framework to examine current practice and inform the development of a core outcome set for early-phase studies of surgical procedures/devices.

Methods: Web of Science and Scopus citation searches were performed to identify author-reported IDEAL/IDEAL-D studies for any surgical procedure/device. Outcomes were extracted verbatim, including contextual information regarding outcome selection and measurement. Outcomes were categorized to inform a conceptual framework of outcome domains relevant to evaluating innovation.

Results: Some 48 studies were identified. Outcome selection, measurement and reporting varied widely across studies in different IDEAL stages. From 1737 outcomes extracted, 22 domains specific to evaluating innovation were conceptualized under seven broad categories: procedure completion success/failure; modifications; unanticipated events; surgeons' experiences; patients' experiences; resource use specific to the innovative procedure/device; and other innovation-specific outcomes. Most innovation-specific outcomes were measured and reported in only a small number of studies.

Conclusion: This review highlighted the need for guidance and standardization in outcome selection and reporting in the evaluation of new surgical procedures/devices. Novel outcome domains specific to innovation have been identified to establish a core outcome set for future evaluations of surgical innovations.

Funding information

National Institute for Health Research Bristol Biomedical Research Centre

Presented to the Society of Clinical Trials 40th Annual Meeting, New Orleans, Louisiana, USA, May 2019, the International Conference for Clinical Trials Methodology Research, Brighton, UK, October 2019, and the National Cancer Research Institute Cancer Conference 2019, Glasgow, UK, November 2019; published in abstract form as *Clinical Trials* 2020; 17(Suppl 1) and *Trials* 2019; 20(Suppl 1): 579.

Paper accepted 20 August 2020

Published online 4 October 2020 in Wiley Online Library (www.bjsopen.com). DOI: 10.1002/bjs5.50358

Introduction

Unlike pharmaceuticals, where the introduction and evaluation of new medicines into clinical practice is highly regulated, the process for introducing new procedures and devices in surgery is inconsistent and lacks standardization.

This has resulted in several high-profile cases of potentially harmful interventions becoming established in clinical practice without robust evaluation, including vaginal mesh and metal-on-metal hip implants^{1,2}.

A key problem in the effective evaluation of new surgical procedures and devices is the lack of standardization in

outcome selection, measurement and reporting. Studies may measure benefits and harms in different ways, so that results cannot be directly compared or combined. Furthermore, investigators can choose which outcomes to measure and report, introducing bias. There is a need for standardized and transparent measurement and reporting of outcomes to improve safe and efficient evaluation of surgical innovation for introduction into clinical practice.

The development and use of a core outcome set (COS) – a minimum set of outcome domains to be measured and reported in all early-phase evaluations of innovative surgical procedures and devices – may be a solution to addressing inconsistent and heterogeneous outcome selection and reporting³. The methodology for the development of a COS in effectiveness studies is well established, including taxonomies to classify outcomes^{4,5}. However, standard methods to identify outcomes relevant to later-phase effectiveness studies, when the intervention under evaluation is no longer being modified, may not identify outcomes specifically relevant to evaluating surgical innovation. Data sources that specifically focus on innovation are required to provide insight into outcome selection and identify potentially relevant and meaningful outcomes for the first step in developing a COS in this setting.

The IDEAL (Idea, Development, Exploration, Assessment and Long-term monitoring) framework describes a pathway for evaluating surgical innovations from first-in-man through to long-term evaluation⁶. The framework has also recently been adapted for evaluation of new medical devices (IDEAL-D)⁷. It is hypothesized that investigators who have engaged with the IDEAL framework when designing and conducting studies may have considered outcome selection and reporting specifically in the context of evaluation of surgical innovation. As such, these studies may provide significant insight into outcomes that have particular relevance to the process of innovation, offering a valuable data source to identify outcomes and contribute to the conceptualization of outcome domains to inform the development of a COS for early-phase studies.

This study aimed to examine outcome selection, measurement and reporting in author-reported IDEAL/IDEAL-D studies to identify current practice, and serve as one of multiple data sources to help conceptualize novel domains relevant to evaluating innovation to inform the development of a COS⁸.

Methods

A systematic review and content analysis of author-reported IDEAL/IDEAL-D studies was

undertaken following PRISMA guidelines⁹, where applicable. This review is one of multiple data sources used to inform the wider study to develop a COS for studies of surgical innovation. A detailed protocol for this systematic review was not registered specifically as methods have been included in the COS development protocol, which has been published previously⁸.

Identification of studies

Electronic searches were performed to identify studies reported as following the IDEAL/IDEAL-D framework. Searches were undertaken in databases with citation tools (Web of Science and Scopus) to identify all publications citing any of ten key IDEAL/IDEAL-D papers^{6,7,10–17} deemed significant in describing the framework by members of the IDEAL collaboration¹⁸. Searches were carried out in April 2019. No restrictions to study design or publication dates were applied. Search results were imported into reference management software and duplicates removed. Results were filtered to select and retain records with the text word IDEAL appearing in the title or abstract.

Selection of studies

Titles and abstracts were screened independently for eligibility by two reviewers. Reports or protocols for studies described as IDEAL or IDEAL-D in the title or abstract were included. Eligible studies were: studies of innovative invasive procedures; studies of innovative devices; or studies where the innovation was something other than an invasive procedure/device but the study involved an invasive procedure, for example radiological imaging for guided biopsy. The latter category of studies was included owing to the potential value of such studies for identifying outcomes relevant to evaluating innovation. Invasive procedures were defined using an existing published definition and included ‘purposeful or deliberate access to the body gained via an incision, percutaneous puncture, where instrumentation is used in addition to the puncture needle, or instrumentation via a natural orifice’¹⁹. Systematic reviews, book chapters, secondary studies and studies that did not involve living human participants (such as cadaver, animal or simulation studies) were excluded. Studies using adaptations of the IDEAL framework that did not include or involve an invasive procedure or device (such as non-surgical complex interventions) were also excluded, along with letters, commentaries, editorials and conference proceedings. Screening results were compared for consistency and discrepancies resolved by discussion. Full-text copies of potentially relevant publications were

obtained and checked for eligibility. Reasons for excluding publications for which full-text copies had been obtained were recorded. Publications of uncertain relevance were discussed within the study team.

Data extraction

A standardized electronic data extraction form (REDCap software²⁰) was developed and piloted by the study team. Data extraction was performed by one reviewer with a second reviewer checking 10 per cent of publications for accuracy and completeness. Discrepancies were resolved by discussion and any additional verbatim outcomes identified during this process were extracted.

Study characteristics were extracted including: year of publication, details of the innovation, author-reported IDEAL stage, number of patients, participating centres and surgeons, and geographical origin of the study. All measured and/or reported outcomes were extracted verbatim together with any contextual information regarding the selection and measurement of outcomes. Rationale for outcome selection was categorized as: detailed, with a hypothesized effect (for example, anticipated improvement in at least one outcome); detailed, but with no hypothesized effect for any outcomes; general rationale (for example 'to examine safety'); or no rationale provided²¹. Contextual information considered potentially relevant to evaluating the process of innovation was extracted, where applicable. This included verbatim text relating to: stopping the innovation or making changes to the procedure/device in future use or application; limitations in study interpretation or conclusion in relation to outcomes; outcome assessment in future studies; and assessment of the surgical learning curve²². In studies where modifications to the procedure or patients selected to undergo the procedure were reported, details of how these were reported (for example, text, graphs or tables) were extracted. Data were recorded in a specifically designed study database²⁰.

Data analysis

Data analysis was undertaken in two stages. Extracted verbatim outcomes were categorized individually into domains to develop a preliminary framework of outcome domains relevant to innovation. Categorization was performed by one reviewer. During this process, the conceptual framework was modified iteratively until data saturation was reached and no new outcome domains were identified (all outcomes had been categorized). A provisional conceptual framework of domains was subsequently attained. A subsample of outcomes (81, from 2 publications selected at random) were independently categorized

by a second reviewer with clinical expertise to ensure methodological rigour. Discrepancies were resolved by discussion with the wider study team.

In the second stage of analysis, all outcomes were recategorized using the provisional conceptual framework by a second independent reviewer. Source documents (IDEAL/IDEAL-D study publications) were referred to for context, for example, to determine whether clinical outcomes were anticipated or unanticipated. Outcomes that were not specific enough to categorize (for example, generic descriptions such as safety or feasibility) were coded as 'too broad for categorization'. Outcomes that did not specifically address the evaluation of the innovation but were routinely measured data variables of interest to the clinical specialty were coded as common data elements²³.

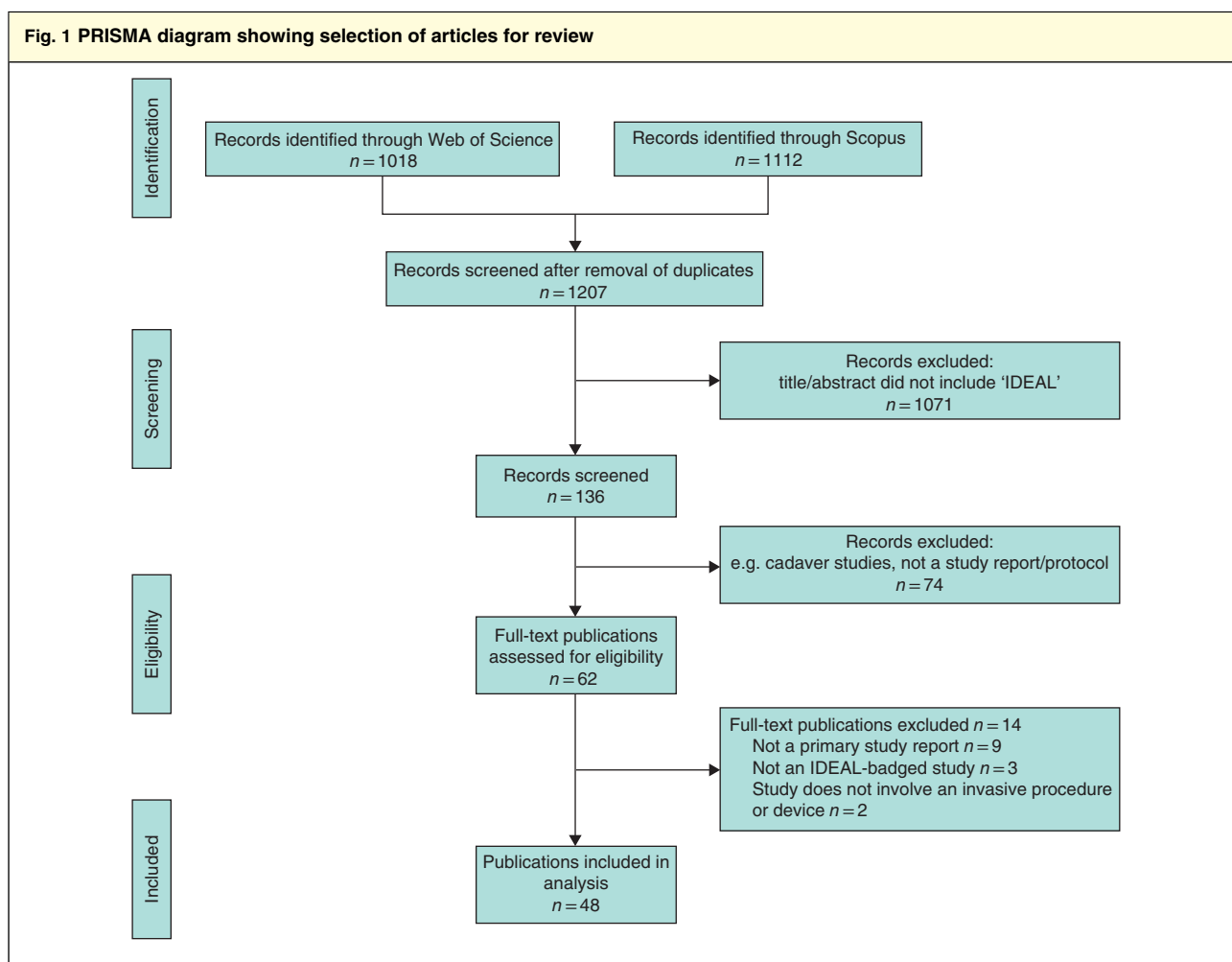
Outcome domains within the provisional conceptual framework were further classified by the authors as being innovation-specific or shared with effectiveness studies. Innovation-specific domains included non-traditional outcome domains (different to outcome domains typically measured in later-phase studies) considered to be conceptually specific to evaluating the ongoing process of innovation when the procedure/device was new or evolving. Examples included modifications to the surgical procedure/device and whether the procedure was completed successfully. Domains shared with effectiveness studies encompassed outcomes that would commonly be measured in later-phase studies⁴ considered not unique to evaluating innovation. Examples included mortality and patient-reported outcomes such as postoperative pain and physical function.

Study characteristics, measured/reported outcomes (based on the derived conceptual framework of domains) and contextual information relevant to selecting or measuring outcomes or evaluating innovation were compared across IDEAL stages specifically reported by the authors. Descriptive statistics were summarized using Microsoft Excel[®] (Microsoft, Redmond, Washington, USA) and STATA[®] statistical software (StataCorp, College Station, Texas, USA).

Assessment of factors relevant to systematic reviews of effectiveness (such as study quality or risk of bias) and data syntheses (for example meta-analyses of outcome data) were not appropriate given the exploratory aims of this review. Findings were synthesized and tabulated using descriptive statistics and a narrative summary directed at the study aims.

Results

After removal of duplicates, the search yielded 1207 records citing any of the ten key IDEAL/IDEAL-D publications.



Of these, 136 (11.3 per cent) included IDEAL as a text word in the title or abstract. Initial screening identified 62 of 136 records (45.6 per cent) for full-text eligibility assessment. Finally, 48 relevant publications (42 study reports, 6 protocols) were included in the analysis²⁴⁻⁷¹ (Fig. 1).

Included publications are detailed in Table 1. The majority were published in 2017 or 2018 (14 and 13 respectively). Thirty-nine publications were from Europe. Twenty-four described studies of innovations in surgical procedures and ten described studies of four different new devices. The remaining 14 publications described non-surgical innovations that included a surgical procedure. The specific IDEAL stage(s) describing the evaluative stage of the innovation was generally reported. Two publications, however, did not specify the IDEAL stage. Frequencies of author-reported IDEAL stages were: stage 1 (Idea, 11); stage 2a (Development, 16); stage 2b (Exploration, 9); stage 3 (Assessment, 1); and multiple stages (9).

Characteristics of IDEAL/IDEAL-D studies by stage

Characteristics of studies across the different IDEAL stages (stated in 46) are summarized in Table 2. The number of centres involved, number of surgeons/operators and number of included patients were broadly in line with the recommendations of the IDEAL framework. For example, early-phase studies (such as IDEAL stages 1 and 2a) typically involved one or two centres, one or two surgeons and a small number of patients, with numbers increasing in stage 2b and stage 3 studies. The exception was one IDEAL-D stage 1 study⁵⁰ describing the use of a type of mesh for cystocele repair that included 37 women. The median number of outcomes per study was relatively similar across IDEAL stages 1, 2a and 2b (median 32, 38 and 36 respectively). The range in number of outcomes described per study, however, varied considerably. For example, in

Table 1 Characteristics of included publications	
	No. of publications (n = 48)
Type of publication	
Study report	42
Protocol	6
Year of publication	
2019	1
2018	13
2017	14
2016	6
2015	4
2014	5
2013	0
2012	2
2011	3
Country of study	
Europe (non-UK)	33
Asia	8
UK	6
North America	1
Type of innovation	
Innovative procedure	24
Device	10
Other innovation that involved surgery*	14
Author-reported IDEAL stage	
1	11
2a	16
2b	9
3	1
Multiple stages	9
Not stated	2

*Such as radiological imaging.

IDEAL stage 1 studies, the number of outcomes described per study ranged from six to 82.

Outcome selection, measurement and reporting in IDEAL/IDEAL-D studies

Most studies (37 of 48) provided a rationale for outcome selection. Often (20 studies) this was a general rationale (for example 'to examine safety'). Fewer studies (12) provided detail on the expected or anticipated effect on outcomes. Instead, outcomes were reported to be measured without explanation or elaboration regarding their selection. There were no observed differences in the level of detail provided in the rationale between studies of different IDEAL stages (Table 2). Studies that mentioned stopping the innovation or making changes to the procedure/device in future applications of its use were broadly in line with the recommendations of the IDEAL framework. For example, 15 of 42 study reports (excluding protocols) included text to this

effect, most of these being IDEAL stage 1 or 2a studies where innovations are less stable.

Limitations to the interpretation of findings or study conclusions in relation to outcomes were discussed in 18 of 42 study reports. These included, for example, the short follow-up time for measuring outcomes and being underpowered to detect potential adverse events with a low incidence. Recommendations/suggestions for outcome assessment in subsequent future studies were discussed in 30 of 42 study reports. These were primarily about generic outcomes, such as measuring efficacy and safety. Four studies provided specific recommendations for outcome assessment in future studies, including two that recommended the inclusion of patient-reported outcome measures. Fewer than half of the studies (20 of 48) mentioned the learning curve. The frequency of mentioning the learning curve did not differ by IDEAL stage.

Types of outcome

A total of 1737 outcomes were extracted from the 48 studies and categorized into outcome domains. The derived conceptual framework comprised 32 domains, 22 of which were considered to be conceptually specific to evaluating innovation and ten to be shared with effectiveness studies (Table 3). The majority of outcomes (1098, 63.2 per cent) were categorized into domains shared with effectiveness studies (Table S1, supporting information), for example, outcomes assessing whether the overall desired effect of the procedure/device had been achieved, such as the number of positive surgical margins in a study of a new surgical procedure in oncology²⁸. Other outcome domains shared with effectiveness studies included anticipated disadvantages, such as adverse events and complications, duration of the procedure, duration of hospital stay, and patient's physical/psychological experiences after the procedure. A smaller number of outcomes (552, 31.8 per cent) were categorized into the 22 domains considered to be conceptually specific to evaluating innovation. These are described in detail below under seven broad subheadings.

Innovation-specific outcome domains

Procedure completion success/failure

Some 107 outcomes (6.2 per cent) were categorized as relating to the success or failure of performing the innovative procedure or using the innovative device (Table S1, supporting information). Examples of success included outcomes assessing whether all the steps of performing the innovative procedure were completed as planned. Examples of failure included the number of patients for which the planned innovative procedure was abandoned or

Table 2 Comparison of study characteristics, outcome selection and reporting in publications by IDEAL stage (n = 46)

	IDEAL stage				
	1 (n = 11)	2a (n = 16)	2b (n = 9)	3 (n = 1)	Multistage (n = 9)
Study characteristics					
No. of centres					
Single centre	11	15	6	0	6
Multicentre (2 centres)	0	1	0	1	1
Multicentre (> 2 centres)	0	0	2	0	1
Not reported	0	0	1	0	1
No. of surgeons/operators					
1	4	6	1	0	1
2	1	2	0	0	1
> 2	0	1	2	1	4
Not reported/unclear	6	7	6	0	3
No. of included patients					
< 10	7	0	0	0	0
10–20	3	11	1	0	2
21–50	1	4	1	0	3
> 50	0	1	7	1	4
Not reported	0	0	0	0	0
No. of verbatim outcomes per study*	32 (6–82)	38 (15–56)	36 (11–74)	17 (17)	30 (10–63)
Reported context for the selection and measurement of outcomes					
Rationale for outcome selection					
Detailed, with hypothesized effect for at least one outcome	3	4	4	0	1
Detailed, but with no hypothesized effect for any outcomes	2	2	0	0	1
General rationale	4	6	4	1	3
No rationale provided	2	4	1	0	4
Reported context relating to outcomes and evaluating innovation					
Text relating to stopping the innovation/making changes to the procedure/device in future applications of use†	5	6	0	n.a.	4
Reported limitations in interpretation/conclusions in relation to outcomes†	5	4	4	n.a.	5
Outcome assessment in future studies†	9	9	5	n.a.	6
Any mention of a learning curve	4	8	5	0	2

*Values are median (range). The analysis includes 46 studies where the authors stated an IDEAL stage. †Excluding protocols. n.a., Not applicable (protocol).

changed to an alternative procedure. Procedure completion success/failure outcomes were measured or reported in most studies (33 of 48) (Table 4). Relative consistency was observed in the proportion of studies reporting success/failure outcomes across the IDEAL stages.

Modifications

Some 92 outcomes (5.3 per cent) were categorized as relating to modifications. This included modifications to the individual steps of the planned innovative procedure (for example, a technical refinement to the procedure, and the number or proportion of patients for whom this was

required). Modification outcomes also included changes to any accompanying intervention related to the innovative procedure. For example, one study⁴² exploring a new method of transvesical suprapubic externalization of ureteral stents reported that modification from local anaesthesia to midazolam sedation was required. Modifications also included changes to patient selection during the course of the study.

Modifications were reported in 24 studies. The number of studies reporting modifications was broadly in line with the IDEAL framework, the majority being IDEAL stage

Broad classification	Domain number	Outcome domain	
Innovation-specific domains	1–2	Procedure completion success/failure	
	3–6	Modifications: to the procedure; to concomitant interventions; and to patient selection during study	
	7–9	Unanticipated advantages: during the procedure; after the procedure – short term; and after the procedure – long term	
	10–12	Unanticipated disadvantages: during the procedure; after the procedure – short term; and after the procedure – long term	
	13	Surgeon/operator's experience of the innovative procedure/device	
	14–15	Patient's experience of the innovative procedure/device, including: physical experiences during procedure, if applicable; and psychological experience of having the innovative procedure/device	
	16–19	Required resource use specific to the innovative procedure: before the procedure; during the procedure; after the procedure during the hospital stay; and after leaving hospital	
	20	Details of patients suitable for the procedure in future	
	21	Details of operator training/expertise necessary to perform the procedure in future	
	22	Mechanical/technical problems with device, if applicable	
	Domains shared with effectiveness studies	23	Overall desired effect of procedure/device achieved
		24–26	Anticipated advantages: during the procedure; after the procedure – short term; and after the procedure – long term
		27–29	Anticipated disadvantages: during the procedure; after the procedure – short term; and after the procedure – long term
20		Duration of procedure	
31		Duration of hospital stay	
32		Patient's physical/psychological experiences after the procedure	

	IDEAL stage				
	1 (n = 11)	2a (n = 16)	2b (n = 9)	3 (n = 1)	Multistage (n = 9)
Outcomes relating to the success or failure of performing the procedure/ using the device	8	12	5	1	7
Outcomes relating to modifications	5	11	3	0	5
Outcomes relating to unanticipated event (advantages/disadvantages)	1	5	0	0	2
Outcomes relating to surgeon/operator experience	5	8	3	0	7
Outcomes relating to patient experience	1	4	1	0	0
Outcomes relating to resource use	9	9	6	0	7

The analysis includes 46 studies where the authors stated an IDEAL stage.

2a (innovations in early development). Most modifications were reported using narrative text alone (11 studies) and less frequently using flow diagrams, figures or graphs.

Unanticipated events (advantages and disadvantages)

Outcomes that related to an unexpected advantage or disadvantage were categorized as unanticipated events. These were further defined as occurring during or after the procedure. At least one unanticipated event was reported in eight of 42 studies (excluding protocols). These were

mostly disadvantages, such as clinical complications (for example, peritonitis due to spontaneous perforation of a sigmoid diverticulum in a study of robot-assisted anterior partial prostatectomy⁷⁰). In one study³⁴, an unanticipated event was reported as advantageous by serving as a useful safety check for the procedure.

Surgeon/operator experience of innovative procedure/device

Some 65 outcomes (3.7 per cent) were categorized as relating to surgeons' experiences of performing the innovative

procedure or experiences of operators using the new device (if not a surgeon). Examples included ergonomics, such as pain or discomfort when carrying out the procedure, or surgeons' reports of how difficult or easy it was to perform the procedure. Further outcomes categorized under this domain included descriptions of problematic points in undertaking the procedure and impressions of the learning experience. Although 23 studies included outcomes relating to surgeon or operator experiences, only three^{33,45,58} reported that experiences were assessed using a formal process (including a rating scale/score, questionnaire and discussion in review meetings between cases). There was a slight trend for surgeon/operator experiences to be measured/reported in studies in the earlier stages of the innovation lifecycle (IDEAL stage 1 and 2a).

Patient experiences of innovative procedure/device

Overall, 215 outcomes (12.4 per cent) were categorized as assessing patients' experiences. The majority (200, 11.5 per cent) were measured after operation and were similar to those measured in effectiveness studies, such as patient-reported quality of life and pain. A minority of outcomes (15, 0.9 per cent; in 6 different studies), however, were specifically relevant to the experiences of undergoing the innovative procedure or receiving treatment with the innovative device. One study³⁹ of percutaneous nephrolithotomy under local anaesthesia, for example, reported patients' experiences specific to undergoing the innovative procedure itself, made possible because the procedure did not require the patient to be under general anaesthetic. Outcomes included patient 'disquietness' (restlessness or uneasiness) and good perioperative cooperation. Further examples of patients' experiences specific to undergoing the innovative procedure or receiving treatment with the innovative device related to psychological or emotional experiences. Details of how these outcomes were measured, however, were lacking. Most studies reporting innovation-specific patient experience outcomes were reported in IDEAL stage 2a (Table 4).

Resource use specific to innovative procedure/device

A small number of outcomes (26, 1.5 per cent) measured resource use specific to the innovative procedure/device. Examples included the cost of the procedure/device or associated costs (such as equipment sterilization costs) and the number of further treatments required owing to complications that had developed as a result of the innovative procedure. Some 31 studies included any kind of outcome relating to resource use (Table 4); however, the majority of outcomes were shared with those typically measured in later-phase studies, such as duration of surgery

(72 outcomes, 4.1 per cent) and duration of hospital stay (23, 1.3 per cent).

Other innovation-specific outcome domains

These domains included: details of patients suitable for the procedure in future; details of operator training/expertise necessary to perform the procedure in future; and mechanical/technical problems with the device. Although few outcomes were categorized into these domains (35, 2.0 per cent), they were considered relevant outcomes important for the evaluation of innovation.

Discussion

Although the majority of outcomes extracted reflected those traditionally measured and reported in effectiveness studies⁴, this review identified several novel innovation-specific outcome domains that reflect the dynamic process of surgical innovation. Broadly, these encompassed outcomes relating to procedure completion success/failure, modifications (to the procedure, concomitant interventions, or patient selection), unanticipated events, and innovation-specific surgeon/operator experience, patient experience and resource use. Although it is recognized that unanticipated events may also occur and are reported in effectiveness studies, they are generally rare and less likely to occur when the procedure has stopped evolving and has stabilized. Unanticipated events may indicate unexpected problems and, as such, be a key driver for modifying the procedure/device, particularly if they are related to patient safety. Unanticipated events were, therefore, considered to be of key importance for evaluating innovation and conceptualized as an innovation-specific domain. These findings served as one of multiple data sources to inform a Delphi survey and consensus study to develop a COS for early-phase studies of surgical procedures and devices⁸.

The heterogeneity observed in the number of studies measuring and reporting outcomes of specific relevance to evaluating innovation highlights the need for more detailed guidance and improved standardization of outcomes that are important to measure when evaluating surgical innovation. The IDEAL framework provides clear guidance regarding the characteristics of studies at each stage of evaluation of surgical innovation, but uses only broad terms to describe the types of outcome that should be measured at different stages of evaluation. The lack of specific guidance around the selection and measurement of recommended outcomes may be one reason why the widespread uptake of IDEAL has been relatively limited^{18,72}.

The review was conducted using robust and established methodology, but has limitations. Methods to identify

IDEAL/IDEAL-D studies selected publications that had included IDEAL as a text word in the title or abstract. This pragmatic filtering step to identify relevant studies was based on the rationale that studies firmly engaging with the IDEAL framework, and therefore most likely to have more thoroughly considered outcomes in the context of evaluating innovation, may be more likely to describe their study as an IDEAL study in the title or abstract. It is accepted that some studies undertaken within the IDEAL framework may have been missed as a result of this search strategy. The number of studies identified (48) may be considered small given that the IDEAL framework was published 10 years ago. This number is, however, consistent with a recent review¹⁸ examining the uptake and use of the IDEAL framework. The high proportion of studies published in recent years (2017–2018) suggests a recent increase in uptake of the IDEAL framework¹⁸. It is acknowledged that publications regarding themselves as IDEAL studies may not necessarily have followed the IDEAL framework from the outset. Authors may have been asked to report the IDEAL stage in their reports by journal editors and reviewers during the publication process. The selection and measurement of outcomes may, therefore, have been chosen independently of the IDEAL framework and recommendations. Analysis of data relied on author-reported information, which may not have always been precise. For example, comparison of studies across the IDEAL stages relied on the author-reported classification of IDEAL stage. It is recognized that classification of studies can be complex and author reports of IDEAL stage may not always be accurate¹⁸. Similarly, categorization of outcomes as unanticipated (rather than anticipated) also relied on author-reported information. If authors did not specifically report that an outcome was unanticipated, it was categorized as anticipated. It is possible, therefore, that the number of unanticipated events in the studies may have been higher. A final limitation is that individual outcomes were categorized into a single domain. In some instances, an outcome could have been included in more than one domain; for example, the overall desired effect of the procedure may have been measured by a patient-reported outcome. Reporting of the number of outcomes within the different outcome domains was, therefore, dependent on the judgement of the reviewers involved in the outcome categorization/checking process. Although exact reporting of the number of outcomes relevant to each domain was not a key objective in the present study, future studies where this is important could follow the methodology recommended for classifying outcomes from effectiveness studies and classify outcomes into multiple relevant domains⁴.

The identification of innovation-specific outcome domains is an important first step in developing a generic COS to standardize outcome selection, measurement and reporting in studies of novel surgical procedures. Such a COS would be appropriate for use before definitive evaluation within an RCT and is intended to support standardized evaluation of surgical procedures and medical devices throughout the stages of the innovation lifecycle, to facilitate safe and transparent introduction. The present review is one of several data sources to inform a conceptual framework for outcome domains for the COS to ensure its validity and comprehensiveness⁸. Given the diversity of procedures that could be performed, the feasibility of developing a generic COS for all innovative procedures could be questioned. The innovation-specific domains identified in this review, however, are sufficiently broad that they could be applied across a wide range of procedures and devices. An international consensus study involving a diverse range of key stakeholders to agree a final core set of domains for early-phase studies has now been completed⁸. This includes the development of clear reporting guidelines to ensure that the COS is adopted and reported in a meaningful way. A further challenge will be to determine how best to measure these key outcome domains. Engagement and creation of appropriate resources with key stakeholders, including surgeons from various disciplines, will be required to optimize the uptake and value of these tools.

Acknowledgements

K.N.L.A. and S.P. are joint senior authors. The authors acknowledge the COHESIVE study group and study steering group for their advice while undertaking this review, and K. Chalmers, Bristol Centre for Surgical Research, University of Bristol, for help with screening abstracts. S.P. is a National Institute for Health Research (NIHR) Clinician Scientist (CS-2016-16-019). A.G.K.M. is an NIHR Clinician Scientist (NIHR-CS-2017-17-010). J.M.B. is an NIHR Senior Investigator. This work was funded by the NIHR Biomedical Research Centre at the University Hospitals Bristol and Weston NHS Foundation Trust and the University of Bristol. This work was supported by the Royal College of Surgeons of England Bristol Surgical Trials Centre and the Medical Research Council (MRC) Collaboration and innovation for Difficult and Complex randomised controlled Trials In Invasive procedures (ConDuCT-II) Hub for Trials Methodology Research (MR/K025643/1). The views and opinions expressed in this publication are those of the authors and not necessarily those of the UK National Health Service, NIHR, Department of Health

and Social Care, Royal College of Surgeons of England or MRC.

Disclosure: The authors declare no conflict of interest.

References

- Heneghan C, Aronson JK, Goldacre B, Mahtani KR, Pluddemann A, Onakpoya I. Transvaginal mesh failure: lessons for regulation of implantable devices. *BMJ* 2017; **359**: j5515.
- Cohen D. How safe are metal-on-metal hip implants? *BMJ* 2012; **344**: e1410.
- Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E *et al.* Developing core outcome sets for clinical trials: issues to consider. *Trials* 2012; **13**: 132.
- Dodd S, Clarke M, Becker L, Mavergames C, Fish R, Williamson PR. A taxonomy has been developed for outcomes in medical research to help improve knowledge discovery. *J Clin Epidemiol* 2018; **96**: 84–92.
- Williamson PR, Altman DG, Bagley H, Barnes KL, Blazeby JM, Brookes ST *et al.* The COMET handbook: version 1.0. *Trials* 2017; **18**: 280.
- McCulloch P, Altman DG, Campbell WB, Flum DR, Glasziou P, Marshall JC *et al.* No surgical innovation without evaluation: the IDEAL recommendations. *Lancet* 2009; **374**: 1105–1112.
- Sedrakyan A, Campbell B, Merino JG, Kuntz R, Hirst A, McCulloch P. IDEAL-D: a rational framework for evaluating and regulating the use of medical devices. *BMJ* 2016; **353**: i2372.
- Avery K, Blazeby J, Wilson N, Macefield R, Cousins S, Main B *et al.* Development of reporting guidance and core outcome sets for seamless, standardised evaluation of innovative surgical procedures and devices: a study protocol for content generation and a Delphi consensus process (COHESIVE study). *BMJ Open* 2019; **9**: e029574.
- Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009; **6**: e1000097.
- Barkun JS, Aronson JK, Feldman LS, Maddern GJ, Strasberg SM, Altman DG *et al.* Evaluation and stages of surgical innovations. *Lancet* 2009; **374**: 1089–1096.
- Ergina PL, Cook JA, Blazeby JM, Boutron I, Clavien PA, Reeves BC *et al.* Challenges in evaluating surgical innovation. *Lancet* 2009; **374**: 1097–1104.
- Cook JA, McCulloch P, Blazeby JM, Beard DJ, Marinac-Dabic D, Sedrakyan A; IDEAL Group. IDEAL framework for surgical innovation 3: randomised controlled trials in the assessment stage and evaluations in the long term study stage. *BMJ* 2013; **346**: f2820.
- Ergina PL, Barkun JS, McCulloch P, Cook JA, Altman DG; IDEAL Group. IDEAL framework for surgical innovation 2: observational studies in the exploration and assessment stages. *BMJ* 2013; **346**: f3011.
- Hirst A, Agha RA, Rosin D, McCulloch P. How can we improve surgical research and innovation?: the IDEAL framework for action. *Int J Surg* 2013; **11**: 1038–1042.
- McCulloch P, Cook JA, Altman DG, Heneghan C, Diener MK; IDEAL Group. IDEAL framework for surgical innovation 1: the idea and development stages. *BMJ* 2013; **346**: f3012.
- Pennell CP, Hirst A, Sedrakyan A, McCulloch PG. Adapting the IDEAL framework and recommendations for medical device evaluation: a modified Delphi survey. *Int J Surg* 2016; **28**: 141–148.
- Pennell CP, Hirst AD, Campbell WB, Sood A, Agha RA, Barkun JS *et al.* Practical guide to the idea, development and exploration stages of the IDEAL framework and recommendations. *Br J Surg* 2016; **103**: 607–615.
- Khachane A, Philippou Y, Hirst A, McCulloch P. Appraising the uptake and use of the IDEAL framework and recommendations: a review of the literature. *Int J Surg* 2018; **57**: 84–90.
- Cousins S, Blencowe NS, Blazeby JM. What is an invasive procedure? A definition to inform study design, evidence synthesis and research tracking. *BMJ Open* 2019; **9**: e028576.
- Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L *et al.* The REDCap consortium: building an international community of software platform partners. *J Biomed Inform* 2019; **95**: 103208.
- McNair AG, Macefield RC, Blencowe NS, Brookes ST, Blazeby JM. 'Trial exegesis': methods for synthesizing clinical and patient reported outcome (PRO) data in trials to inform clinical practice. A systematic review. *PLoS One* 2016; **11**: e0160998.
- Subramonian K, Muir G. The 'learning curve' in surgery: what is it, how do we measure it and can we influence it? *BJU Int* 2004; **93**: 1173–1174.
- Karaa A, Rahman S, Lombes A, Yu-Wai-Man P, Sheikh MK, Alai-Hansen S *et al.* Common data elements for clinical research in mitochondrial disease: a National Institute for Neurological Disorders and Stroke project. *J Inherit Metab Dis* 2017; **40**: 403–414.
- Kroeze SGC, Huisman M, Verkooijen HM, van Diest PJ, Bosch J, van den Bosch M. Real-time 3D fluoroscopy-guided large core needle biopsy of renal masses: a critical early evaluation according to the IDEAL recommendations. *Cardiovasc Intervent Radiol* 2012; **35**: 680–685.
- Swaan A, Mannaerts CK, Scheltema MJV, Nieuwenhuijzen JA, Savci-Heijink CD, de la Rosette J *et al.* Confocal laser endomicroscopy and optical coherence tomography for the diagnosis of prostate cancer: a needle-based, *in vivo* feasibility study protocol (IDEAL phase 2A). *JMIR Res Protoc* 2018; **7**: e132.
- Valerio M, Dickinson L, Ali A, Ramachandran N, Donaldson I, Freeman A *et al.* A prospective development study investigating focal irreversible electroporation in men with localised prostate cancer: Nanoknife Electroporation Ablation Trial (NEAT). *Contemp Clin Trials* 2014; **39**: 57–65.
- Joukhadar R, Wockel A, Herr D, Paulus V, Radosa J, Hamza A *et al.* Challenges of longevity: safety of vaginal and laparoscopic urogynecological procedures in septuagenarians and older patients. *Biomed Res Int* 2016; **2016**: 1–9.

- 28 Kaouk JH, Haber GP, Autorino R, Crouzet S, Ouzzane A, Flamand V *et al.* A novel robotic system for single-port urologic surgery: first clinical investigation. *Eur Urol* 2014; **66**: 1033–1043.
- 29 Liem E, Freund JE, Baard J, de Bruin DM, Pes MPL, Savci-Heijink CD *et al.* Confocal laser endomicroscopy for the diagnosis of urothelial carcinoma in the bladder and the upper urinary tract: protocols for two prospective explorative studies. *JMIR Res Protoc* 2018; **7**: e34.
- 30 Lim R, Liang ML, Leong WS, Khan NAK, Yuen KH. Effect of pulsed magnetic stimulation on quality of life of female patients with stress urinary incontinence: an IDEAL-D stage 2b study. *Int Urogynecol J* 2018; **29**: 547–554.
- 31 Menon M, Abaza R, Sood A, Ahlawat R, Ghani KR, Jeong W *et al.* Robotic kidney transplantation with regional hypothermia: evolution of a novel procedure utilizing the IDEAL guidelines (IDEAL phase 0 and 1). *Eur Urol* 2014; **65**: 1001–1009.
- 32 Paleri V, Fox H, Coward S, Ragbir M, McQueen A, Ahmed O *et al.* Transoral robotic surgery for residual and recurrent oropharyngeal cancers: exploratory study of surgical innovation using the IDEAL framework for early-phase surgical studies. *Head Neck J Sci Special Head Neck* 2018; **40**: 512–525.
- 33 Saglam R, Muslumanoglu AY, Tokath Z, Caskurlu T, Sarica K, Tasci AI *et al.* A new robot for flexible ureteroscopy: development and early clinical results (IDEAL stage 1–2b). *Eur Urol* 2014; **66**: 1092–1100.
- 34 Solari V, Jawaid W, Jesudason EC. Enhancing safety of laparoscopic vascular control for neonatal sacrococcygeal teratoma. *J Pediatr Surg* 2011; **46**: E5–E7.
- 35 Sood A, Ghani KR, Ahlawat R, Modi P, Abaza R, Jeong W *et al.* Application of the statistical process control method for prospective patient safety monitoring during the learning phase: robotic kidney transplantation with regional hypothermia (IDEAL phase 2a–b). *Eur Urol* 2014; **66**: 371–378.
- 36 Kroeze SGC, Agenant M, Jonges GN, Stein T, Bosch J. Clinical efficacy of bipolar radiofrequency ablation of small renal masses. *World J Urol* 2015; **33**: 1535–1540.
- 37 del Val ID, Loureiro C, McCulloch P. The IDEAL prospective development study format for reporting surgical innovations. An illustrative case study of robotic oesophagectomy. *Int J Surg* 2015; **19**: 104–111.
- 38 Douglas AJW, Kyzas PA. A new autologous block-bone prefabricated flap concept based on the supraclavicular artery island flap (SCAIF) for reconstruction of a neo-mandibular osteoradionecrosis (ORN) defect, IDEAL stage 1 report. *JPRAS Open* 2017; **12**: 19–24.
- 39 Ecke TH, Barski D, Weingart G, Lange C, Hallmann S, Ruttloff J *et al.* Presentation of a method at the exploration stage according to IDEAL: percutaneous nephrolithotomy (PCNL) under local infiltrative anesthesia is a feasible and effective method – retrospective analysis of 439 patients. *Int J Med Sci* 2017; **14**: 302–309.
- 40 Gerullis H, Barski D, Georgas E, Borós M, Ramon A, Ecke TH *et al.* Protocol for a randomized phase II trial for mesh optimization by autologous plasma coating in prolapse repair: IDEAL stage 3. *Adv Ther* 2017; **34**: 995–1006.
- 41 Gerullis H, Ecke TH, Barski D, Bantel C, Weyland A, Uphoff J *et al.* Retrospective analysis of a surgical innovation using the IDEAL framework: radical cystectomy with epidural anaesthesia. *J Int Med Res* 2017; **45**: 714–722.
- 42 Gerullis H, Schwartmann K, Ecke TH, Georgas E, Herberg P, Winter A *et al.* Transvesical suprapubic externalization of ureteral stents – introduction of a surgical innovation at the development stage. *Urol Int* 2017; **99**: 69–76.
- 43 Hallmann S, Petersein J, Ruttloff J, Ecke TH. Successful evacuation of large perirenal hematoma after extracorporeal shock wave lithotripsy (ESWL) – step 1 of the IDEAL recommendations of surgical innovation. *Clin Case Rep* 2017; **5**: 123–125.
- 44 Heikens JT, Gooszen HG, Rovers MM, van Laarhoven C. Stages and evaluation of surgical innovation: a clinical example of the ileo neorectal anastomosis after ulcerative colitis and familial adenomatous polyposis. *Surg Innov* 2013; **20**: 459–465.
- 45 Inoue Y, Saiura A, Sato T, Ishizawa T, Arita J, Takahashi Y *et al.* Laparoscopic pancreatoduodenectomy combined with a novel self-assessment system and feedback discussion: a phase 1 surgical trial following the IDEAL guidelines. *Langenbecks Arch Surg* 2016; **401**: 1123–1130.
- 46 Bus MTJ, de Bruin DM, Faber DJ, Kamphuis GM, Zondervan PJ, Laguna-Pes MP *et al.* Optical coherence tomography as a tool for *in vivo* staging and grading of upper urinary tract urothelial carcinoma: a study of diagnostic accuracy. *J Urol* 2016; **196**: 1749–1755.
- 47 Diepstraten SCE, Verkooijen HM, van Diest PJ, Veldhuis WB, Fernandez-Gallardo AM, Duvivier KM *et al.* Radiofrequency-assisted intact specimen biopsy of breast tumors: critical evaluation according to the IDEAL recommendations. *Cancer Imaging* 2011; **11**: 247–252.
- 48 Buijs M, van Lienden KP, Wagstaff PGK, Scheltema MJV, de Bruin DM, Zondervan PJ *et al.* Irreversible electroporation for the ablation of renal cell carcinoma: a prospective, human, *in vivo* study protocol (IDEAL phase 2b). *JMIR Res Protoc* 2017; **6**: e21.
- 49 Baekelandt J, Bosteels J. Hysterectomy through the looking glass: iHysterectomy frugal by iPhone. *BMJ Innov* 2017; **3**: 71–75.
- 50 Barski D, Arndt C, Gerullis H, Yang J, Boros M, Otto T *et al.* Transvaginal PVDF-mesh for cystocele repair: a cohort study. *Int J Surg* 2017; **39**: 249–254.
- 51 Barski D, Gerullis H, Ecke T, Varga G, Boros M, Pintelon I *et al.* Repair of a vesico-vaginal fistula with amniotic membrane – step 1 of the IDEAL recommendations of surgical innovation. *Central Eur J Urol* 2015; **68**: 459–461.
- 52 Barentsz MW, Verkooijen HM, Pijnappel RM, Fernandez MA, van Diest PJ, van der Pol CC *et al.* Sentinel lymph node localization with contrast-enhanced ultrasound and an I-125

- seed: an ideal prospective development study. *Int J Surg* 2015; **14**: 1–6.
- 53 Banagala S, Jayarajah U, Almeida I, Samarasekera DN. Efficacy of irrigation tubes in the management of para rectal cavities associated with complex fistula-*in-ano*. *BMC Surg* 2018; **18**: 95.
- 54 Blazeby JM, Blencowe NS, Titcomb DR, Metcalfe C, Hollowood AD, Barham CP. Demonstration of the IDEAL recommendations for evaluating and reporting surgical innovation in minimally invasive oesophagectomy. *Br J Surg* 2011; **98**: 544–551.
- 55 Barski D, Gerullis H, Ecke T, Kranz J, Schneidewind L, Leistner N *et al*. Registry of implants for the reconstruction of pelvic floor in males and females: a feasibility case series. *Int J Surg* 2017; **42**: 27–33.
- 56 Chandak P, Byrne N, Lynch H, Allen C, Rottenberg G, Chandra A *et al*. Three-dimensional printing in robot-assisted radical prostatectomy – an idea, development, exploration, assessment, long-term follow-up (IDEAL) phase 2a study. *BJU Int* 2018; **122**: 360–361.
- 57 Chen J, Li Y, Wang Z, McCulloch P, Hu L, Chen W *et al*. Evaluation of high-intensity focused ultrasound ablation for uterine fibroids: an IDEAL prospective exploration study. *BJOG* 2018; **125**: 354–364.
- 58 Gaboardi F, Pini G, Suardi N, Montorsi F, Passaretti G, Smelzo S. Robotic laparoendoscopic single-site radical prostatectomy (R-LESS-RP) with daVinci Single-Site® platform. Concept and evolution of the technique following an IDEAL phase 1. *J Robot Surg* 2019; **13**: 215–226.
- 59 Graff P, Portalez D, Lusque A, Brun T, Aziza R, Khalifa J *et al*. IDEAL 2a phase II study of ultrafocal brachytherapy for low- and intermediate-risk prostate cancer. *Int J Radiat Oncol Biol Phys* 2018; **102**: 903–911.
- 60 Kidane B, Lung K, McCreery G, El-Khatib C, Ott MC, Hernandez-Alejandro R *et al*. Early rescue from acute severe *Clostridium difficile*: a novel treatment strategy. *Surg Infect* 2018; **19**: 78–82.
- 61 Mannaerts CK, Wildeboer RR, Postema AW, Hagemann J, Budaus L, Tilki D *et al*. Multiparametric ultrasound: evaluation of greyscale, shear wave elastography and contrast-enhanced ultrasound for prostate cancer detection and localization in correlation to radical prostatectomy specimens. *BMC Urol* 2018; **18**: 98.
- 62 Minervini A, Vanacore D, Vittori G, Milanesi M, Tuccio A, Siena G *et al*. Florence robotic intracorporeal neobladder (FloRIN): a new reconfiguration strategy developed following the IDEAL guidelines. *BJU Int* 2018; **121**: 313–317.
- 63 Breda A, Territo A, Guttilla A, Sanguedolce F, Manfredi M, Quaresima L *et al*. Correlation between confocal laser endomicroscopy (Cellvizio®) and histological grading of upper tract urothelial carcinoma: a step forward for a better selection of patients suitable for conservative management. *Eur Urol Focus* 2018; **4**: 954–959.
- 64 Greco F, Pini G, Alba S, Altieri VM, Verze P, Mirone V. Minilaparoscopic single-site pyeloplasty: the best compromise between surgeon's ergonomics and patient's cosmesis (IDEAL phase 2a). *Eur Urol Focus* 2016; **2**: 319–326.
- 65 Rischmann P, Gelet A, Riche B, Villers A, Pasticier G, Bondil P *et al*. Focal high intensity focused ultrasound of unilateral localized prostate cancer: a prospective multicentric hemiablation study of 111 patients. *Eur Urol* 2017; **71**: 267–273.
- 66 Springer C, Kawan F, La Rocca R, Mohammed N, Fornara P, Mirone V *et al*. New hybrid mini-laparoscopic single-site partial nephrectomy with early unclamped technique for renal tumors with intermediate PADUA score (IDEAL phase 2a). *Urology* 2018; **111**: 104–109.
- 67 Stenstra M, van Workum F, van den Wildenberg FJH, Polat F, Rosman C. Evolution of the surgical technique of minimally invasive Ivor-Lewis esophagectomy: description according to the IDEAL framework. *Dis Esophagus* 2019; **32**: doy122.
- 68 Versteeg AL, van der Velden JM, Hes J, Eppinga W, Kasperts N, Verkooijen HM *et al*. Stereotactic radiotherapy followed by surgical stabilization within 24 h for unstable spinal metastases; a stage I/IIa study according to the IDEAL framework. *Front Oncol* 2018; **8**: 626.
- 69 Sood A, McCulloch P, Dahm P, Ahlawat R, Jeong WJ, Bhandari M *et al*. Ontogeny of a surgical technique: robotic kidney transplantation with regional hypothermia. *Int J Surg* 2016; **25**: 158–161.
- 70 Villers A, Puech P, Flamand V, Haber GP, Desai MM, Crouzet S *et al*. Partial prostatectomy for anterior cancer: short-term oncologic and functional outcomes. *Eur Urol* 2017; **72**: 333–342.
- 71 van den Hoven AF, Prince JF, de Keizer B, Vonken E, Bruijnen RCG, Verkooijen HM *et al*. Use of C-arm cone beam CT during hepatic radioembolization: protocol optimization for extrahepatic shunting and parenchymal enhancement. *Cardiovasc Intervent Radiol* 2016; **39**: 64–73.
- 72 Hirst A, Philippou Y, Blazeby J, Campbell B, Campbell M, Feinberg J *et al*. No surgical innovation without evaluation: evolution and further development of the IDEAL framework and recommendations. *Ann Surg* 2019; **269**: 211–220.

Supporting information

Additional supporting information can be found online in the Supporting Information section at the end of the article.