Current status of validation for robotic surgery simulators – a systematic review

Hamid Abboudi, Mohammed S. Khan, Omar Aboumarzouk*, Khurshid A. Guru†, Ben Challacombe, Prokar Dasgupta and Kamran Ahmed

*MRC Centre for Transplantation, King's College London, King's Health Partners, Department of Urology, Guy's Hospital, London, *Department of Urology, Aberdeen Royal Infirmary, Aberdeen, UK, and †Department of Urology, Roswell Park Center for Robotic Surgery, Roswell Park Cancer Institute, Buffalo, New York, USA*

What's known on the subject? and What does the study add?

- Little is known on how best to train the future generation of robotic surgeons. It has been postulated that virtual reality (VR) simulators may aid the progression along the learning curve for this rapidly developing surgical technique within a safe training environment. There are several simulators available on the market, the best known is that developed by Intuitive Surgical Inc.
- The present study provides the first systematic review of all the trails of the various VR robotic platforms. It explores the evidence supporting the effectiveness of the various platforms for feasibility, reliability, validity, acceptability, educational impact and cost-effectiveness. This article also highlights the deficiencies and future work required to advance robotic surgical training.

To analyse studies validating the effectiveness of robotic surgery simulators. The MEDLINE®, EMBASE® and PsycINFO® databases were systematically searched until September 2011. References from retrieved articles were reviewed to broaden the search. The simulator name, training tasks, participant level, training duration and evaluation scoring were extracted from each study. We also extracted data on feasibility, validity, cost-effectiveness, reliability and educational impact. We identified 19 studies investigating simulation options in robotic surgery. There are five different robotic surgery simulation platforms available on the market. In all, 11 studies sought opinion and compared performance between two different groups; 'expert' and 'novice'. Experts ranged in experience from 21–2200 robotic cases. The novice groups consisted of participants with no prior experience on a robotic platform and were often medical students or junior doctors. The Mimic dV-Trainer®, ProMIS®, SimSurgery Educational Platform® (SEP) and Intuitive systems have shown face, content and construct validity. The Robotic Surgical SimulatorTM system has only been face and content validated. All of the simulators except SEP have shown educational impact. Feasibility and cost-effectiveness of simulation systems was not evaluated in any trial. Virtual reality simulators were shown to be effective training tools for junior trainees. Simulation training holds the greatest potential to be used as an adjunct to traditional training methods to equip the next generation of robotic surgeons with the skills required to operate safely. However, current simulation models have only been validated in small studies. There is no evidence to suggest one type of simulator provides more effective training than any other. More research is needed to validate simulated environments further and investigate the effectiveness of animal and cadaveric training in robotic surgery.

Keywords

robotics, urology, surgery, training, simulation, education

Introduction

Since its first application in 2000 by Binder and Kramer [1], the uptake of robotic surgery has been rapid [2]. This ever evolving surgical technique has cemented itself as the 'gold standard' operative procedure for the removal of the prostate gland. In 2007 it was estimated that 68% of radical prostatectomies in the USA were performed using robotic assistance [3]. The potential benefits of robotic surgery are multiple including; shorter recovery time, less postoperative pain, lower blood loss and improved cosmesis. The most commonly used system is the da Vinci® Surgical System (dVSS) from Intuitive Surgical, Inc., and consists of two main components; master console and a slave robot [2,4]. A surgeon provides input through manipulation of the master console which, in turn, controls a slave robot to perform the necessary motions on the patient.

Robotic training poses several unique challenges to educators, trainees and training programme directors alike. During conventional open and laparoscopic surgery the mentoring surgeon is adjacent to the trainee and has the same view of the procedure, as well as being able to take over at any given moment where patient safety may be compromised. This is currently not the case in robotic assisted procedures as only one surgeon can be at the operating console at any given time thus competency before embarking on robotic procedures is paramount. From the trainees' perspective, with limits in working hours, fear of litigation and financial constraints, the prospect of training in robotic surgery seems a daunting task given the individual nature of the surgery. Trainees and programme directors have recognised that 'on the job' training will be difficult in this context and are therefore turning to alternative methods to solve the robotic training conundrum, namely robotic fellowships and simulation training.

Surgical simulation has advanced tremendously over the last two decades with the development of laparoscopic and now robotic surgery. This novel approach to surgical training has been validated as a training and assessment tool and has been shown to improve a surgeon's performance in the operating room [5–7].

Surgical simulator training can be separated into two broad categories: physical (mechanical) simulators, in which the task is performed under videoscopic guidance within usually a box trainer and 'virtual reality' (VR) simulators, where the task is performed on a computer-based platform and artificially generated virtual environments. Improvements in computer processing have led to more realistic and sensitive VR simulators, which are now capable of providing statistical feedback on the surgeons performance, a quality that is not shared by mechanical or cadaveric simulator trainers.

Before a surgical simulator can be used to assess the competency of surgeons, the simulator must undergo initial testing across a variety of parameters. This would include the assessment of face validity, which examines the realism of the simulator; construct validity, is it able to differentiate novice from experienced operators; context validity, examines whether the device can teach what it is supposed to teach; concurrent validity, the extent to which the results of the test correlate with the 'gold standard' tests known to measure the same domain; and predictive validity, the extent to which an assessment will predict future performance [8–10] (Fig. 1). The validity of mechanical and VR simulators in the context of laparoscopic surgery has been established. However, their effectiveness in training surgeons on robotic surgical systems is less clear.

In this systematic review we identified available robotic surgery simulators, explored the evidence supporting the effectiveness of the various platforms in terms of feasibility, reliability, validity, acceptability, educational impact and cost-effectiveness. This article also highlights the deficiencies and future work required to advance robotic surgical training.

Materials and Methods

This study was performed following guidelines defined in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement (http://www.prisma-statement.org/) [11].

Study Eligibility Criteria

Empirical studies describing the types, development, validation and use of simulation for training and assessment purposes in robotic surgery were included. Review articles, studies describing models, letters, bulletins, comments and studies describing nontechnical skills were excluded from analysis.

Information Sources and Search

A broad search of the English language literature was performed in September 2011 using MEDLINE® (1950 to September 2011), EMBASE® (1980 to September 2011) and PsychINFO® (1966 to September 2011) databases. The following key words were used during the search: 'robotics', 'robotic surgery', 'computer assisted surgery', 'simulation', 'computer simulation', 'virtual reality', 'surgical training', and 'surgical education'. The Cochrane database and the Database of Abstracts of Reviews of Effectiveness were reviewed. Search terms included a combination of simulation, robotic surgery, learning, training or assessment. References of published review articles were checked to supplement the mentioned searches. We also reviewed the EAU and AUA conference abstracts from 2000 to 2011.

Study Selection and Data Collection

Two reviewers (K.A. and H.A.) independently identified potentially relevant articles. The full text of each article was obtained and further screened for inclusion if it had at least one of certain categories of information, including simulator, training, education, assessment, animal model, human cadaver or VR. Conflicts between reviewers were

Competence
The habitual and judicious use of communication, knowledge, technical skills, clinical reasoning, emotions, values and reflection in daily practice for the benefit of the individual and community being served. Competence builds on a foundation of basic clinical skills, scientific knowledge and moral development

Feasibility
Measure of whether an assessment process is capable of being done or carried out.

Validity
Determines whether an exam or a test actually succeeds in testing the competencies that it is designed to test. valid assessment method covering all the facets of clinical competence needs to have following attributes:

- Face validity: the extent to which the examination resembles the situation in the real world.
- Content validity: the extent to which the intended content domain is being measured by the assessment exercise (e.g. while trying to assess technical skills we may actually be testing knowledge)
- Construct validity: the extent to which a test measures the trait that it purports to measure. One inference of construct validity is the extent to which a test discriminates between various levels of expertise.
- Concurrent validity: the extent to which the results of the test correlate with the gold standard tests known to measure the same domain.
- Predictive validity: the extent to which an assessment will predict future performance.

Reliability
A measure of the reproducibility or consistency of performance, and is affected by factors such as examiner judgments, cases selection, candidate nervousness, and test conditions. reliability is a measure of a test to generate similar results when applied at two different points (test and retest) or consistency of marking among examiners (inter-rater).

Educational Impact
The ability of a training item to improve performance.

Acceptability
The extent to which an assessment tool is accepted by the subjects involved in the assessment.

Cost Effectiveness
Does the assessment tool provide maximum value for money.

Fig. 1 Definitions of terms related to competence, training and assessment [8–10].

subsequently discussed until there was 100% agreement on the final studies to be included (Fig. 2).

Data Items

Certain information was extracted from each study, including model trade name, training tasks, participant level, training duration and evaluation scoring. Each simulator study was evaluated for feasibility, acceptability, face validity, construct validity, content validity, reliability, educational impact and cost-effectiveness (Table 1) [12–29].

Results

Study Selection

In all, 459 potentially relevant publications were identified by the search, of which 406 were excluded from analysis after the abstract review. Another five studies were identified after reviewing AUA and EAU conference abstracts. Of the remaining 58 studies we excluded a further 39 after reviewing the full text. Thus, 19 studies were finally included in the systematic review.

The following simulators were identified:

Fig. 2 Study selection.

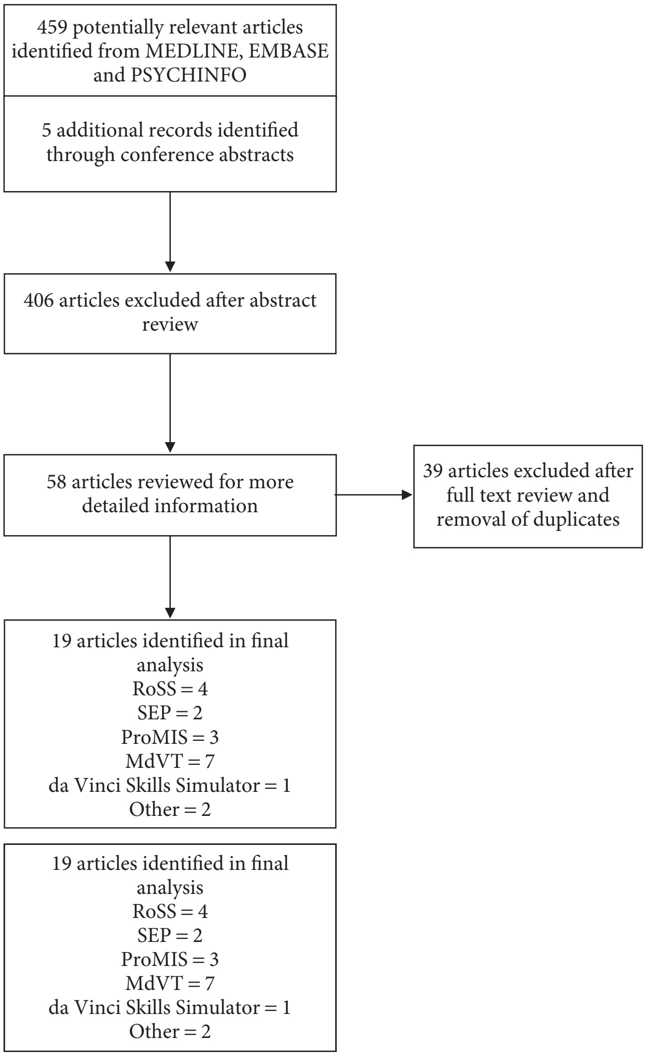


Table 1 Overview of surgical simulation studies in robotic surgery.

Reference	Simulator	Expert no. of robotic cases performed (n)	Novice no. of robotic cases performed (n)	Feasibility	Face validity	Construct validity	Content validity	Reliability	Educational impact	Cost-effectiveness
Seixas-Mikelus et al. [12]	RoSS	200–1350 (24)	0 (6)	N	Y	N	N	N	N	N
Seixas-Mikelus et al. [13]	RoSS	160–2200 (31)	0 (11)	N	N	N	Y	N	N	N
Kesavades et al. [14]	RoSS	0 (0)	0 (46)	N	N	N	N	N	Y	N
Guru et al. [15]	RoSS	–	– (10)	N	N	N	N	N	Y	N
Gavazzi et al. [16]	SEP	30–500 (12)	0 (18)	N	Y	Y	Y	N	N	N
van der Meijden et al. [17]	SEP	>50 MIS (7)	<50 MIS (9)	N	N	N	N	N	N	N
McDonough et al. [18]	ProMIS	– (8)	0 (10)	N	Y	Y	Y	N	N	N
Feifer et al. [19]	ProMIS	– (0)	0 (20)	N	N	N	N	N	Y	N
Jonsonn et al. [20]	ProMIS	21 to >100 (5)	0 (19)	N	N	Y	N	N	N	N
Lendvay et al. [21]	MdVT	– (4)	0 (11)	No	Yes	Yes	Yes	No	No	No
Sethi et al. [22]	MdVT	>50 (5)	0–1 (15)	No	Yes	Yes	Yes	No	No	No
Kenney et al. [23]	MdVT	30–320 (7)	1.3 ≤ 2.2 h (19)	No	Yes	Yes	Yes	No	No	No
Korets et al. [24]	MdVT	55–170 (x)	0–15 (x)	No	Yes	Yes	No	No	No	No
Korets et al. [25]	MdVT	– (0)	x (16)	No	No	No	No	No	Yes	No
Lerner et al. [26]	MdVT	– (0)	0 (12)	No	No	No	No	No	Yes	No
Hung et al. [27]	da Vinci Skills Simulator	100–800 (15)	0 (16)	No	Yes	Yes	Yes	No	Yes	No
Fiedler et al. [28]	University of Nebraska simulator	– (0)	0 (5)	No	No	No	No	No	Yes	No
Katsavelis et al. [29]	University of Nebraska simulator	– (0)	0 (8)	No	No	No	No	No	Yes	No

n, Number of participants in study group; x, figures unavailable.

Robotic Surgical Simulator (RoSS®)

We identified four studies assessing the use of the RoSS system for training surgeons on the dVSS. At the AUA meeting in 2009, Seixas-Mikelus et al. [12] recruited 30 subjects, 24 experienced robotic surgeons and six surgeons with no experience of performing robotic surgery. In all, 77% of the study group had performed an average of 340 robotic cases on the dVSS as the primary console surgeon. To prove face validity they invited the subjects to an orientation session with the RoSS followed by two modules; a basic object acquisition and placement task and a more advanced module that required object acquisition, precision control, and positioning of objects. This was followed by a questionnaire where all subjects indicated that RoSS was realistically close to the dVSS console for virtual simulation and instrumentation. For the pinch device 84% found RoSS 'somewhat close' or 'very close' to the dVSS. While, 90%

described the movement arm as 'somewhat close', 'very close' or 'just like' the dVSS and 89% reported the camera movement and clutch functions as 'somewhat' or very close to the dVSS.

At the International Robotic Surgery Symposium in 2010 the same research group reported RoSS had content validity [13]. In all, 42 subjects were recruited, 31 experienced surgeons and 11 novices. The experienced group was further subdivided into 'expert' (17) those who have performed ≥150 robotic cases and 'intermediate' (14) who had performed 1–149 robotic cases. The expert group comprised surgeons with a mean (range) of 881 (160–2200) robot-assisted cases. Experts rated the 'clutch control' virtual simulation task as a good (71%) or excellent (29%) teaching tool. In all, 78% rated the ball place task as 'good' or 'excellent' but 22% rated it as 'poor'. While, 27% rated the needle removal task as an excellent teaching tool, 60% rated it as 'good' and 13% rated it 'poor'. Overall, 91% rated the fourth-arm tissue removal task as 'good' or 'excellent' and 94% responded that RoSS would be useful for training purposes. RoSS was considered to be an appropriate training and testing format before operating room experience for residents (88%) and 79% indicated that RoSS could be used for privileging or certifying in robotic surgery.

Kesavades et al. [14] showed that training on RoSS had a positive impact on the time taken to complete a procedure on the dVSS. Three groups of subjects performed two exercises, ball drop and needle capping. Group 1 (20) was a control group and were provided no training on the RoSS before performing the tasks on the dVSS. Group 2 (15) was given 40 min of training on the RoSS and group 3 (11) were trained on the dVSS for 40 min. Compared with no training, novices trained on RoSS significantly reduced the time taken to complete tasks on the dVSS ($P = 0.002$).

Guru et al. [15] assessed the benefit of introducing a new cognitive skill software integrated into RoSS to improve procedure specific anatomical landmark recognition during surgery while using the dVSS. In all, 10 participants were recruited. Group 1 (five, no training) and Group 2 (five, RoSS trained with augmented videos). Both groups were assessed on their time taken to complete and the number of correct answers in an anatomical landmark identification test during a robotic cystectomy. Both groups had $P <$ 0.001, indicating significant training and uniform performance with mean times to completion of: Group 1 142.8s and Group 2 118.4s; and number of correct answers, Group 1 2.9 and Group 2 4.2.

Simsurgery Educational Platform (SEP)

We identified two studies that assessed validity of the SEP robot simulator (SimSurgery, Oslo, Norway). In the study by Gavazzi et al. [16], 30 participants (12 experts and 18 novices) were recruited to perform two tasks on the SEP robot simulator. Face and content validity was achieved by asking all participants to complete a questionnaire of their experience. In all, 90% rated the trainer realistic and easy to use, 87% considered it generally useful for training and 90% agreed that the simulator was useful for hand–eye co-ordination and suturing. Construct validity was achieved by analysing the performance of the experts compared with the novice participants. The novice group required more time to complete the tasks than the expert group; arrow placement (mean time 86.94 s for novice vs 57.42 s for experts; $P = 0.03$) and suturing (mean time 78.8 s for experts vs 198.3 s for novices; $P = 0.001$). During the surgeon's knot task, experts significantly outperformed novices in maximum tightening stretch (72.62 vs 164.81; $P = 0.002$), instruments dropped (0.33 vs 2.33; $P = 0.002$), maximum winding stretch (85.28 vs 229.97; $P = 0.027$) and tool collisions (4.42 vs 13.56; $P = 0.001$).

A study group in the Netherlands attempted but failed to show face and construct validity of the SEP robot [17]. For construct validity they compared the performance of an 'expert' group who consisted of surgeons with >50 minimally invasive surgery (MIS) procedures completed to a 'novice' group who had performed <50 MIS procedures. There was a significantly shorter instrument tool tip trajectory in the MIS-expert group as compared with the MIS-novice group. There were no differences in the remaining correlations analysed, e.g. total procedure time.

ProMIS®

Three studies investigating the role of ProMIS (Haptica, Ireland) in the context of robotic simulation were found. McDonough et al. [18] showed face, content and construct validity. The group recruited participants into two groups; experienced (eight surgeons) and novice (10). After a standardised orientation and practice session, all subjects completed three tasks (peg transfer, precision cutting, and intracorporeal suture/knot) on the ProMIS laparoscopic simulator using the dVSS. The expert group outperformed the novice group in peg transfer time (68.4 vs 235.5 s; $P =$ 0.001), precision cutting time (93.6 vs 283.9 s; $P = 0.001$), and intracorporeal suture/knot time (93.2 vs 380.3 s; $P \leq$ 0.001). Subjects rated ProMIS as easy to use, relevant to robotic surgical training and as an accurate measure of their robotic surgical proficiency. The experts described the simulator platform as useful for training and agreed with incorporating it into a robotic surgery training curriculum.

In a randomised controlled trial by Feifer et al. [19], ProMIS was used in conjunction with LapSim® (Surgical Science Sweden AB) to show that training on both of these conventional laparoscopic simulators could lead to improved performance on tasks performed on the dVSS. Novice participants were randomised to receive no training, training on either ProMIS or LapSim alone or training on both simulator systems. Cannulation score (96.6 vs 102.8; $P = 0.08$) and total score (300.4 vs 262.8; $P = 0.03$) were significantly improved in the ProMIS-alone group before and after training. Interestingly the LapSim-only group showed no overall score improvement. However, the study showed that training on both simulators was better than no training or single simulator system use across all the tasks assessed.

Jonsson et al. [20] showed construct validity of the ProMIS system by comparing the performance of five experienced robotic surgeons against a novice group which consisted of 13 consultants and six residents. Each participant performed four tasks: pull and loosen elastic bands, cutting a circle, suturing and tying, and vesicourethral anastomosis. The ProMIS simulator registered objective data concerning how the surgeon performed in the box environment related to time, path, and smoothness. There was a statistically significant difference between experts and novices in all tasks concerning time and smoothness ($P < 0.002$).

Mimic dV-Trainer (MdVT)

We found seven studies investigating the efficacy and validity of the Mimic dV-Trainer (MdVT, mimic Technologies, Seattle, WA, USA). The first study performed to investigate the validity of the MdVT system was in 2008 [21]. In all, 27 robotic surgery learners were randomly divided into three groups: didactics, VR simulation and robot dry laboratory; with the instructors 'blinded' to the experience of the learners. A survey was conducted before and after task completion on the MdVT. All of the participants felt that robots were a valuable tool for surgery, 87% thought there was a role for computerised simulation of robotic surgery to maintain skills after training and 93% thought that an offline simulator, such as the MdVT, would be a useful training aid. Those participants with dVSS experience performed better than novice users and showed statistically significant differences when it came to task time completion in three ring transfer, economy of motion, and time the master controls were out of centre.

A year later in 2009 another study involving five experts and 15 novices looked at the validity of the MdVT [22]. All of the participants rated the device as 'average' to 'easy' to use and 'above average' and 'high' in all parameters of realism including exercises, visual, hardware and movement realism. Construct validity for time to completion of tasks and time instruments were out of view was only statistically significant in one out of the three tasks that were tested between expert and novice participants ($P = 0.04$ and 0.03 in the 'letterboard' exercise vs $P = 0.59$, 0.98 and 0.08 'ring/cone' and 'string walk' tasks).

Kenney et al. [23] conducted a study to assess face, content and construct validity of the MdVT. Medical students, residents, and attending surgeons were prospectively categorised as novice (19) or experienced (seven). Each subject completed two EndoWrist modules and two needle-driving modules followed by a questionnaire. Experienced robotic surgeons outperformed novices in nearly all variables, including total score [mean (SD) 82 (11) vs 74 (18), $P \leq 0.01$), total task time [mean (SD) 193 (56) vs 269 (196) s, $P = 0.01$), total instrument motion [mean (SD) 1674 (712) vs 2635 (1682) cm, $P = 0.01$) and number of instrument collisions (mean (SD) 0.9 (1.3) vs 4.9 (7.8), $P = 0.001$). All experienced surgeons ranked the simulator as useful for training and agreed with incorporating the simulator into a residency curriculum. The VR and instrumentation achieved acceptability. The needle-driving modules did not exceed the acceptability threshold. This study showed that the MdVT has face, content, and construct validity as a virtual reality simulator for the dVSS.

Korets et al. [24] were able to show face and construct validity by comparing the performance of eight urology residents and two endourology fellows. Each trainee completed 15 exercises from four domains. The participants were allocated into either a novice group if they had performed 0–15 cases and an expert group who had performed 55–170 cases. The expert group had less instrument collisions (6.3 vs 2.5; $P \leq 0.01$), better 'instrument out of view' scores (81.0 vs 90.3; $P = 0.02$) and missed less targets within the 'needle driving' domain (6.6 vs 1.9; $P \leq 0.01$). Both groups rated the MdVT as 'easy to use' and 'useful' in improving robotic surgery expertise with the novice group reporting improved confidence in robotic surgery skills after completing the training set. The expert group deemed the MdVT as 'somewhat realistic' in the 'arm manipulation' and 'camera movement' domains, but judged 'needle control' and 'needle driving' domains as 'not realistic'.

The same research group used the MdVT to show that curriculum-based training using this simulator improved performance on the dVSS compared with no training at all [25]. They only looked at a limited number of domains including; EndoWrist manipulation, camera movement, needle control and robotic suturing. However, they found no difference in improvement during secondary evaluation for exercises shared between the MdVT and dVSS groups implying that MdVT is equivalent to dVSS for improving robotic aptitude in EndoWrist manipulation and camera movements.

Lerner et al. [26] performed a prospective study to determine if training on the MdVT provides acquisition of skills on the dVSS. After completing a baseline set of tasks on the dVSS, 12 novice medical students completed four training sessions on the MdVT. After the four sessions they repeated the initial tasks on the dVSS and their results were compared with 10 urology residents (no robotic experience) and one fellow who all received four to six sessions on the dVSS only. The MdVT group showed a statistically significant improvement in the 'peg board' (MdVT group; initial time 153 s and final 91 s, $P = 0.002$) and 'pattern cutting' times (MdVT group; initial time 547 s and final 385 s, $P = 0.004$) from the initial testing compared with the final test. There were no other statistically significant improvements in the other tasks. They also investigated the level of improvement between the dVSS group and MdVT group in each task and found that there were statistically significant improvements in the MdVT group in the 'peg board' and 'pattern cutting' times ($P = 0.008$ and $P = 0.02$, respectively). There were no statistically significant differences between the two groups in their levels of improvement for the remaining dVSS exercises. Across the whole study there was an overall trend toward improvement despite a lack of statistically significant results.

Da Vinci Skills Simulator®

Hung et al. [27] evaluated the face, content and construct validity of the da Vinci Skills Simulator. Participants were categorised as either novice (16) with no prior robotic experience, intermediate (32) with <100 robotic cases and experienced (15) >100 primary surgeon robotic cases. Each participant completed 10 virtual reality exercises with three repetitions and a questionnaire with a 1–10 visual analogue scale (VAS) to assess face and content validity. The performance of experts, intermediates and novices was compared to determine construct validity. Participants rated the overall virtual reality and console experience as 'very realistic' (median VAS 8/10). Expert surgeons further scored the visual field (median VAS 9/10), movement (median VAS 9/10) and precision of the virtual reality platform (median VAS 9/10) as 'very realistic'. Expert surgeons also rated the simulator as a 'very useful training tool' for residents (median 10/10) and fellows (9/10), although less so for experienced robotic surgeons (6/10). Experts significantly outperformed novices in almost all metrics, including overall score (median 88.3 vs 62.1%; $P < 0.001$), economy of motion (6983.5 vs 10 554.5 cm; $P < 0.001$), time with excessive instrument force (11 vs 151 s; $P < 0.001$), instrument collisions (26 vs 117; $P < 0.001$), instruments out of view (38.4 vs 120.1 cm; $P < 0.001$), master controller range (277.9 vs 317.5 cm; $P = 0.002$), missed targets (71 vs 152; $P < 0.001$), time to completion (3612 vs 7534 s; $P < 0.001$) and misapplied energy (40 vs 90 s; $P = 0.02$). Interestingly, there was no difference between experts and novices in broken vessel and blood loss although these two parameters are only represented in one out of the 10 exercises and may not be a fair representation of actual level of skill. Experts outperformed intermediates and novices (median overall score 88.3 vs 75.6 vs 62.1%, respectively; $P < 0.001$). Intermediates also outperformed novices although less significantly in most measurements compared with the comparison between experts and novices or experts and intermediates.

Other Virtual Reality Simulators

The biomechanics laboratory at the University of Nebraska at Omaha developed their own VR simulator [28]. Five students performed two tasks, bimanual carrying and needle passing. Each task was performed on the dVSS and then repeated using their department designed VR simulator. This was followed by a questionnaire. Data analysis showed statistically significant results for time to task completion and distance travelled between the two environments for both tasks. The subjects partially agreed that they would like to have VR as part of their regular training. This simulator was tested once again, taking into account more complex parameters such as wrist flexor and extensor movements using electromyography (EMG) [29]. Six medical students and two medical research fellows with no prior experience on the dVSS were recruited to participate. There were no significant differences for the bimanual-carrying task in all parameters between the actual and the VR environment, with moderate correlations for spatiotemporal parameters and high correlations for most EMG parameters. For the needle-passing task, results showed significant differences in most parameters including time to task completion ($P = 0.003$), total travelling distance ($P < 0.001$), wrist flexion/extension range of movement ($P = 0.002$) and elbow flexion/extension range of movement ($P = 0.002$) with moderate correlations for spatiotemporal parameters and high correlations for most EMG parameters.

Discussion

To our knowledge this is the first systematic review of all the simulation options available to robotic surgeons. They provide a safe environment for trainees to develop their skills. This generation of robotic simulators has provided more questions than answers. Firstly, there is a lack of standardisation upon which the metrics of simulator quality are tested in each of the different platforms. For example, there is no agreed definition upon which to assess face validity, whether it is the 'very close/somewhat close' [15] scale devised for the RoSS platform or the visual analogue scale described for the dVSS [29]. Similarly, there are no consistent definitions as to what qualifies as a 'novice' or an 'expert'. These concerns apply to each of the metrics used to measure simulators, and until these fundamental issues are addressed, we will never be able to rigorously study and compare these simulators and provide evidence-based solutions on how best to train the current and future generations of robotic surgeons.

Unfortunately, most of the exercises currently available on VR simulators are generic tasks testing hand–eye co-ordination, tissue manipulation, dissection, suturing and knot tying. There is no evidence to suggest which exercises lead to improved real-setting performance. Training scenarios for specific procedures incorporating challenging scenarios and complications are under development and will be much welcomed [19].

Further questions remain about the use of simulation training in the context of different skill levels. It has been shown that simulation models are valid and reliable for the initial phase of training and assessment in urological procedures; however, this is not the case for advanced and specialist level skill learning [30]. In a training report by Davis et al. [31], the trainers were successful in teaching the introductory steps to robotic prostatectomy but their exposure to advanced steps were more limited, and often incomplete. Consequently, we advocate the use of robotic simulation in the early phase of robotic training. Further studies investigating its effectiveness in more complex situations and skills levels are required.

We used the criteria proposed by van de Vlueten [32] and Ahmed et al. [9] to evaluate the quality of each study. All of the simulators except RoSS have demonstrated face, content and construct validity but the numbers in these studies remain small. Educational impact was shown in eight studies and in all commercially available simulators except SEP. Evidence of criterion validity, such as predictive or concurrent validity, was very sparse. Other parameters, such as inter-rater and inter-item reliability, feasibility, acceptability, and cost-effectiveness of the simulation platforms were not evaluated by any of the studies. Similarly no group has validated the use of animal models and freshly frozen cadavers, and structured skills training based on observation for robotic surgery.

Data analysis was conducted using classic test theory. There was a wide variability and inconsistency of statistical methods used to evaluate data.

Given the lack of comparative studies between the different simulators the current body of evidence does not identify any one simulator being more effective in training the next generation of robotic surgeons than another. Each platform has the capability to train and assess a range of different robotic skills fundamental to the technique (Table 2). Unlike the dVSS the MdVT and RoSS platforms feature user interfaces that are similar to but not exact duplicates of the dVSS console used in clinical practice. The ProMIS simulator enables virtual and physical reality to be used together and has been investigated in the laparoscopic setting previously [33,34]. The randomised control trial by Feifer et al. [21] represents the highest level of evidence for any of the simulators currently available. Their study showed that the use of ProMIS and LapSim simulators in conjunction with each other could improve robotic console performance. Interestingly, the LapSim group showed no improvement, and it was therefore not clear what

Table 2 Simulator properties.

Simulator Name	RoSS	SEP	ProMIS	MdVT	dVSS
Developer	Simulated surgical systems	Sim surgery	CAE healthcare	Mimic	Intuitive surgical
Endowrist manipulation	Yes	Yes	No	Yes	Yes
Camera and clutching	Yes	No	No	Yes	Yes
Fourth arm integration	Yes	No	No	Yes	Yes
System settings	Yes	No	Yes	Yes	Yes
Needle control and driving	Yes	No	Yes	Yes	Yes
Energy and dissection	Yes	No	No	Yes	Yes
Performance feedback	Yes	Yes	Yes	Yes	Yes
Developed for robotic surgery	Yes	Yes	No	Yes	Yes
Cost, USA dollars	120 000	62 000	35 000	158 000	89 000

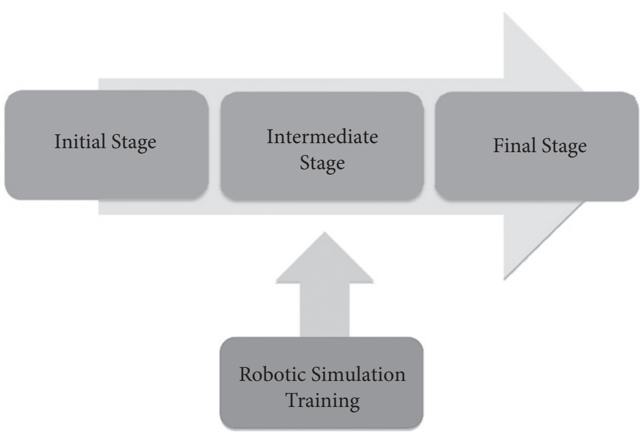
- ***EndoWrist™ manipulation*** *– EndoWrist instruments are designed to provide surgeons with natural dexterity and a wide range of motion.*
- ***Camera and clutching***
- ***Fourth Arm Integration*** *[41] – For more advanced instrument control skills, some exercises include a fourth instrument arm that must be used. This is designed to promote instrument skill, and encourages users to think strategically about instrument placement during tasks.*
- ***System settings*** *[41] – The surgeon console features a comprehensive set of controls for user settings. Quiz exercises on the simulator focus on basic setting topics, e.g. icons, ergonomics and instrument scaling.*
- ***Needle control and driving*** *[41] – These scenarios are designed to help users develop skill when manipulating needles, including a focus on how to effectively hand off and position needles while practicing with a range of geometries.*
- ***Energy and dissection*** *[41] – The footswitch panel enables users to perform a range of tasks, e.g. swapping between different types of energy instruments. These exercises allow users to gain familiarity with the footswitch panel by letting them practice applying monopolar and bipolar energy while working on dissection tasks.*
- ***Performance feedback*** *[41] – automatically stores data on performance measures, e.g. economy of motion, instrument out of view time and time to completion of task.*

contribution LapSim had on the overall improvement seen when both simulators were used in conjunction. Despite SEPs level of validation its face validity must come into question given that the participants in the van der Meijden et al. [17] study commented so negatively on the hardware, coupled with the MIS experts being so highly critical of the overall ergonomics of the training apparatus. Its biggest disadvantage lies in the fact that the images are not three-dimensional (3D), a fundamental concept pertaining to robotic compared with laparoscopic surgery. Further studies or perhaps even hardware upgrades to convert the 2D simulator into a 3D platform are therefore warranted.

More studies have been conducted using the MdVT platform than any other, with three out of the four showing face, content and construct validity [23–25]. The Intuitive simulator has the distinct advantage that the same company who has developed the dVSS manufactures it. However, only one study has validated its use as a training tool [29].

With current level of validation of the available robotic simulators can be integrated as an adjunct to the basic phase of robotic training (Fig. 3). Until further studies can evaluate these simulators in greater detail this integration is likely to be on a local level in centres with significant funding and research capabilities. The choice of simulator currently is also likely to be department specific. This will require committees of surgeons with special expertise to assess robotic competence. National implementation of robotic surgical simulation training requires directives from national organisations to ensure that a structured, standardised approach is used. It is essential that competence can be defined in accordance with proficiency levels and that validated assessment tools are developed [9]. Formal assessments of robotic surgeons in training have been attempted but the evaluation tools used were

Fig. 3 A suggested time frame for when simulation training can be implemented within a urology residency programme.



subjective and had not been validated [31,35]. Such evaluation tools need to be reproducible and objective to accurately examine a surgeon's technical and non-technical skills.

Within the robotic operating theatre non-technical factors, e.g. communication, team working, decision making, and judgment are key domains that must also be honed to ensure one possesses the ability for independent and competent practice [36,37]. As yet there are very few if any team-based robotic simulation environments that have been able to encompass these important non-technical domains.

Recent years have witnessed trainees, trainers and more experienced robotic surgeons alike embracing VR training for robotic surgery with great optimism [30] (Fig. 4) [18,38–41]. However, the cost of the robotic system alone is in the order of several million dollars, therefore it is most cost effective to devote as much of the surgical robot's time to performance of actual procedures. Therefore, the availability of such expensive equipment for training is usually low. With Intuitive Surgical, Inc., developing new robotic simulators, such as the latest six-arm robot, an advance on the older three-arm device and further developments in the pipeline, institutions with dated systems can donate old systems to their robotic training programmes. Funding from universities, charities and registered health organisations can aid in the development of simulation-training programmes and in the acquisition of the simulators themselves. With the ever increasing market competition between the different simulator manufacturers the cost of the simulators may decrease in the near future.

To date there is only one randomised control trial investigating the simulators available, and this looked at educational impact alone [25]. In order to justify the costs, VR simulators will require further validation studies with greater sample sizes. Despite these issues robotic surgical simulators hold the greatest potential for robotic surgical training in the 21st century.

This article has some limitations. First, we may have missed a few relevant studies. We reviewed various databases with free text and Medical Subject Headings (MeSH®) terms to overcome this. Second, we could not use formal meta-analytical methods to pool results, as the included studies used different measuring tools and outcome measures for all metrics of simulator quality. There are several components to the simulators that were not investigated, e.g. concurrent and predictive validity, inter-rater and inter-item reliability, feasibility, acceptability, and cost-effectiveness. However this reflects paucity in the available data of these factors in published studies.

Fig. 4 Images of the VR robotic simulators. **A**, SEP [18]; **B**, RoSS [38]; **C**, ProMIS [39]; **D**, MdVT [40]; **E**, dVSS [41].

Conclusions

Simulation training holds the greatest potential to be used as an adjunct tool to train the next generation of robotic surgeons. Its implementation in the initial phase of training has been validated. Its role lies alongside traditional training and is unlikely to ever supersede the benefits of real life operative experience. Cost is undoubtedly the major burden in the current economic climate. Existing robotic simulators can be used for initial phase of training. Keeping up with technological advances and organising a proficiency-based curriculum are essential to driving simulation into robotic surgery. Further studies with larger cohorts are required to assess cost-effectiveness and the transferability of skills from simulation to real patients.

Acknowledgements

Professor Prokar Dasgupta acknowledges financial support from the Department of Health via the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy's & St Thomas' NHS Foundation Trust in partnership with King's College London and King's College Hospital NHS Foundation Trust. He also acknowledges the support of the Medical Research Council Centre for Transplantation, London Deanery, London School of Surgery and Olympus.

Authors have been involved in development and validation of the SEP robot and RoSS System. The conclusions from this article are based on the evidence from the existing literature and are not influenced by the author's views.

Conflict of Interest

Khurshid A. Guru has been involved in the development and validation of the ROSS system. Mohammed S. Khan, Prokar Dasgupta and Kamran Ahmed were involved in the development and validation of the ROSS system and SEP.

Source of funding: Prokar Dasgupta acknowledges financial support from the Department of Health via the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy's & St Thomas' NHS Foundation Trust in partnership with King's College London and King's College Hospital NHS Foundation Trust. He also acknowledges the support of the MRC Centre for Transplantation, London Deanery, London School of Surgery and Olympus.

References

1 **Binder J, Kramer W.** Robotically-assisted laparoscopic radical prostatectomy. *BJU Int* 2001; 87: 408–10

2 **Ahmed K, Khan MS, Vats A et al.** Current status of robotic assisted pelvic surgery and future developments. *Int J Surg* 2009; 7: 431–40

3 **Zorn KC, Gofrit ON, Steinberg GD, Shalhav AL.** Evolution of robotic surgery in the treatment of localized prostate cancer. *Curr Treat Options Oncol* 2007; 8: 197–210

4 **Maeso S, Reza M, Mayol JA et al.** Efficacy of the da Vinci surgical system in abdominal surgery compared with that of laparoscopy: a systematic review and meta-analysis. *Ann Surg* 2010; 252: 254–62

5 **Torkington J, Smith SG, Rees BI, Darzi A.** Skill transfer from virtual reality to a real laparoscopic task. *Surg Endosc* 2001; 15: 1076–9

6 **Andreatta PB, Woodrum DT, Birkmeyer JD et al.** Laparoscopic skills are improved with LapMentor training: results of a randomized, double-blinded study. *Ann Surg* 2006; 243: 854–63

7 **Sedlack RE, Kolars JC.** Computer simulator training enhances the competency of gastroenterology fellows at colonoscopy: results of a pilot study. *Am J Gastroenterol* 2004; 99: 33–7

8 **Wass V, Van der Vleuten C, Shatzer J, Jones R.** Assessment of clinical competence. *Lancet* 2001; 357: 945–9

9 **Ahmed K, Miskovic D, Darzi A, Athanasiou T, Hanna GB.** Observational tools for assessment of procedural skills: a systematic review. *Am J Surg* 2011; 202: 469–80

10 **Ahmed K, Ashrafian H, Hanna GB, Darzi A, Athanasiou T.** Assessment of specialists in cardiovascular practice. *Nat Rev Cardiol* 2009; 6: 659–67

11 **Liberati A, Altman DG, Tetzlaff J et al.** The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009; 339: b2700

12 **Seixas-Mikelus SA, Kesavadas T, Srimathveeravalli G, Chandrasekhar R, Wilding GE, Guru KA.** Face validation of a novel robotic surgical simulator. *Urology* 2010; 76: 357–60

13 **Seixas-Mikelus SA, Stegemann AP, Kesavadas T et al.** Content validation of a novel robotic surgical simulator. *BJU Int* 2011; 107: 1130–5

14 **Kesavadas T, Kumar A, Srimathveeravalli G et al.** Efficacy of Robotic Surgery Simulator (RoSS) for the da Vinci® Surgical System. *J Urol* 2009; 181 (Suppl.): 823

15 **Guru K, Baheti A, Kesavadas T, Kumar A, Srimathveeravalli G, Butt Z.** In-vivo videos enhance cognitive skilla for da Vinci® Surgical System. *J Urol* 2009; 181 (Suppl.): 823

16 **Gavazzi A, Bahsoun AN, Van Haute W et al.** Face, content and construct validity of a virtual reality simulator for robotic surgery (SEP Robot). *Ann R Coll Surg Engl* 2011; 93: 146–50

17 **van der Meijden OA, Broeders IA, Schijven MP.** The SEP 'Robot': a valid virtual reality robotic simulator for the da vinci surgical system? *Surg Technol Int* 2010; 19: 51–8

18 **McDonough P, Peterson A, Brand T.** Initial validation of the ProMIS surgical simulator as an objective measure of robotic task performance. *J Urol* 2010; 183 (Suppl.): e515

19 **Feifer A, Al-Ammari A, Kovac E, Delisle J, Carrier S, Anidjar M.** Randomized controlled trial of virtual reality and hybrid simulation for robotic surgical training. *BJU Int* 2011; 108: 1652–6

20 **Jonsson MN, Mahmood M, Askerud T et al.** ProMIS™ can serve as a da Vinci® simulator – a construct validity study. *J Endourol* 2011; 25: 345–50

21 **Lendvay TS, Casale P, Sweet R, Peters C.** VR robotic surgery: randomized blinded study of the dV-Trainer robotic simulator. *Stud Health Technol Inform* 2008; 132: 242–4

22 **Sethi AS, Peine WJ, Mohammadi Y, Sundaram CP.** Validation of a novel virtual reality robotic simulator. *J Endourol* 2009; 23: 503–8

23 **Kenney PA, Wszolek MF, Gould JJ, Libertino JA, Moinzadeh A.** Face, content, and construct validity of dV-trainer, a novel virtual reality simulator for robotic surgery. *Urology* 2009; 73: 1288–92

24 **Korets R, Graversen JA, Mues A, Gupta M, Landman J, Badani KK.** Face and construct validity assessment of 2nd generation robotic surgery simulator. *J Urol* 2011; 185 (Suppl.): e488

25 **Korets R, Mues AC, Graversen J, Gupta M, Landman J, Badani KK.** Comparison of robotic surgery skill acquisition between DV-Trainer and da Vinci surgical system: a randomized controlled study. *J Urol* 2011; 185 (Suppl.): e593

26 **Lerner MA, Ayalew M, Peine WJ, Sundaram CP.** Does training on a virtual reality robotic simulator improve performance on the da Vinci surgical system? *J Endourol* 2010; 24: 467–72
27 **Hung AJ, Zehnder P, Patil MB et al.** Face, content and construct validity of a novel robotic surgery simulator. *J Urol* 2011; 186: 1019–24
28 **Fiedler MJ, Chen SJ, Judkins TN, Oleynikov D, Stergiou N.** Virtual reality for robotic laparoscopic surgical training. *Stud Health Technol Inform* 2007; 125: 127–9
29 **Katsavelis D, Siu KC, Brown-Clerk B et al.** Validated robotic laparoscopic surgical training in a virtual-reality environment. *Surg Endosc* 2009; 23: 66–73
30 **Ahmed K, Jawad M, Abboudi M et al.** Effectiveness of procedural simulation in urology: a systematic review. *J Urol* 2011; 186: 26–34
31 **Davis JW, Kamat A, Munsell M, Pettaway C, Pisters L, Matin S.** Initial experience of teaching robot-assisted radical prostatectomy to surgeons-in-training: can training be evaluated and standardized? *BJU Int* 2010; 105: 1148–54
32 **van der Vleuten C.** The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ Theory Pract* 1996; 1: 41–67
33 **Pellen MG, Horgan LF, Barton JR, Attwood SE.** Construct validity of the ProMIS laparoscopic simulator. *Surg Endosc* 2009; 23: 130–9
34 **Neary PC, Boyle E, Delaney CP, Senagore AJ, Keane FB, Gallagher AG.** Construct validation of a novel hybrid virtual-reality simulator for training and assessing laparoscopic colectomy; results from the first course for experienced senior laparoscopic surgeons. *Surg Endosc* 2008; 22: 2301–9
35 **Schroeck FR, de Sousa CA, Kalman RA et al.** Trainees do not negatively impact the institutional learning curve for robotic prostatectomy as characterized by operative time, estimated blood loss, and positive surgical margin rate. *Urology* 2008; 71: 597–601
36 **Moorthy K, Munz Y, Adams S, Pandey V, Darzi A.** A human factors analysis of technical and team skills among surgical trainees during procedural simulations in a simulated operating theatre. *Ann Surg* 2005; 242: 631–9
37 **Moorthy K, Munz Y, Forrest D et al.** Surgical crisis management skills training and assessment: a simulation[corrected]-based approach to enhancing operating room performance. *Ann Surg* 2006; 244: 139–47
38 **Simulated Surgical Systems.** RoSS™ Hardware Features. 2010. Available at: http://www.simulatedsurgicals.com/hardware.html. Accessed January 2012
39 **Healthcare CAE.** Why train with the CAE ProMIS simulator? 2008. Available at: http://www.cae.com/en/healthcare/why.promis.simulator.asp. Accessed January 2012
40 **Mimic Technologies.** dV Trainer, Surgical Skills Development for the da Vinci® System. 2011. Available at: http://www.mimic.ws/products/. Accessed January 2012
41 **Intuitive Surgical.** da Vinci skills simulator. 2012. Available at: http://www.intuitivesurgical.com/products/skills_simulator/. Accessed January 2012

Correspondence: Kamran Ahmed, Guy's and St. Thomas' Hospitals, Urology Centre, 1st Floor Southwark Wing, Great Maze Pond, London SE1 9RT, UK.

e-mail: k.ahmed@imperial.ac.uk

Abbreviations: dVSS, da Vinci Surgical System; EAU, European Association of Urology; MdVT, Mimic dV-Trainer; MIS, minimally invasive surgery; RoSS, Robotic Surgical Simulator; SEP, SimSurgery Educational Platform; VAS, visual analogue scale; VR, virtual reality; (3)(2)D, (three-)(two-)dimensional.