Utilization and Surgical Skill Transferability of the Simulator Robot to the Clinical Robot for Urology Surgery

Introduction

Virtual reality is a surgical training tool of increasing popularity. The advantages of this tool include its low-risk environment and the ability of simulating multiple assignments in various situations and scenarios. Currently, commercially available simulators include the Mimic dV-Trainer (MdVT, Mimic Technologies, Seattle, Wash., USA), da Vinci Skills Simulator (dVSS, Intuitive Surgical, Sunnyvale, Calif., USA), and Robotic Surgery Simulator (RoSS, Simulated Surgical Systems, Williamsville, N.Y., USA).

A number of studies have been performed to investigate the practicality of these simulators in the field of urology surgical training, which showed varying degrees of educational impacts.

Simulation training is on the rise in Australia, where the first virtual reality simulation machine was introduced at Liverpool Hospital, Sydney. This literature review aims at identifying gaps in previous studies, and in providing a comprehensive review of future studies to be further developed in Australia.

Method

A multi-field research was performed, combining the key terms 'uro*' and 'virtua*' and 'simula*' and 'robo*'. Academic search engines used in this literature review included 'Medline', 'Scopus', and 'Sciencedirect'. Studies with laparoscopic skills as a focused investigation but not robotic skills were excluded. Critical appraisal of each of the article was conducted with a discussion of key topic involving urologists with expert skills on robotic surgery.

Result: A thorough literature review discovered 3 main types of studies in this area. These are: (1) validity studies; (2) studies specific for urology procedures; (3) studies on skill transfer, in general. Cohort study and randomized control trial are the 2 dominant forms of research designs. Conclusion: Future studies need to focus more around the investigation of operation-specific training, in conjunction with skills-based teaching. Also, it is important that these studies incorporate teamwork, decision-making, and communication skills.

© 2016 S. Karger AG, Basel
Simulation robotic training in general or specific to urology practice. Included study types were randomized control trials (RCT) or cohort studies. Studies with laparoscopic skills as a focused investigation, but not robotic skills were excluded. Critical appraisal of each article was conducted, in accordance with PRISMA methodology. Key topic discussion also involved urologists with expert skills on robotic surgery.

Result

Ninety eight articles were found, and after careful selection, 25 articles remained to be critically appraised. A thorough literature review discovered 3 main types of studies in this area. These were: (1) validity studies; (2) studies specific for urology procedures; (3) studies on skill transfer, in general. Cohort study and RCTs were the 2 dominant forms of research designs. A summary of the reviewed literatures is outlined in Table 1.

Discussion

Validity Studies

Currently, there are 3 virtual simulators that have been studied extensively in the literatures in relation to their validity. These are MdVT, dVSS, and RoSS.

Validity can be categorized as either subjective or objective [1]. Subjective validation includes face and content validity. To define face validity, unofficial evaluation of realism needs to be provided by the person manipulating the simulator, whereas content validity is defined as the official assessment of suitability as a means for teaching [1].

Objective validation includes construct, concurrent, and predictive validities. Construct validity is defined as the ability of a simulator to differentiate experts from beginners [2]. Beginners are candidates without any experience in robotic skills, whereas the term 'expert' is used in reference to a urology trainee or specialist. Concurrent validity is defined as the ability to compare performance on a simulator with gold standard tests known to measure the same domain, such as a tissue or animal lab [2]. Predictive validity is defined as the ability to predict future performance based on performance on the simulator [1].

Being one of the most globally validated simulators, the MdVT is a tabletop unit with foot pedals imitating the real robotic console [3]. A scoring system called MScore is attached within the simulator to provide the trainee with performance tracking and individualized grades, making it superior for training purposes [1].

The objective validity of the MdVT has been highlighted by a number of studies. One study proved that training with the MdVT could improve the skills of a trainee when performing the same inanimate tasks using a clinical robot, verifying its construct validity [4]. Furthermore, it was also discovered in 2 later studies that there was an apparent correlation between the performance on the MdVT and that on the gold standard clinical robot. This provides an indication that the MdVT

Table 1. Different types of study designs and focus research areas

<table>
<thead>
<tr>
<th>Study type</th>
<th>Cohort study</th>
<th>Randomized control study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Validity study</td>
<td>Lyons et al. [7], 2013</td>
<td>Brinkman et al. [9], 2013</td>
</tr>
<tr>
<td></td>
<td>Perrenot et al. [2], 2012</td>
<td>Hung et al. [11], 2013</td>
</tr>
<tr>
<td></td>
<td>Lee et al. [5], 2012</td>
<td>Liss et al. [8], 2012</td>
</tr>
<tr>
<td></td>
<td>Hung et al. [6], 2011</td>
<td>Hung et al. [10], 2012</td>
</tr>
<tr>
<td></td>
<td>Lerner et al. [4], 2010</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Seixas-Mikelus et al. [12], 2010</td>
<td></td>
</tr>
<tr>
<td>Specific urology skills study</td>
<td>Kang et al. [13], 2014</td>
<td>Chowriappa et al. [18], 2015</td>
</tr>
<tr>
<td></td>
<td>Hung et al. [14], 2016</td>
<td>Whitehurst et al. [17], 2015</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Alzahrani et al. [15], 2013</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Finnegan et al. [16], 2012</td>
</tr>
<tr>
<td>General skills transfer study</td>
<td>Lerner et al. [4], 2010</td>
<td>Kiely et al. [19], 2014</td>
</tr>
<tr>
<td></td>
<td>Lawson Health Research Institute [25], 2016</td>
<td>Stegemann et al. [20], 2013</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vaccaro et al. [23], 2013</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Korets et al. [22], 2011</td>
</tr>
</tbody>
</table>
was equipped with an advanced concurrent validity [2, 5].

Another virtual reality training platform, the dVSS, is literally an attachment of the da Vinci Surgical System from the clinical robot. Thus, it can be employed as a virtual simulator outside of operating hours. Superior face, content, and construct validity have been reported for this system compared to the MdVT in various studies [6–9].

The concurrent and predictive validities of the dVSS were explored by executing ex vivo experiments in one study, in which the finding suggested significant concurrent and predictive validities of the dVSS [10]. Furthermore, a later study demonstrated that the dVSS also possessed sound concurrent validity when performing in vivo exercises [11].

The RoSS is another simulator created to mimic the clinical robot, similar to the MvDT. It holds complete features of virtual reality [1]. However, no studies of construct validation for this modality have been published to date. Only the face and content validities of this simulator has been reviewed [12].

Till date, no standardized definitions of face, content, construct, concurrent, and predictive validities exist in the literatures. This needs to be considered and formalized for future studies. Very few RCTs actually compared the objective validity across the 3 different robotic simulators. Therefore, whether or not one simulator possessed a superior validity over the other had not been established, owing to the deficiency of such RCTs.

**Studies Specific to Simulated Procedures for Urology**

A cohort study assessed urethral anastomosis (UA) based on the most recent Tube group model called the ‘Tube Three’ model [13]. In this study, the Tube Three model was considered realistic and useful for training by expert urologists, indicated by its high numeric scores [13]. The study concluded with excellent face validity and superior content validity of this model. In terms of construct validity, the experts outperformed the beginners, yielding a statistically significant difference for all metrics and the overall score [13].

In addition to the fundamental skills that the simulator could offer to train for familiarizing with the real clinical robot interface and control, some virtual simulator programs were also tailored to provide urology training with specific augmented reality, for instance, partial nephrectomy by the MdVT and prostatectomy by the RoSS.

The augmented simulation of partial nephrectomy by the MdVT comprises of 3-dimensional videos of robotic partial nephrectomy, superimposed with virtual means to instruct anatomy, surgical skills, and most importantly each step of the operation.

One cohort study investigated this program in particular, by categorizing the cohort into experts, intermediate, and beginners [14]. Likewise, a numeric score was used to assess the subjective validities in this study. The MdVT was deemed highly realistic with a median score of 8 (total score of 10) for face validity. It was also appraised as a useful teaching tool for every section, including mobilization of the colon, kidney, duodenum kocherization, hilar dissection, and tumor resection, proving its content validity [14].

In addition, this study compared specialists with intermediate candidates for the renorrhaphy simulation task. As expected, the surgeons outperformed the intermediate trainees for all domains of the Global Evaluative Assessment of Robotic Skills (GEARS) score, highlighting the construct validity [14]. The concurrent validity was examined using the Spearman’s test in this study. As a result, the extent of correlation from a virtual reality renorrhaphy task to a live porcine robotic partial nephrectomy performance was analyzed in this study [14]. It was one of the few studies in which in vivo experiment was performed to explore the actual transferability of skills from the virtual simulator compared to live tissue operation. This study revealed a high correlation (r = 0.8) for all GEARS components [14].

Nonetheless, it is crucial to understand that the conclusion on the construct validity given by these observational cohort studies may not be accurate, as the research itself did not involve the comparison between an exposure and a control group. This could potentially create confounding biases. For instance, an expert urologist may already have better skills from own experience, and simply perform better with the virtual model. Furthermore, these cohort studies mainly focused on the subjective validity of the virtual simulators. Therefore, such discussions in these studies could merely give the audience an understanding of the skill transferability of the simulators into clinical practice.

In contrast, skill transferability is more likely to be better reflected by RCT, where a control group and an experimental group are compared. The control group is not exposed to simulator training, whereas the experimental group is tested after being exposed to simulator training. The comparators in these studies are the same, however, they differ in the training curriculum on which the simu-
lators is based, for instance, the dVSS or the MdVT. GEARS score is used by RCTs for the assessment of performance. It is composed of 6 domains, namely, depth perception, bimanual dexterity, efficiency, autonomy, force sensitivity, and robotic control.

Among robotic-assisted surgeries, prostatectomy is the most common urology surgery [3–8]. Available in programs of the MdVT and the dVSS, one series of tasks called the ‘Tubes group’ is designed specifically for UA in prostatectomy.

To date, 4 RCTs have examined UA by performing the Tubes group tasks [6, 7, 15, 16]. Their conclusions are consistent, showing statistically significant improvements in the experimental group compared to the control group. Accordingly, it can be summarized that the utilization of the virtual simulator to practice the ‘Tubes group’ series tasks could lead to a better performance of trainees, that is, effective construct validity.

Two RCTs assessed the candidate’s performance on the execution of the actual UA procedure on an inanimate model and cystotomy on an animal (swine) model [17, 18]. One was conducted on the MdVT while the other was on the RoSS. However, these studies showed contradicting results.

The MdVT study found no significant difference between the experimental and control groups in the overall GEARS score when performing cystotomy closure and overall GEARS score [17].

On the contrary, the RoSS study found that the experimental group, trained on an augmented reality module, outperformed the control group on an inanimate model for UA, with significant differences using both GEARS criteria and their exclusive UA score [18].

The objective UA score was developed as part of the study, advancing other research counterparts that only relied on GEARS score. The UA score is unique from the GEARS score in that it was tailored towards the practical aspects of performing an UA, including assessing needle positioning, needle driving, suture placement, and tissue manipulation [18]. Therefore, it added value to the literature and provided an alternative insight into future research designs.

Skills Transfer Studies in General Surgical Techniques

The discussion of these studies is equally significant, owing to the fact that urology procedures share many other similar general surgical techniques when using the robot.

A cohort study from 2010 was designed to investigate whether the MdVT could enhance surgical skills compared to the dVSS [4]. It was found that the MdVT was able to teach beginners better for the acquisition of essential robotic skills, such as picking up and transferring pegs, pattern cutting, string running, and intracorporeal knot tying [4].

The result analysis in this study was more standardized than other peer cohort studies, as it involved beginners practice sessions and subsequently compared their performance post-practice. Still, the performance assessment was achieved by observation from a surgical technician, whose title was left undefined, for instance, an expert surgeon, an anatomist, or a company representative. It is worth pondering the outcome of the score when assessed by an expert surgeon, who may not necessarily find some candidate accurate on task performance, compared to a technician.

Four RCTs assessed skill transferability in general on inanimate models [19–22], one RCT on animal tissues [23], and one cohort study on real patients [24].

The results across these studies were inconsistent. Five RCTs assessing the MdVT found that the experimental group trained on the simulator outperformed the control group [12, 19, 23, 24]. In one RCT that compared the performance by using the MdVT and the real da Vinci robot, equivalent performance was reported for both groups [22].

The inconsistent finding could be attributed to the use of different criteria in proficiency assessment in these studies. For instance, the assessment of proficiency was arbitrarily defined in 3 RCTs, for instance, using an overall score of 80% in 2 studies [19, 22], and 60% in the other study [23]. Furthermore, the assessment in one RCT was only based on time, rather than proficiency-based outcomes [21]. Last but not least, the major disadvantage of these studies is in relation to the small numbers of participants, which ranged from 12 to 53.

Thus far, the evaluation on real patients was only reported in one RCT that included 14 subjects in the experimental group and 4 in the control group for hysterectomy procedures [24]. The simulator-trained group outperformed the control group during hysterectomy with a da Vinci robot on the overall score, with significant differences for time and estimated blood loss [24]. Nonetheless, the randomization was unclear in this study and hysterectomy was not a primary urological procedure. Hence, the result may not be of significant reference for urology.

Currently, there is one ongoing RCT comparing different training modalities for robotic surgery involving virtual simulators. Assessment is conducted using GEARS
score for each task [25]. However, tasks in this study are cardiothoracic-related, which may not be transferable to urological procedures.

## Conclusion

A thorough literature review has identified a need for future studies to investigate operation-specific training, in conjunction with skills-based teaching in a more realistic augmented reality environment [26]. Moreover, it is important to incorporate teamwork, decision-making, and communication skills as these skills directly influence the overall performance of the trainees [26, 27]. Last but not least, when defining the validity of a virtual machine, more standardized definitions should be established across the different validities, such as face validity or concurrent validity. RCTs have the advantage of providing a more precise judgment to the audience than cohort studies, and therefore this should be the future trend for study design in this area [28]. Last but not least, given the number of RCTs incorporated in this review, it may be possible to perform a meta-analysis from the results of those studies as a follow-up study, to achieve the supreme level of evidence [28].

## Compliance with Ethical Standards

Both authors have no conflicts of interest to declare. This article does not contain any studies with human participants or animals performed by any of the authors.

## References


27 Porcaro AB, Siracusano S, De Luyk N, et al: Low-risk prostate cancer and tumor upgrading to higher patterns in the surgical specimen. Analysis of clinical factors predicting tumor upgrading to higher Gleason patterns in a contemporary series of patients who have been evaluated according to the modified Gleason score grading system. Urol Int 2016;97:32–41.