EUROPEAN UROLOGY 69 (2016) 526–535

available at www.sciencedirect.com
journal homepage: www.europeanurology.com

European Association of Urology

Education

Structured and Modular Training Pathway for Robot-assisted Radical Prostatectomy (RARP): Validation of the RARP Assessment Score and Learning Curve Assessment

Catherine Lovegrove[a], *Giacomo Novara*[b], *Alex Mottrie*[c], *Khurshid A. Guru*[d], *Matthew Brown*[e], *Ben Challacombe*[a], *Richard Popert*[a], *Johar Raza*[d], *Henk Van der Poel*[f], *James Peabody*[g], *Prokar Dasgupta*[a,*], *Kamran Ahmed*[a]

[a] *King's College London, Guy's Hospital, London, UK;* [b] *University of Padua, Padua, Italy;* [c] *OLV Clinic, Aalst, Belgium;* [d] *Roswell Park Cancer Institute, Buffalo, NY, USA;* [e] *Fiona Stanley Hospital, Perth, Australia;* [f] *Netherlands Cancer Institute, Amsterdam, The Netherlands;* [g] *Henry Ford Hospital, Detroit, MI, USA*

Article info

Article history:
Accepted October 23, 2015

Associate Editor:
James Catto

Keywords:
Assessment
Learning curve
Patient safety
Robot-assisted radical prostatectomy
Training

Abstract

Background: Use of robot-assisted radical prostatectomy (RARP) for prostate cancer is increasing. Structured surgical training and objective assessment are critical for outcomes.
Objective: To develop and validate a modular training and assessment pathway via Healthcare Failure Mode and Effect Analysis (HFMEA) for trainees undertaking RARP and evaluate learning curves (LCs) for procedural steps.
Design, setting, and participants: This multi-institutional (Europe, Australia, and United States) observational prospective study used HFMEA to identify the high-risk steps of RARP. A specialist focus group enabled validation. Fifteen trainees who underwent European Association of Urology robotic surgery curriculum training performed RARP and were assessed by mentors using the tool developed. Results produced LCs for each step. A plateau above score 4 indicated competence.
Outcome measurements and statistical analysis: We used a modular training and assessment tool (RARP Assessment Score) to evaluate technical skills. LCs were constructed. Multivariable Kruskal-Wallis, Mann-Whitney *U*, and κ coefficient analyses were used.
Results and limitations: Five surgeons were observed for 42 console hours to map steps of RARP. HFMEA identified 84 failure modes and 46 potential causes with a hazard score ≥8. Content validation created the RARP Assessment Score: 17 stages and 41 steps. The RARP Assessment Score was acceptable (56.67%), feasible (96.67%), and had educational impact (100%). Fifteen robotic surgery trainees were assessed for 8 mo. In 426 RARP cases (range: 4–79), all procedural steps were attempted by trainees. Trainees were assessed with the RARP Assessment Score by their expert mentors, and LCs for individual steps were plotted. LCs demonstrated plateaus for anterior bladder neck transection (16 cases), posterior bladder neck transection (18 cases), posterior dissection (9 cases), dissection of prostatic pedicle and seminal vesicles (15 cases), and anastomosis (17 cases). Other steps did not plateau during data collection.
Conclusions: The RARP Assessment Score based on HFMEA methodology identified critical steps for focused RARP training and assessed surgeons. LCs demonstrate the experience necessary to reach a level of competence in technical skills to protect patients.
Patient summary: We developed a safety and assessment tool to gauge the technical skills of surgeons performing robot-assisted radical prostatectomy. Improvement was monitored, and measures of progress can be used in future to guide mentors when training surgeons to operate safely.

* Corresponding author. King's College London, Guy's Hospital, Great Maze Pond, London SE 1 9RT, UK. Tel. +44 0 20 71885906.
E-mail address: prokarurol@gmail.com (P. Dasgupta).

http://dx.doi.org/10.1016/j.eururo.2015.10.048
0302-2838/

1. Introduction

Robot-assisted surgery training presents new challenges that must be addressed to minimise the risk of errors, patient harm, and adverse outcomes [1]. Effective training and assessment is fundamental to ensuring that surgeons operate safely [2–4]. Given changes in technology, reduced resources, and training time as a consequence of the European Working Time Directive, it is imperative that training and assessment adapt to reflect challenges faced by surgeons in the modern era [5–7]. At present, a significant gap in surgical education places patients at risk of adverse events. Modular training curricula have been developed, but they do not consider learning curves (LCs) or when technical competence may be attained by trainees [3]. Healthcare Failure Mode and Effect Analysis (HFMEA) is a form of human risk analysis validated for use in health care [8]. Initially, the technique was used in high-risk settings such as aviation and engineering. It has since been adapted for use in a medical setting [9,10]. Training tools should be developed using validated measures to ensure efficacy.

This study had two objectives: (1) Develop a modular training and assessment tool to examine the technical skills in robot-assisted radical prostatectomy (RARP) using HFMEA and validate them in terms of content validity, construct validity, interrater reliability, feasibility, acceptability, educational impact, and cost effectiveness, and (2) evaluate the LC for individual substeps of RARP.

2. Materials and methods

2.1. Study design

This multi-institutional observational prospective longitudinal study was conducted between September 2013 and August 2014. The HFMEA method of risk assessment was used to develop the RARP Assessment Score. Once developed, the RARP Assessment Score went through a process of validation. The RARP Assessment Score was then distributed to 15 surgeons and their mentors undertaking the European Association of Urology (EAU) Robotic Urology Section Fellowship Curriculum Pilot Study II, 14 in Europe and 1 in Australia (Fig. 1).

The study was conducted in two phases: (1) development and content validation of the RARP Assessment Score by evaluating reliability, validity, feasibility, acceptability, educational impact, and cost effectiveness and (2) construction and evaluation of the RARP LC.

2.1.1. Development and validation

For the development of the RARP Assessment Score, the HFMEA method of risk assessment was used as shown in Figure 2. A structured approach mapped the procedural steps with expert consensus. Recommendations of the Pasadena Consensus Panel were reviewed and included [11]. Following mapping, high-risk steps that could result in an adverse event were identified using hazard scoring. The hazard score is a product of severity and probability (Supplementary Fig. 1) [8,11]. Hazard scores $\geq$8 and single-point weaknesses were taken forward to implement actions and outcome measures in the RARP Assessment Score. An HFMEA hazard analysis was constructed and an international teleconference held with five experts (B.C., K.A., K.G., J.R., C.L.) allocating severity and probability scores (Supplementary Table 1).

Content validation occurred during HFMEA, and on completion the hazardous steps identified by HFMEA were included in an initial assessment for RARP. This list of crucial steps was circulated internationally among surgeons and operating theatre staff to confirm that all important steps had been included and for content validation. It was also circulated and revised through international content validation by experts in the United States, United Kingdom, Italy, and Belgium. This process finalised the RARP Assessment Score.

The RARP Assessment Score was distributed for use to 17 trainee surgeons and their mentors. These trainees had undergone training in accordance with the EAU robotic curriculum [3,6]. Mentors observed trainees as they performed steps of RARP in subsequent cases during training and used the assessment to score their technical performance for individual steps. At the close of the study, trainees submitted a final video for an objective video analysis. Mentors (experts) provided their videos of RARP for comparison. The videos were also evaluated by independent blinded reviewers using the RARP Assessment Score.

Construct validity was examined by a comparison of experts and trainees' outcomes on the RARP Assessment Score. Then κ coefficients were used to study interrater reliability (agreement between two assessors who used the RARP Assessment Score for evaluation). A questionnaire addressing feasibility, acceptability, and educational impact was distributed to 17 mentors and trainees who had ever used the RARP Assessment Score. Cost effectiveness was appraised by considering the financial cost involved in using the assessment.

2.1.2. Learning curve evaluation

Fifteen mentors trained and assessed 15 trainees between January and August 2014 using the RARP Assessment Score. Median previous console experience was 0 mo (interquartile range: 0–5.5 mo). These trainees went through modular training in the operating theatres. Modular training entails dissection of procedural steps into modules. Each module is graded according to requisite skills level from the lowest level of difficulty to the highest level [12]. Each module in the RARP Assessment Score was assigned a difficulty level during the checklist development process.

LCs were analysed using mean scores for substeps (modules) and plotting a scatter graph with score 3 ("average") as a reference value. Competence was defined as consistently scoring $\geq$4 ("good") of 5. The primary outcome measure was a score (1–5) for each step. Information on surgeon and patient variables was collected.

2.2. Statistical analysis

Scores were collated using Microsoft Excel (Redmond, WA, USA); statistical analysis used SPSS v.21 (IBM Corp., Armonk, NY, USA). Construct validity and interrater reliability used Mann-Whitney *U* tests and κ coefficient analysis, respectively. Comparison between professional position and previous experience was made using Kruskal-Wallis and Mann-Whitney *U* tests. LCs were constructed from a quadratic-derived curve of group scores. Multivariable analysis examined surgeons and patient variables. The p values <0.05 were statistically significant. SurveyMonkey.com was used for the questionnaire.

3. Results

3.1. Development and validation of the RARP Assessment Score

After involving the multidisciplinary team of anaesthetists, operating theatre assistants, scrub nurses, and theatre staff, seven surgeons participated in mapping the steps of RARP. After 42 h of console observation with five expert surgeons,