ELSEVIER

Urologic Oncology: Seminars and Original Investigations ▮ (2017) ▮▮▮–▮▮▮

UROLOGIC ONCOLOGY

Original article

A decade of robot-assisted radical prostatectomy training: Time-based metrics and qualitative grading for fellows and residents

Muammer Altok, M.D., Mary F. Achim, B.S., Surena F. Matin, M.D., Curtis A. Pettaway, M.D., Brian F. Chapin, M.D., John W. Davis, M.D.*

Department of Urology, MD Anderson Cancer Center, Houston, TX

Received 13 July 2017; received in revised form 25 August 2017; accepted 30 August 2017

Abstract

Objectives: As modern urology residency and fellowship training in robot-assisted surgery evolves toward standardized curricula (didactics, dry/wet-laboratory exercises, and surgical assistance), additional tools are needed to evaluate on-console performance. At the start of our robotics program in 2006, we set-up a time- and quality-based evaluation program and aim to consolidate this data into a simple set of metrics for self-evaluation.

Materials and methods: Using our index procedure of robot-assisted radical prostatectomy (RARP), we prospectively collected data on 2,215 cases over 10 years from 6 faculty surgeons and 94 trainees (43 urologic oncology fellows and 51 urology residents). The steps of the operation were divided into 11 consistent steps, and the metrics included time to completion and quality using a 6-level grading system. Time metrics were consolidated into quartiles for benchmarking.

Results: The median times for trainees to complete each step were 15% to 120% higher than those of the staff ($P < 0.001$). Each step can be presented with quartile-based time metrics by pooled trainee and staff results. Steps performed by trainees were carefully chosen for a high success rate, and on our Likert-like scale were graded 4 to 5 in more than 95% of cases. There were no grade 0 (very poor) cases, and grades 1 (multiple technical errors) and 2 (could not be completed but without safety issues) were rare ($<1\%$).

Conclusions: RARP training can be evaluated with a time-based metric that allows a quartile-based comparison to a large experience of trainees and staff. As a trainee progress through a rotation, these benchmarks can assist in prioritizing the need for more attention to a basic step vs. progression to more advanced steps.

Keywords: Education; Learning curve; Robot-assisted radical prostatectomy; Robotics; Training

1. Introduction

Over the last 2 decades, robot-assisted radical prostatectomy (RARP) has become an accepted minimally invasive surgical option for prostate cancer [1]. The cancer-control and quality-of-life outcomes achieved with radical prostatectomy are highly dependent on the surgeon's technique and skill [2–4]. Therefore, training fellows and residentsin RARP is of the utmost importance.

Considering the absence of validated training programs for RARP, in 2006 we started prospectively collecting data on training experiences and published a pilot study evaluating the time required for staff vs. trainees to perform each step of the procedure, as well as a subjective grading of the quality of execution [4]. After a decade of training 4 clinical fellows and up to 12 residents per year, we aimed to assess their step-wise time metrics and transform them into a simple table to use to benchmark performance. We also assessed the faculty surgeons' qualitative feedback on the trainees' performance.

Supported in part by the NIH/NCI under Award no. P30CA016672 and used the Clinical Trials Support Resource. Scholarship support (M.A.) from Association of Urooncology and Society of Urological Surgery (Turkey).

Dr. Davis is a consultant with Intuitive Surgical.

* Corresponding author. Tel.: +1 713 792-3250; fax: +1 713 794-4824.
E-mail address: johndavis@mdanderson.org (J.W. Davis).

http://dx.doi.org/10.1016/j.urolonc.2017.08.028
1078-1439/

2. Materials and methods

From July 2006 to January 2016, training data were prospectively collected from 2,215 patients undergoing RARP at a single, tertiary referral cancer hospital. Training data collection was voluntary (nonprotocol based) and represents approximately 1/3 of all cases (open/robotic) performed in this era by the department. IRB permission was obtained using a waiver of consent method to allow collection of the clinical data and training data from these cases. The cohort of patients included cases from 6 staff and 94 trainees, including 43 urological oncology fellows and 51 residents. This total includes the 4 fellows, additional residents (pooled), and 178 cases described in our pilot study, in which we tested and described our evaluation method [4]. A detailed description of the RARP technique taught to trainees is given [4]. Pelvic lymph node dissections done during this era were obturator-zone dissections, although we have subsequently adopted extended zones [5,6]. Additional training methods are detailed in the Supplemental methods.

There were no validated training tools for RARP before those we developed and described in the previous pilot study [4]. The training program was designed to breakup the procedure into discrete steps, time the steps, and ensure patient safety with a simple Likert-like quality metric. Over the course of a rotation and a clinical year, trainees would have metrics on overall console time and step-by step metrics, distribution, and progress. The 11 procedure steps were proposed in the pilot study and detailed: bladder drop, endopelvic fascia, dorsal vein ligation, anterior bladder neck, posterior bladder neck (including bladder base pedicles), SV/Vas, neurovascular bundles, apex, pelvic lymph nodes, posterior anastomosis, and anterior anastomosis. We evaluated the trainees independently for time to complete a procedure step (objective evaluation) and quality of results (objective and subjective evaluations).

The quality grading system was a simple Likert-like metric, and the staff intentionaly selected steps for the trainee with a high likelihood of success based upon their level of experience and previously demonstrated skills. In most cases, the staff directly first assisted their console time and "stopped the line" if any quality or excess time concerns arose. Therefore, the quality ratings were designed more for safety measurement and not expected to have a "distribution" effect. The Likert descriptors were grade 5, the trainees' quality of execution equal to that of staff with minimal coaching (verbal or visual); grade 4, mistakes requiring correction, additional steps, or hemostasis but easy to correct; grade 3, mistakes requiring correction, difficult to correct; grade 2, trainee could not complete the step (not due to case difficulty), but there were no safety issues; grade 1, multiple technical errors; and grade 0, very poor execution. Time per step did not factor into grading unless incomplete. As shown in our pilot study, specific comments on quality were recorded for any critique in performance, and collated into a step by step list of potential "pitfalls" to consider in training.

Clinical data on patients was abstracted including age, race/ethnicity, preoperative PSA level and prostate volume (assessed by transrectal ultrasound), BMI, biopsy Gleason score, clinical stage, risk group according to the NCCN guideline [1], nerve-sparing status, the extent of lymph node dissection.

Statistical analysis was performed using the SPSS 22.0 software program for Windows (SPPS Inc., Chicago, IL). The operative times (in minutes) for the staff and trainees (as a group) were reported as the medians. Data for steps that were only partially performed by the trainee or staff were excluded from the analysis, and the median times for steps performed solely by the staff and solely by the trainee were compared. Comparison of the numerical variables between the 2 groups was performed using the Kruskal-Wallis test. All $P < 0.05$ was considered statistically significant.

Table 1
Patient and surgical characteristics of patients who underwent RARP

Characteristics	N = 2,215
Age (y)	
Median (IQR)	60 (55–66)
Race/ethnicity, *n* (%)	
White	1730 (78)
African American	264 (12)
Hispanic	144 (7)
Asian	52 (2)
Other	25 (1)
PSA (ng/ml)	
Median (IQR)	5.4 (4.0–7.9)
TRUS prostate volume (ml)	
Median (IQR)	30 (24–40)
BMI (kg/m^2)	
Median (IQR)	28.7 (26.1–31.7)
Biopsy grade, *n* (%)	
3 + 3 or lower	463 (20.9)
3 + 4	981 (44.3)
4 + 3	389 (17.6)
4 + 4 or higher	380 (17.2)
Clinical stage, *n* (%)	
cT1	1481 (66.9)
cT2	673 (30.4)
cT3–4	60 (2.7)
Risk group, *n* (%)	
Low	431 (19.5)
Intermediate	1344 (60.7)
High-locally advanced	440 (19.9)
Nerve sparing, *n* (%)	
None	104 (4.7)
Unilateral	259 (11.7)
Bilateral	1852 (83.6)
PLND, *n* (%)	
None	593 (26.8)
Standard	535 (24.2)
Extended	1087 (49.1)

BMI = body mass index; PLND = pelvic lymph node dissection; PSA = prostate-specific antigen; RARP = robot-assisted radical prostatectomy; TRUS = transrectal ultrasonography.