J Robotic Surg
DOI 10.1007/s11701-011-0293-4

ORIGINAL ARTICLE

Surgeon perception is not a good predictor of peri-operative outcomes in robot-assisted radical prostatectomy

Joshua Stern · Saurabh Sharma · Pierre Mendoza · Mary Walicki · Rachel Hastings · Kelly Monahan · Baber Sheikh · Alexei Wedmid · David I. Lee

Received: 10 February 2011 / Accepted: 13 June 2011

Abstract Surgeons have always used their cognitive intuition for the execution of skilled tasks and real-time perception of intra-operative outcomes. We attempted to measure the overall accuracy of intra-operative surgeon perception on the functional outcome of early continence after robot-assisted radical prostatectomy (RARP). A single experienced surgeon (D.I.L.) used a scoring sheet to prospectively capture his subjective opinion of how well a particular portion of the RARP procedure was completed. Surgeon perception of factors affecting post-operative continence such as quality of bladder neck preservation, nerve sparing, urethral length, anastomosis, striated sphincter thickness, quality of Rocco repair and bladder neck plication suture (total 7 variables) were graded as "poor", "average" or "good". Urinary continence was graded as either total continence [0 pads per day (PPD) or social continence (security pad or one PPD)]. A total of 273 (39 patients × 7 variables) responses were recorded: 58.6% were rated as "good", 32.2% as "average" and 8.4% as "poor". A log-rank test for all perception variables showed no significant differences in subsequent achievement of continence (either 0 or 1 PPD) ($P > 0.05$) at both the 1- and 3-month time points. In the case of some perception variables, patients with "bad" scores gained continence a median of 3 weeks sooner than patients with "good" scores. Surgeon perception of intra-operative performance during RARP is a poor predictive indicator of subsequent functional outcome in terms of urinary continence. Inter-surgeon variability of perception may vary and needs further investigation.

Keywords Radical prostatectomy · Robotic surgery · Robotic prostatectomy · Prostate cancer

Abbreviations

RARP	Robot-assisted radical prostatectomy
PPD	Pads per day
VAS	Visual analogue scale

Introduction

Surgeons can evaluate their own performance in a surgery during the post-operative visit using objective measures such as retained or recovered erectile function, urinary continence, occurrence of post-operative complications and long-term development of urethral strictures and bladder neck contracture. There are, however, limited objective clinical criteria for measuring intra-operative success. Patients are becoming increasingly knowledgeable about prostate cancer [1], particularly since the 2002 US Preventive Services Task Force (USPSTF) recommendations on informed consent for prostate-specific antigen screening have encouraged primary care physicians to discuss with patients the potential risks of surgical intervention on the prostate [2]. Early in their post-operative course, patients commonly ask questions regarding their surgery such as, "Did you spare the nerves?", "How was the sewing together of the bladder neck to the urethra?" and "How soon will I be continent?"

J. Stern · S. Sharma · P. Mendoza · M. Walicki · R. Hastings · K. Monahan · B. Sheikh · A. Wedmid · D. I. Lee (✉)
Division of Urology, Penn Presbyterian Medical Center, University of Pennsylvania, 51 North 39th Street, Medical Office Building Suite 300, Philadelphia, PA 19104, USA
e-mail: david.i.lee@uphs.upenn.edu

Published online: 08 July 2011

Long-term continence rates are >90% in several large studies, including our own robot-assisted radical prostatectomy (RARP) database. Short-term continence may be confounded by patient-related factors. Many surgeons have data on overall continence rates in their practice and can present those statistics to patients. Based upon our own 3-month continence data, 75% of patients use 1 pad per day (PPD) or a security pad, with total continence of 0 PPD in 35%. In order to personalize the rates and predictions, surgeons answer these questions based on their perception of how well the surgery went. Perception, however, is a subjective criterion [3]. To assess the accuracy of a surgeon's perception, we created a data sheet that captured real-time assessment of 7 key steps during the RARP procedure. This data was compared to eventual early continence recovery.

Materials and methods

This study was conducted under an Institutional Review Board approved study protocol. A single experienced surgeon (D.I.L.) used a scoring sheet to prospectively document intra-operative self-assessments of specified portions of the RARP procedure. This surgeon had completed in excess of 2000 RARPs at the time of the study. Surgeon perception was graded on a ternary scale as "bad", "average" or "good". Seven key operative steps thought to affect continence were graded immediately following each respective step: (1) quality of bladder neck preservation, (2) cavernous nerve sparing, (3) urethral length, (4) quality of anastomosis, (5) striated sphincter thickness, (6) quality of posterior reconstruction, and (7) quality of bladder neck plication stitch (total 7 variables). Performance of these key steps is by definition subjective to the surgeon's perception of that individual task. Elements perceived as "good" technique focus on maintaining bloodless surgical planes and minimizing electrocautery and mechanical trauma. Quality of the bladder neck preservation focused on identifying the plane between the bladder neck and prostate using a bladder neck sparing technique while maintaining a consistent circumferential bladder wall thickness. Cavernous nerve sparing focused on athermal dissection of the neurovascular bundle, with "good" sparing indicating 90% nerve sparing bilaterally, "average" indicating 75–90% nerve sparing and "poor" indicating <75% nerve sparing. Urethral length was a subjective consideration of the spared apical urethra to which the anastomosis would be performed. Striated sphincter thickness focused on perceived thickness of the urethra and surrounding supporting muscles.

Quality of the anastomosis and posterior reconstruction focused on creating a secure, tension-free reconstruction. We use a 3-0 monocryl suture for both reconstructions, with a running double-armed Van Velthoven stitch for the anastomosis and a single 6-inch suture for the Rocco posterior reconstruction. We also utilize an anterior bladder neck plication stitch, to help funnel the distal bladder just proximal to the anastomosis, using a figure-of-8 stitch at the lateral bladder margins using a 3-0 monocryl suture. An 18-French catheter is then placed and the balloon is inflated with 10 ml of sterile water. We then hand irrigate with a bulb syringe to test the integrity of the anastomosis. The catheter remains for 1 week and no cystogram is performed prior to its removal.

Follow-up data for continence was available in 39 consecutive patients. Recovery of continence was measured in terms of social continence (security pad or 1 PPD) and total continence (0 PPD) at the 1- and 3-month time points post-operatively by using a structured questionnaire. Outcomes were then correlated using Kaplan–Meier survival analysis. The log-rank test was used to assess differences between curves. A P value of <0.05 was considered statistically significant. Statistical analysis was performed using SPSS version 16.0 (SPSS, Inc. Chicago, IL, USA).

Results

Thirty-nine patients were scored with a total of 273 responses (Table 1) and an average median follow-up of 3 months. There were no patient dropouts for analysis of short-term continence. Patient-related factors were typical

Table 1 Continence factors

Procedure	Average	Bad	Good
Bladder neck sparing ($n = 39$)	39% ($n = 15$)	28% ($n = 11$)	33% ($n = 13$)
Nerve sparing ($n = 39$)	31% ($n = 12$)	15% ($n = 6$)	54% ($n = 21$)
Urethral length ($n = 39$)	38% ($n = 15$)	3% ($n = 1$)	59% ($n = 23$)
Sphincter thickness ($n = 39$)	64% ($n = 25$)	–	36% ($n = 14$)
Rocco (posterior rhabdosphincter) repair ($n = 39$)	18% ($n = 7$)	8% ($n = 3$)	74% ($n = 29$)
Anastomosis ($n = 39$)	15% ($n = 6$)	–	85% ($n = 33$)
Bladder neck plication ($n = 39$)	21% ($n = 8$)	5% ($n = 2$)	69% ($n = 27$)

for our institution's population. Mean age was 58.8 years (range 42–72) with a mean body mass index (BMI) of 29.3 kg/m^2 (range 20.2–46.3). The pre-operative mean American Urological Association (AUA) symptom score was 7.6. Twenty-seven patients had mild symptoms (1–7), 8 had moderate symptoms (8–19), and 4 had severe symptoms (20–35). Mean prostate volume was 62.6 ml (range 29–169). Intra-operative conditions were relatively constant with a mean estimated blood loss of 138 ml (50–300). No urologic or other surgical complications were noted in the study group. Kegel exercises were strictly recommended at the initial pre-operative visit, to be continued indefinitely. Subjects all claimed compliance with the regimen and were reinforced to continue at each visit.

Our continence variability within the study at 1 month ranged from 0 pads up to 3-4 PPD, with one outlier at 12 PPD. At 3 months, pad use variability ranged from 0 to 3 PPD, with the same outlier now essentially dry using only a security pad. This outlier patient did have "severe symptoms" on pre-operative AUA score. Pre-operative AUA score alone was not a good predictor for this outlier, as the other 3 patients with "severe symptoms" on pre-operative AUA score varied in their post-operative continence, with 2–3 PPD at 1 month and 0–3 PPD at 3 months.

Patient or anatomic factors such as BMI and prostate gland size did not affect our key surgical step variables. High BMI may affect urethral length and repair, but the 5 most obese patients with BMI $\geq$35 only accounted for 2 of the 11 "poor" bladder neck preservations. The heaviest outlier at BMI 46 was found to have "good" urethral length and posterior repair and "average" anastomosis and bladder neck dissection. Large prostate volumes may make surgical dissection more difficult, but for our 8 patients with prostate volumes $\geq$80 ml, 1 had "good", 4 "average" and 3 "poor" bladder neck dissections, with no predilection for any particular outcome in any of the 7 key step variables.

Interestingly, a favorable surgeon assessment of urethral length and posterior repair—two steps commonly associated with continence—did not correlate with patient outcomes. For the assessment of preservation of urethral length, 15 were scored "average", 23 "good" and 1 "bad" (Table 1). At 3 months, 50% of those with an "average" score for urethral length had reached 0 PPD while only 30% of those scored "good" had reached 0 PPD. Similarly, the median time to reach 0 PPD was 13 weeks in the "average" group and not reached in the "good" group ($P = 0.81$). There was no correlation seen with any of the 7 variables measured (Tables 2, 3) for either total or social continence.

Only perception of the quality of the bladder neck dissection (Fig. 1) demonstrated a significant relation between our grading score and patients' return to continence. After the bladder neck dissection, 13 patients were perceived to have "good" bladder neck sparing, 15 "average" and 11 had what was considered "poor" bladder neck preservation. The median time to reach total continence, when analyzed using this single variable, was 12 weeks for those with a perceived good dissection while for the remaining 26 patients the median time was not reached ($P = 0.10$). Similarly, 62% of those scored to have a good dissection were pad-free at 3 months while 43% were pad-free with a poor dissection. However, only 16% of those with an average dissection were pad-free (Table 2), thus confounding this correlation.

Table 2 Reached total continence

Procedure	1 Month (%)	3 Months (%)	Median time (weeks)	*P* value[†]
Bladder neck				
Average	7	16	NR	0.1
Bad	0	43	NR	
Good	23	62	12	
Nerves				
Average	25	44	NR	0.57
Bad	0	25	NR	
Good	5	33	NR	
Urethral length				
Average	0	50	13	0.81
Good	17	30	NR	
Sphincter thickness				
Average	8	45	NR	0.28
Good	14	14	NR	
Rocco				
Average	29	64	13	0.19
Bad	0	0	*	
Good	7	33	NR	
Anastomosis				
Average	0	33	NR	0.8
Good	12	34	NR	
Bladder neck plication				
Average	12	34	NR	0.9
Good	12	37	NR	

[†] Log-rank test; * Cannot be calculated; all patients are censored

Discussion

In this study, we found that a surgeon's intra-operative perception of seven possible continence-related steps was not a good predictor of the post-operative recovery of continence after RARP. Karlizek et al. [4], in a study of 191 patients undergoing colorectal surgery, evaluated the surgeon's predictive accuracy for anastomosis leakage

Table 3 Reached social continence

Procedure	1 Month (%)	3 Month (%)	Median time (weeks)	*P*-value*
Bladder neck				
Average	40	85	7	0.32
Bad	18	73	12	
Good	54	77	2	
Nerve sparing				
Average	50	92	4	0.35
Bad	33	100	7	
Good	33	67	10	
Urethral length				
Average	27	90	10	0.91
Good	48	74	7	
Sphincter thickness				
Average	40	80	8	0.63
Good	36	79	11	
Rocco				
Average	29	52	12	0.45
Bad	67	67	1	
Good	38	87	8	
Anastomosis				
Average	17	83	8	0.66
Good	42	78	8	
Bladder neck plication				
Average	50	83	4	0.49
Good	37	79	10	

* Log-rank test

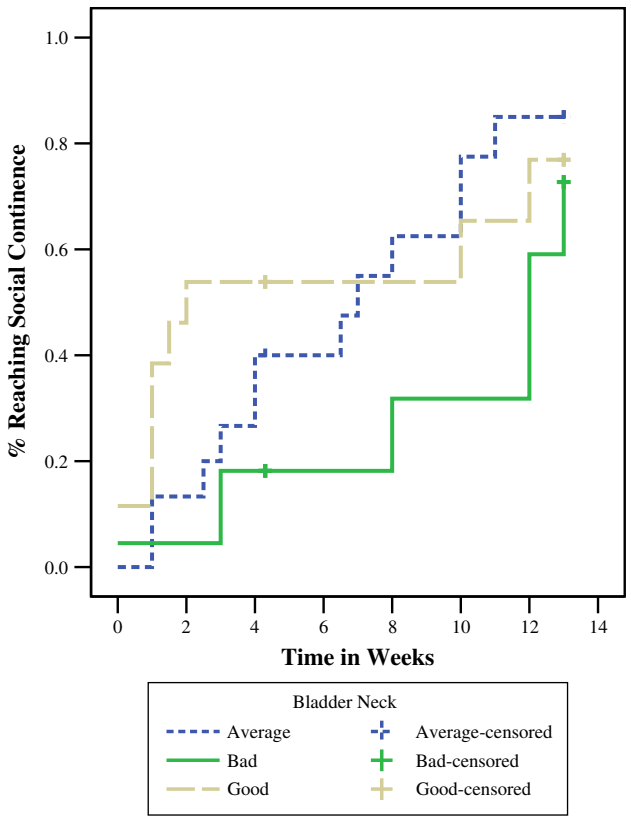

Fig. 1 Kaplan–Meier curves showing the difference in recovery of social continence (1 PPD) in groups with "good", "average" and "bad/poor" bladder neck preservation

using a visual analogue scale (VAS). They concluded that surgeons' clinical risk assessment appeared to have a low predictive value for anastomotic leakage. Additionally, the diagnostic accuracy was not influenced by the surgeons' training level (attending vs. resident). This is in agreement with our findings for continence outcomes. Others have sought to further understand surgical perception and decision-making (Fig. 2) [5]. Way et al. [6] analyzed 252 cases of bile duct injury and concluded that errors leading to bile duct injury stem principally from misperception and not errors of knowledge or judgment. This underscores the fact that perception can be misleading.

In contrast to our findings, several studies have found surgeon assessment to be a strong predictor of patient outcomes. Hartley et al. [7] investigated the predictive power of the surgeon's "gut feeling" in gastrointestinal surgery patients using a similar ternary grading schema and found that surgeon perception was a good indicator of postoperative course, and even found it to be as predictive as or better than the Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity (*POSSUM*) [8]. Markus et al. [9] in a similar study described grading of surgeon assessment via a percentage rather than a ternary scale and obtained the same result. In a prospective observational study, Woodfield et al. [10] used a VAS for surgeons to estimate post-operative risk of complications before and after performing abdominal surgery. In this study, pre-operative VAS scores were strongly correlated with incidence of post-operative complications; furthermore, the authors found that an increase in the VAS score in the post-operative assessment was correlated with significantly higher complication and mortality rates. Arvidsson et al. [3] found that a VAS for "intuitively appreciated global risk" was the most efficient predictor of post-operative adverse events. The emphasis here was on global risk and not on any specific event.

Sandblom et al. [11], in a prospective study, measured the outcomes of hernia repair using a ternary scale, and found that neither the grade of difficulty nor the surgeon's perception of the quality of repair significantly predicted the final outcome. Similarly, we found that "surgical technique" in isolation is very difficult to assess solely on the basis of perception. Only a handful of studies have examined the predictive capacities of surgeon perception for patient outcome; the studies discussed above represent

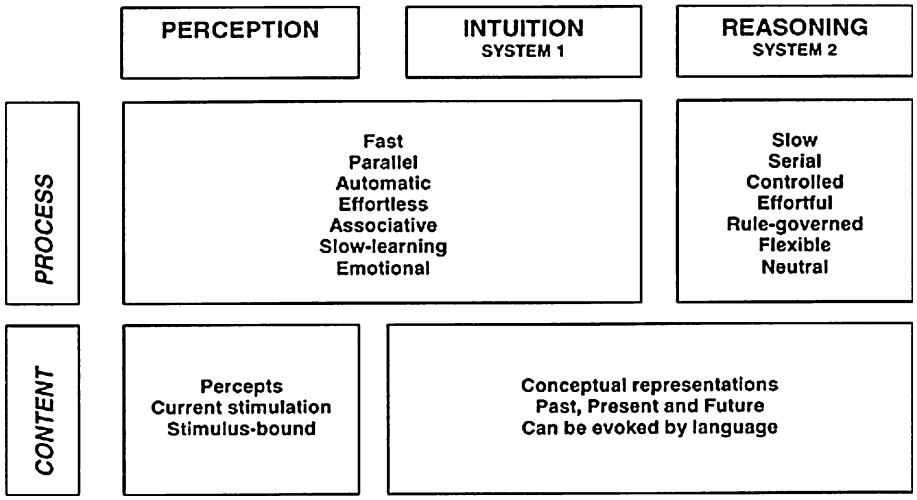

Fig. 2 Process and content in two cognitive systems. Reproduced with permission from American Psychological Association

a large proportion of the available literature. To our knowledge, our study is the first such study involving robot-assisted radical prostatectomy.

We do recognize several limitations of our study which should be addressed in future studies. Our study involved a single and highly experienced surgeon who had performed more than 2,000 RARP operations. At this level, it is our belief that performance is very consistent, with minimal variability in execution of skilled tasks and intra-operative maneuvers. What is perceived as a difference by a very experienced surgeon may not even be appreciated by a surgeon very early in their learning curve. The terms "best"/"average"/"poor" may also be subjective depending on the stage and level of training of the surgeon. Likewise, with experience, the true difference between the "best" and the "poor" performances may be negligible and may not have a large effect on outcomes. Increased variability in surgical performance is more likely early in one's learning curve and thus may translate to more clinical relevance and impact. Inter-surgeon variability in assessing an identical operation (i.e. video review of another's case) must be also further investigated [12].

Though our study did not find "surgeon perception" to be a useful tool for predicting continence outcomes, still we cannot deny the importance of this factor. Certainly surgical outcomes improve with increasing experience. This likely results in large part from the repetition of performing the procedure, the subtle refinements of dissection technique, and improved visual recognition of key anatomic landmarks. If these key improvements in technique and perception can somehow be identified and taught, more rapid improvement in teaching radical prostatectomy might occur. However, cognition is a complex process and improved study designs, finer criteria of assessment, or improved procedural review tools may help in elucidating its implication. This underscores the importance of further studies investigating new technologies that offer real-time, objective feedback during robotic surgery to further enhance the perception that has been gained through surgical experience. Surgeon perception, in conjunction with an objective predictive model, may be a helpful tool in predicting outcomes [10].

Conclusion

Overall, we found that surgeon perception of intra-operative performance during RARP lacks predictive power with respect to continence. Many interesting questions remained to be explored in the field of surgical judgment and perception.

Conflict of interest None.

References

1. Fitzpatrick JM, et al (2009) Awareness of prostate cancer among patients and the general public: results of an international survey. Prostate Cancer Prostatic Dis 12:347-354.
2. Harris R, Lohr KN (2002) Summaries for patients. Screening for prostate cancer: a recommendation from the US Preventive Services Task Force. Ann Intern Med 137(11):I48
3. Arvidsson S et al (1996) Predicting postoperative adverse events. Clinical efficiency of four general classification systems. The project perioperative risk. Acta Anaesthesiol Scand 40(7):783–791
4. Karliczek A et al (2009) Surgeons lack predictive accuracy for anastomotic leakage in gastrointestinal surgery. Int J Colorectal Dis 24(5):569–576
5. Kahneman D (2003) A perspective on judgment and choice: mapping bounded rationality. Am Psychol 58(9):697–720
6. Way LW et al (2003) Causes and prevention of laparoscopic bile duct injuries: analysis of 252 cases from a human factors and cognitive psychology perspective. Ann Surg 237(4):460–469
7. Hartley MN, Sagar PM (1994) The surgeon's 'gut feeling' as a predictor of post-operative outcome. Ann R Coll Surg Engl 76(6 Suppl):277–278

8. Copeland GP, Jones D, Walters M (1991) POSSUM: a scoring system for surgical audit. Br J Surg 78(3):355–360
9. Markus PM et al (2005) Predicting postoperative morbidity by clinical assessment. Br J Surg 92(1):101–106
10. Woodfield JC et al (2007) Accuracy of the surgeons' clinical prediction of perioperative complications using a visual analog scale. World J Surg 31(10):1912–1920
11. Sandblom G, Sevonius D, Stael von Holstein C (2009) Impact of operative time and surgeon satisfaction on the long-term outcome of hernia repair. Hernia 13(6):581–583
12. Ouriel K et al (1990) Factors determining survival after ruptured aortic aneurysm: the hospital, the surgeon, and the patient. J Vasc Surg 11(4):493–496