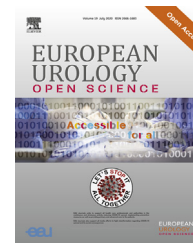


available at www.sciencedirect.com
journal homepage: www.eu-openscience.europeanurology.com



Prostate Cancer

Surgeon Automated Performance Metrics as Predictors of Early Urinary Continence Recovery After Robotic Radical Prostatectomy—A Prospective Bi-institutional Study

Andrew J. Hung^{a,*}, Runzhuo Ma^a, Steven Cen^b, Jessica H. Nguyen^a, Xiaomeng Lei^b, Christian Wagner^c

^a Center for Robotic Simulation & Education, Catherine & Joseph Aresty Department of Urology, USC Institute of Urology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA; ^b Department of Radiology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA; ^c Department of Urology, Pediatric Urology and Urologic Oncology, St. Antonius-Hospital, Gronau, Germany

Article info

Article history:

Accepted March 3, 2021

Associate Editor:

Guillaume Ploussard

Keywords:

Prostatectomy
Urinary continence
Automated performance metrics
Surgeon performance

Abstract

Background: During robotic surgeries, kinematic metrics objectively quantify surgeon performance.

Objective: To determine whether clinical factors confound the ability of surgeon performance metrics to anticipate urinary continence recovery after robot-assisted radical prostatectomies (RARPs).

Design, setting, and participants: Clinical data (patient characteristics, continence recovery, and treatment factors) and surgeon data from RARPs performed between July 2016 and November 2018 were prospectively collected. Surgeon data included 40 automated performance metrics (APMs) derived from robot systems (instrument kinematics and events) and summarized over each standardized RARP step. The data were collected from two high-volume robotic centers in the USA and Germany. Surgeons from both institutions performed RARPs. The inclusion criteria were consecutive RARPs having both clinical and surgeon data.

Intervention: RARP with curative intent to treat prostate cancer.

Outcome measurements and statistical analysis: The outcome was 3- and 6-mo urinary continence recovery status. Continence was defined as the use of zero or one safety pad per day. Random forest (SAS HPFOREST) was utilized.

Results and limitations: A total of 193 RARPs performed by 20 surgeons were included. Of the patients, 56.7% (102/180) and 73.3% (129/176) achieved urinary continence by 3 and 6 mo after RARP, respectively. The model anticipated continence recovery (area under the curve = 0.74, 95% confidence interval [CI] 0.66–0.81 for 3-mo, and area under the curve = 0.67, 95% CI 0.58–0.76 for 6 mo). Clinical factors, including pT stage, confounded APMs during prediction of continence recovery at 3 mo after RARP ($\Delta\beta$ median –13.3%, interquartile range [–28.2% to

* Corresponding author. University of Southern California Institute of Urology, 1441 Eastlake Avenue Suite 7416, Los Angeles, CA 90089, USA. Tel. +1 323-865-3700; Fax: +1 323-865-0120. E-mail address: Andrew.Hung@med.usc.edu (A.J. Hung).



–6.5%]). After adjusting for clinical factors, 11/20 (55%) top-ranking APMs remained significant and independent predictors (ie, velocity and wrist articulation during the vesicourethral anastomosis). Limitations included heterogeneity of surgeon/patient data between institutions, although it was accounted for during multivariate analysis.

Conclusions: Clinical factors confound surgeon performance metrics during the prediction of urinary continence recovery after RARP. Nonetheless, many surgeon factors are still independent predictors of early continence recovery.

Patient summary: Both patient factors and surgeon kinematic metrics, recorded during robotic prostatectomies, impact early urinary continence recovery after robot-assisted radical prostatectomy.

© 2021 The Authors. Published by Elsevier B.V. on behalf of European Association of Urology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Surgeon performance impacts clinical outcomes [1]. Evidence continues to demonstrate that superior surgeon performance is associated with superior outcomes after surgery [1,2].

To date, the gold standard method for assessing surgeon performance has been manual evaluation based upon validated assessment tools. This feedback, while informative, is not consistent between evaluators, nor is it scalable unless done through crowd-sourced evaluation [3,4].

Surgeon automated performance metrics (APMs), derived from instrument kinematic and system events data during robot-assisted surgery, are validated and truly objective measurements of surgical performance [2,5,6]. Since the data collection and resulting metrics are automatically derived, they represent an opportunity to evaluate surgeons in a scalable manner that was unavailable previously. We have found that APMs can be utilized to anticipate short- [6] and long-term outcomes (ie, continence recovery) [2]. However, while representing surgeon performance, APMs are also impacted by patient factors [7]. We have yet to quantify how clinical factors may impact the ability of APMs to predict outcomes.

Herein, we have utilized prospectively collected data from two high-volume robotic centers to evaluate the relationship between surgeon performance (APMs) and urinary continence recovery after a robot-assisted radical prostatectomy (RARP), adjusting for patient confounders. We also examined APMs' predictive ability in different patient factor subgroups (effect modification).

2. Patients and methods

Surgeons from two high-volume institutions contributed RARP case data from July 2016 to November 2018. All RARPs were performed using the anterior, non-Retzius-sparing approach. Consecutive cases that had the required data were included: (1) recorded robot system data during surgery to derive APMs, (2) baseline clinical characteristics (Table 1), and (3) 3- and 6-mo continence recovery status. Data collection protocols were standardized at both institutions.

APMs derived directly from the da Vinci surgical system data (instrument kinematics and system events) were computed as

previously described [2,5]. APMs were reported during each of 11 standardized steps. As both contributing institutions are teaching hospitals, step completion by the *primary surgeon* (faculty) or *secondary surgeon* (trainee involved in case) was also noted in all cases, and such designation was included in the predictive model.

Clinical data consisted of both patient and treatment factors (Table 1). The endpoint for prediction was postoperative urinary continence recovery status at 3 and 6 mo after RARP. Continence was defined as zero pads or one safety pad per day [8]. These data were prospectively collected by an independent research coordinator at either center utilizing patient-reported outcomes. In total, 454 features per case were utilized for the predictive modeling, including 14 clinical features (nine patient factors + five treatment factors) and 440 APMs (40 APMs × 11 steps).

Demographics and clinical characteristics were presented in a descriptive table by status of continence. Data were examined for normality. For normally distributed data, independent *t* test was used for descriptive analysis, while the Wilcoxon rank sum test was used for data with poor normality. Chi-square test was used for categorical data. Spearman's correlation was used to examine the association between clinical features and surgeon APMs.

Random forest (SAS HPFOREST) with 1000 trees utilized APMs and clinical factors to anticipate urinary continence recovery at 3 and 6 mo after RARP. Of original observations, 60% were used for bootstrapping each tree. Each candidate variable must have met the $p \leq 0.05$ threshold for a node to be split. The maximal depth of tree was set to 50. Of the data, 10% were reserved as the independent testing sample. We repeated this 90% versus 10% validation procedure ten times, with ten mutually exclusive 10% testing datasets. The area under the curve (AUC) was used to assess model performance. We compared the AUCs when using APMs alone, clinical factors alone, or APMs and clinical factors jointly by *z* test.

Variables of importance (VOIs) were selected by the out-of-bag Gini index. Ranking of important features was generated based on the frequency of appearance of the features in the top 20 for each of the 90% learning versus 10% testing procedures (reiterated ten times). They were then ranked on the average Gini index if tied on frequency count.

The top 20 VOIs from APMs were then examined for confounder and modifier effect from clinical and demographical measurements. A Poisson regression with generalized estimation equation (GEE) estimate was used to incorporate the nested data structure where patients clustered within surgeons. Model fitting was tested by the goodness-of-fit chi-square test. If $p > 0.05$ for the goodness-of-fit test, we concluded that the model was well fit; otherwise, we investigated both linear and overdispersion assumption. Each candidate confounder was examined by a univariate model with a single APM and a single candidate confounder. If >10% change was made in the association between an APM and the outcome after adjusted by the confounder, and the

Table 1 – Demographic characteristics of the cohort with continence recovery status at 3 mo after RARP

Characteristics	No continence at 3 mo Median (IQR)/mean \pm SD ^a (N = 78)	Continence at 3 mo Median (IQR)/mean \pm SD ^a (N = 102)	p value
<i>Patient factors</i>			
Age (yr)	66.6 \pm 7.1	64.2 \pm 6.7	0.02
BMI (kg/m ²)	28.9 (25.6–31.4)	26.9 (25.4–29.4)	0.02
ASA	3 (2–3)	2 (2–3)	<0.01
PSA (ng/ml)	7.3 (5.4–10.6)	7.5 (5.9–10.3)	0.90
Preop ISUP grade groups, n (%)			0.56
1	13 (16.7)	23 (22.6)	
2–3	48 (61.5)	61 (59.8)	
4–5	17 (21.8)	18 (17.7)	
Postop ISUP grade groups, n (%)			0.26
1	5 (6.4)	14 (13.7)	
2–3	55 (70.5)	69 (67.7)	
4–5	18 (23.1)	19 (18.6)	
Pathological tumor stage, n (%)			<0.01
pT2	27 (34.6)	55 (53.9)	
\geq pT3	51 (65.4)	47 (46.1)	
Prostate weight (g)	51.0 (40–67)	43.5 (36–55)	<0.01
Positive surgical margin, n (%)			0.46
No	65 (83.3)	89 (87.3)	
Yes	13 (16.7)	13 (12.7)	
<i>Treatment factors</i>			
Nerve sparing, n (%)			0.04
No nerve sparing	12 (15.4)	6 (5.9)	
Partial or full nerve sparing	66 (84.6)	96 (94.1)	
Bladder neck reconstruction, n (%)			0.55
No	71 (91.0)	90 (88.2)	
Yes	7 (9.0)	12 (11.8)	
Posterior reconstruction, n (%)			<0.01
No	23 (29.5)	60 (58.8)	
Yes	55 (70.5)	42 (41.2)	
Urethropexy, n (%)			0.69
No	29 (37.2)	35 (34.3)	
Yes	49 (62.8)	67 (65.7)	
Radiation after surgery, n (%)			0.07
No	63 (80.8)	92 (90.2)	
Yes	15 (19.2)	10 (9.8)	
ASA = American Society of Anesthesiology physical status classification system; BMI = body mass index; IQR = interquartile range; ISUP = International Society of Urological Pathology; PSA = prostate-specific antigen; RARP = robotic-assisted radical prostatectomy; SD = standard deviation.			
^a Continuous variable with normal distribution were compared by <i>t</i> test and reported as mean \pm SD; when not normally distributed, they were compared by Wilcoxon rank sum test and reported as median (IQR). Categorical variables were compared by chi-square test.			

confounder had $p < 0.05$, the confounder would enter the multivariate model for the given APM. We repeated this procedure for all APMs identified by the Random forest model as a VOI. A similar procedure was applied to the effect modifier test, but the impact of the effect modifier was examined by each candidate effect modifier individually, and not jointly. SAS 9.4 was used for all data analyses.

3. Results

In total, we accrued 193 RARP cases. From the University of Southern California, we included 116 patients by 11 faculty surgeons (robotic experience: median 375 cases [interquartile range {IQR} 250–1900]). From St. Antonius-Hospital, Gronau, we included 77 patients from nine faculty surgeons (median 397 cases [IQR 167–2822]).

From the combined cohort, 56.7% (102/180) of patients achieved urinary continence by 3 mo, while 73.3% (129/176) of patients achieved continence by 6 mo.

Overall, there were weak-moderate yet statistically significant correlations between APMs and clinical char-

acteristics ($\rho = -0.59$ to 0.40 , $p < 0.05$). The specific strength of these correlations varied across different APM categories. For example, a positive correlation was observed between frequency of energy usage during neurovascular bundle (NVB) dissection and prostate volume ($\rho = 0.40$, $p < 0.01$), while a negative correlation was observed between the idle time of the third arm during NVB dissection and patient body mass index (BMI; $\rho = -0.23$, $p = 0.03$). These overall correlations warranted that we put both clinical factors and APMs into the prediction model.

Utilizing the available clinical factors and APMs, the Random forest model achieved an AUC of 0.74 (95% confidence interval [CI] 0.66–0.81) for 3-mo and 0.67 (95% CI 0.58–0.76) for 6-mo continence predictions (Fig. 1). For the 3-mo prediction, we observed a trend of having superior performance utilizing the combined clinical factors and APM datasets to that utilizing the clinical factors or APM datasets alone. However, we did not achieve statistical significance in these comparisons ($p > 0.05$).

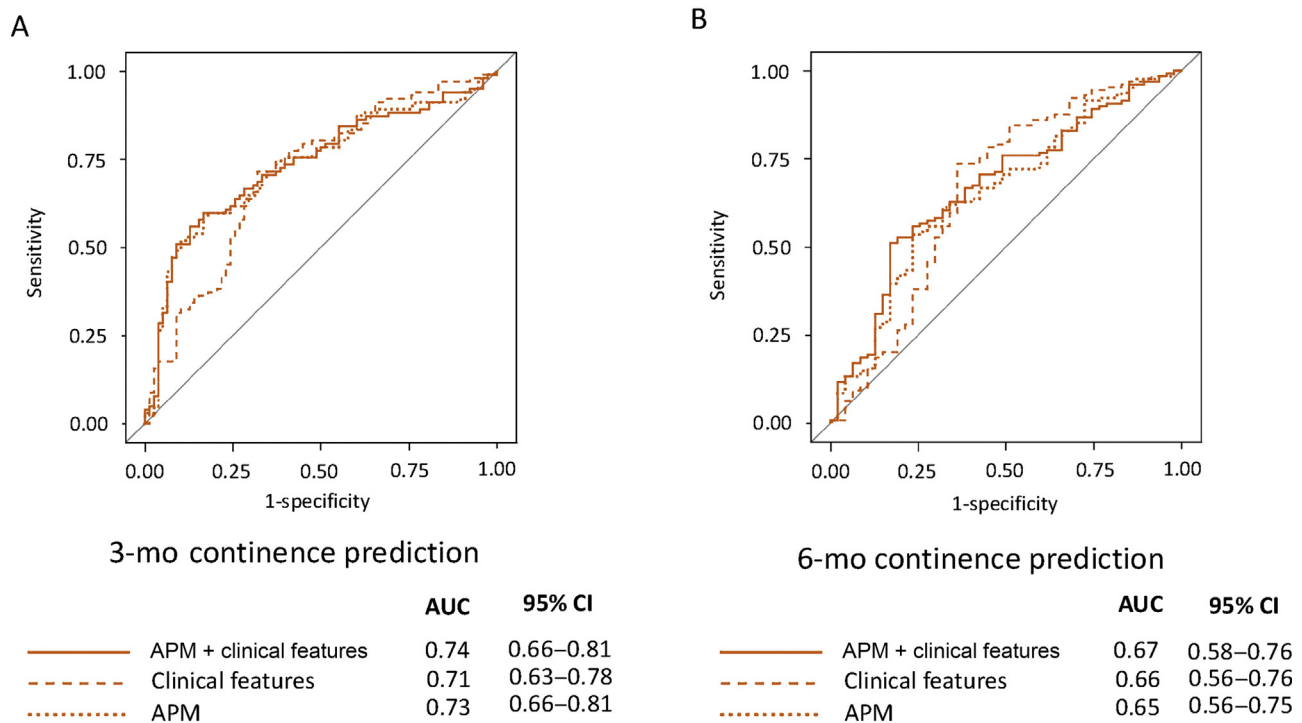


Fig. 1 – Receiver operating characteristic curves of predictive models for 3- and 6-mo continence recovery. APM=automated performance metric; AUC=area under curve; CI=confidence interval.

Given the higher performance for 3-mo continence prediction, we further report detailed results for the 3-mo data.

Patients who achieved or did not achieve urinary continence at 3 mo after RARP differed in clinical characteristics (Table 1). Patient factors were as follows:

those who achieved early continence were younger (mean 64.2 vs 66.6 yr; $p=0.02$), had lower BMI (median 26.9 vs 28.9 kg/m²; $p=0.02$), had fewer comorbidities (American Society of Anesthesiologists median score 2 vs 3; $p<0.01$), and had smaller prostate sizes (43.5 vs 51.0 g; $p<0.01$). Further, patients who achieved 3-mo continence recovery

Table 2 – Top 20 predictive features ranking by random-forest model^a

Ranking	Surgical step	APMs	Operating robotic instrument/system event	Frequency	Average out-of-bag Gini score × 1000	OR (95% CI) in multivariate analysis
1	Right NVBD	Linear velocity	Nondominant	9	0.25	1.22 (1.10–1.36)
2	Posterior VUA	Total wrist articulation	Dominant	7	0.50	0.84 (0.69–1.02)
3	Posterior VUA	Axis 1 wrist translation	Nondominant	7	0.39	0.87 (0.72–1.05)
4	Posterior VUA	Axis 2 wrist translation	Nondominant	7	0.35	0.83 (0.67–1.03)
5	Anterior VUA	Shaft rotation	3rd arm	7	0.34	1.08 (1.00–1.17)
6	Left PLND	Linear velocity	Camera	7	0.27	0.69 (0.57–0.84)
7	Posterior VUA	Mean time of each camera movement	Camera	6	0.48	1.17 (1.01–1.36)
8	Posterior VUA	Shaft rotation	3rd arm	6	0.38	1.10 (0.99–1.23)
9	Anterior VUA	Linear velocity	Camera	6	0.35	0.87 (0.72–1.04)
10	Anterior BND	Shaft rotation	Dominant	6	0.30	1.15 (1.07–1.23)
11	Apical dissection	Idle time	Dominant	6	0.25	0.80 (0.69–0.93)
12	Anterior BND	Moving time	Camera	5	0.33	1.25 (1.08–1.44)
13	Anterior BND	Total wrist articulation	Dominant	5	0.27	1.14 (1.04–1.26)
14	Posterior VUA	Axis 2 wrist translation	3rd arm	5	0.22	1.10 (0.98–1.23)
15	Posterior VUA	Angular velocity	Nondominant	4	0.36	0.80 (0.72–0.90)
16	Posterior VUA	Shaft rotation	Nondominant	4	0.34	0.87 (0.73–1.05)
17	Posterior plane dissection	Path length ratio	Dominant/nondominant	4	0.33	1.05 (0.96–1.14)
18	Anterior VUA	Axis 1 wrist translation	3rd arm	4	0.29	1.08 (1.02–1.15)
19	Right PLND	Linear velocity	Camera	4	0.21	0.88 (0.74–1.05)
20	Left NVBD	Master clutch usage	System event	4	0.19	1.10 (1.03–1.18)

APMs = automated performance metrics; BND = bladder neck dissection; CI = confidence interval; NVBD = neurovascular bundle dissection; OR = odds ratio; PLND = pelvic lymph node dissection; VUA = vesicourethral anastomosis.

^a Features were ranked by frequency of appearance during individual iterations of ten-fold cross validation, and then by average out-of-bag Gini score.

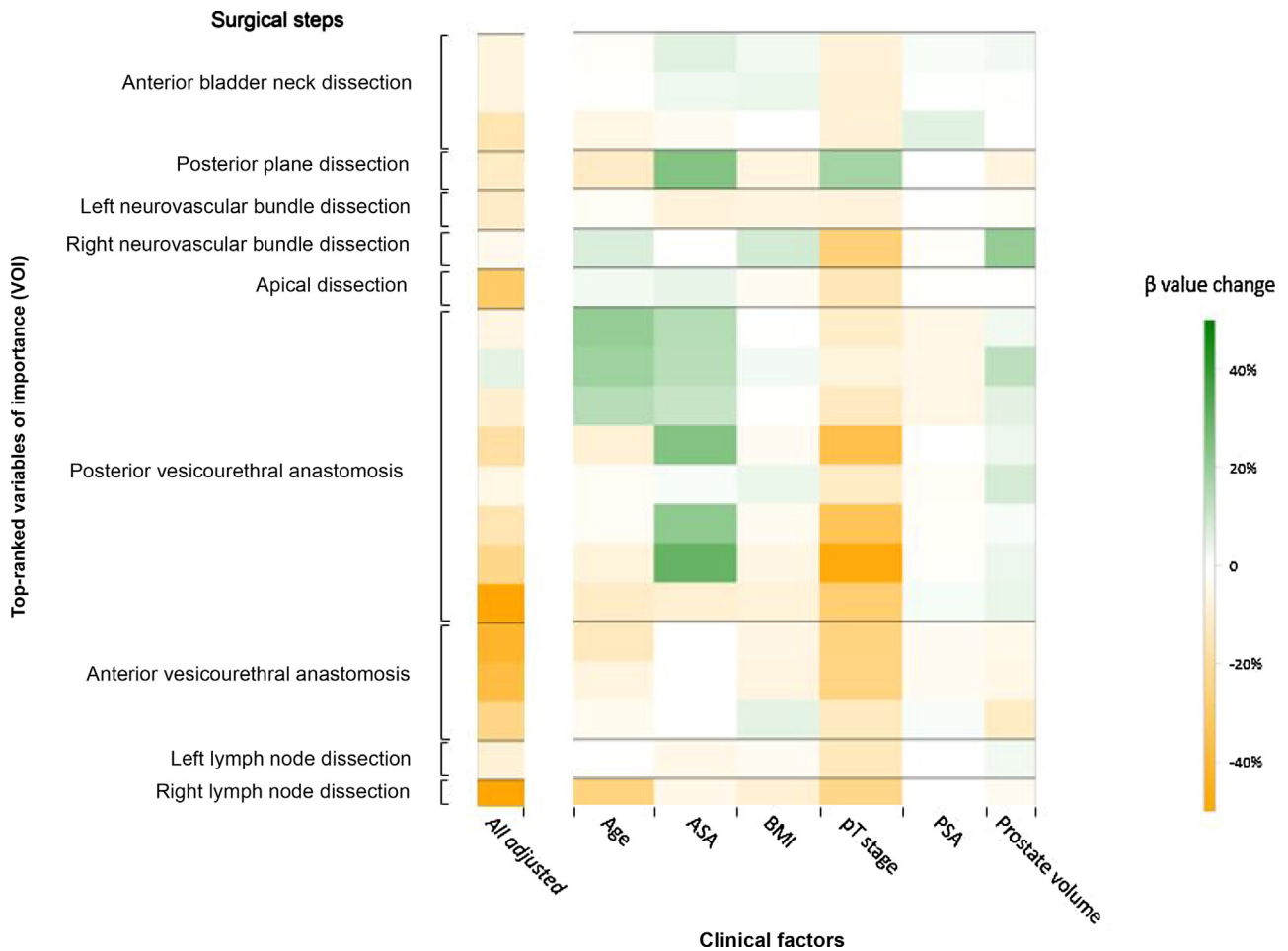


Fig. 2 – Heatmap showing percentage of β value change after adjusting for confounders during 3-mo continence prediction. All top 20 VOIs are automated performance metrics (APMs). ASA=American Society of Anesthesiologists; BMI=body mass index; PSA=prostate-specific antigen; VOI=variable of importance.

also had a greater proportion of organ-confined (pT2) disease (53.9% vs 34.6%; $p < 0.01$). Treatment factors were as follows: A greater proportion of patients who recovered continence at 3 mo had at least a partial nerve-sparing operation (94.1% vs 84.6%; $p = 0.04$). A greater proportion who recovered by 3 mo did *not* have a “posterior reconstruction” performed (58.8% vs 29.5%; $p < 0.01$). A smaller proportion of patients who had 3- versus 6-mo continence recovery underwent adjuvant radiotherapy during the 12 mo after RARP (9.8% vs 19.2%; $p = 0.07$).

Using the gradient of frequency count and average out-of-bag Gini index from the ten iterations of cross validation, we selected the top 20 VOIs that contribute to an accurate prediction of 3-mo continence status (Table 2). Notably, these features were all APMs (no clinical factors) and all belonged to steps performed by the primary surgeon. Of these 20 metrics, 11 (55%) were belonging to the vesicourethral anastomosis (VUA), the critical reconstruction (suturing) step of an RARP. Ten of these 11 (91%) VUA metrics were wrist articulation APMs. Other metrics occurred during bladder neck dissection, NVB dissection,

pelvic lymph node dissection, and posterior plane dissection.

We next evaluated for the confounding effect of clinical factors (Fig. 2). We confirmed that clinical factors have a confounding effect on the ability of APMs to predict continence recovery at 3 mo after RARP ($\Delta\beta$ median -13.3% IQR $[-28.2\% \text{ to } -6.5\%]$; Fig. 2). Prior to adjusting for confounding effect, 14/20 (70%) top VOI APMs were significant predictors of urinary continence status. After adjusting for clinical factors, 11/20 (55%) VOI APMs remained significant and independent predictors of early (3-mo) continence recovery, particularly the wrist articulation metrics during the VUA.

In certain circumstances, the effect on APMs' ability to predict continence recovery depends on the variability of patient factors (effect modification; Fig. 3). For example, when patients have organ-confined disease (pT2), increasing nondominant instrument articulation (sum of all angles) does not impact 3-mo continence recovery after RARP with a rate ratio of 1.07 (95% CI: 0.89, 1.29). However, for locally advanced cases ($\geq pT3$), the increase of 1 standard

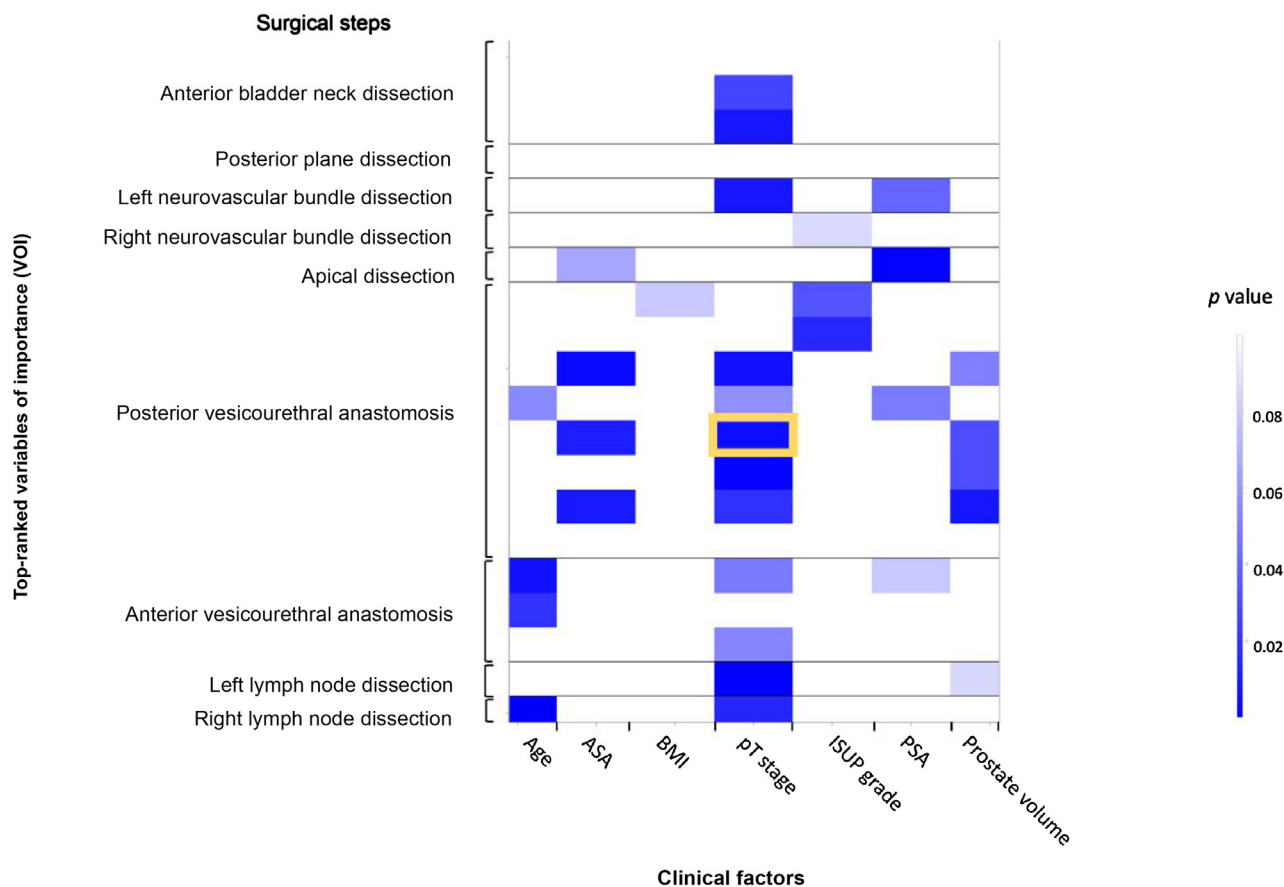


Fig. 3 – Heatmap showing effect modification of clinical characteristics on APMs' predictive ability of 3-mo continence recovery. All top 20 VOIs are APMs. Color scheme represents the p value of interaction test ranging from 0.1 to <0.01 , with darker blue representing a smaller p value; white represents $p \geq 0.1$. The square highlighted by yellow is the specific example mentioned in the manuscript (Fig. 4). APMs=automated performance metrics; ASA=American Society of Anesthesiologists; BMI=body mass index; ISUP=International Society of Urological Pathology; PSA=prostate-specific antigen; VOI=variable of importance.

deviation in nondominant instrument articulation is associated with a significantly lower rate of continence recovery with a rate ratio of 0.73 (95% CI: 0.58, 0.91, interaction $p < 0.01$; Fig. 4).

4. Discussion

Previously, we demonstrated that APMs can anticipate postoperative urinary continence recovery [2]. In the present study, we have combined surgeon performance data and patient perioperative data from two different high-volume institutions, established similar trends from our previous work, and improved the C-index of our predictive model. We have further discovered that many of the top-ranked APMs are independent predictors of continence recovery after adjusting for the confounding effect of patient and treatment factors.

Our bi-institutional data confirm that several patient and treatment factors may impact continence recovery, with most following the trend seen in the existing literature (ie, early continence recovery associated with younger [9] and lower BMI patients [10]). The only unexpected trend was that fewer patients with early recovery had a posterior

reconstruction performed; this may be due to the general variability of how this was performed and institutional variability of practice. While the posterior reconstruction

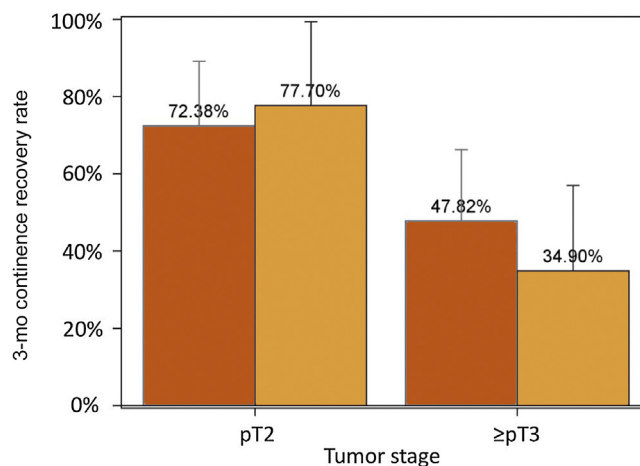


Fig. 4 – Effect modification of pT stage on the ability of “APM-posterior VUA nondominant instrument articulation” to predict 3-mo continence recovery. APM=automated performance metric; VUA=vesicourethral anastomosis.

was originally intended, in part, to improve early continence recovery, the evidence is conflicted in the literature [11,12]. At one institution of the present study, 18.2% (14/77) patients received a dedicated posterior reconstruction step during RARP. At the other, 77.6% (90/116) patients had the dedicated posterior reconstruction.

Our data suggest that clinical factors confound the ability of APMs to anticipate urinary continence recovery. This is a major finding of our paper, as this relationship between surgeon and patient factors, and postoperative clinical outcomes has not been quantified previously despite the logical inferences that we can make. We also demonstrate that clinical factors and APMs retain some independency from each other, and we are not merely evaluating the same factors with different labels.

The top 20 VOIs were all APMs during steps performed by the primary surgeon, suggesting that the outcome of continence recovery comes from the actions of the primary surgeon. Incidentally, consistent with our previous deep learning paper on APMs and urinary continence recovery, the majority of these top-ranked features were APMs during the VUA, the critical reconstructive (suturing) step of the RARP [2]. As we noted in the previous work, this result does not imply that the VUA directly impacts continence recovery, but rather it is possible that superior surgeon performance, as perhaps best captured during the VUA, leads to a superior outcome. While our prior work was with a single institution, the present work with an entirely retrained model is to some degree confirmatory with additional data from a second institution.

Importantly, after adjusting for potential confounders, the majority of APMs (55%) ranked in the top 20 in importance to predict early continence recovery remained as independent predictors of recovery. We further investigated the “modification effect” of patient factors on the impact of APMs on continence recovery. Our findings suggest that surgery quality may play a more important role in continence recovery of difficult cases, such as patients with locally advanced disease, as our example in Figure 4 illustrates.

From a methodological perspective, our present study represents a hybrid combination of machine learning and traditional statistics, which helps us process a tremendous amount of data while simultaneously overcoming the “black box phenomenon” of machine learning.

We share a few limitations of this study. There was heterogeneity of clinical data between institutions; nonetheless, we accounted for it in our multivariate (adjusted) analysis with mixed-effect model. There was also surgeon heterogeneity, with most cases performed at both institutions by multiple surgeons; we meticulously assigned procedure step performance by the primary or secondary surgeon. Finally, the predictive performance for 3-mo continence recovery is high to moderate (AUC=0.74), but still has room for improvement. Further inclusion of continence-related factors, such as preoperative International Prostate Symptom Score/American Urological Association symptom score, use of α -blocker medical therapy,

history of benign prostate hyperplasia or transurethral resection of the prostate, and membranous urethral length, may help improve future prediction performance.

5. Conclusions

Surgeon factors, represented by APMs from certain steps (eg, velocity and wrist articulation during the VUA and bladder neck dissection), are independent predictors of urinary continence recovery after RARP, while patient factors, especially tumor stage, can modify the impact of surgery performance on continence recovery.

Author contributions: Andrew J. Hung had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Hung, Cen, Wagner.

Acquisition of data: Ma, Nguyen.

Analysis and interpretation of data: Ma, Cen, Lei.

Drafting of the manuscript: Hung, Ma, Cen, Nguyen.

Critical revision of the manuscript for important intellectual content: Hung, Cen.

Statistical analysis: Cen, Lei.

Obtaining funding: Hung.

Administrative, technical, or material support: Hung.

Supervision: Hung, Wagner.

Other: None.

Financial disclosures: Andrew J. Hung certifies that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (eg, employment/affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), are the following: Andrew J. Hung is a consultant for Mimic Technologies, Quantgene, and Johnson & Johnson. Christian Wagner has a financial relationship with Intuitive Surgical (speaker, proctor, and consultant).

Funding/Support and role of the sponsor: Research reported in this publication was supported in part by the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under Award Number K23EB026493 and an Intuitive Surgical Clinical Research grant.

Acknowledgments: The authors acknowledge Anthony Jarc (Intuitive Surgical) for decryption of da Vinci systems data and processing of APMs, Swetha Rajkumar and Katarina Urbanova for data abstraction, and Inderbir Gill and J.H. Witt for their departmental leadership and project support.

CRedit authorship contribution statement

Andrew J. Hung: Conceptualization, Methodology, Resources, Supervision, Writing - original draft, Funding acquisition. **Runzhuo Ma:** Formal analysis, Investigation, Visualization, Writing - original draft. **Steven Cen:** Methodology, Software, Validation, Writing - review & editing. **Jessica H. Nguyen:** Investigation, Data curation, Writing - review & editing. **Xiaomeng Lei:** Software, Validation, Visualization. **Christian Wagner:** Conceptualization, Resources, Writing - review & editing.

References

- [1] Birkmeyer JD, Finks JF, O'Reilly A, et al. Surgical skill and complication rates after bariatric surgery. *N Engl J Med* 2013;369:1434–42.
- [2] Hung AJ, Chen J, Ghodoussipour S, et al. A deep-learning model using automated performance metrics and clinical features to predict urinary continence recovery after robot-assisted radical prostatectomy. *BJU Int* 2019;124:487–95.
- [3] Lendvay TS, White L, Kowalewski T. Crowdsourcing to assess surgical skill. *JAMA Surg* 2015;150:1086–7.
- [4] Ghani KR, Miller DC, Linsell S, et al. Measuring to improve: peer and crowd-sourced assessments of technical skill with robot-assisted radical prostatectomy. *Eur Urol* 2016;69:547–50.
- [5] Hung AJ, Chen J, Jarc A, Hatcher D, Djaladat H, Gill IS. Development and validation of objective performance metrics for robot-assisted radical prostatectomy: a pilot study. *J Urol* 2018;199:296–304.
- [6] Hung AJ, Chen J, Gill IS. Automated performance metrics and machine learning algorithms to measure surgeon performance and anticipate clinical outcomes in robotic surgery. *JAMA Surg* 2018;153:770–1.
- [7] Chen J, Chu T, Ghodoussipour S, et al. Effect of surgeon experience and bony pelvic dimensions on surgical performance and patient outcomes in robot-assisted radical prostatectomy. *BJU Int* 2019;124:828–35.
- [8] Ficarra V, Novara G, Rosen RC, et al. Systematic review and meta-analysis of studies reporting urinary continence recovery after robot-assisted radical prostatectomy. *Eur Urol* 2012;62:405–17.
- [9] Matsushita K, Kent MT, Vickers AJ, et al. Preoperative predictive model of recovery of urinary continence after radical prostatectomy. *BJU Int* 2015;116:577–83.
- [10] Wolin KY, Luly J, Sutcliffe S, Andriole GL, Kibel AS. Risk of urinary incontinence following prostatectomy: the role of physical activity and obesity. *J Urol* 2010;183:629–33.
- [11] Joshi N, de Blok W, van Muilekom E, Van Der Poel H. Impact of posterior musculofascial reconstruction on early continence after robot-assisted laparoscopic radical prostatectomy: results of a prospective parallel group trial. *Eur Urol* 2010;58:84–9.
- [12] Sutherland DE, Linder B, Guzman AM, et al. Posterior rhabdosphincter reconstruction during robotic assisted radical prostatectomy: results from a phase II randomized clinical trial. *J Urol* 2011;185:1262–7.