The Virtual Reality Transurethral Prostatic Resection Trainer: Evaluation of Discriminate Validity

Hani H. Rashid, Timothy Kowalewski, Peter Oppenheimer, Ann Ooms, John N. Krieger and Robert M. Sweet*

From the Department of Urological Surgery (HHR, RMS) and College of Education and Human Development (AO), University of Minnesota, Minneapolis, Minnesota, and Department of Urology (TK, JNK) and Human Interface Technology Laboratory (PO), University of Washington, Seattle, Washington

Purpose: To understand how urologists acquire resection skills we analyzed factors correlating with favorable resection metrics in groups defined as experts, residents and novices. We then evaluated discriminate validity by determining factors correlating with proficiency among individuals in the expert, resident and novice groups.

Materials and Methods: A total of 136 subjects completed the protocol, including 72 urologists, 45 residents and 19 novices. After a pre-task questionnaire and training video subjects performed a standardized 5-minute resection task. Primary metrics were gm resected, blood loss, irrigant volume used, foot pedal use and differential time spent with orientation, cutting or coagulation.

Results: Among experts larger resection correlated with more time spent cutting ($p < 0.001$). In contrast, increased coagulation time correlated with gm resected in the novice group ($p = 0.001$). The number of transurethral prostate resections that residents reported having done in the real operating room correlated with gm resected ($p = 0.043$), use of more irrigating fluid ($p = 0.024$) and less time spent coagulating ($p = 0.027$) on the simulator. In residents and experts exclusively primary resection efficiency metrics, fluid use and blood loss correlated with cuts at tissue and correlated inversely with coagulation and orientation time ($p < 0.05$).

Conclusions: Different factors determine transurethral prostate resection performance metrics among experts, residents and novices. These correlations reinforce discriminate validity and provide insight into specific factors that likely determine success at different training levels. Such data could be used to isolate and train skill subsets in the curriculum and they may elucidate the safest and most efficient approach to train resection skills.

Key Words: prostate; transurethral resection of prostate; computer simulation; education, medical; task performance and analysis

Transurethral resection of the prostate remains the gold standard surgical therapy for lower urinary tract symptoms related to benign prostatic hyperplasia. However, training residents and urologists in TURP skills is difficult. With the recent increase in alternative treatments for benign prostatic hypertrophy the number of cases available for residents to learn this technically challenging procedure has decreased.[1] The dilemma is how to best teach TURP.

Simulation offers an appealing adjunct for learning TURP skills because TURP is difficult to train, relies primarily on visual cues and proves highly amenable to virtual reality simulation techniques.[2,3] Simulation is increasingly important with the recent trend toward minimally invasive surgery and the decrease in resident work hours. Properly constructed simulation models may provide trainees with increased opportunities to learn standard and new techniques in a safe and effective manner. Simulator ability to train is greatly enhanced when it can provide realistic simulation of the actual procedure. The fundamental property of any measuring instrument, device or test is that it "measures what it purports to measure."[4] Simulators should provide accurate detail, and be anatomically precise and highly interactive. Numerous TURP simulation models have tried to follow this creed with varying success.[3,5,6]

"Validity is an integrated judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inference and actions based on test scores or other modes of assessment."[7] "The process of validation involves accumulating evidence to provide a sound scientific basis for the proposed score interpretations."[8] Validation is a continuing process and it is always incomplete. Thus, validation is "a matter of making the most reasonable case, on the basis of the balance of available evidence."[9] In previous studies we noted that our virtual reality model had face validity. Experts believed that at face value the TURP simulator seemed to adequately simulate and train TURP. Content validity was also established since experts judged that simulator content, including realism and interactivity, was appropriate on a 5-point Likert scale. Prior experience with TURP also predicted performance

0022-5347/07/1776-2283/0
THE JOURNAL OF UROLOGY®
Copyright © 2007 by AMERICAN UROLOGICAL ASSOCIATION

2283

Vol. 177, 2283-2286, June 2007
Printed in U.S.A.
DOI:10.1016/j.juro.2007.01.120

since experts did better than novices on the first try, which is an aspect of construct validity.[3,4]

We present additional evidence of discriminate validity. Discriminate validity represents a more intricate form of construct validity, that is the ability to differentiate ability levels in groups with similar experience. We determined which factors correlated with levels of proficiency between subjects in groups defined as experts, residents and novices. In other words, we asked what made some experts better than other experts and what made some residents better than other residents. We then determined if these proficiency correlations differed among the groups.

MATERIALS AND METHODS

Version 1.0 of the TURP simulator integrates novel 3-dimensional virtual anatomy and force feedback technology (Mimic Technology, Seattle, Washington) with a physical model (Simulab, Seattle, Washington).[3,10] Simulation data were generated at the 2002 annual meeting of the American Urological Association.[3] Without another benchmark to assess proficiency experts were defined as board certified urologists. Residents were defined as trainees in accredited urology training programs who had done at least 1 TURP. Novices were defined as participants with no TURP experience. Briefly, 136 subjects completed the study protocol, including 72 experts, 45 residents and 19 novices with at least a master's degree education level. None of the subjects had previously used the simulator.

The database was generated with a pre-task questionnaire, which provided demographic data, training status and TURP related questions. Subjects viewed an introductory training video. Each participant was then given a 100 gm prostate with 5 minutes to "resect as much tissue as possible, most efficiently, with the least amount of blood loss, using the least amount of irrigant and coagulation current." This statement defined the PPMs. The 4 PPMs for the 5-minute task were defined as 1) gm resected, 2) blood loss per gm resected, 3) gm resected per cut and 4) amount of irrigation fluid used. These PPMs were deemed acceptable retrospectively by a consensus conference of 9 TUR experts participating in a multicenter predictive validity study of the trainer.

The simulator logged all features of instrument interaction in the virtual environment. Only data on subjects who completed the task were included in the analyses. For study purposes demographic data on training status and primary metrics consisted of operative errors, gm resected, blood loss, irrigant volume, foot pedal use, and differential time spent with orientation, cutting and coagulation. Three trained technicians administered the task to the participants. Technicians were coached to give consistent, predetermined responses to anticipated questions or comments.

Pearson's correlation analyses were used to evaluate factors associated with favorable PPMs in each of the 3 groups. We then looked at whether the factors correlating with success differed in and among participant groups. The study was approved by the American Urological Association and University of Washington Institutional Review Board.

RESULTS

Demographics and Clinical Experience

Subjects were 23 to 68 years old (mean ± SD age 40.5 ± 10.6). Of the 72 experts 46% were in academic practice, 19% were in solo private practice, 18% were in small group practice, 10% were in large group practice and 4% worked for a health maintenance organization. Approximately half of the expert participants completed training before 1994. Of the 45 resident participants the median number of years of residency completed was 3.5 with 15% having completed 1 year or less of residency and 22% having completed 5 years or greater.

The median number of TUR procedures performed by experts in the preceding month was 6.79 (range 0 to 30). The median number performed by the expert group in the preceding year was 52.6 (range 0 to 500). The median number of TUR procedures performed by residents in the preceding month was 4.93 (range 0 to 30). The median number of procedures performed by residents in the preceding year was 33.6 (range 0 to 360). Of the residents 75% who performed 20 or greater procedures in the last year had at least 3 years of urological training. With regard to video game experience 47% of participants reported never playing, 25% reported playing once yearly, 18% played monthly, 9% played weekly and approximately 1% played daily. As expected, video game experience correlated negatively with participant age (Pearson $r = -0.212$, $p = 0.011$). The table lists group correlations with PPMs.

Novice Strategies

In the novice group only there was a correlation between gm resected and total time spent coagulating. Blood loss also correlated with the amount of tissue resected ($p = 0.001$). Novices but not residents or experts demonstrated a positive correlation between video game use, total gm resected and total coagulation time ($p < 0.05$).

Correlations with PPMs			
	p Value		
	Novices	Trainees	Experts
Increased gm resected:			
Coagulation time	0.001	0.05	0.32
Orientation time	0.001 (less)	0.17	<0.001 (less)
Video game use	0.014	0.17	0.09
Blood loss	<0.0001	0.08	<0.0001
No. TURPs in last 2 yrs	Not applicable	0.04	0.80
More tissue cuts	0.10	0.001	<0.0001
Cutting time	0.06	0.001	<0.0001
Coagulation pedal	0.19	0.01	0.32
Decreased blood loss/gm resected:			
More tissue cuts	0.10	0.01	0.04
Cutting time	0.08	0.01	0.002
Orientation time	0.09	0.005	0.003
Increased gm resected/cut:			
Coagulation time	0.40	0.006 (less)	0.001 (less)
Cutting time	0.74	0.60	<0.001
Less fluid use:			
Postgraduate yr level	Not applicable	0.005 (lower)	Not applicable
Age	0.605	0.024	0.661
Coagulation time	0.650	0.027	0.606

Resident Strategies

There was an expected correlation between residents who were closer to completing the residency program and the reported number of TURPs performed in the last month and last year (p = 0.029). Residents in the group with more years of training correlated with more fluid use (p = 0.005), more cuts at tissue (p = 0.035) and more hits on the cutting pedal (p = 0.015). As a group, residents reporting more TURPs in the last 2 years tended to resect more tissue (p = 0.043). Residents who resected more tissue also tended to use more fluid (p = 0.024) and they spent less time with coagulation (p = 0.027). Similar correlations were not observed in the novice or the expert group. In residents and experts exclusively primary efficiency metrics for resection, fluid use and blood loss correlated with more cuts at tissue along with less coagulation and orientation time (p <0.05).

Expert Strategies

Larger resections correlated with more time cutting and, therefore, less time coagulating and orienting in the expert group only (p <0.001). Additionally, as a group, experts who resected more gm per cut at tissue tended to spend less time coagulating and orienting. In the expert group age positively correlated with more frequent coagulation pedal presses (p = 0.031), although without a correlation noted with total blood loss. In the expert group more TURPS reported being performed in the last 2 years correlated only with a tendency to hit the cutting pedal more frequently (p = 0.029). No differences in any primary metrics were noted. Experts who spent more time cutting tended to use less fluid (p = 0.045). Cutting frequency (total cuts at tissue) and the cut pedal presses count were the only metrics that correlated with video game use in the expert category.

DISCUSSION

Continued improvements in medical management and minimally invasive technologies combined with curtailed resident working hours make it challenging to train residents in surgical fields. Maintaining surgical skills represents an ongoing challenge for practitioners who have completed training. In theory simulation represents an attractive approach to provide reality based training without harming patients.[11] Simulation provides intricate performance metrics that highlight areas to improve. For the first time teaching surgeons can gain quantitative insight into exactly how trainees learn procedural skills. Virtual reality models for simulation also hold the possibility for patient specific modeling and rehearsal.

Important disadvantages of simulation include limited availability and development costs. A critical issue with any simulator is whether it accurately represents reality. Simulators should provide visual reality or sufficiently high resolution to look like a patient.[12] Effective simulators must also be interactive and react with reality when manipulated, eg bleed when cut.[11,12] With that said, realism in simulation is simply a means by which to achieve the true overlying objective of whether the simulator actually trains and whether it can accurately assess skill.

In this study PPMs were associated with specific factors in each participant group (expert, residents and novice) and these factors differed among the groups. One would expect a valid simulator to show a great range of skill among the groups and among individuals in each group, as documented in this study. Understanding why certain groups followed different TURP strategies may help improve the training process.

As a group, novices most proficient in total gm resected cut until they lost more blood and consequently they spent more time coagulating and less time with orientation. Novices with video game experience performed better than other novices. In effect, they treated the TURP simulator like a video game. We hypothesized that, because they had not developed coagulation skills, they worked on coagulation only after bleeding became too active. This finding is consistent with the increased error rate among novices in our earlier analysis.[3] Among novices there was no correlation with the more sophisticated PPM. These observations may reflect absent knowledge and a systemic approach to TURP in the novice group.

Residents appeared to follow a different strategy than the other groups. As we would expect for a valid trainer, residents who resected more tissue on the simulator had more real life TURP experience. The most experienced residents spent more time cutting, had more cuts at the tissue and used less coagulation between cuts. In the resident group less orientation time and more cuts at tissue correlated with less blood loss per gm resected. They appeared to take a more systematic approach to resection, whereby they learned from experience and established a methodology by which they could cut the most amount of tissue in the most effective manner. This was similar to the expert group compared to the novice group, whose success correlated mainly with video games, suggesting that experience and establishing an effective approach to resection are critical to mastering TURP.

Residents were the only group that showed a correlation between more fluid use and the resection of more prostate tissue during simulation. Less experienced residents may have been more cautious about fluid management because they are constantly told about the driving factor for fluid management, that is TUR syndrome. They possibly understood that we were measuring less fluid as a metric and thought of it as having the same importance as the other metrics measured. They dealt with visualization difficulties by actually addressing bleeding, as opposed to using fluid to manage it. On the other hand, experts showed a negative correlation between cutting and fluid use. We hypothesized that with time and experience residents gained understanding of resection and hemostasis as more important metrics and focused on those aspects in their succession toward establishing their approach to TURP.

The defining characteristic of the expert group was that they appeared to bridge all performance metrics using an established resection method. Experts who proved most adept at the simulator systematically spent more time cutting and less time orienting, and they appeared to worry less about blood loss because they knew they could control bleeding. This theory was consistent with the gm resected per cut, in that proficient experts were able to cut more tissue while using less coagulation and irrigation fluid. Like residents, experts spent less time with orientation and had more cuts at tissue, which correlated with less blood loss per gm resected. The primary goal of TURP is to remove prostatic tissue to relieve obstruction in the safest most effective manner. As one would expect, the most proficient experts

appear to achieve this goal by using less fluid when cutting without an effect on blood loss. When bleeding obscured vision, there were 2 strategies, including increasing fluid and coagulating the bleeding vessel. We noted that subjects who had visibility problems early on seemed to turn the irrigation inflow all the way up and leave it at maximum flow for the duration of the task.

This study shows correlations but it has important limitations. We had a limited number of subjects, particularly in the novice group, which limited statistical power for some desired multivariate subset analyses. The appropriate way to do this would be to identify the important variables via univariate modeling, as we did, and then construct a multivariate model to find independent predictors. Because subjects were tested in the American Urological Association exhibit hall, they may have approached the model as more of a video game than a training exercise. Participants may have deviated from their standard approach. The simulation was only 5 minutes in duration. PPMs might change with longer simulation times. We chose 5 minutes following a consensus of experts and because we did not wish to keep subjects waiting. We did not ask whether experts used a camera in their practice. It is possible that using a camera may have modified the standard approach of the older generation of resectionists.

The composite validation results with version 1.0 of the TURP trainer led to successful licensing of the trainer. We have also initiated a multi-institutional study to determine if training on the simulator improves resident performance in the operating room.

CONCLUSIONS

The virtual reality TURP trainer distinguished differences in achieving success with the PPMs for TURP among novices, trainees and expert urologists, contributing to the discriminate validity of the trainer. We gained new insight into how urologists learn TURP skills. These promising results support a multi-institutional study to determine whether skills acquired in the virtual environment translate to the operating room.

ACKNOWLEDGMENTS

Abbreviations and Acronyms		
PPM	=	primary proficiency metric
TUR	=	transurethral resection
TURP	=	transurethral resection of the prostate

REFERENCES

1. Holtgrewe HL: Surgical management of benign prostatic hyperplasia in 2001—a pause for thought. J Urol 2001; **166:** 177.
2. Manyak MJ, Santangelo K, Hahn J, Kaufman R, Carleton T and Hua XC: Virtual reality surgical simulation for lower urinary tract endoscopy and procedures. J Endourol 2002; **16:** 185.
3. Sweet R, Kowalewski T, Oppenheimer P, Weghorst S and Satava R: Face, content and construct validity of the University of Washington virtual reality transurethral prostate resection trainer. J Urol 2004; **172:** 1953.
4. Gallagher AG, Ritter EM and Satava RM: Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. Surg Endosc 2003; **17:** 1525.
5. Michel MS, Knoll T, Kohrmann KU and Alken P: The URO Mentor: development and evaluation of a new computer-based interactive training system for virtual life-like simulation of diagnostic and therapeutic endourological procedures. BJU Int 2002; **89:** 174.
6. Kallstrom R, Hjertberg H, Kjolhede H and Svanvik J: Use of a virtual reality, real-time, simulation model for the training of urologists in transurethral resection of the prostate. Scand J Urol Nephrol 2005; **39:** 313.
7. Messick S: Validity, 3rd ed. New York: American Council on Education and Macmillan Publishing 1989; pp 13–103.
8. Aera A: Standards for Educational and Psychological Testing. Washington, D. C.: National Council on Measurement in Education 2002.
9. Messick S: Validity of Test Interpretation and Use. Report No. ETS-RR-90-11. Princeton, New Jersey: Educational Testing Service 1990.
10. Sweet R, Porter J, Oppenheimer P, Hendrickson D, Gupta A and Weghorst S: Simulation of bleeding in endoscopic procedures using virtual reality. J Endourol 2002; **16:** 451.
11. Shah J, Mackay S, Vale J and Darzi A: Simulation in urology—a role for virtual reality? BJU Int 2001; **88:** 661.
12. Stacey R: Marketing medical simulation-what industry needs from the clinical community. Min Invasive Ther Allied Tech 2000; **9:** 357.