Author's Accepted Manuscript

Effect of Level of Urology Training on Gleason Score and Prostate Volume Estimation Agreement Between Transrectal Ultrasound-Guided Biopsy and Radical Prostatectomy Specimen

Alessandro Morlacco, Christopher R. Murphy, Laureano J. Rangel, Lance A. Mynderse, Robert H. Thompson, R. Jeffrey Karnes

PII: S2352-0779(16)30277-1
DOI: 10.1016/j.urpr.2016.11.010
Reference: URPR 258

To appear in: *Urology Practice*
Accepted Date: 29 November 2016

Please cite this article as: Morlacco A, Murphy CR, Rangel LJ, Mynderse LA, Thompson RH, Karnes RJ, Effect of Level of Urology Training on Gleason Score and Prostate Volume Estimation Agreement Between Transrectal Ultrasound-Guided Biopsy and Radical Prostatectomy Specimen, *Urology Practice* (2017), doi: 10.1016/j.urpr.2016.11.010.

Effect of Level of Urology Training on Gleason Score and Prostate Volume Estimation Agreement Between Transrectal Ultrasound-Guided Biopsy and Radical Prostatectomy Specimen

Authors: Alessandro Morlacco[1], Christopher R. Murphy[1], Laureano J. Rangel[2], Lance A. Mynderse[1], Robert H. Thompson[1], R. Jeffrey Karnes[1 *].

Affiliations: [1]Department of Urology; [2] Health Science Research; Mayo Clinic Rochester, MN, USA

Running title – Effect of level of training on TRUS biopsy accuracy

Keywords: prostate biopsy, residency, training, TRUS, Gleason grading, prostatic cancer, ultrasonography

Abstract word count: 240

Total word count (excluding abstract): 2485

Tables: 5

Supplementary tables: 4

Figures: 0

Corresponding Author:
R. Jeffrey Karnes, MD, FACS
Department of Urology, Mayo Clinic
200 First Street SW Rochester, MN 55905
Phone 507-284-3981 Fax: 507-284-4951
Karnes.r@mayo.edu

Conflict of interest: none

Financial support: none relevant

Abstract

Introduction: TRUS-guided prostatic biopsy may be performed by operators with different level of training. Little is known about the impact of training level on biopsy results. This study aims to evaluate the effect of level of training on the accuracy of TRUS-guided prostate biopsy findings.

Methods: We retrospectively reviewed 500 consecutive patients who underwent TRUS-guided prostate biopsy and subsequent radical prostatectomy (RP).

TRUS operators were stratified based on level of training: junior, senior, chief, fellow, or staff. Linear regression was performed to analyze the effect of training level on volume estimates. A weighted Kappa statistic evaluated agreement between biopsy and pathologic Gleason scores, while an adjusted cumulative logistic regression model analyzed effects of training level.

Results: 482 patients were included in the final analysis. TRUS biopsy was performed by staff in 78 (16%) patients, by fellows in 18 (4%), chief residents in 48 (10%), senior residents in 126 (26%), and by junior residents in 212 (44%). There was no significant difference between TRUS and RP specimen volume estimates between the training levels. Level of training was not significantly associated with pathologic features including Gleason score, primary Gleason grade, highest single Gleason grade, and estimated tumor volume. Limitations include the retrospective design, and the variability between members of the same group.

Conclusions: Agreement between biopsy and pathologic Gleason scores is high for all levels of training. Level of training has no impact on prostate volume estimations or prediction of pathologic features.

Introduction:

Transrectal ultrasound (TRUS) was first introduced in the field of urology approximately 40 years ago, and became one of the most important diagnostic tools available to the urologist[1]. Today, its most common use remains imaging of the prostate especially in the setting of prostate core needle biopsies: despite being increasingly challenged by MRI-guided techniques, TRUS-guided biopsy still remains the gold standard for prostate cancer diagnosis. A wide range of cancer detection rates have been reported for TRUS biopsy, ranging from 25-49%, the variation depending on the indications used for biopsy as well number of cores taken and various templates used over the years[2-8].

Numerous studies evaluated templates for both the optimum number and location of cores to be sampled. Overall detection rates were shown to increase with the number of cores taken[9]. Use of extended 10-12 core systematic biopsy templates, which include samples of the lateral peripheral zone, in addition to sampling suspicious lesions was shown to improve cancer detection rates[9-13]. In addition, an extended template biopsy has been shown to decrease the probability of a positive repeat biopsy following an initial negative extended biopsy[14]. Furthermore, the extended 12 core biopsy template has been shown to have no significant difference in quality of life and return to daily activity, work, or exercise compared to a sextant biopsy[15]. The constellation of these findings, along with other similar studies, led to the adoption of a 12 core template including sextant and lateral peripheral zone sampling at base, mid-gland and apex bilaterally for TRUS-guided prostate biopsy in addition to biopsy of any palpable

abnormalities on DRE or suspicious lesion on TRUS as the standard of care according to NCCN and EAU guidelines[16, 17].

In addition to guiding prostate biopsy, prostate volume estimation is another tool that TRUS offers to the urologist; this is commonly calculated using an ellipsoid formula. Volume measurements may be helpful in surgical planning and are necessary in calculating PSA density which may be used adjunctively in determining whether to recommend a biopsy or in prognostic models to predict surgical outcomes[18-20].

TRUS-guided prostate biopsy is one of the first urologic procedures taught to urology residents. Whether experienced urologists better sample the prostate during biopsy compared to urology residents has not been extensively studied: quality standards for prostate biopsy have been established and include length of the core and percentage of cores with no prostatic tissues. Accuracy of prostate cancer detection has been shown to be influenced by the length of biopsy cores[21]. Benchikh et al studied average length of biopsy cores taken by residents and reported significant improvement in average core length from the first to twelfth biopsies performed, after which point average core length plateaued[22]. A few studies have examined the performance of urology residents in TRUS prostate biopsies with regards to cancer detection, and have shown residents at all levels of training perform equally well[22, 23] . In our program, prior to autonomy, residents have formal didactics and supervised training during their internship (PGY-1) year.

The objective of this study is to evaluate the performance of residents at various levels of training to accurately biopsy the prostate and estimate volume, using the RP specimen as a gold standard.

Materials and Methods:

We retrospectively reviewed 500 consecutive patients who underwent TRUS-guided transrectal prostate biopsy and subsequent RP for definitive treatment of prostate cancer at our institution from 2005-2007. The minimum standard was a 12-core biopsy template. TRUS operators were stratified based on level of training : junior resident (PGY2), senior resident (PGY3-4), chief resident (PGY5-6), fellow, or staff.

Prostate volume:

Prostate dimensions in terms of length (craniocaudal), width, and height (anteroposterior) were measured using TRUS. Prostate volume was estimated by applying an ellipsoid equation ($\pi/6 * l * w * h$).

Volume of the RP specimen was calculated from the true dimensions of the fresh specimen applying the same ellipsoid equation. Differences between TRUS-estimated and RP specimen-estimated volumes were compared within each level of training as were differences between the various levels of training. Logarithmic transformation of both volume estimates was performed to facilitate linear regression analysis to assess the value of TRUS volume and level of training as predictors of RP specimen-estimated volume.

Gleason Score: For practical purposes, Gleason scores were categorized into three different groups – low (Gleason 6), intermediate (Gleason 7), and high (Gleason 8-10) – in accordance with prostate cancer risk stratification. Both biopsy and RP specimen examinations were performed by experienced uropathologists at our institution. Analysis of agreement between biopsy and pathologic Gleason scores for each level of training was done using a weighted Kappa statistic. Kappa values <0.2 indicate poor to slight agreement while values of 0.2-0.4, 0.4-0.6, 0.6-0.8 and >0.8 indicate fair, moderate, substantial, and near perfect agreement, respectively. An adjusted cumulative logistic regression model was used to evaluate how prediction of pathologic Gleason score from the biopsy Gleason score was affected by level of training. Sub-analyses evaluating the highest single (3, 4, or 5) clinical and pathologic Gleason grades as well as the primary clinical and pathologic Gleason grades (3, 4, or 5) were done using the same methods.

The percentage of positive core from each individual biopsy core was summed and divided by the sum of each individual biopsy core length to calculate percentage of positive cores for the entire biopsy. Tumor volume was estimated using the ellipsoid formula and the pathologic specimen. Logarithmic transformation of estimated tumor volume was performed to facilitate linear regression analysis to evaluate percentage of positive cores in the biopsy and level of training as predictors of estimated tumor volume.

Results:

This IRB approved study included 482 patients who provided their consent for research purposes. TRUS biopsy was performed exclusively by staff in 16.2% (78/482), fellows in 3.7% (18/482), chief residents in 10% (48/482), senior residents in 26.1% (126/482), and junior residents in 44% (212/482).

Mean and median TRUS-estimated and RP specimen-estimated volumes stratified by level of training are displayed in Table 2. Overall, there was no significant difference between TRUS-estimated volume and RP specimen-estimated volume (p=0.33) nor were there any significant differences between volume estimations within each experience level (Table 1). The differences between the various experience levels were not significant (p=0.24).

Overall, level of training did not significantly affect prediction of the RP specimen-estimated volume (p=0.26) nor did the TRUS volume slope vary by level of training (p=0.58). There was a significant difference in the prediction of RP specimen-estimated volume by junior residents compared to staff (p=0.045). Preoperative PSA and TRUS volume were found to be significantly associated with RP specimen-estimated volume (Table 2).

Using the three different classifications of Gleason score, agreement between clinical and pathologic Gleason scores for junior residents, senior residents, chief residents, fellows and staff was 76.7%, 72.2%, 72.3%, 83.3%, and 87.0%, respectively. Using the weighted Kappa statistic, the highest concordance indices were observed for staff and fellows (both $\kappa = 0.72$) and the lowest for senior residents ($\kappa = 0.46$); however, this method does not allow adjustment for covariates. When clinical and pathologic Gleason scores differed, biopsy Gleason score was

more frequently lower than pathologic Gleason score, a trend consistent across all levels of training (Table 3).

On logistic regression, level of training had no significant effect on prediction of pathologic Gleason score by clinical Gleason score using staff as a reference ($p = 0.30$). Clinical/biopsy Gleason score and estimated tumor volume were found to be significant predictors of pathologic Gleason score (Table 4). Additionally, there was no evidence that the effect of training varied by clinical Gleason score ($p = 0.13$).

Similar sub-analyses evaluating agreement between single highest clinical and pathologic Gleason grades (Supplementary Tables 1-2), and primary clinical and pathologic Gleason grades (Supplementary Tables 3-4) were performed. Results from both these focused analyses mirrored those of the total clinical and pathologic Gleason scores in that level of training did not significantly impact prediction of these pathologic features.

Finally, the prediction of estimated tumor volume by the percentage of positive cores in the biopsy was not affected by level of training overall ($p=0.08$) nor did its effect vary significantly by maximum clinical Gleason grade ($p=0.50$). The percentage of positive cores in the biopsies and preoperative PSA were significant predictors of estimated tumor volume (Table 5).

Discussion:

To our knowledge, the findings of resident-performed TRUS-guided prostate biopsies – specifically Gleason score and volume estimation have not been compared to radical prostatectomy (RP) specimens to determine the potential effect the level of training may have on their accuracy.

Experience has previously been studied for potential effect on cancer detection for TRUS-guided prostate biopsy. Lawrentschuk et al evaluated the difference in prostate cancer detection in over 4500 TRUS biopsies performed by four different uro-radiologists[3]. The authors reported cancer detection ranged from 43.8-52.4% among the four operators and noted the operator to be a significant predictor of cancer detection; however, no learning curve was detected. Rather, differences in cancer detection rates between operators were concluded to be the result of difference in expertise and/or technique.

Previous studies have examined how level of training impacts the rate of cancer detection when performing TRUS-guided prostate biopsy and found no significant differences between resident training levels in overall cancer detection[22-24]. Benchikh et al described a learning curve for residents on the basis of improvement of the average biopsy core length, which plateaued after 12 procedures[22]. However, this learning curve did not seem to affect cancer detection rates, which were stable throughout the study period[22].

Our study evaluates the impact level of training may have on the accuracy of prostate biopsy by comparing the Gleason score from the RP specimen to that from the TRUS-biopsy. Obviously, even in the most experienced hands there is not 100% concordance between clinical/biopsy and

pathologic/RP-specimen Gleason scores. At all levels, clinical variables appropriately predicted the corresponding pathological variables. Using staff performance as a reference, we found level of training to have no significant effect on predicting the pathologic Gleason score.

In sub-analyses of the both the highest single Gleason grade and the primary Gleason grade, both biopsy Gleason grades remained predictive of their pathologic counterpart and no significant differences were observed between levels of training. It should be noted that when biopsy Gleason score was not in agreement with pathologic Gleason score, biopsy Gleason score was more frequently lower than pathologic Gleason score. This was true for overall, maximum, and primary Gleason scores, with no relationship to the level of training. When examining individual contrasts with staff, it was noted that senior residents differ from staff in prediction of primary Gleason grade (OR 5.530; p=0.015).

We also evaluated the effect of level of training on the accuracy of TRUS prostate volume estimations. TRUS, in the hands of experienced operators, yields precise volume estimations such that there is good agreement when TRUS is either repeated by the same operator or performed by different, experienced operators[25]. Sech et al evaluated the effect level of training had on TRUS-estimated prostate volumes having a junior resident, senior resident, and attending physician perform TRUS on the same 121 patients[26]. The study authors reported an intraexaminer correlation of 0.96 (0.95-0.97) for total volume[26].

However, while TRUS has been shown to be precise in its volume estimations, its accuracy has been questioned as studies have shown TRUS generally underestimates prostate volume by as

much as 20% when compared to prostate specimen weight following RP[27-29]. More recently, Rodriguez et al suggested the ellipsoid formula is the primary source for inconsistency in TRUS prostate volume estimations as they showed volume estimations using an ellipsoid formula and measurements from RP specimens also underestimated actual gland weight[29].

We compared TRUS-generated volume estimates to volume estimates from RP specimens calculated using the same ellipsoid formula and dimensions of the RP specimen, similar to Rodriguez et al. Since both estimations use the same ellipsoid formula, the formula is not a confounder in our evaluation of level of training's potential effect on accuracy of the TRUS-volume estimation.

Within each level of training, there was no significant difference between TRUS-estimated volume and RP specimen-estimated volume. Furthermore, the differences between TRUS-estimations and RP specimen estimations were not significantly different for the various levels of training. Trends were noted for both junior residents and chief residents to overestimate prostate volume with TRUS by a median of 0.9 cc and 3.0 cc ($p=0.06$ and $p=0.09$), respectively.

Linear regression analysis confirmed that prediction of RP specimen-estimated volume, as expected, is affected by changes in TRUS volume and preoperative PSA (Table 2). However, level of training overall did not significantly affect the prediction of RP specimen-estimated volume. When examining individual contrasts with staff, junior residents differed from staff in their prediction of RP specimen-estimated volume ($p=0.045$). However, given the modest p-value and the number of comparisons, the actual relevance of this finding is uncertain.

Additionally, the smaller magnitude of the regression coefficient (0.09) for this difference from staff may carry little clinical significance.

While this study is novel in assessing accuracy of TRUS biopsy findings among different levels of training, it is not devoid of limitations beginning with its retrospective nature. We also did not utilize a strictly prospective and standardized biopsy protocol. However, all residents are required to complete the same ultrasonography training course in which they are taught appropriate technique for prostate biopsy prior to performing biopsies in clinic, thus there is minimal variation in their education regarding how to perform this procedure. Of interest, no staff members were regularly present in the room when residents/fellows performed the biopsy, thus minimizing the possible exteral influence on the operator. However, use of a standardized biopsy protocol is likely to increase both inter-group and interobserver agreement, further reducing the differences among different training level groups. Additionally, there was no distinction made between office-based biopsies and those conducted in an outpatient surgical center setting; in spite of that, in our clinical practice we do not observe any difference in core number among different training grups.

Finally, there are some aspects that could limit the applicability of our findings. First of all, the present study was carried out at an acedemic institution with a structured training program, therefore it might only partially apply to institutions without a well-organized training program. Moreover, while the use of RP pathology as a reliable gold standard is an advantage of the present work, on the other hand, this choice could theoretically limit the generalization of our findings to other groups of patients such as those managed with active surveillance of radiation therapy.

Nonetheless, this novel study provides evidence to support the idea that prostate cancer management is not compromised when residents perform TRUS-guided prostate biopsy. However, confirmation of these training level and achievement of appropriate skills will need to be re-assessed as more advanced biopsy techniques are introduced into our biopsy schemes.

In conclusion, our findings show that the level of training does not significantly impact the accuracy of TRUS-guided prostate biopsy findings and the procedure can be safely performed by relatively inexperienced urology residents without compromising patient care.

References:

1. Watanabe, H., Igari, D., Tanahashi, Y. et al.: Transrectal ultrasonotomography of the prostate. The Journal of urology, **114:** 734, 1975
2. Ellis, W. J., Chetner, M. P., Preston, S. D. et al.: Diagnosis of prostatic carcinoma: the yield of serum prostate specific antigen, digital rectal examination and transrectal ultrasonography. The Journal of urology, **152:** 1520, 1994
3. Lawrentschuk, N., Toi, A., Lockwood, G. A. et al.: Operator is an independent predictor of detecting prostate cancer at transrectal ultrasound guided prostate biopsy. The Journal of urology, **182:** 2659, 2009
4. Letran, J. L., Meyer, G. E., Loberiza, F. R. et al.: The effect of prostate volume on the yield of needle biopsy. The Journal of urology, **160:** 1718, 1998
5. Uzzo, R. G., Wei, J. T., Waldbaum, R. S. et al.: The influence of prostate size on cancer detection. Urology, **46:** 831, 1995
6. Orozco, R., O'Dowd, G., Kunnel, B. et al.: Observations on pathology trends in 62,537 prostate biopsies obtained from urology private practices in the United States. Urology, **51:** 186, 1998
7. O'Dowd G, J., Miller, M. C., Orozco, R. et al.: Analysis of repeated biopsy results within 1 year after a noncancer diagnosis. Urology, **55:** 553, 2000
8. Roehl, K. A., Antenor, J. A., Catalona, W. J.: Serial biopsy results in prostate cancer screening study. The Journal of urology, **167:** 2435, 2002

9. Presti, J. C., Jr., O'Dowd, G. J., Miller, M. C. et al.: Extended peripheral zone biopsy schemes increase cancer detection rates and minimize variance in prostate specific antigen and age related cancer rates: results of a community multi-practice study. The Journal of urology, **169:** 125, 2003

10. Babaian, R. J., Toi, A., Kamoi, K. et al.: A comparative analysis of sextant and an extended 11-core multisite directed biopsy strategy. The Journal of urology, **163:** 152, 2000

11. Norberg, M., Egevad, L., Holmberg, L. et al.: The sextant protocol for ultrasound-guided core biopsies of the prostate underestimates the presence of cancer. Urology, **50:** 562, 1997

12. Eskew, L. A., Bare, R. L., McCullough, D. L.: Systematic 5 region prostate biopsy is superior to sextant method for diagnosing carcinoma of the prostate. The Journal of urology, **157:** 199, 1997

13. Gore, J. L., Shariat, S. F., Miles, B. J. et al.: Optimal combinations of systematic sextant and laterally directed biopsies for the detection of prostate cancer. The Journal of urology, **165:** 1554, 2001

14. Mian, B. M., Naya, Y., Okihara, K. et al.: Predictors of cancer in repeat extended multisite prostate biopsy in men with previous negative extended multisite biopsy. Urology, **60:** 836, 2002

15. Naughton, C. K., Miller, D. C., Yan, Y.: Impact of transrectal ultrasound guided prostate biopsy on quality of life: a prospective randomized trial comparing 6 versus 12 cores. The Journal of urology, **165:** 100, 2001

16. Kawachi, M. H., Bahnson, R. R., Barry, M. et al.: NCCN clinical practice guidelines in oncology: prostate cancer early detection. Journal of the National Comprehensive Cancer Network : JNCCN, **8:** 240, 2010

17. Heidenreich, A., Bellmunt, J., Bolla, M. et al.: EAU guidelines on prostate cancer. Part 1: screening, diagnosis, and treatment of clinically localised disease. European urology, **59:** 61, 2011

18. Taneja, S. S., Hsu, E. I., Cheli, C. D. et al.: Complexed prostate-specific antigen as a staging tool: results based on a multicenter prospective evaluation of complexed prostate-specific antigen in cancer diagnosis. Urology, **60:** 10, 2002

19. Naya, Y., Fritsche, H. A., Cheli, C. D. et al.: Volume indexes of total, free, and complexed prostate-specific antigen enhance prediction of extraprostatic disease extension in men with nonpalpable prostate cancer. Urology, **62:** 1058, 2003

20. Bianco, F. J., Jr., Mallah, K. N., Korets, R. et al.: Prostate volume measured preoperatively predicts for organ-confined disease in men with clinically localized prostate cancer. Urology, **69:** 343, 2007

21. Bostwick, D. G., Qian, J., Drewnowska, K. et al.: Prostate needle biopsy quality in reduction by dutasteride of prostate cancer events study: worldwide comparison of improvement with investigator training and centralized laboratory processing. Urology, **75:** 1406, 2010

22. Benchikh El Fegoun, A., El Atat, R., Choudat, L. et al.: The learning curve of transrectal ultrasound-guided prostate biopsies: implications for training programs. Urology, **81:** 12, 2013

23. Karam, J. A., Shulman, M. J., Benaim, E. A.: Impact of training level of urology residents on the detection of prostate cancer on TRUS biopsy. Prostate cancer and prostatic diseases, **7:** 38, 2004

24. Nguyen, C. T., Gao, T., Hernandez, A. V. et al.: Can residents perform transrectal ultrasound-guided prostate biopsy with patient comfort comparable to biopsy performed by attending staff urologists? Prostate cancer and prostatic diseases, **13:** 52, 2010

25. Collins, G. N., Raab, G. M., Hehir, M. et al.: Reproducibility and observer variability of transrectal ultrasound measurements of prostatic volume. Ultrasound in medicine & biology, **21:** 1101, 1995

26. Sech, S., Montoya, J., Girman, C. J. et al.: Interexaminer reliability of transrectal ultrasound for estimating prostate volume. The Journal of urology, **166:** 125, 2001

27. Loeb, S., Han, M., Roehl, K. A. et al.: Accuracy of prostate weight estimation by digital rectal examination versus transrectal ultrasonography. The Journal of urology, **173:** 63, 2005

28. Myschetzky, P. S., Suburu, R. E., Kelly, B. S., Jr. et al.: Determination of prostate gland volume by transrectal ultrasound: correlation with radical prostatectomy specimens. Scandinavian journal of urology and nephrology. Supplementum, **137:** 107, 1991

29. Rodriguez, E., Jr., Skarecky, D., Narula, N. et al.: Prostate volume estimation using the ellipsoid formula consistently underestimates actual gland size. The Journal of urology, **179:** 501, 2008

Table 1: Differences between TRUS-estimated and RP specimen-estimated volumes for level of training and comparison between these differences by level of training

	Junior mean (SD) median	**Senior** mean (SD) median	**Chief** mean (SD) median	**Fellow** mean (SD) median	**Staff** mean (SD) median	**Total** mean (SD) median	
TRUS volume	41.5 (20.2) 36.8	41.9 (22.1) 37.3	42.2 (21.6) 36.9	36.4 (15.7) 33.4	34.9 (16.2) 30.8	40.4 (20.2) 35.9	
RP volume	39.6 (20.4) 34.0	41.2 (23.9) 34.0	38.8 (18.4) 33.5	38.2 (13.0) 39.6	37.0 (16.2) 33.0	39.5 (20.3) 33.9	
Difference mean (SD)	(n=185) -1.9 (13.3)	(n=114) +0.1 (19.3)	(n=42) -3.4 (12.6)	(n=18) +1.8 (15.8)	(n=67) +2.1 (12.5)	(n=426) -0.7 (15.1)	**p value** 0.2428
p value	0.0606	0.9686	0.0907	0.6307	0.1740	0.3335	

Table 2: : Linear regression of logarithm of RP specimen-estimated volume by logarithm of TRUS volume controlling by level of training, its interaction with logarithm of TRUS volume and preoperative PSA

Variable	**Coefficient (SE)**	**p value**
Log(TRUS volume)	0.73 (0.03)	<0.001
Junior (reference is staff, junior = 1, staff = 0)	-0.09 (0.04)	0.0450
Senior (reference is staff, senior = 1, staff = 0)	-0.05 (0.05)	0.3280
Chief (reference is staff, chief = 1, staff = 0)	-0.11 (0.06)	0.0663
Fellow (reference is staff, fellow = 1, staff = 0)	-0.00 (0.08)	0.9981
Preoperative PSA	0.06 (0.01)	<0.0001

Table 3 : Kappa statistic and percent agreement between clinical and pathologic Gleason scores (grouped as 2-6, 7, and 8-10) for each level of training

Level of training	**N**	**κ (SE)**	**% agreement**	**% Bx < RP**	**% Bx > RP**
Junior	210	0.58 (0.05)	76.7 %	18.6 %	4.7 %
Senior	126	0.46 (0.07)	72.2 %	22.2 %	5.6 %
Chief	47	0.50 (0.12)	72.3 %	14.9 %	12.8 %
Fellow	18	0.72 (0.14)	83.3 %	16.7 %	0.0 %
Staff	77	0.72 (0.09)	87.0 %	13.0 %	0.0 %
Total	478				

Table 4: Cumulative logistic regression odds ratios of pathologic Gleason score (2-6, 7, 8-10) by clinical Gleason score, controlling by level of training, clinical stage, logarithm of RP specimen-estimated volume, logarithm of estimated tumor volume, and preoperative PSA

	OR	95% CI	p-value
Clinical Gleason score (unit increase in group)	16.360	9.480 - 28.232	<0.0001
Junior (reference is Staff)	1.096	0.552 - 2.177	0.7924
Senior (reference is Staff)	1.541	0.739 - 3.215	0.2488
Chief (reference is Staff)	0.716	0.272 - 1.887	0.4999
Fellow (reference is Staff)	0.982	0.272 - 3.547	0.9777
Clinical stage (reference is T1ab, T1c)	1.246	0.749 - 2.071	0.3975
Log (RP specimen-estimated volume)	0.774	0.435 - 1.377	0.3836
Log (estimated tumor volume)	1.523	1.308 - 1.773	<0.0001
Preoperative PSA	1.028	0.990 - 1.067	0.1506

Table 5: Linear regression of logarithm of estimated tumor volume by percentage of positive cores controlling by level of training, interaction with percentage of positive cores, clinical stage, logarithm of RP specimen-estimated volume and preoperative PSA

	Coefficient (SE)	**p value**
% positive cores	3.94 (0.46)	<0.0001
Junior (reference is staff, junior = 1, staff = 0)	0.24 (0.26)	0.3468
Senior (reference is staff, senior = 1, staff = 0)	0.06 (0.29)	0.8276
Chief (reference is staff, chief = 1, staff = 0)	0.71 (0.37)	0.0564
Fellow (reference is staff, fellow = 1, staff = 0)	0.81 (0.54)	0.1321
Clinical stage (reference is T1ab, T1c)	0.31 (0.20)	0.1179
Log (RP specimen-estimated volume)	-0.42 (0.22)	0.0560
Preoperative PSA	0.03 (0.01)	0.0044

Abbreviations:

DRE – digital rectal examination

PGY – post-graduate year

PSA – Prostate-specific antigen

PSAD – prostate-specific antigen density

RP – radical prostatectomy

TRUS – transrectal ultrasound

Supplementary Table 1: Kappa statistic and percent agreement between clinical and pathologic highest Gleason grade for each level of training

Level of training	N	κ (SE)	% agreement	% Bx < RP	% Bx > RP
Junior	210	0.59 (0.05)	79.0 %	18.1 %	2.9 %
Senior	126	0.47 (0.07)	73.0 %	22.2 %	4.8 %
Chief	47	0.51 (0.13)	76.6 %	17.0 %	6.4 %
Fellow	18	0.61 (0.07)	77.8 %	2.2 %	0.0 %
Staff	77	0.70 (0.09)	87.0 %	13.0 %	0.0 %

Supplementary Table 2: Cumulative logistic regression odds ratios of maximum pathologic Gleason grade by maximum clinical Gleason grade, controlling by level of training, clinical stage, logarithm of RP specimen-estimated volume, logarithm of estimated tumor volume and preoperative PSA

	OR	95% CI	p-value
Maximum clinical Gleason grade (unit increase)	23.718	12.510 - 44.971	<0.0001
Junior (reference is Staff)	1.106	0.543 - 2.251	0.7818
Senior (reference is Staff)	1.633	0.765 - 3.487	0.2052
Chief (reference is Staff)	1.003	0.375 - 2.682	0.9957
Fellow (reference is Staff)	1.200	0.331 - 4.353	0.7821
Clinical stage (reference is T1ab, T1c)	1.208	0.718 - 2.032	0.4767
Log (RP specimen-estimated volume)	0.861	0.479 - 1.549	0.6180
Log (estimated tumor volume)	1.553	1.330 - 1.814	<0.0001
Preoperative PSA	1.031	1.000 - 1.063	0.0504

Supplementary Table 3: Kappa statistic and percent agreement between clinical and pathologic primary Gleason grade for each level of training

Level of training	**N**	**κ (SE)**	**% agreement**	**% Bx < RP**	**% Bx > RP**
Junior	210	0.78 (0.07)	95.2 %	2.4 %	2.4 %
Senior	126	0.47 (0.11)	87.3%	10.3 %	2.4 %
Chief	47	0.38 (0.19)	85.1 %	6.4 %	8.5 %
Fellow	18	0.63 (0.17)	88.9 %	11.1%	0.0 %
Staff	77	0.64 (0.19)	96.1 %	1.3 %	2.6 %

Supplementary Table 4: Cumulative logistic regression odds ratios of primary pathologic Gleason grade by primary clinical Gleason grade, controlling by level of training, clinical stage, logarithm of RP-specimen-estimated volume, logarithm of estimated tumor volume and preoperative PSA

	OR	95% CI	p-value
Primary Gleason grade (unit increase)	40.930	16.585 - 101.012	<.0001
Junior (reference is Staff)	1.753	0.452 - 6.800	0.4173
Senior (reference is Staff)	5.530	1.396 - 21.903	0.0149
Chief (reference is Staff)	1.414	0.254 - 7.865	0.6921
Fellow (reference is Staff)	5.038	0.739 - 34.361	0.0988
Clinical stage (reference is T1ab, T1c)	0.936	0.421 - 2.079	0.8711
Log (RP specimen-estimated volume)	0.736	0.299 - 1.809	0.5042
Log (estimated tumor volume)	1.356	1.070 - 1.717	0.0117
Preoperative PSA	1.006	0.973 - 1.040	0.7256