



Published in final edited form as:

Urology. 2011 May ; 77(5): 1155–1160. doi:10.1016/j.urology.2010.08.031.

Interobserver Variability in Histologic Evaluation of Radical Prostatectomy Between Central and Local Pathologists: Findings of TAX 3501 Multinational Clinical Trial

George J. Netto, Mario Eisenberger, Jonathan I. Epstein, and for the TAX 3501 Trial Investigators

Departments of Pathology, Urology, and Oncology, Johns Hopkins University, Baltimore, Maryland

Abstract

OBJECTIVES—To determine the agreement between the local pathologist findings and central pathologist findings using data from the TAX 3501 trial. TAX 3501 was a randomized, multinational trial comparing the outcomes of patients with high-risk prostate cancer treated with androgen deprivation with or without docetaxel after radical prostatectomy (RP). Patient eligibility was determined by a minimal 5-year progression-free survival estimate of 60% using Kattan's nomogram.

METHODS—The pathologic findings were reassessed in 257 consecutive RP specimens by 2 central pathologists and compared with the local pathologist data.

RESULTS—For the Gleason score, agreement was found in 181 (70%) of 257 cases, upgrading in 57 (75%), and downgrading in 25% of the RP specimens. The most frequent upgrade was from Gleason score 7 to 8 or 9 and downgrading from Gleason score 8 to 7. Of the upgrades and downgrades, 37% and 21% were of 2 Gleason score points, respectively. For the tumor extent, agreement was found in 179 (70%) of 256 specimens, with upstaging in 70 (91%) and downstaging in 9%. The most frequent upstage was from focal to extensive extraprostatic extension (45%). For seminal vesicle invasion, agreement was found for 238 (93%) of 256 RP specimens. Almost equal rates of underdiagnosing and overdiagnosing seminal vesicle invasion was observed. For margin status, agreement was present for 229 (89%) of 256 cases. The central pathologist review led to reclassification as a positive margin in 17 cases and a negative margin in 10. For lymph node status, 2 (1%) of 210 RP specimens had positive nodes identified only by the central pathologist. Agreement was observed in 154 negative and 54 positive cases.

CONCLUSIONS—Significant interobserver variations were found between the central and local pathologists. From the central pathologist review, the progression-free survival estimates were altered in 31 patients (13%), including 22 who were reassigned a greater risk estimate, rendering them study eligible. Thus, interobserver variability affected prognostication and trial accrual.

More than one third of patients who undergo retropubic radical prostatectomy (RP) or radiotherapy for localized prostate cancer will experience disease recurrence within 3–5 years.¹ Hormonal deprivation treatment, chemotherapy, and biologic modifiers are being studied in both the adjuvant and the neoadjuvant settings to lower the treatment failure rates

© 2011 Elsevier Inc.

Reprint requests: Jonathan I. Epstein, M.D., Department of Pathology, Johns Hopkins University, 401 North Broadway, Weinberg Building 2242, Baltimore, MD 21231. jepstein@jhmi.edu.

Presented in part at the 2009 United States and Canadian Academy of Pathology Meeting, Boston, Massachusetts, and the American Urologic Association 2009 Meeting, Chicago, Illinois

for this particular population. Although chemotherapy has historically been regarded as only modestly effective against castration-resistant prostate cancer, more recent studies have demonstrated that treatment with a cytotoxic agent effectively reduces the risk of death in castration-resistant prostate cancer.² The current criteria for identifying patients at high risk of recurrence after RP have included a combination of clinical and pathologic features.^{3,4} The accurate assessment of recurrence risk could improve the design and evaluation of clinical trials of localized prostate cancer and facilitate the selection of appropriate adjuvant therapy. Kattan's postoperative nomogram predicts the probability of prostate cancer progression within 5 years of RP by combining multiple prognostic variables, including prostate-specific antigen level before RP, Gleason score, disease extent, and margin, lymph node, and seminal vesicle status).^{3,4}

TAX 3501 was a randomized Phase III multinational trial comparing the outcomes of post-RP immediate versus deferred adjuvant androgen deprivation with or without docetaxel in high-risk patients. Eligibility was determined by a 5-year biochemical progression-free survival (PFS) probability of $\geq 60\%$ calculated using Kattan's postoperative nomogram to define the eligible high-risk prostate cancer patient population. The pathologic data from central pathologist (CP) review of RP specimens were used for TAX 3501 eligibility, providing a unique opportunity to compare the interobserver agreement (local pathologist [LP] vs CP assessments) on histologic findings in a group of patients who agreed to be considered for enrollment in the present study.

The present study aimed at defining the role and effect of central pathology review on prostate cancer grading, staging, and risk assessment in the setting of multicenter clinical trials to ensure the consistency of analysis and validity of future study findings.

MATERIAL AND METHODS

TAX 3501 was a multicenter, multinational study. Eligibility was determined by a predicted 5-year PFS of $\geq 60\%$, as determined by Kattan's postoperative nomogram using the CP review findings. Eligible patients were randomized to either deferred or immediate adjuvant androgen-deprivation therapy, with or without docetaxel.

After the initial LP review, all hematoxylin-eosin sections of the RP specimen and accompanying the LP laboratory report were mailed to the Department of Pathology at the Johns Hopkins Hospital to be blindly reviewed again by either of the 2 central urologic pathologists participating in the trial (J.I.E. and G.J.N.). None of the cases were reviewed by both CPs.

The Gleason score (GSc) was assigned according to the World Health Organization/International Society of Urologic Pathologist 2005 consensus conference criteria.^{5,6} Extraprostatic extension (EPE) was categorized as either focal EPE (FEPE, 1 or 2 microscopic foci involving extraprostatic fibroadipose tissue and located immediately beyond the prostatic perimeter) or extensive EPE (EEPE). Tumors with invasion into, but not through, the "prostatic capsule" were categorized as organ confined (OC).^{7,8}

Data on the GSc, presence of EPE (OC vs FEPE vs EEPE), margin status, seminal vesicle (SV) status, and lymph node status was independently assessed by the CP and a LP for each specimen. A checklist, including all the previous pathologic RP findings, was provided by the trial administrator and was filled by both the CP and LP for 257 consecutive RP specimens. The pathologic findings were compared between the CP and LP.

RESULTS

Our findings are summarized in Table 1 and graphically illustrated in Figure 1A, B.

Interobserver Agreement on Evaluation of GSc

Agreement was found in 181 (70%) of 257 RP specimens. Of the 76 cases in which GSc variance was present, a CP upgrade occurred in 57 (75%) and a downgrade in 19 (25%) cases. The most frequent upgrade was from GSc 7 to 8 or 9 (34 [60%] of 57 cases), and the most frequent downgrade was from GSc 8 to 7 (10 [53%] of 19 cases). Of the GSc upgrades, 37% (22 of 57) and 2% (1 of 57) were of 2 and 3 GSc increments, respectively. Of the downgrades, 21% (4 of 19) were of 2 GSc increments, with all remaining GSc changes (30 [40%] of 76 cases%) of 1 GSc change from the originally assigned LP Gleason score.

Interobserver Agreement on Assessment of EPE (OC/FEPE/EEPE)

Agreement between the CP and LP was achieved in 179 (70%) of 256 of the RP specimens. Of the 77 cases (30%) with stage variance, upstaging by the CP occurred in 70 (91%) and downstaging in 7 cases (9%). The most frequent upstage was from FEPE to EEPE (31 of 69, 45%) followed by OC to EEPE (19 of 69, 28%). Of the 69 RP specimens originally staged as capsular invasion by the LP, 18 (26%) were restaged as EEPE by the CP. The reasons that led to understaging of EEPE by the LP included ambiguity at the apical section, misinterpreting the presence of extraprostatic tumor associated with a desmoplastic response lacking surrounding fat, and ambiguity of the definition of “capsular invasion” (Fig. 2).

Interobserver Agreement on Evaluation of SV Status

Agreement was encountered in 238 (93%) of 256 cases. Of the 18 RP specimens with disagreement (7%), an almost equal number of cases were undercalled (8 of 18) and overcalled (10 of 18) by the LP compared with the CP. Overcalling SV involvement by the LP primarily resulted from misinterpreting the presence of tumor in the soft tissue surrounding SV proper without tumor invasion into the muscularis layer of the SV. Other reasons that led to overcalling of SV involvement included tumor invasion of the ejaculatory duct and the presence of tumor in the prostate tissue adjacent to the SV without actual invasion of the SV.

Interobserver Agreement on Evaluation of Margin Status

Agreement was present in 229 (89%) of 256 RP specimens. Of the 27 cases (11%) with margin status evaluation variance, the CP review led to reclassification of the margin status from negative to positive in 17 cases (62%) and from positive to negative in 10 cases (38%). Overcalling a margin as positive by the LP primarily resulted from falsely interpreting an artifactual tissue tear at the inked prostatic surface. Undercalling primarily resulted from missing an area of cauterized or crushed tumor cells at the inked surface (Fig. 3).

Interobserver Agreement on Evaluation of Lymph Node Status

The pelvic lymph nodes were not removed in 47 patients. Positive lymph nodes were identified on CP review in 2 (1%) of 210 RP specimens that were originally missed on the LP review. A single involved node was identified in each of the 2 cases missed on the original review. Agreement was observed in all remaining 208 lymph node dissection specimens (154 negative and 54 positive cases).

Effect of CP Review on Estimates of PFS

From the CP review, 31 (13%) of 257 patients were reclassified in the terms of their PFS risk estimate and subsequent study eligibility. Of the 31 patients, 22 (71%) were reassigned

to a greater PFS risk estimate on CP review (predicted 5-year rate PFS of 60%) placing them within the eligibility criteria. Of the 31 patients, 9 (29%) initially deemed eligible were reassigned to a lower risk estimate (predicted 5-year PFS rate >60%), making them ineligible for inclusion.

COMMENT

The precise histologic assessment of RP specimens in patients with prostate cancer is of paramount importance for achieving an accurate risk assessment of disease recurrence and hence implementation of proper additional management.^{3,4,7-9} TAX 3501 was uniquely designed to use a postoperative nomogram to calculate the risk estimates to define patient eligibility. The pathologic data used in the nomogram was determined using the RP findings from the CP and not the LP to ensure consistency. Therefore, the TAX 3501 trial presented us with an opportunity to evaluate the effect of central pathology review on the risk assessment and assess interobserver variability between the CPs and LPs in RP Gleason grading, tumor extent (OC vs EPE), margin status, seminal vesicle, and lymph node involvement.

Several studies have previously assessed interobserver variability in Gleason grading in both the setting of needle biopsy and RP specimens.¹⁰⁻¹⁵ In contrast, the studies evaluating interobserver variation of the histologic elements of RP staging and margin status have been limited to only 4 previous analyses.¹⁵⁻¹⁸ The 70% Gleason score agreement rate found in the present study is within the 36%–73% range cited by previous studies¹¹ and is almost identical to the rates encountered in more recent studies.^{12,19} Similar to previous reports, our study showed that most disagreements in GSc were within 1 point.

Among the 4 previous studies evaluating interobserver variability in pathologic staging and margin status in RP,¹⁵⁻¹⁷ the study by van der Kwast et al¹⁷ shares the closest design similarities with our present study. It involved a single CP review of 552 RP specimens obtained as a part of the European Organization for Research and Treatment of Cancer 22911 trial. Our 70% rate of agreement between the CP and LP in evaluating organ confinement versus EPE is virtually identical to the 69% rate found by van der Kwast et al.¹⁷ In contrast to that study, however, we found the greater proportion of the differences resulted from CP upstaging, rather than downstaging (27% and 3%), and van der Kwast et al¹⁷ encountered 4% CP upstaging and 27% downstaging in their European Organization for Research and Treatment of Cancer 22911 cohort. The study by Ekici et al,¹⁵ similar to ours, found a relatively greater likelihood of upstaging from OC to EPE (13%) compared with their 3% rate of downstaging on expert review.

Of the reasons we found responsible for understaging of EEPE by LPs, ambiguity of the prostate confines at the apex deserves further discussion. Variation in interpreting EPE at this location continues to exist among expert urologic pathologists, which was highlighted recently at the latest International Society of Urologic Pathologist Consensus Conference in 2009. Some of the cases in which tumor involved the apex and extended to the margin of resection were designated as organ confined by the LP and, at RP analysis, were designated as EEPE by the CP. Currently, the recommendation is to stage these cases as pT2+ (ie, uncertain whether the cases are OC or EPE) owing to the ambiguity of the boundaries of the prostate at this region, although disagreement among expert urologic pathologists continues to exist on the nuances of how to interpret EPE at the apical region. Additional studies to address the true biologic effect of tumor presence beyond the level of benign glands and tumor extension to the ink at the apical margin are needed and could help pathologists in the future reach a better consensus regarding EPE versus OC staging at the apex.

As with the previous 2 studies by Ekici et al¹⁵ and van der Kwast et al,¹⁷ we found a very high rate of concordance in the evaluation of SV and lymph node involvement (93% and 99%, respectively) corresponding to 95% and 99% in the study by Ekici et al¹⁵ and a 94% concordance rate of SV status assessment in the study by van der Kwast et al.¹⁷ Compared with the results from the present study, the recent study by Kuroiwa et al.¹⁸ found a slightly lower rate of SV evaluation concordance (83%) but a similarly high rate of agreement in the lymph node assessment.

Our finding of an 11% rate of margin status disagreement between the CP and LP was almost identical to the 12% rate found by Kuroiwa et al,¹⁸ was lower than the 26% rate found by van der Kwast et al,¹⁷ and was slightly greater than the 8% rate found by Ekici et al.¹⁵ Most discrepancies in our study resulted from undercalls. Attention to the presence of tumor cells at the inked areas of cautery and crush artifact will help improve the margin status interobserver variability.

From the CP review, 31 (13%) of our initially consented patients who underwent RP had their disease reclassified in terms of their PFS risk estimate and TAX 3501 study eligibility. More than two thirds of the reclassified patients were reassigned to a greater PFS risk estimate, placing them within the eligibility criteria for the study using Kattan's nomogram (predicted 5-year PFS of 60%). The latter further illustrates the value of CP review in ensuring accuracy in adherence to multi-institutional clinical trials' eligibility criteria and providing consistency in the initial PFS estimates that might ultimately affect the trial outcomes and jeopardize the validity of final results.

CONCLUSIONS

We found significant interobserver variation between the CP and LP evaluation of the RP specimens from patients considered for the TAX 3501 trial. The variation was most pronounced in the assessment of tumor EPE and GSc. These findings suggest that CP review might be an important factor to ensure the uniformity of pathologic analysis and prevent interobserver variations from affecting the rate of patient accrual to multi-institutional clinical trials.

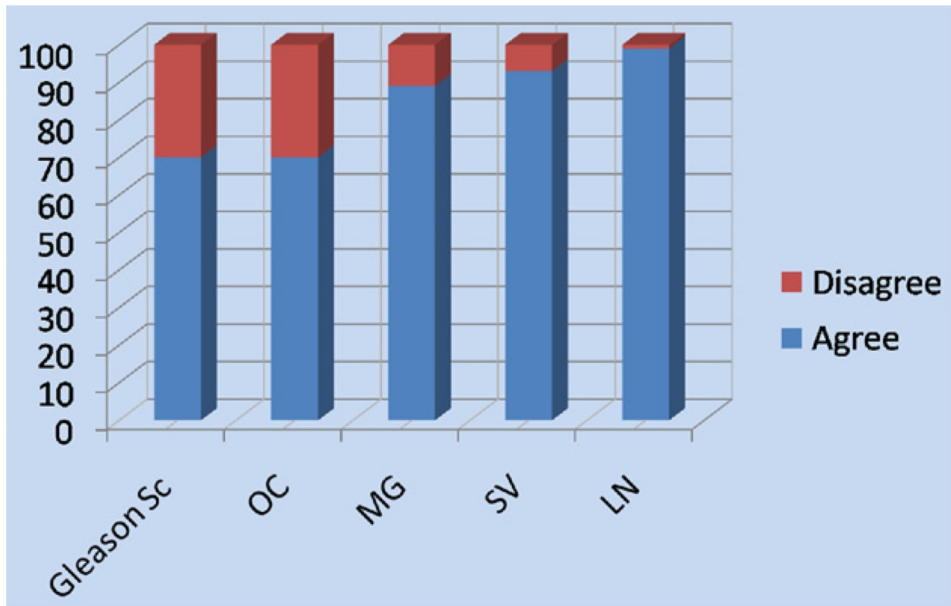
Acknowledgments

Sponsored in part by Sanofi-Aventis, Paris, France

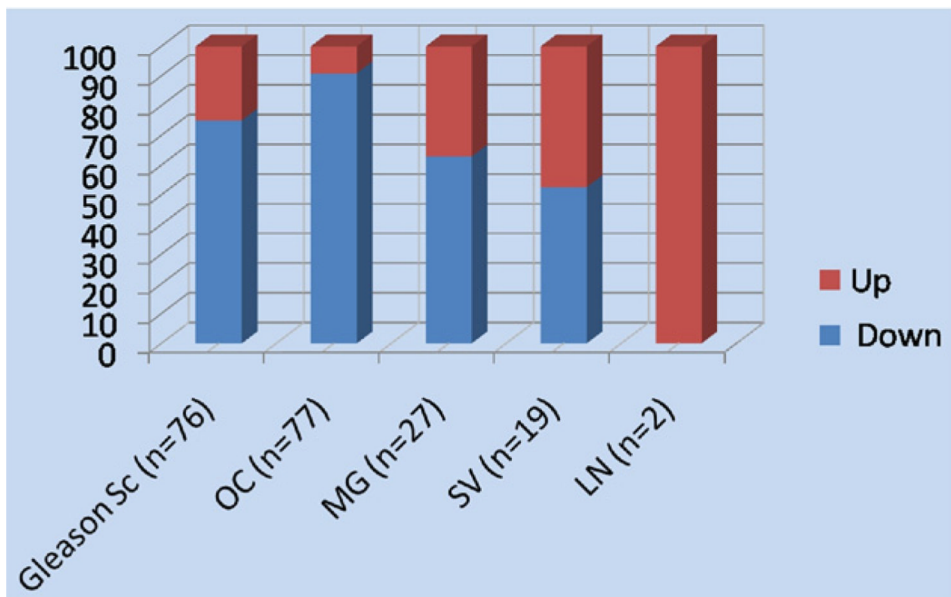
References

1. Hull GW, Rabbani F, Abbas F, et al. Cancer control with radical prostatectomy alone in 1,000 consecutive patients. *J Urol*. 2002; 167:528–534. [PubMed: 11792912]
2. Petrylak DP, Tangen CM, Hussain MH, et al. Docetaxel and estramustine compared with mitoxantrone and prednisone for advanced refractory prostate cancer. *N Engl J Med*. 2004; 351:1513–1520. [PubMed: 15470214]
3. Kattan MW, Wheeler TM, Scardino PT. Postoperative nomogram for disease recurrence after radical prostatectomy for prostate cancer. *J Clin Oncol*. 1999; 17:1499–1507. [PubMed: 10334537]
4. Kattan MW, Eastham J. Algorithms for prostate-specific antigen recurrence after treatment of localized prostate cancer. *Clin Prostate Cancer*. 2003; 1:221–226. [PubMed: 15040880]
5. Epstein JI, Allsbrook WC Jr, Amin MB, et al. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason grading of prostatic carcinoma. *Am J Surg Pathol*. 2005; 29:1228–1242. [PubMed: 16096414]
6. Epstein JI, Allsbrook WC Jr, Amin MB, et al. Update on the Gleason grading system for prostate cancer: results of an international consensus conference of urologic pathologists. *Adv Anat Pathol*. 2006; 13:57–59. [PubMed: 16462155]

7. Epstein JI, Amin M, Boccon-Gibod L, et al. Prognostic factors and reporting of prostate carcinoma in radical prostatectomy and pelvic lymphadenectomy specimens. *Scand J Urol Nephrol Suppl.* 2005; 216:34–63. [PubMed: 16019758]
8. Epstein JI, Srigley J, Grignon D, et al. Recommendations for the reporting of prostate carcinoma. *Hum Pathol.* 2007; 38:1305–1309. [PubMed: 17707261]
9. Kausik SJ, Blute ML, Sebo TJ, et al. Prognostic significance of positive surgical margins in patients with extraprostatic carcinoma after radical prostatectomy. *Cancer.* 2002; 95:1215–1219. [PubMed: 12216087]
10. Allsbrook WC Jr, Mangold KA, Johnson MH, et al. Interobserver reproducibility of Gleason grading of prostatic carcinoma: urologic pathologists. *Hum Pathol.* 2001; 32:74–80. [PubMed: 11172298]
11. Allsbrook WC Jr, Mangold KA, Johnson MH, et al. Interobserver reproducibility of Gleason grading of prostatic carcinoma: general pathologist. *Hum Pathol.* 2001; 32:81–88. [PubMed: 11172299]
12. Glaessgen A, Hamberg H, Pihl CG, et al. Interobserver reproducibility of modified Gleason score in radical prostatectomy specimens. *Virchows Arch.* 2004; 445:17–21. [PubMed: 15156317]
13. Glaessgen A, Hamberg H, Pihl CG, et al. Interobserver reproducibility of percent Gleason grade 4/5 in prostate biopsies. *J Urol.* 2004; 171:664–667. [PubMed: 14713782]
14. Oyama T, Allsbrook WC Jr, Kurokawa K, et al. A comparison of interobserver reproducibility of Gleason grading of prostatic carcinoma in Japan and the United States. *Arch Pathol Lab Med.* 2005; 129:1004–1010. [PubMed: 16048389]
15. Ekici S, Ayhan A, Erkan I, et al. The role of the pathologist in the evaluation of radical prostatectomy specimens. *Scand J Urol Nephrol.* 2003; 37:387–391. [PubMed: 14594686]
16. Evans AJ, Henry PC, Van der Kwast TH, et al. Interobserver variability between expert urologic pathologists for extraprostatic extension and surgical margin status in radical prostatectomy specimens. *Am J Surg Pathol.* 2008; 32:1503–1512. [PubMed: 18708939]
17. van der Kwast TH, Collette L, van Poppel H, et al. Impact of pathology review of stage and margin status of radical prostatectomy specimens (EORTC trial 22911). *Virchows Arch.* 2006; 449:428–434. [PubMed: 16941153]
18. Kuroiwa K, Shiraishi T, Ogawa O, et al. Discrepancy between local and central pathological review of radical prostatectomy specimens. *J Urol.* 2010; 183:952–957. [PubMed: 20083260]
19. Burchardt M, Engers R, Muller M, et al. Interobserver reproducibility of Gleason grading: evaluation using prostate cancer tissue microarrays. *J Cancer Res Clin Oncol.* 2008; 134:1071–1078. [PubMed: 18392850]



A



B

Figure 1. (A, B) Graphic depiction of CP and LP agreement/disagreement on interpretations of 5 pathologic parameters in 257 consecutive RP specimens from patients considered for TAX 3501 trial.

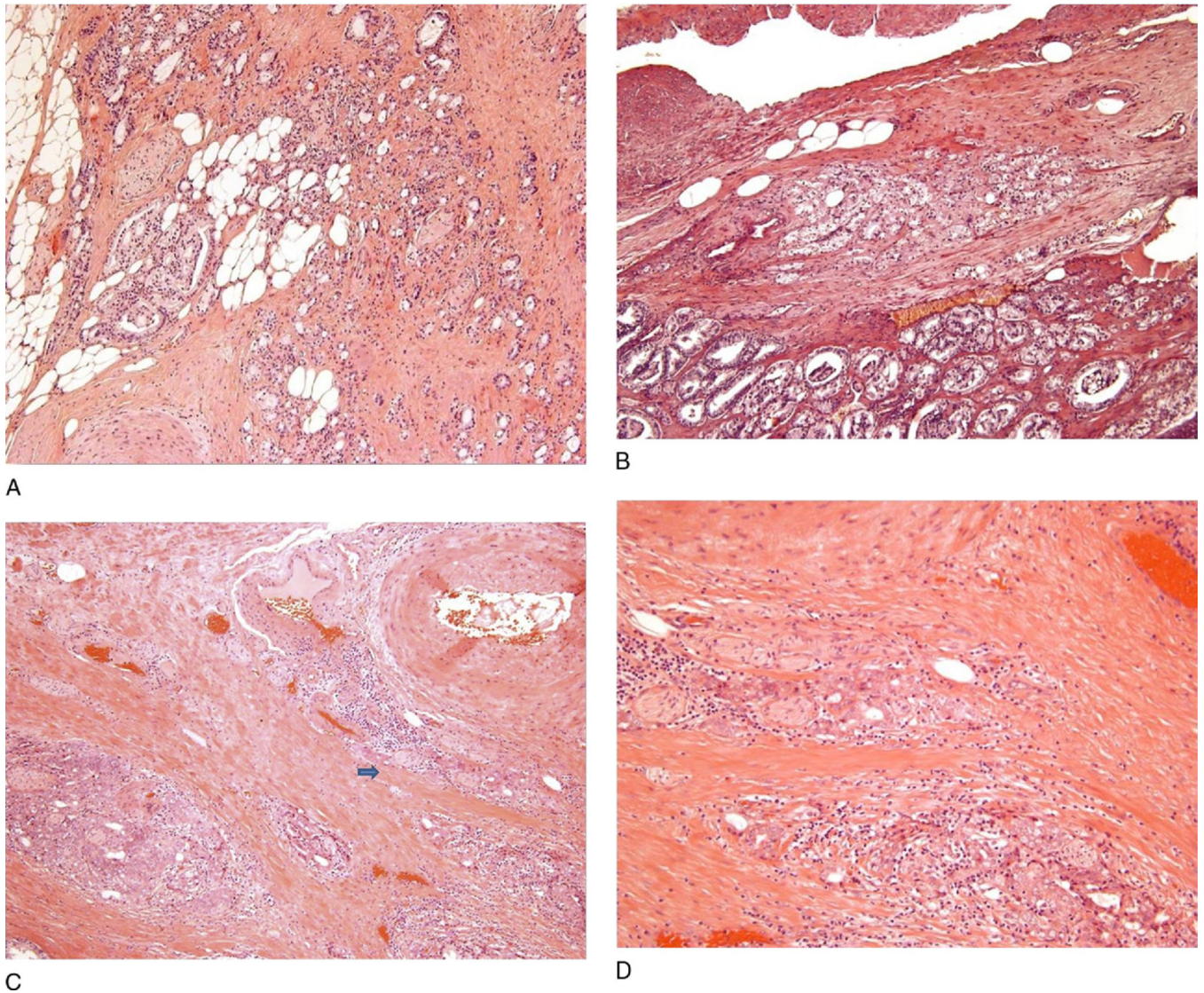
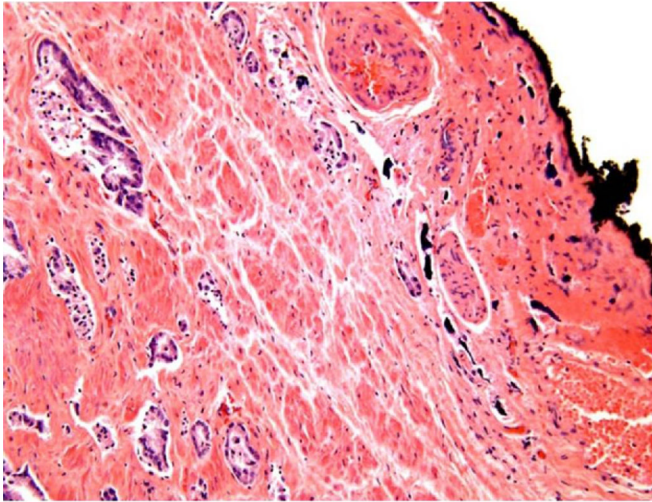
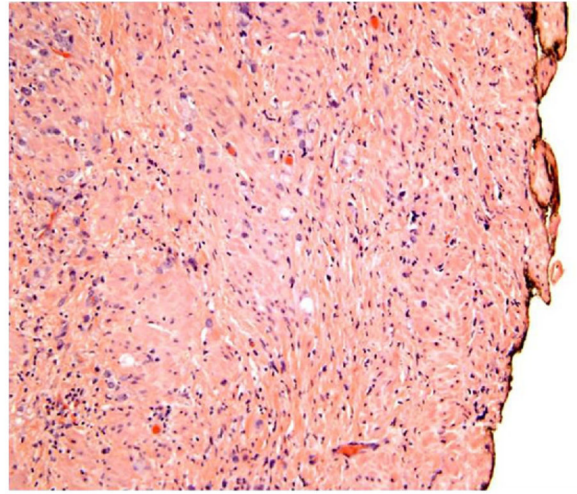


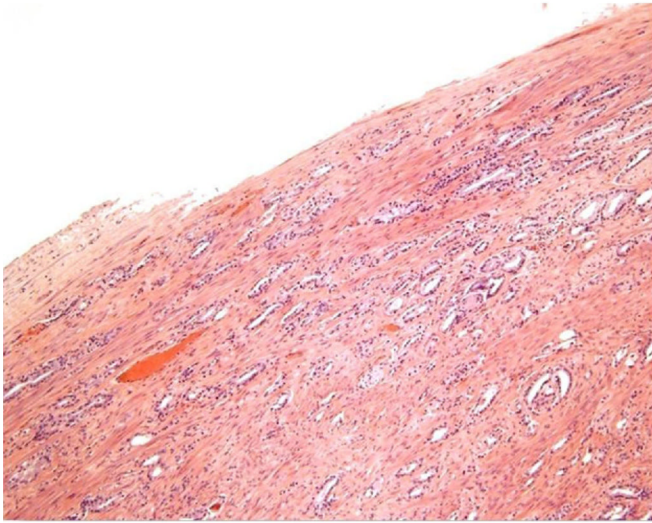
Figure 2. (A, B) EEPE misinterpreted by LP as “invasion into capsule.” Tumor glands infiltrating adipose tissue indicating EPE. (C, D) EEPE misinterpreted by LP as OC disease. (C) Tumor glands infiltrating fibrovascular extraprostatic tissue (arrow) lacking organized smooth muscle bundles of prostatic stroma. (D) Higher magnification showing cancer in fibrous tissue.



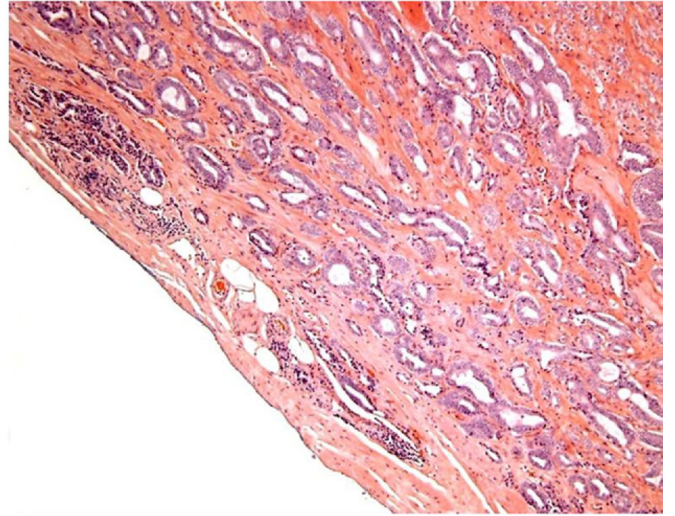
A



B



C



D

Figure 3.

(A, B) Positive peripheral prostatic margin misinterpreted by LP as negative. Margin considered positive because of presence of (A) cauterized tumor or (B) crushed individual cancer cells at inked surface. (C, D) Negative margins misinterpreted by LP as positive. (C) Margin could not be considered positive owing to artifactual absence of evaluable true inked surface. (D) presence of even a thin fibrous rim separating tumor glands from smooth inked border indicates negative margin.

Table 1

Rate of agreement and disagreement between central pathologist and local pathologist review of 5 pathologic parameters assessed in 257 prostatectomy specimens of TAX 3501 trial

Parameter	CP/LP Agreement (%)	CP/LP Disagreement (%)	Upgrade/Upstage/MG (Negative to Positive) (%)	Downgrade/Downstage/MG (Positive to Negative) (%)
GSc	181/257 (70)	76/257 (30)	57/76 (75)	19/76 (25)
OC/FEPE/EEPE	179/256 (70)	77/256 (30)	70/76 (90)	7/76 (9)
MG status	229/256 (89)	27/256 (11)	17/27 (62)	10/27 (38)
LN status	208/210 (99)	2/210 (1)	2/2 (100)	0/2 (0)
SV status	238/256 (93)	18/256 (7)	8/18 (44)	10/18 (66)

CP, central pathologist; LP, local pathologist; MG, margin; GSc, Gleason score; OC, organ confined; FEPE, focal extraprostatic extension; EEPE, extensive EPE; LN, lymph node; SV, seminal vesicle.