Running head: Assessment of Renal Mass Biopsies

Interobserver Variability in Assessment of Renal Mass Biopsies

Łukasz Nyk[1], Wojciech Malewski[1], Krystian Kaczmarek[2], Piotr Kryst[3], Michał Pyźlak[4], Aneta Andrychowicz[5], Tomasz Ząbkowski[6]

1. Department of Urology, European Health Center, Otwock, Poland, II Urology Clinic, Centre of Postgraduate Medical Education, Warsaw, Poland
2. Department of Urology and Urological Oncology, Pomeranian Medical University, Szczecin, Poland
3. Department of Urology, Bielański Hospital, Warsaw, Poland, II Urology Clinic, Centre of Postgraduate Medical Education, Warsaw, Poland
4. Department of Pathology and Laboratory Medicine. Maria Sklodowska-Curie Institute - Cancer Center. Roentgena 5, 02-781 Warsaw, Poland
5. Urological Clinic, Warsaw, Poland
6. Department o Urology, Military Institute of Medicine, Warsaw, Poland

ABSTRACT

Purpose: The main goal of this study was to assess the histopathological efficacy of renal mass biopsy and to check the concordance between pathological results and biopsy of the final specimen, as well as interobserver variability in the assessment of biopsy cores.

Materials and Methods: A hundred sets of core biopsies of postoperative specimens (renal masses) have been performed. Three core biopsies of the intact specimen had been performed once the kidney with the tumor, or the tumor alone were resected. The urologist aimed to obtain two cores from the peripheral sides of the tumor and one core from its center.

The surgical specimen was evaluated by a single pathologist, whereas biopsy samples were referred to three independent pathologists who were blinded to the final results of the renal mass biopsy.

Results: Nondiagnostic biopsy rates ranged from 13% to 22%. Sensitivity and specificity ranged 83-97% and 97-99% by excluding nondiagnostic results. The concordance between assessment of surgical specimen and biopsy in the Fuhrman grading system ranged 36.5-77.0%, respectively. Interobserver agreement between the three pathologists was substantial or moderate, depending on the tumor subtype. The Krippendorff's alpha coefficient, calculated by excluding the nondiagnostic results, was 0.28 (moderate agreement) for the Fuhrman grading system.

Conclusion: The agreement regarding grading of biopsies between three pathologists ranged from moderate to substantial. Therefore, a team of dedicated uropathologists should be engaged in final diagnosis of renal mass biopsy rather than single one before implementing the proper treatment.

Keywords: renal mass biopsy; interobserver variability; assessment; efficacy; treatment

INTRODUCTION

Over the past decades, the detection rate of renal cell carcinoma (RCC) has increased. Availability of ultrasound diagnostics has contributed to frequent diagnoses of small renal masses (SRMs) as well as larger asymptomatic tumors [(1)]. Because up to 33% of SRMs present as benign lesions on the final pathological examination, preoperative diagnosis is of significant value [(2)]. Currently, only angiomyolipomas (AMLs) can be confirmed with cross-sectional imaging without histopathological examination [(3)]. Although techniques of partial nephrectomy have been refined through robotic assistance, nephron-sparing surgery still carries a risk of complications [(4)]. Consequently, SRM surveillance poses an interesting management modality, especially in the elderly and/or comorbid patients [(5)]. Moreover, a large multi-institutional study by *Pierorazio* confirmed the safety and uncompromised cancer-specific survival of patients with SRM managed with active surveillance (AS). Renal mass biopsy, pathological proof of benignancy or relatively low-risk pathology, with regular radiological follow-up, are essential parts of such management [(6)].

EAU guidelines recommend performing biopsies with at least two cores, avoiding necrotic areas in the tumor. Biopsy of cystic masses is questionable. On the other hand, obtaining reliable pathology from Bosniak III lesions preoperatively would be valuable as most of them are benign or have low malignant potential [(7)]. The present metanalysis confirmed high sensitivity and specificity of renal mass biopsy in the diagnosis of malignancy. Concordance of biopsy results and final specimen for histotype is lower. Correct assessment of tumor grade seems to be the most challenging [(6)]. As the diagnosis of malignancy is of the highest importance for active surveillance, variability of assessments between different pathologists is intriguing. This study focuses on the accuracy and interobserver variability of histopathological results of renal mass biopsy performed in ideal non-real life conditions. Even computerized tomography guidance may result in insufficient material for analysis [(8)].

As biopsies were performed “in-bench” postoperatively, samples were most representative for this kind of study as the tumor was sampled directly without imaging guidance. To best of our knowledge this is the second study assessing histopathological interobserver variability of renal mass biopsies performed “in-bench” with a large number of cases.

MATERIALS AND METHODS

A hundred sets of core biopsies of postoperative specimens (renal masses) have been performed. All patients provided written informed consent before the procedure, to allow the use of the specimen for this study. Therefore a team of dedicated uropathologists should be engaged in final diagnosis of renal mass biopsy rather than single one before implementing the proper treatment, especially active surveillance. It was used an 18-G core needle for each biopsy. The urologist aimed to obtain two cores from the peripheral sides of the tumor and one core from its center. After the biopsy, the surgical specimen was processed as previously described.

Biopsy samples were fixed in formalin, embedded in paraffin, and stained with hematoxylin and eosin dye. The surgical specimen was evaluated by a single pathologist, whereas biopsy samples were referred to three independent pathologists who were blinded to the final results of the renal mass biopsy. All three pathologists are trained in genitourinary pathology with at least ten years of work experience. Their task was to subclassify biopsy samples into one of the following tumor types: clear cell RCC (ccRCC), chromophobe RCC (chRCC), papillary RCC (pRCC), urothelial carcinoma, collecting duct carcinoma, neuroendocrine tumor, renal oncocytoma, and angiomyolipoma. Furthermore, they were asked to identify the ccRCC grade according to the Fuhrman grading system.

Samples without tumor patterns were classified as non-diagnostic, whereas samples in which the pathologist could not decide between malignant or benign were classified as nonconclusive.

Statistical analysis

The diagnostic accuracy was calculated for each pathologist. The results obtained with the index test were compared with those of the reference standard, which was the complete surgical specimen. Analysis of the diagnostic accuracy included assessment of the following measures: sensitivity/specificity, positive predictive value (PPV), and negative predictive value. For each measure, 95% confidence intervals (CIs) were calculated. Additionally, overall accuracy was calculated by the sum of correctly scored core biopsies. Since there were four possible results of the index test (nondiagnostic, nonconclusive, malignant tumor, and benign tumor), diagnostic accuracy was calculated in two different ways: with and without exclusion of nondiagnostic results from the index test. The diagnostic accuracy to classify a malignant or benign tumor was calculated by excluding nondiagnostic samples.

The generalized kappa was calculated to measure the agreement between the three pathologists in classification of subtypes of renal tumors, and Krippendorff's alpha coefficient was used to measure agreement in the ccRCC grade (interobserver variability). The generalized kappa and Krippendorff's alpha coefficient were calculated by excluding the nondiagnostic results. The following interpretation of agreement was used: fair, 0.00-0.20; moderate, 0.21-0.45; substantial, 0.46-0.75; almost perfect, 0.76-0.99; and perfect, 1.00 [(9)]. Negative value indicates nonstoichastic

agreement. An unpaired (two-sample) t-test was performed to evaluate differences between means. Statistica software, version 13.5 (StatSoft, Inc., Tulsa, OK) was used for all statistical analyses. A p-value <0.05 was considered significant and all p-values were two-sided.

RESULTS

Nondiagnostic biopsy rates ranged from 13% to 22%. Seven sets of cores were recognized as nondiagnostic by all pathologists, of which, six were derived from nephrectomy specimens and one from nephron-sparing surgery of multi-cystic RCC lesions. The mean tumor size of diagnostic and nondiagnostic CBs (for at least one pathologist) was 44.6 mm (SD ±22.5) and 40.6 mm (SD ±17.5), respectively. No differences between the groups were observed (p=0.380). There were no nonconclusive samples. The summary of the scoring results of nondiagnostic, nonconclusive, correctly and incorrectly scored CBs, and overall accuracy of the three pathologists is presented in Table 1.

The diagnostic accuracy of renal core biopsies, calculated by excluding nondiagnostic results, was high in the assessments performed by all pathologists. Sensitivity and specificity ranged 83-97% and 97-99%, respectively. High diagnostic accuracy was also estimated for malignant tumors (sensitivity 74-86%, and specificity 100%). All the above-mentioned measures had narrow 95% CIs. The lowest diagnostic accuracy was calculated for benign tumors, with sensitivity ranging 66.7-83.3% and specificity ranging 88.5-100% and wide 95%CIs. Correspondingly, PPV for benign tumors varied across pathologists and estimated 95%Cls were wide (Table 2, 3).

Malignant tumors dominated in the analyzed populations (93%). In addition, ccRCC was the most representative group (74 cases). The concordance between surgical specimen and biopsy for ccRCC ranged between 75% and 87%. In two cases, ccRCC was mistaken as a benign tumor in biopsy. Further, 100% concordance with biopsy results was found for RO and UCC. Perfect interobserver agreement was estimated for AML and UCC, whereas only fair agreement was estimated for CDC and cRCC (Table 4, 5).

Distribution of the ccRCC grade in the Fuhrman grading system was 23% (Grade 1), 66.2% (Grade 2), 5.4% (Grade 3), and 5.4% (Grade 4). The concordance between assessment of surgical specimen and biopsy in the Fuhrman grading system ranged 36.5-77.0%, respectively. Interobserver agreement between the three pathologists was substantial or moderate, depending on the subtype (Table 5). The Krippendorff's alpha coefficient, calculated by excluding the nondiagnostic results, was 0.28 for the Fuhrman grading system.

DISCUSSION

RMB plays a pivotal role in the active surveillance of renal tumors. Proper assessment of biopsy cores is crucial in the final decision making. The main goal of this study was to assess the histopathological efficacy of RMB and to check the concordance between pathological results and biopsy of the final specimen, as well as interobserver variability in the assessment of biopsy cores.

The number of nondiagnostic biopsy results (13-22%) in the current study is comparable with other series (10-20%) [(10)].

Meta-analysis provided the highest level of evidence available on RMB performance [11, 12]. Although the biopsies were performed after the resection of the specimen, we expected higher diagnostic yield. In a similar study with a lower number of cases by Kummerlin et al., nondiagnostic biopsy rate ranged from 8-16% [13]. The reason for this might be the performance of biopsies by a few different surgeons. Inconclusive results of the biopsies do not exclude further repeat RMBs. Diagnostic yield of secondary RMB may reach up to 83% [14].

The most significant role of RMB is to differentiate malignant tumors from benign lesions. Including only diagnostic cores, sensitivity and specificity in diagnosing malignancy were similar to those reported in a large meta-analysis by Marconi et al. in which sensitivity and specificity reached 99.1% and 99.7%, respectively. However, direct comparison of these two studies is not possible as that meta-analysis mentioned excluded studies with *ex vivo* biopsies [12].

Currently, the largest study on diagnostic accuracy of "in bench" biopsies was published in 2007. Sensitivity ranged between 79-91% and specificity was 100% in malignancy diagnosis. This analysis focused on interobserver variability in tumor subtyping, which ranged from substantial to almost perfect. However, it did not include assessment of interobserver variability in tumor grade based on biopsy cores. To our knowledge, our study is the first to evaluate this issue. In real life situations, decisions regarding introducing active surveillance are not only based on diagnosing malignancy. The crucial issue is also proper assessment of the tumor grade. Interobserver agreement in tumor grade was moderate and substantial. Therefore, in our opinion, the final diagnosis should be provided by the team of pathologists rather than an individual one [13].

In our study, three cases of chromophobe carcinoma were erroneously diagnosed by two pathologists as oncocytoma based on biopsy cores. The

diagnostic challenge of differentiating low grade chromophobe and hybrid oncocytoma-chromophobe RCCs from oncocytic lesions is well known. However, using additional immunohistochemical staining limits this problem. Moreover, the course of disease in low grade chromophobe and hybrid oncocytoma-chromophobe RCCs is rather benign.

Study limitations

First of all, the biopsies were performed "in bench" therefore the study does not reflect real life conditions. Study material was collected prospectively irrespective of tumor size and imaging suspicion of tumor type. Consequently, it does not reflect the biopsy potential within active surveillance setting. Moreover, operations and postoperative biopsies were performed by several different surgeons, which may justify lower than expected diagnostic yield.

CONCLUSIONS

The agreement regarding grading of biopsies between three pathologists ranged from moderate to substantial. Therefore, a team of dedicated uropathologists should be engaged in final diagnosis of renal mass biopsy rather than single one before implementing the proper treatment, especially active surveillance. Further analysis of larger cohort of cases should be performed to confirm our results.

CONFLICT OF INTEREST

There is no conflict of interest.

REFERENCES

(1) Ljungberg B, Campbell SC, Choi HY, Jacqmin D, Lee JE, Weikert S, et al. The epidemiology of renal cell carcinoma. Eur Urol. 2011;60:615-21.

(2) Corcoran AT, Russo P, Lowrance WT, Asnis-Alibozek A, Libertino JA, Pryma DA, et al. A review of contemporary data on surgically resected renal masses-benign or malignant? Urology. 2013;81:707-13.

(3) Flum AS, Hamoui N, Said MA, Yang XJ, Casalino DD, McGuire BB, et al. Update on the diagnosis and management of renal angiomyolipoma. J Urol. 2016;195(4P1):834-46.

(4) Choi JE, You JH, Kim DK, Rha KH, Lee SH. Comparison of perioperative outcomes between robotic and laparoscopic partial nephrectomy: a systematic review and meta-analysis. Eur Urol. 2015;67:891-01.

(5) Ljungberg B, Bensalah K, Canfield S, Dabestani S, Hofmann F, Kuczyk MA, et al. EAU guidelines on renal cell carcinoma: 2014 update. Eur Urol. 2015;67:913–24.

(6) Pierorazio PM, Johnson MH, Ball MW, Gorin MA, Trock BJ, Chang P, et al. Five-year analysis of a multi-institutional prospective clinical trial of delayed intervention and surveillance for small renal masses: the DISSRM registry. Eur Urol. 2015;68:408-15.

(7) Schoots IG, Zaccai K, Hunink MG, Verhagen PCMS. Bosniak classification for complex renal cysts reevaluated: a systematic review. J Urol. 2017;198:12-21.

(8) Neuzillet Y, Lechevallier E, Andre M, Daniel L, Coulange C. Accuracy and clinical role of fine needle percutaneous biopsy with computerized tomography guidance of small (less than 4.0 cm) renal masses. J Urol. 2014;171(5):1802-05.

(9) Munoz SR, Bangdiwala SI. Interpretation of kappa and b statistics measures of agreement. J Appl Statistics. 1997;24:105-11.

(10) Patel HD, Johnson MH, Pierorazio PM, Sozio SM, Sharma R, Iyoha E, et al. Diagnostic accuracy and risks of biopsy in the diagnosis of a renal mass suspicious for localized renal cell carcinoma: systematic review of the literature. J Urol. 2016;195:1340-7.

(11) Richard PO, Jewett MA, Tanguay S, Saarela O, Liu ZA, Pouliot F, et al. Safety, reliability and accuracy of small renal tumour biopsies: results from a multi-institution registry. BJU Int. 2016;119:543-9.

(12) Marconi L, Dabestani S, Lam TB, Hofmann F, Stewart F, Norrie J, et al. Systematic review and meta-analysis of diagnostic accuracy of percutaneous renal tumour biopsy. Eur Urol. 2016;69:660-73.

(13) Kummerlin I, ten Kate F, Smedts F, Horn T, Algaba F, Trias I, et al. Core biopsies of renal tumors: a study on diagnostic accuracy, interobserver, and intraobserver variability. Eur Urol. 2008; 53(6):1219-27.

(14) Leveridge MJ, Finelli A, Kachura JR, Evans R, Chung H, Shiff DA, et al. Outcomes of small renal mass needle core biopsy, nondiagnostic percutaneous biopsy, and the role of repeat biopsy. Eur Urol. 2011;60:578-84.

Corresponding author:

MD, PhD Tomasz Ząbkowski, Department o Urology, Military Institute of Medicine, Warsaw, Poland, e-mail: urodent@wp.pl, phone number: 0048 791 533 555

Table 1. Diagnostic accuracy of renal core biopsies for the individual pathologists, calculated by excluding nondiagnostic results and nonconclusive results

Diagnostic accuracy of renal core biopsies for the individual pathologists, calculated by excluding nondiagnostic results and nonconclusive results									
	1 – Pathologist 1			TD – Pathologist 2			MP – Pathologist 3		
	Estimated Value	Lower Limit	Upper Limit	Estimated Value	Lower Limit	Upper Limit	Estimated Value	Lower Limit	Upper Limit
Sensitivity (%)	97.7%	91.1%	99.6%	83.3%	72.8%	90.5%	85.5%	75.7%	92.0%
Specificity (%)	99.7%	98.7%	99.9%	97.6%	95.6%	98.7%	97.8%	96.0%	98.8%
PPV (%)	97.7%	91.1%	99.6%	85.5%	75.1%	92.2%	86.6%	76.8%	92.8%
NPV (%)	99.7%	98.7%	99.9%	97.2%	95.1%	98.4%	97.6%	95.7%	98.7%

Table 2. Diagnostic accuracy of renal core biopsies to classify a malignant tumor for the individual pathologists, calculated by including the nondiagnostic results

Diagnostic accuracy of renal core biopsies to classify a malignant tumor for the individual pathologists, calculated by including the nondiagnostic results									
	1 – Pathologist 1			TD – Pathologist 2			MP – Pathologist 3		
	Estimated Value	Lower Limit	Upper Limit	Estimated Value	Lower Limit	Upper Limit	Estimated Value	Lower Limit	Upper Limit
Sensitivity (%)	86.2%	77.1%	92.1%	74.5%	64.2%	82.6%	79.8%	70.0%	87.1%
Specificity (%)	100.0%	51.7%	100.0%	100.0%	51.7%	100.0%	100.0%	51.7%	100.0%
PPV (%)	100.0%	94.4%	100.0%	100.0%	93.5%	100.0%	100.0%	93.9%	100.0%
NPV (%)	31.6%	13.6%	56.5%	20.0%	8.4%	39.1%	24.0%	10.2%	45.5%

Table 3. Diagnostic accuracy of renal core biopsies to classify a benign tumor for the individual pathologists, calculated by including the nondiagnostic results

Diagnostic accuracy of renal core biopsies to classify a benign tumor for the individual pathologists, calculated by including the nondiagnostic results									
	1 – Pathologist 1			TD – Pathologist 2			MP – Pathologist 3		
	Estimated Value	Lower Limit	Upper Limit	Estimated Value	Lower Limit	Upper Limit	Estimated Value	Lower Limit	Upper Limit
Sensitivity (%)	83.3%	36.5%	99.1%	66.7%	24.1%	94.0%	66.7%	24.1%	94.0%
Specificity (%)	100.0%	95.1%	100.0%	96.8%	90.3%	99.2%	95.7%	88.8%	98.6%
PPV (%)	100.0%	46.3%	100.0%	57.1%	20.2%	88.2%	50.0%	17.4%	82.5%
NPV (%)	98.9%	93.4%	99.9%	97.8%	91.7%	99.6%	97.8%	91.6%	99.6%

Table 4. Concordance between the surgical specimen and renal core biopsies for the individual pathologists (%)

Concordance between the surgical specimen and renal core biopsies for the individual pathologists (%)			
	1 – Pathologist 1	TD – Pathologist 2	MP – Pathologist 3
RCC (74)	87.7	75.7	81.1
pRCC (10)	60.0	30.0	40.0
cRCC (5)	80.0	0.0	0.0
RO (3)	100.0	100,0	100.0
XGO (1)	0.0	0.0	0.0
AML (2)	50.0	50.0	50.0
UCC (2)	100.0	100.0	100.0
CDC (1)	100.0	0.0	0.0
NET (2)	100.0	0.0	50.0

Table 5. Interobserver variability for the renal subtypes

Interobserver variability for the renal subtypes	
RCC	0.6
.pRCC	0.5
cRCC	0.1
RO	0.7
unRCC	0.5
AML	1.0
UCC	1.0
CDC	0.0
NET	0.3