

**Inter-observer and intra-observer agreement of Bosniak classification of cystic renal masses: Comparison between original version to version 2019 and effect of an online support calculator**

Heba Osman, MD<sup>1</sup>; Jin Hui Yan, MD<sup>1</sup>; Jason Chan, MBChB<sup>1</sup>; Javeria Munir, MD<sup>1</sup>; Sumaya Alrasheed, MD<sup>1</sup>; Satheesh Krishna, MD<sup>2</sup>; Nicola Schieda, MD<sup>1</sup>

<sup>1</sup>Department of Medical Imaging, The Ottawa Hospital, University of Ottawa, Ottawa, ON, Canada; <sup>2</sup>Joint Department of Medical Imaging, University Health Network, University of Toronto, Toronto, ON, Canada

**Cite as:** Osman H, Yan JH, Chan J, et al. Inter-observer and intra-observer agreement of Bosniak classification of cystic renal masses: Comparison between original version to version 2019 and effect of an online support calculator. *Can Urol Assoc J* 2021 June 22; Epub ahead of print. <http://dx.doi.org/10.5489/cuaj.7369>

Published online June 22, 2021

**Correspondence:** Dr. Nicola Schieda, Department of Medical Imaging, The Ottawa Hospital, Ottawa, ON, Canada; [nschieda@toh.ca](mailto:nschieda@toh.ca)

\*\*\*

**Introduction**

The Bosniak classification of cystic renal masses version 2019 (Bosniak.v2019) presents a major revision to the original Bosniak classification (Bosniak.original). Among many changes, Bosniak.v2019 aims to improve inter-observer agreement compared to Bosniak.original through rigorous definitions of imaging features and Bosniak classes (1). Preliminary studies comparing inter-observer agreement between the two classification systems published to date have shown either no difference or only modest improvement (2-4). Recently published support aids for use of Bosniak.v2019, including diagrammatic and online decision support aim to simplify use of Bosniak.v2019 (5) but the impact of these decision support aids have not yet been determined. The purpose of this study was to compare inter-observer and intra-observer agreement of Bosniak.v2019 to Bosniak.original and the effect of decision support aids on inter-observer agreement using an online support calculator developed for Bosniak.v2019.

## Methods

This IRB approved cross-sectional study evaluated 103 consecutively identified cystic ( $\leq 25\%$  enhancing (1)) masses: 70.9% (73/103) with pathological confirmation (80.8% [59/73] malignant, 19.2% [14/73] benign), and 29.1% (30/103) Bosniak 2/2F masses classified by an expert genitourinary radiologist each with minimum 5 years stability on follow-up (1); the latter group added to balance distribution of cystic masses among the Bosniak categories, Figure 1. Cystic masses were identified through a PACS and pathology database search for the term ‘Bosniak’ between 2010-2020. All masses were imaged with renal protocol CT or MRI. For patients with both CT and MRI, MRI was evaluated since MRI is known to result in a higher Bosniak class (6). A portion of the study population has been evaluated previously [BLINDED]; however, the purpose of specifically comparing agreement between the two systems over time with and without use of a support calculator has not been previously studied.

Three blinded fellowship-trained abdominal radiologists (R2-R4, years of experience 5-8) independently assigned Bosniak.original Class first and then Bosniak.v2019 Class second during one session. After a  $\geq 2$ -month washout period, R2-R4 re-assigned Bosniak.original and Bosniak.v2019 Class during a second session. Before the second session, radiologists were provided with a pictorial review of the Bosniak.v2019 system, decision support flow diagrams (5) and an on-line support calculator developed for Bosniak.v2019 (bosniak-calculator.herokuapp.com). The radiologists jointly trialed the calculator using 5 test cases in conjunction with R1. Inter-observer and intra-observer agreement were calculated using Cohen’s Kappa and agreement between Bosniak.original and Bosniak.v2019 was also calculated for both sessions. Kappa values of 0-2.0 indicated slight agreement, 2.1-4.0 fair agreement, 4.1-6.0 moderate agreement, 6.1-8.0 substantial agreement and 8.1-1.0 almost perfect agreement.

## Results

A summary of inter-observer and intra-observer agreement, using Bosniak.original and Bosniak.v2019 is provided in Table 1. In the first session, marginally higher moderate inter-observer agreement was noted comparing Bosniak.v2019 ( $K=0.44$ ) to Bosniak.original ( $K=0.39$ ). There was no change in inter-observer agreement comparing Bosniak.v2019 classification with ( $K=0.42$ ) or without the on-line calculator ( $K=0.44$ ). Comparatively, inter-observer agreement for Bosniak.original decreased in the second session ( $K=0.25$ ). Within reader intra-observer agreement was similar for both Bosniak.orginal and Bosniak.v2019 and within reader level of agreement comparing class designation using Bosniak.original and Bosniak.v2019 was also similar for both sessions, Table 2.

## Discussion

Bosniak.v2019 presents a major revision to the original Bosniak classification and aims, through rigorously defined imaging features and classes, to improve upon inter-observer agreement which is a known limitation of Bosniak.original (1). Studies to date have either shown only a modest improvement in inter-observer agreement using Bosniak.v2019 or no difference (2-4). A challenge of Bosniak.v2019, compared to Bosniak.original is new definitions for existing imaging features which theoretically should improve inter-observer agreement but may be difficult for inexperienced users. On-line decision support tools are now available for many standardized radiological reporting systems. These calculators aim to improve uptake, ease and fidelity of use or inter-observer agreement of reporting systems. We aimed to compare agreement between Bosniak.v2019 and Bosniak.original and explore the impact of decision support aids.

Our results indicate that Bosniak.v2019 has higher inter-observer agreement compared to Bosniak.original, though agreement was at best moderate. The difference was more pronounced during the second session, where inter-observer agreement for Bosniak.original dropped substantially. The observation of improved stability in inter-observer agreement when comparing Bosniak.v2019 to Bosniak.original is, to our knowledge, previously unreported. In our study, the use of an on-line calculator developed for Bosniak.v2019 did not improve inter-observer agreement among three radiologists.

Studies evaluating the value of support calculators for radiology classification systems are lacking, despite their proliferation and promotion by various groups. One study evaluating PI-RADS scoring with and without a calculator, reported increased speed of reporting with no difference in diagnostic accuracy or inter-observer agreement (7). We similarly showed no improvement in Bosniak.v2019 agreement when using a calculator. It is possible the calculator did not improve agreement in our study because its' intended use is for inexperienced readers unfamiliar with Bosniak.v2019. Thus, the calculator may have limited impact on radiologists in our study, since they had already used Bosniak.v2019 in the first interpretation session. In our study, inter-observer agreement for Bosniak.original classification decreased substantially between sessions, which differed from Bosniak.v2019 where inter-observer agreement was stable. In our study, intra-observer agreement was similar for both systems within readers and the degree of agreement between the two systems did not substantially change at either session for any reader. A moderate level of agreement using Bosniak classification has been reported previously using both the original and v2019 systems (3, 8). This level of agreement is not optimal; however, implications on clinical care are not directly correlated to agreement. For example, Yan et al. recently demonstrated that disagreement in Bosniak Classification occurs mainly in Class 2 versus 2F and 2F versus 3 masses (9). Since Class 2, 2F and increasingly many Class 3 masses are managed with surveillance in modern urological practices, disagreement in Bosniak Class may not therefore directly impact management (10).

This study is limited by the combined use of both CT and MRI and assignment of Bosniak.original before Bosniak.v2019 class during the same session, which may bias the outcome towards higher Bosniak.v2019 agreement. Our study population included an increased number of pathologically proven cystic renal masses, which is not reflective of the true prevalence of cystic masses in clinical practice; a consequence of a retrospective sample of cystic masses with adequate imaging reference standard.

### Conclusions

Bosniak.v2019 inter-observer agreement is improved compared to Bosniak.original and levels of inter-observer agreement are more stable with Bosniak.v2019 over serial interpretation sessions but do not improve when using an on-line support calculator.

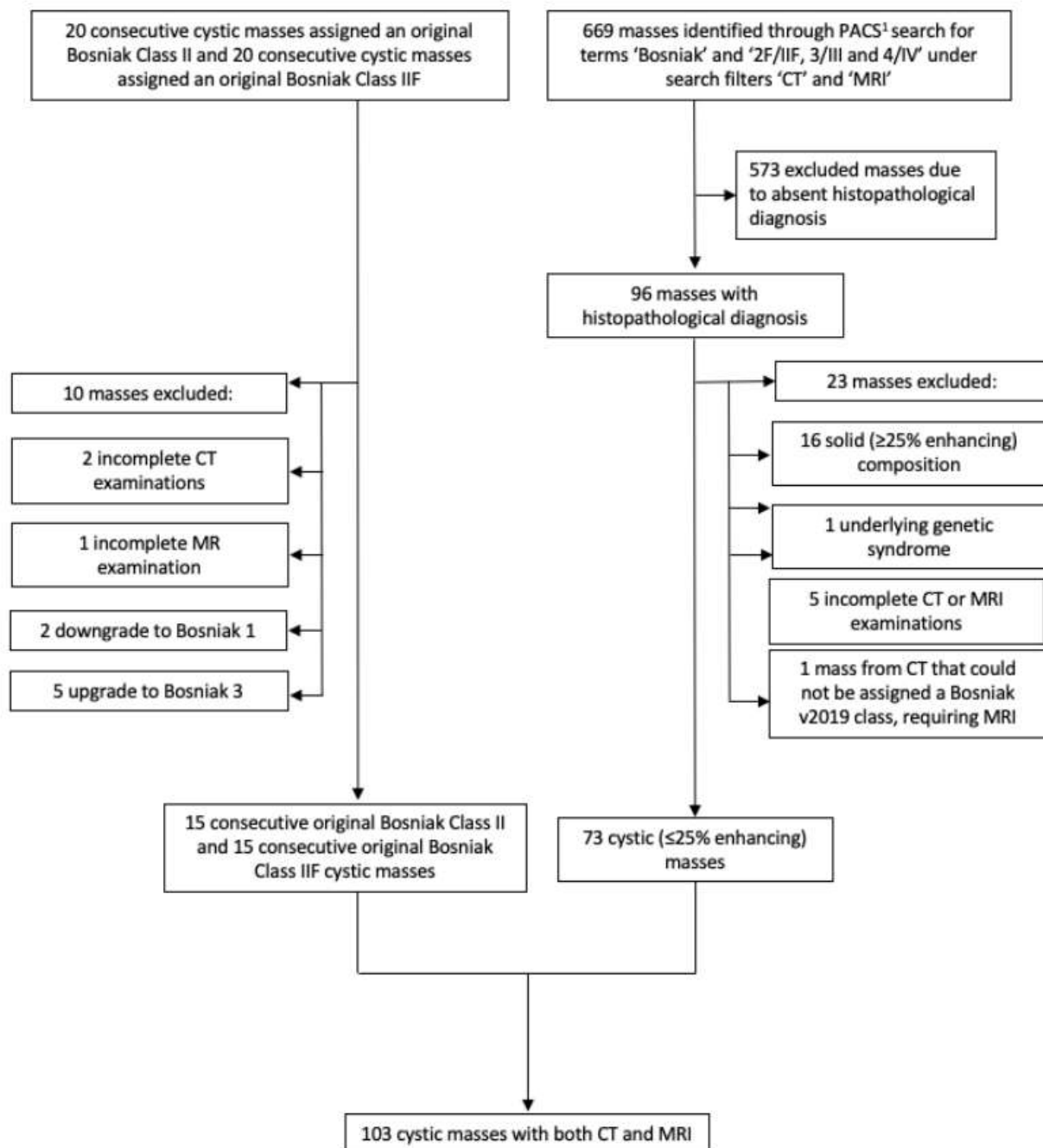
DRAFT

## References

1. Silverman SG, Pedrosa I, Ellis JH, et al. Bosniak Classification of Cystic Renal Masses, Version 2019: An Update Proposal and Needs Assessment. *Radiology*. 2019;182646.
2. Tse JR, Shen J, Yoon L, Kamaya A. Bosniak Classification Version 2019 of Cystic Renal Masses Assessed With MRI. *AJR Am J Roentgenol*. 2020;215(2):413-9.
3. Bai X, Sun SM, Xu W, et al. MRI-based Bosniak Classification of Cystic Renal Masses, Version 2019: Interobserver Agreement, Impact of Readers' Experience, and Diagnostic Performance. *Radiology*. 2020:200478.
4. Pacheco EO, Torres US, Alves AMA, Bekhor D, D'Ippolito G. Bosniak classification of cystic renal masses version 2019 does not increase the interobserver agreement or the proportion of masses categorized into lower Bosniak classes for non-subspecialized readers on CT or MR. *Eur J Radiol*. 2020;131:109270.
5. Schieda N, Davenport MS, Krishna S, et al. Bosniak Classification of Cystic Renal Masses, Version 2019: A Pictorial Guide to Clinical Use. *Radiographics*. 2021:200160.
6. Israel GM, Hindman N, Bosniak MA. Evaluation of cystic renal masses: comparison of CT and MR imaging by using the Bosniak classification system. *Radiology*. 2004;231(2):365-71.
7. Barth BK, Martini K, Skawran SM, et al. Value of an online PI-RADS v2.1 score calculator for assessment of prostate MRI. *Eur J Radiol Open*. 2021;8:100332.
8. Tse JR, Shen J, Shen L, Yoon L, Kamaya A. Bosniak Classification of Cystic Renal Masses Version 2019: Comparison of Categorization using CT and MRI. *AJR Am J Roentgenol*. 2020.
9. Yan JH, Chan J, Osman H, et al. Bosniak Classification version 2019: validation and comparison to original classification in pathologically confirmed cystic masses. *Eur Radiol*. 2021.
10. Couture F, Finelli A, Tetu A, et al. Management of complex renal cysts in Canada: results of a survey study. *BMC Urol*. 2020;20(1):47.

## Figures and Tables

**Fig. 1.** Flow diagram of patient inclusion for the current study. CT: computed tomography; MRI: magnetic resonance imaging.



<b>Table 1. Inter-observer and intra-observer agreement (Cohen's Kappa value) for Bosniak class assignment in 103 cystic renal masses imaged with CT and MRI between three radiologists who evaluated each mass independently using the original Bosniak and Bosniak version 2019 classification without and with a Bosniak version 2019 digital calculator support aid</b>				
<b>Inter-observer agreement</b>				
	<b>Interpretation Session #1</b>		<b>Interpretation Session #2 (Using on-line calculator for Bosniak v2019)</b>	
	<b>Original Bosniak Classification</b>	<b>Bosniak v2019 Classification</b>	<b>Original Bosniak Classification</b>	<b>Bosniak v2019 Classification</b>
Reader 2 vs. Reader 3	0.40	0.42	0.20	0.41
Reader 2 vs. Reader 4	0.42	0.48	0.34	0.40
Reader 3 vs. Reader 4	0.39	0.43	0.27	0.44
Overall	<b>0.39</b>	<b>0.44</b>	<b>0.25</b>	<b>0.42</b>
<b>Intra-observer agreement</b>				
	<b>Original Bosniak Classification</b>		<b>Bosniak v2019 Classification</b>	
Reader 2	0.43		0.42	
Reader 3	0.38		0.38	
Reader 4	0.66		0.55	

CT: computed tomography; MRI: magnetic resonance imaging.

<b>Table 2. Agreement (Cohen's Kappa value) for Bosniak class assignment in 103 cystic renal masses imaged with CT and MRI within three radiologists comparing class assigned using the original Bosniak and Bosniak version 2019 classification systems</b>		
	<b>Interpretation Session #1</b>	<b>Interpretation Session #2</b>
<b>Reader 2</b>	0.79	0.71
<b>Reader 3</b>	0.43	0.42
<b>Reader 4</b>	0.72	0.68

CT: computed tomography; MRI: magnetic resonance imaging.