

Comparación entre los scores de Guy y S.T.O.N.E. para clasificar la litiasis renal ¿Es determinante la experiencia del evaluador?

Comparison between Guy's and S.T.O.N.E. scores to classify kidney stones. Is the experience of the evaluator decisive?

Leandro Padial Tagliapietra, Horacio Sanguinetti, Lucas Yirula, Jorge Aguilar, Raúl Montiel, Nicolás Rosenfeld, Norberto Bernardo

Servicio de Urología. Hospital de Clínicas "José de San Martín", Buenos Aires Argentina

RESUMEN

Introducción: Los sistemas de clasificación más utilizados en litiasis renal son el de Guy y el S.T.O.N.E. Ambos fueron validados estadísticamente y se basan en hallazgos de imágenes pre operatorios, sin embargo, todos tienen algún grado de dificultad en su aplicación.

Objetivo: El objetivo principal fue evaluar el grado de concordancia de los scores Guy y S.T.O.N.E. realizado por un grupo de residentes de urología, urólogos no especialistas en litiasis y urólogos de la sección Endourología y como segundo objetivo conocer la preferencia del uso de los scores por los evaluadores.

Materiales y método: Se realizó un estudio observacional y transversal. Se incluyeron 25 de casos con diagnóstico de litiasis renal. Todos fueron clasificados según el score de Guy y S.T.O.N.E. por tres grupos de médicos, el de referencia, por especialistas en Endourología, urólogos de otras subespecialidades y un tercero compuesto por residentes de urología.

Resultados: Para el score de Guy realizado por los urólogos no especialistas la concordancia que la concordancia del Modo fue del 44% (11 casos) y los médicos residentes tuvieron una concordancia del 56% (14 casos). Para el S.T.O.N.E, el grupo de urólogos no especialistas en Endourología tuvieron una concordancia en 22 casos (88%) y el grupo de residentes, en 21 (84%). El 95% de los médicos refirió que el Score de Guy le resultó más sencillo de realizar.

Conclusión: La concordancia entre evaluadores especialistas en Endourología, urólogos no especialistas y residentes de urología en los Scores de Guy y S.T.O.N.E. es baja. Los residentes tienen mejor concordancia que los urólogos no especialistas en el score de Guy, pero no en el de S.T.O.N.E. El Score de Guy es el más elegido entre los médicos no especialistas.

ABSTRACT

Introduction: The most widely used classification systems in kidney stones are Guy's and S.T.O.N.E. Both were statistically validated and are based on pre-operative imaging findings, however, all have some degree of difficulty in their application.

Objective: The main objective was to evaluate the degree of concordance of the Guy and S.T.O.N.E scores. carried out by a group of urology residents, urologists not specialists in lithiasis and urologists from the Endourology section and as a second objective to know the preference of the use of the scores by the evaluators.

Materials and method: An observational and cross-sectional study was carried out. Twenty-five cases with a diagnosis of kidney stones were included. All were classified according to Guy's score and S.T.O.N.E. by three groups of doctors, the one of reference, by specialists in Endourology, urologists of other subspecialties and a third made up of urology residents.

Results: For the Guy's score performed by non-specialist urologists, the concordance that the Mode concordance was 44% (11 cases) and the resident physicians had a 56% concordance (14 cases). For the S.T.O.N.E, the group of urologists not specialists in Endourology had a concordance in 22 cases (88%) and the group of residents, in 21 (84%). 95% of the doctors reported that the Guy Score was easier for them to perform.

Conclusion: The concordance between endourology specialist evaluators, non-specialist urologists and urology residents in Guy's Scores and S.T.O.N.E. it is low. Residents have better agreement than non-specialist urologists in the Guy's score, but not in the S.T.O.N.E. Guy's Score is the most chosen among non-specialist physicians.

INTRODUCCIÓN

La litiasis renal es una enfermedad con alta tasa de recidiva cuya frecuencia aumentó en los últimos años. Los sistemas de clasificación sirven tanto en el ámbito clínico, donde nos asisten a la hora de tomar decisiones terapéuticas, como en el ámbito académico, donde nos permiten comparar resultados de trabajos. Para la litiasis renal existen varios scores validados entre los que sobresalen el de Guy y el S.T.O.N.E. Ambos se basan en hallazgos de imágenes pre operatorios^{1,2}.

Un método ideal para clasificar una patología debería ser rápido, simple y reproducible³. Sin embargo, todos tienen algún grado de dificultad en su aplicación, en particular, cuando son realizados por médicos no especializados.

El objetivo de este trabajo fue evaluar el grado de concordancia de los scores Guy y S.T.O.N.E. realizado por urólogos de la Sección Endourología con un grupo de médicos residentes de urología y otro grupo de urólogos no especialistas en litiasis. Secundariamente, se indagó cuál de los dos scores preferían utilizar los profesionales.

MATERIALES Y MÉTODO

Se realizó un estudio observacional y transversal en el Servicio de Urología del Hospital de Clínicas “José de San Martín”.

Se seleccionaron 25 casos clínicos con diagnóstico tomográfico de litiasis renal. Las tomografías fueron evaluadas por un grupo de médicos urólogos especialistas en Endourología quienes asignaron un valor, considerado de referencia, a los scores de Guy y de S.T.O.N.E. El Score de Guy evalúa alteraciones ana-

tómicas, ubicación y cantidad de litiasis (tabla 1) y el de S.T.O.N.E. evalúa el tamaño de la litiasis, longitud del tracto, obstrucción, números de cálices involucrados y la densidad de la litiasis (tabla 2). La escala incluye una puntuación de 5 a 13 donde una puntuación de 5 sería de baja complejidad, de 6 a 8 de mediana complejidad y de 9 a 13 de alta complejidad².

A dos grupos de médicos, uno de 6 residentes de urología y otro de 6 urólogos de secciones diferentes a Endourología (andrología, urodinamia, oncología y cirugía reconstructiva) se les proporcionó instrucciones estandarizadas sobre la aplicación de cada score, se les solicitó que evaluaran los 25 casos clínicos seleccionados y, finalmente, se les consultó cuál de los dos scores era de su preferencia.

Los resultados se presentan como mediana y rango para el score de STONE y número de observaciones para cada categoría del score de Guy. Para evaluar la concordancia de ambos scores con el valor de la referencia (especialista en Endourología) se calculó el modo (la respuesta más frecuente) de cada caso en ambos grupos (residentes y urólogos de secciones diferentes a Endourología). Se presenta el gráfico de Bland & Altman, la diferencia de medias con sus intervalos de confianza (IC) para el 95% y el coeficiente de correlación de la concordancia (CCC) de Lin para el score de STONE. Se informa la concordancia absoluta y el índice Kappa para el score de Guy y el score de STONE con 3 categorías.

RESULTADOS

Las 25 tomografías fueron clasificadas por los especialistas en Endourología quienes informaron para STONE una mediana de 7 (rango=5-11) y para

Tabla 1: Score de Guy. Identifica localización, número de litos en cálices renales y alteraciones anatómicas.

Grado		Descripción	
I	Lito en polo medio/inferior con anatomía normal	Lito solitario en pelvis con anatomía normal	
II	Lito solitario en polo superior con anatomía normal	Múltiples litos con anatomía normal	Lito con anatomía anormal
III	Múltiples litos con anatomía normal	Lito en divertículo calicial	Lito coraliforme incompleto
IV	Lito coraliforme completo	Lito en pacientes con espina bífida o lesión medular	

Tabla 2: Score S.T.O.N.E. Toma en cuenta 5 variables que se escalan en 4 niveles en litiasis renal.

Variable	1	2	3	4
Tamaño (mm ²)	0-399	400-799	800-1599	>1600
Tracto (mm)	<100	>100		
Obstrucción	Ausente	Presente		
Cálices	1-2	3	Litos coraliformes	
Densidad (UH)	<950	>950		

el score de Guy, el modo fue la categoría 1 con 9 casos (36%), categoría 2 con 8 (32%) casos, la categoría 3 con 5 (20%) casos y la categoría 4 con 3 (12%) casos. En la tabla 3 se muestran los resultados para el score de Guy, donde podemos observar que la concordancia del Modo de Staff es del 44% (11 casos) y los médicos residentes tuvieron una concordancia del 56% (14 casos).

Respecto al análisis del scores de STONE, el modo de los grupos con respecto al valor de referencia se observó que el grupo de urólogos no especialistas en endourología concordaron en 22 casos (88%) y el grupo de residentes, en 21 (84%). En la tabla 4 y 5 se presentan los resultados con el Score STONE.

Tabla 3. Resultados Score de Guy.

Evaluador	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Concordancia absoluta n (%)	Índice kappa
REFERENCIA	9	8	5	3	-	-
Staff 1	11	8	1	5	15 (60)	0,43
Staff 2	4	13	6	2	10 (40)	0,16
Staff 3	4	14	6	1	9 (36)	0,09
Staff 4	5	12	6	2	11 (44)	0,21
Staff 5	13	9	2	1	15 (60)	0,40
Staff 6	14	7	3	1	13 (52)	0,29
Modo Staff	5	12	6	2	11 (44)	0,21
Residente 1	12	7	3	3	14 (56)	0,37
Residente 2	11	7	5	2	20 (80)	0,71
Residente 3	10	2	8	5	11 (44)	0,24
Residente 4	13	6	4	2	16 (64)	0,48
Residente 5	8	9	4	4	18 (72)	0,61
Residente 6	14	7	2	2	15 (60)	0,41
Modo Residentes	14	5	3	3	14 (56)	0,36

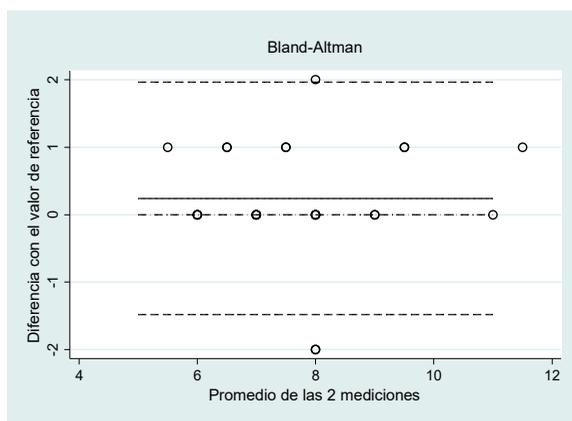
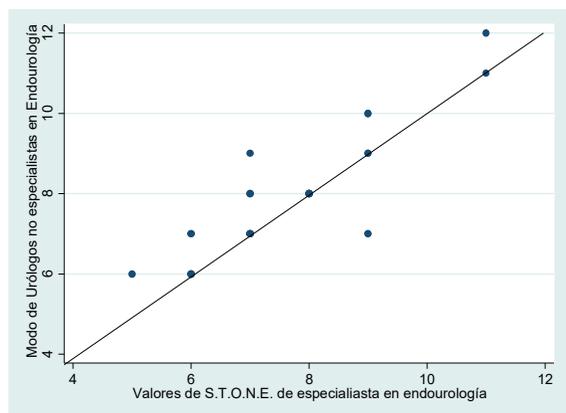
La figura 1 presenta la concordancia del grupo de urólogos no especialistas en endourología con los valores de referencia de los médicos especialistas para el Score S.T.O.N.E. La figura 2 presenta la concordancia del grupo de residentes de urología con los valores de referencia de los médicos especialistas para el Score S.T.O.N.E. Aunque tendieron a otorgar valores más altos, se observó que la concordancia fue aceptable, la mediana de la diferencia con el valor

de referencia fue cero y el rango de la diferencia fue similar en ambos grupos.

Se categorizo el score STONE en 3 grupos de acuerdo a la escala de clasificación en baja (5 puntos), mediana (6-8 puntos) y alta complejidad (9-13 puntos) y en el análisis estadístico se concluyó para el modo de los médicos de staff y residentes los mismos resultados para mediana (8) y rango (6-12) con un valor de media para los médicos no endourologos de

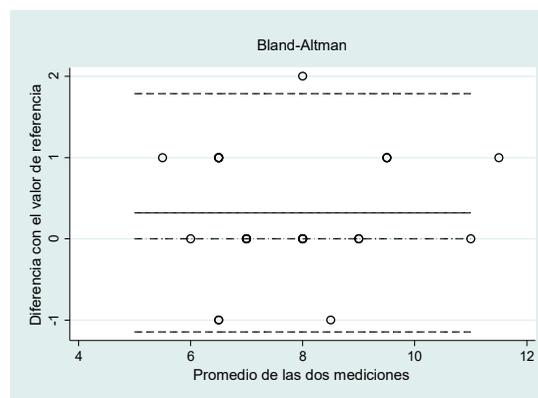
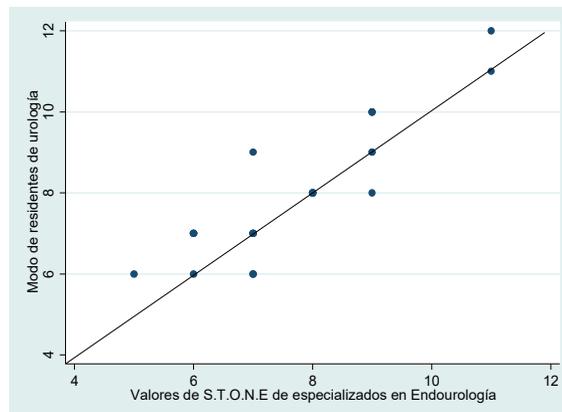
7,92 CCC= 0.831 y para los médicos residentes una media de 8 y CCC=0.871. El 95% de los médicos refirió que preferían realizar el Score de Guy.

Figura 1. Concordancia del modo en el grupo de Urólogos no especialistas en Endourología respecto al valor de referencia.



El score de Guy tiene en cuenta el número de cálculos involucrados, la anatomía anormal (renal o de la vía excretora) y los pacientes con espina bífida o lesión medular con litiasis renal. La escala incluye 4 grados, siendo el primero el de mejor tasa libre de cálculos^{1,8}.

Figura 2. Concordancia del modo en el grupo de Residentes de Urología respecto al valor de referencia.



DISCUSIÓN

Los sistemas de puntuación, en general, para cualquier patología son importantes ya que permiten planificar y prevenir complicaciones⁴. Un sistema óptimo debe ser rápido, simple, reproducible y fácil de implementar, y debe tener una buena correlación con la tasa libre de litiasis y las complicaciones¹. Estas herramientas ampliamente aplicables mejoran en gran medida la elección de los pacientes, la toma de decisiones clínicas, la evaluación de los resultados y el estudio académico de los mismos^{5,6}. Esto puede permitir comparaciones confiables y precisas de la seguridad y eficacia del tratamiento, y facilitar la comparación significativa de estudios clínicos⁷.

El score S.T.O.N.E. es un acrónimo que hace referencia a las cinco variables que se tienen en cuenta, (S) tamaño de la piedra, (T) tracto (distancia piel-litio), (O) obstrucción (presencia de uronefrosis), (N) número de cálices involucrados y (E) esencia de la piedra (densidad) medido en unidades Hounsfield². La escala incluye una puntuación de 5 a 13 donde una puntuación de 5 sería de baja complejidad, de 6 a 8 de mediana complejidad y de 9 a 13 de alta complejidad según la descripción del trabajo original⁹.

Ambas clasificaciones demostraron ser rápidas en su aplicación, simples y reproducibles y fueron validadas estadísticamente, estableciéndose una exce-

Tabla 4. resultados con el Score STONE.

Evaluador	Score STONE			Diferencia de medias respecto al valor de referencia	IC 95% de la diferencia de medias		CCC
	Mediana	Rango	Media		Límite inferior	Límite superior	
Referencia	7	5-11	7,68	-	-	-	-
Staff 1	7	4-10	7	-0,68	-1,50	0,14	0,2526
Staff 2	8	6-12	7,92	0,87	-0,12	0,60	0,8319
Staff 3	8	6-12	7,92	0,87	-0,12	0,60	0,8319
Staff 4	8	6-11	7,88	0,2	-0,13	0,53	0,8445
Staff 5	8	6-10	7,8	0,12	-0,41	0,65	0,6267
Staff 6	8	4-10	7,36	-0,32	-0,74	0,10	0,7669
Modo Staff	8	6-12	7,92	0,24	-0,12	0,60	0,8319
Residente 1	8	6-12	8,08	0,4	0,10	0,69	0,8773
Residente 2	8	6-11	8,04	0,36	-0,06	0,78	0,7411
Residente 3	9	6-12	8,6	0,92	0,50	1,33	0,7052
Residente 4	8	6-12	7,92	0,24	-0,10	0,58	0,8513
Residente 5	8	5-12	8	0,32	-0,13	0,77	0,7814
Residente 6	8	5-12	7,76	0,08	-0,36	0,52	0,7781
Modo Residentes	8	6-12	8	0,32	0,01	0,62	0,8717

*CCC=Coefficiente de correlación de concordancia de Lin.

lente correlación con distintas variables como la tasa libre de cálculos, las complicaciones y la pérdida sanguínea^{10, 11, 12, 13, 14}. Sin embargo, no existen trabajos publicados que evalúen la aplicación de estos scores en grupos de médicos no especializado.

Por otro lado, la variación interobservador es un tema ampliamente analizado en la literatura. En una publicación de Harbias y colaboradores se demuestra que la variación interobservador en la clasificación anatomopatológica de Gleason produce cambios significativos en el tratamiento que deberá recibir posteriormente el paciente¹⁵. Por su parte, Hong y colaboradores en su trabajo sobre interpretación de imágenes tomográficas en pacientes con cólico renal observan un menor acuerdo interobservador en tomografías de baja dosis^{15,16}.

En nuestro estudio hemos observado que existe una pobre concordancia entre el grupo de referencia, conformado por los médicos especialistas en Endourológica y los demás médicos para ambos scores. Sin embargo, esta concordancia fue mucho mayor para el score de S.T.O.N.E. lo cual lo hace más reproducible que el de Guy.

Al comparar los dos grupos estudiados, encontramos una diferencia en la concordancia entre los urólogos no especialistas en Endourológica y residentes para el Score de Guy a favor de los últimos. Es posible que esto se deba a que la comparación se hizo con un grupo de médicos que no utilizan el score en su práctica habitual, en cambio los residentes dentro de un sistema de formación continua, utilizan frecuentemente estos scores en la sección de Endourológica. Además, en los casos donde hubo discordancia, encontramos que los urólogos no especialistas tuvieron la tendencia a sobrevalorar las respuestas, mientras que los residentes la subvaloraron. Es posible que esto se deba a que los médicos en formación tiendan a subestimar los casos lo que hace subvalorar los mismos.

Es de notar que esta diferencia no se aprecia en el score S.T.O.N.E., lo que puede deberse a que esta clasificación cuenta con más variables que la de Guy, permitiendo que aumente la concordancia interobservador por un motivo estadístico. Esta situación avala al Score de S.T.O.N.E. como más reproducible al compararlo con el de Guy.

Tabla 5. Resultados Score de Guy.

Evaluador	Grupo 1	Grupo 2	Grupo 3	Concordancia absoluta n (%)	Índice kappa	-
REFERENCIA	1	22	2	-	-	-
Staff 1	3	20	2	20 (80)		0,29
Staff 2	0	22	3	23 (92)		0,62
Staff 3	0	22	3	23 (92)		0,62
Staff 4	0	22	3	23 (92)		0,62
Staff 5	0	21	4	20 (80)		0,19
Staff 6	2	20	3	21 (84)		0,43
Modo Staff	0	21	4	22 (88)		0,51
Residente 1	0	19	6	20 (80)		0,35
Residente 2	0	22	3	23 (92)		0,62
Residente 3	0	19	6	20 (80)		0,35
Residente 4	0	20	5	21 (84)		0,42
Residente 5	1	19	5	20 (80)		0,36
Residente 6	1	20	4	19 (76)		0,14
Modo Residentes		20	5	21 (84)		0,42

De manera interesante, al ser consultados sobre qué Clasificación preferían utilizar, la mayoría de los encuestados optaron por la de Guy. En un trabajo de Vicentini y colaboradores donde se analizó cuanto tardaba el mismo observador en realizar distintas clasificaciones, se observó que la de Guy era significativamente más rápida de hacer que la de STONE (27,5 vs 300,6 segundos)¹⁷. Es posible que, al tener menos variables y no requerir el uso de un software para el análisis de las imágenes, el Guy sea más fácil de administrar y más amigable, en parte por el mismo motivo que la hace menos reproducible.

Si bien ambos scores en sus trabajos originales reportan que el porcentaje de concordancia entre di-

ferentes evaluadores es alto^{1,18}, en nuestro trabajo no hemos podido corroborar esos resultados, por lo que creemos que la experiencia del evaluador es determinante en la aplicación de estos scores.

CONCLUSIÓN

La concordancia entre evaluadores especialistas en Endourología, urólogos no especialistas y residentes de urología en los Scores de Guy y S.T.O.N.E. es baja. Los residentes tienen mejor concordancia que los urólogos no especialistas en el score de Guy, pero no en el de S.T.O.N.E. El Score de Guy es el más elegido entre los médicos no especialistas.

BIBLIOGRAFÍA:

1. Thomas K, Smith NC, Hegarty N, et al. The Guy's stone score—grading the complexity of percutaneous nephrolithotomy procedures. *Urology*. 2011; 78(2):277-81.
2. Okhunov Z, Friedlander JI, George AK et al: S.T.O.N.E. nephrolithometry: novel surgical classification system for kidney calculi. *Urology* 2013; 81:1154-9.
3. J. Gómez-Sánchez et al. Resultados obtenidos con nefrolitotomía percutánea y su correlación con las escalas de Guy y S.T.O.N.E. en un hospital de la Ciudad de México. *Rev Mex Urol*. 2016; 76(4):203-206.
4. Muzaffer Akçay, Muhammed Tosun, Fatih Gevher, Senad Kalkan, Cevper Ersöz, Yunus Kayalı, Abdulkadir Tepeler. Comparison of Scoring Systems in Predicting Success of Percutaneous Nephrolithotomy. *Balkan Med J* 2019; 36:32-6.
5. Allen DOBT, Tiptaft R, Glass J. Defining the learning curve for percutaneous nephrolithotomy. *J Endourol*. 2005; 19(3):279-82.
6. de la Rosette JJ, Laguna MP, Rassweiler JJ, Conort P. Training in percutaneous nephrolithotomy—a critical review. *Eur Urol*. 2008; 54(5):994-1001.
7. Autorino RQG, Sio MD, Lima E, Quarto E, Damiano R, Oliviero R, Osorio L, Marcelo F, D'Armiento M. Fate of abstracts presented at the world congress of Endourology: are they followed by publication in peer-reviewed journals? *J Endourol*. 2006; 20(12):996-1001.
8. Vernez SLOZ, Motamedinia P, Bird V, Okeke Z, Smith A. Nephrolithometric scoring systems to predict outcomes of percutaneous Nephrolithotomy. *Rev Urol*. 2016; 18(1):15-27.
9. Withington J, Armitage J, Finch W, Wiseman O, Glass J, Burgess N. Assessment of stone complexity for PCNL: a systematic review of the literature, how best can we record stone complexity in PCNL? *J Endourol*. 2016; 30(1):13-23.
10. Labadie K, Okhunov Z, Akhavein A, Moreira DM, Moreno-Palacios J, Del Junco M, et al. Evaluation and comparison of urolithiasis scoring systems used in percutaneous kidney stone surgery. *J Urol* 2015; 193:154-9.
11. Sfoungaristos S, Gofrit ON, Pode D, Landau EH, Duvdevani M. Percutaneous nephrolithotomy for staghorn stones: Which nomogram can better predict postoperative outcomes? *World J Urol* 2016; 34:1163-8.
12. Mandal S, Goel A, Kathpalia R, Sankhwar S, Singh V, Sinha RJ, et al. Prospective evaluation of complications using the modified Clavien grading system, and of success rates of percutaneous nephrolithotomy using Guy's Stone Score: a single-center experience. *Indian J Urol* 2012; 28:392-8.
13. Vicentini FC, Marchini GS, Mazzucchi E, Claro JF, Srougi M. Utility of the Guy's stone score based on computed tomographic scan findings for predicting percutaneous nephrolithotomy outcomes. *Urology* 2014;83:1248-53.
14. Ingimarsson JP, Dagrosa LM, Hyams ES et al (2014) External validation of a preoperative renal stone grading system: reproducibility and inter-rater concordance of the Guy's stone score using preoperative computed tomography and rigorous postoperative stone-free criteria. *Urology* 83(1):45-49
15. Hong JY, et al. Inter-observer Agreement between Urologists and Radiologists in Interpreting the Computed Tomography Images of Emergency Patients with Renal Colic. *Urol J*. 2018 Mar 18;15(2):6-9.
16. Smith A, Averch TD, Shahrour K et al: A nephrolithometric nomogram to predict treatment success of percutaneous nephrolithotomy. *J Urol* 2013; 190: 149.
17. Vicentini FC, Serzedello FR, Thomas K et al. What is the quickest scoring system to predict percutaneous nephrolithotomy outcomes? A comparative study among S.T.O.N.E score, guy's stone score and croes nomogram. *Int Braz J Urol* Nov-Dec 2017;43(6):1102-1109.
18. Harbias A, et al. Implications of Observer Variation in Gleason Scoring of Prostate Cancer on Clinical Management: A Collaborative Audit. *Gulf J Oncolog*. 2017 Sep;1(25):41-45.