Recent scoring systems predicting stone-free status after retrograde intrarenal surgery; a systematic review and meta-analysis

Oktay Özman¹, Hacı Murat Akgül², Cem Başataç³, Eyüp Burak Sancak⁴, Önder Çınar⁵, Hakan Çakır⁶, Cenk Murat Yazıcı⁷, Haluk Akpınar⁸, Bülent Önal⁹, RIRSearch Study Group

¹Gaziosmanpasa Training and Research Hospital, Urology Clinic, Istanbul, Turkey
²Tekirdağ Namık Kemal University, School of Medicine, Department of Urology, Tekirdağ, Turkey
³Group Florence Nightingale Hospitals, Department of Urology, Istanbul, Turkey
⁴Çanakkale Onsekiz Mart University, School of Medicine, Department of Urology, Çanakkale, Turkey
⁵Zonguldak Bülent Ecevit University, School of Medicine, Department of Urology, Zonguldak, Turkey
⁶Fulya Acıbadem Hospital, Urology Clinic, Istanbul, Turkey
⁷Istanbul University-Cerrahpaşa, Cerrahpaşa Medical Faculty, Department of Urology, Istanbul, Turkey

Article history
Submitted: Oct. 8, 2021
Accepted: March 15, 2022
Published online: March 24, 2022

Introduction

Several scoring systems and nomograms have been developed to predict the success of retrograde intrarenal surgery. But no meta-analysis for the performance of scoring systems has yet been performed. The aim of this study was to compare predictive ability of recent scoring systems for stone-free rate of retrograde intrarenal surgery.

Material and methods

PubMed and Web of Science databases were searched systematically between April and May 2021. The scoring systems which were validated externally or studied at least by two different researchers groups were selected for further analysis. Of 59 records, 14 studies met the inclusion criteria (n = 4137). Area under curve (AUC) values of selected scoring systems were pooled in random or fixed effects. The I² test was used to quantify heterogeneity.

Results

Eight, 5, 4 and 3 studies included in meta-analyses for the modified Seoul National University Renal Stone Complexity Score (S-ReSC), R.I.R.S., Resorlu-Unsal Score (RUS), S.T.O.N.E., and Ito’s Nomogram, respectively. We found pooled AUC values 0.709 (95% CI 0.670–0.748), 0.704 (95% CI 0.668–0.739), 0.669 (95% CI 0.646 to 0.692), and 0.771 (95% CI 0.724 to 0.818), for first four of them, respectively. Heterogeneity was very high to pool AUC values for Ito’s nomogram.

Conclusions

Although S.T.O.N.E. score showed higher pooled AUC value, this systematic review and meta-analysis has not revealed superiority of any scoring system. High heterogeneity between studies and dependencies between scoring systems make it difficult to design a comparative statistical model to generalize the findings. Also, limitations aside, neither scoring system has demonstrated good predictive/discriminative performance.

Key Words: kidney stone † retrograde intrarenal surgery † scoring system † nomogram † stone

INTRODUCTION

Advances in surgical techniques and urological devices have made minimal invasive surgery for kidney stones more preferred in recent years. Shock wave lithotripsy, retrograde intrarenal surgery (RIRS), and percutaneous nephrolithotomy (PNL) have replaced open kidney stone surgery in many cases [1]. Which treatment should be chosen for which stone in the kidney is of critical importance. The key is
to use the right weapon at the right time. Retrograde intrarenal surgery stands out with its low complication rate and high stone-free rate (SFR), especially for stones up to 2 cm [2]. It is known that the success rate of RIRS depends on multiple factors, such as stone burden, number, localization, renal calyceal anatomy [3]. Several scoring systems and nomograms have been developed with these factors to predict the success of RIRS [3]. Emerging data suggest that scoring systems might provide a preoperative prediction for the outcome of RIRS but none of them gained popularity and were not widely used. Similar findings have been reported from different authors for several scoring systems [4, 5]. To date, no meta-analysis for the performance of scoring systems has yet been performed. We therefore aimed to compare predictive ability of the scoring systems for SFR of RIRS.

**MATERIAL AND METHODS**

**Search strategy and selection criteria of studies and scoring systems**

This systematic review and meta-analysis was conducted following the Preferred Reporting Items for Systematic Review and Meta-analysis (PRISMA) guidelines [6]. Study protocol was registered with PROSPERO (number: CRD42021252873). PubMed and Web of Science databases were searched systematically between April and May 2021. The search strategy was developed with keywords in accordance with Population Intervention Comparison Outcomes (PICO) strategy. All keywords and details of the search strategy can be found in Table 1–2. For each database, the search strategy, removal of duplicates, title-abstract screening were performed by 2 different authors independently (OÖ and ÖÇ). After the resolution of discrepancies by head-to-head meeting, the remaining articles were evaluated by OÖ based on full text. Full text versions of studies were requested from the authors if needed. Finally, reference lists of all full texts were reviewed for further relevant studies. The scoring systems which were validated externally or studied at least by two different researcher groups were selected for further analysis. First research strategy was validated by second research strategy including names of selected scoring systems in ‘Comparison’ components instead of general relevant keywords. But no additional articles were found. Only original studies which have area under curve (AUC) metadata of ROC curve analysis and written in English were included in the meta-analysis. Studies including patients who underwent bilateral RIRS or had a different simultaneous surgical procedure (for kidney stone or other indication) in the same session, and patients who underwent RIRS to remove encrusted ureteral stent or other foreign bodies were excluded. Also studies including only specific patient groups (pediatric, elderly etc.) or only patients with renal abnormalities (horseshoe kidney, solitary kidney, transplanted kidney etc.) were excluded.

**Quality assessment**

Included studies were assessed by two reviewers (HÇ and OÖ) independently. The risk of bias assessments were done according to modified version of the Quality In Prognostic Studies (QUIPS) tool’s six criteria; study participation, study attrition, prognostic factor measurement, outcome measurement, confounding, and statistical analysis and reporting [7]. Final decisions about discrepancies were made after a head-to-head discussion between the scorers. Surgical outcome evaluations by kidney, ureter, and bladder (KUB) imaging, ultrasonography (US), computed tomography (CT), or second-look flexible ureterorenoscopy were considered to be appropriate. Because of consensus lack on the exact diameter of significant residual fragments in the literature, any cut-off value was not took account during quality assessment of reported outcomes.

**Table 1. First research strategy before selection of candidate scoring systems for review**

| P (Population) | I (Intervention) | C (Comparison) | O (Outcome) |
|---------------|-----------------|---------------|-------------|
| kidney stone  | OR               | retrograde    | scoring      |
| kidney calculi| OR               | intrarenal    | system OR    |
| renal stone   | OR               | surgery OR    | nomogram     |
| renal calculi | OR               | flexible      | ureterorenoscopy OR |
| urolithiasis  | OR               | OR            | flexible     |
| nephrolithiasis| OR               | OR            | ureteroscopy |
| upper ureter stone | OR | upper | ureter calculus OR |
| OR            | proximal        | ureter stone  |
| proximal     | OR               | ureter calculus |
relevant studies included in the systematic review. (T)allness, (O)ccupied lesion and (HO)unsfield units (T.O.HO) score and 2 nomograms were excluded due to the lack of external validation [9, 10, 11]. Also one nomogram which developed for pediatric patients and one nomogram predicting perioperative complications were excluded [12, 13]. After exclusion of 7 article in the full-text evalua-
tion process, remaining 14 studies were included in systematic review. Thirteen studies providing meta-data for quantitative analyses were included in one or more meta-analyses. The characteristics of the selected studies were shown in Table 4. There were 3, 1, and 4 articles which studied 4, 3, and 2 different scoring systems at the same time. Five articles had evaluated only single scoring system. All scoring systems had been included in 3 or more studies. Most studied scoring systems were RUS and S-ReSC scores (8 times for each). Parameters which included in scoring

### Table 4. Characteristics of studies included in systematic review

| First author/year | Country | Design | Stone localization and stone burden | Number of patients | Outcome (SFR) | Postoperative imaging method | Stone-free status definition | Studied scoring systems |
|------------------|---------|--------|------------------------------------|-------------------|--------------|-----------------------------|---------------------------|-----------------------|
| Selmi 2020       | Turkey  | P      | Kidney 924 mm^3                   | 110               | (81/110) 73.6% | Not indicated               | Not having residual stone fragments greater than 4 mm | S-ReSC RUS R.I.R.S. S.T.O.N.E Ito’s |
| Bozkurt 2021     | Turkey  | R      | Kidney 103 mm^3                   | 949               | (743/949) 78.3% | CT                          | Residual fragments <2 mm | + + + + |
| Richard 2020     | France  | R      | Kidney and/or upper ureter 11 mm  | 800               | (593/800) 74.1% | Radioscopic imaging or CT   | Total absence of residual stone     | + + + + |
| Ozbek 2020       | Turkey  | R      | Kidney 13 mm                      | 280               | (215/280) 76.7% | CT                          | Complete clearance          | + + + |
| Erbin 2016       | Turkey  | R      | Kidney 145 mm^3                   | 339               | (238/339) 70.1% | KUB, US or CT               | No evidence of residual stones or fragments at 1 month follow-up | + + |
| Karsiyakali 2020 | Turkey  | R      | Kidney and/or upper ureter 140 mm^3 | 81                | (60/81) 74.1% | KUB or CT                   | Clinically insignificant residual stones <4 mm | + + |
| Jung 2014        | S. Korea| R      | Kidney 12 mm                      | 88                | (75/88) 85.2% | CT                          | No evidence of residual stone on post-operative CT for 1 month | + |
| Park 2015        | S. Korea| R      | Kidney 1.6/2.5 cm^3               | 159               | (116/159) 73%  | CT                          | No evidence of a stone or with clinically insignificant residual fragments less than 2 mm | + |
| Xiao 2017        | China   | R      | Kidney 14 mm                      | 382               | (281/382) 73.6% | KUB or CT (if KUB showed any high-densities or radiolucent stones) | No detectable stone on KUB, and fragments of less than 2 mm | + + |
| Wang 2021        | China   | R      | Kidney 13 mm                      | 147               | 105/147 (71.4%) | KUB or CT                   | No detectable stone on KUB or non-contrast CT, or fragments of less than 2 mm | + + |
| Sfoungaristos 2016 | Israel | P      | Kidney 10 mm                      | 85                | 63/85 (74.1%) | CT                          | The absence of any residual fragment | + |
| Molina 2014      | USA     | R      | Kidney and/or ureter 9 mm         | 200               | 164/200 (82%)  | Intraoperative endoscopic inspection with fluoroscopy and/or CT | The absence of stone fragments or fragments ≤ 2 mm | + |
| Ito 2014         | Japan   | R      | Kidney 679/3035 mm^3              | 310               | 185/310 (59.7%) | CT                          | The strict absence of visible stones on imaging | + |
| Resorlu 2012     | Turkey  | R      | Kidney 16 mm                      | 207               | 178/207 (86%)  | Intraoperative endoscopic inspection, US, CT | CIRF ≤1 mm | + |

P – prospective; R – retrospective; CT – computed tomography; KUB – kidney ureter bladder; SRF – stone-free rate; CIRF – clinically insignificant residual fragment; US – ultrasonography; RUS – Resorlu-Unsal Score; SReSC – Seoul National University Renal Stone Complexity
systems were summarized in Table 3. These can be classified in three categories; stone related parameters (stone burden, localization, density etc.), anatomic parameters (infundibular measurements, hydronephrosis, abnormal anatomy etc.), and surgeon related parameter (experience).

Quality assessment

The risks of bias assessments were displayed in Table 5. Prognostic factor and outcome measurements were the most poorly rated domains among six criteria of QUIPS tool. Except two of them, retrospective design of included studies caused a scoring system calculation bias risk. There weren’t clear statements indicating exact calculation time of scoring systems in validation studies (preoperatively or in the retrospective study period). More importantly, most of the studies didn’t report inter-/intraobserver variability analyses. Thus, most of the studies showed poor quality in terms of prognostic factor measurement criteria. Another common lack that reduces the quality of the studies was the non-standard outcome measurement.

Data analysis

Relevant metadata which extracted from selected studies for each scoring system and results of meta-analyses with pooled data can be seen in Table 6. All scoring systems' AUC values showed statistically significant heterogeneity in first inclusive meta-analyses (S-ReSC and RUS; moderate, R.I.R.S. and S.T.O.N.E.; high and Ito's nomogram; very high).

**S-ReSC score.** Eight studies reported AUCs for S-ReSC score [4, 5, 15, 20–24]. Meta-analysis yielded a pooled AUC of 0.716 (95% CI 0.669 to 0.762). Heterogeneity was moderate ($I^2 = 74\%$). It reduced to $I^2 = 51\%$ ($p = 0.09$) after exclusion of three studies by Jung, Erbin, and Ozbek with new pooled AUC of 0.709 (95% CI 0.670 to 0.748, random effects) [15, 20, 21].

**R.I.R.S. score.** Five studies reported UACs for R.I.R.S. score [4, 5, 16, 20, 25]. Meta-analysis yielded a pooled AUC of 0.781 (95% CI 0.711 to 0.851). Heterogeneity was high ($I^2 = 89\%$). It reduced to $I^2 = 0\%$ ($p = 0.43$) after exclusion of Xiao’s and Ozbek’s studies with new pooled AUC of 0.704 (95% CI 0.668 to 0.739, fixed effects) [16, 20].

**RUS score.** Eight studies reported AUCs for RUS score [4, 5, 16, 20, 21, 22, 25, 26]. Meta-analysis yielded a pooled AUC of 0.711 (95% CI 0.668 to 0.754). Heterogeneity was moderate ($I^2 = 75\%$). It reduced to $I^2 = 6\%$ ($p = 0.4$) after exclusion of two studies by Xiao and Ozbek with new pooled AUC of 0.669 (95% CI 0.646 to 0.692, fixed effects) [16, 20].

**S.T.O.N.E. score.** Four studies reported AUCs for S.T.O.N.E [5, 18, 20, 22]. Meta-analysis yielded a pooled AUC of 0.728 (95% CI 0.647 to 0.809). Heterogeneity was high ($I^2 = 88\%$). It reduced to $I^2 = 22\%$ ($p = 0.3$) after exclusion of Richard’s study with new pooled AUC of 0.771 (95% CI 0.724 to 0.818, fixed effects) [22].

### Table 5. Risk of bias rating

| Study          | Study participation | Study attrition | Prognostic Factor Measurement | Outcome Measurement | Study Confounding | Statistical Analysis and Reporting |
|---------------|---------------------|-----------------|-------------------------------|---------------------|-------------------|-----------------------------------|
| Selmi 2020    | +                   | +               | +                             | -                   | +                 | +                                 |
| Bozkurt 2021  |                     | +               | ?                             | -                   | +                 | +                                 |
| Richard 2020  | +                   | ?               | ?                             | +                   | ?                 | +                                 |
| Ozbek 2020    | +                   | +               | ?                             | +                   | +                 | +                                 |
| Erbin 2016    | +                   | +               | ?                             | -                   | +                 | +                                 |
| Karsiyakali 2020 | +               | +               | -                             | -                   | ?                 | +                                 |
| Jung 2014     | +                   | +               | ?                             | +                   | ?                 | +                                 |
| Park 2015     | ?                   | +               | +                             | +                   | +                 | +                                 |
| Xiao 2017     | +                   | +               | -                             | ?                   | +                 | +                                 |
| Wang 2021     | +                   | +               | -                             | -                   | +                 | +                                 |
| Sfoungaristos 2016 | +               | +               | -                             | +                   | +                 | +                                 |
| Molina 2014   |                     | -               | ?                             | -                   | ?                 | +                                 |
| Ito 2014      | +                   | +               | ?                             | +                   | +                 | +                                 |
| Resorlu 2012  | ?                   | +               | ?                             | +                   | +                 | ?                                 |

Key: +; low risk bias; -; high risk bias, ?; unclear risk of bias
### Table 6. Meta-data extracted from selected studies for all scoring systems and results of meta-analyses

| Studies including S-ReSC | AUC   | SE    | 95% CI          | Weight (% random) |
|--------------------------|-------|-------|-----------------|-------------------|
| Bozkurt et al.           | 0.657 | 0.0220| 0.614 to 0.700  | 32.20             |
| Richard et al.           | 0.651 | 0.0230| 0.692 to 0.778  | 32.20             |
| Selmi et al.             | 0.755 | 0.0480| 0.661 to 0.849  | 13.10             |
| Park et al.              | 0.732 | 0.0420| 0.650 to 0.814  | 15.90             |
| Karsiya et al.           | 0.687 | 0.0730| 0.544 to 0.830  | 6.60              |
| Total (random effects)   | 0.709 | 0.0200| 0.670 to 0.748  | 100.00            |

| Studies including R.I.R.S. | AUC   | SE    | 95% CI          | Weight (% fixed)  |
|---------------------------|-------|-------|-----------------|-------------------|
| Wang et al.               | 0.737 | 0.0480| 0.643 to 0.831  | 14.26             |
| Bozkurt et al.            | 0.690 | 0.0210| 0.649 to 0.731  | 74.48             |
| Selmi et al.              | 0.752 | 0.0540| 0.646 to 0.858  | 11.26             |
| Total (fixed effects)     | 0.704 | 0.0181| 0.668 to 0.739  | 100.00            |

| Studies including RUS     | AUC   | SE    | 95% CI          | Weight (% fixed)  |
|---------------------------|-------|-------|-----------------|-------------------|
| Erbin et al.              | 0.655 | 0.0330| 0.590 to 0.720  | 12.66             |
| Bozkurt et al.            | 0.689 | 0.0210| 0.648 to 0.730  | 31.27             |
| Richard et al.            | 0.644 | 0.0180| 0.609 to 0.679  | 42.56             |
| Selmi et al.              | 0.735 | 0.0520| 0.633 to 0.837  | 5.10              |
| Sfoungaristos et al.      | 0.707 | 0.0690| 0.572 to 0.842  | 2.90              |
| Wang et al.               | 0.700 | 0.0500| 0.602 to 0.798  | 5.52              |
| Total (fixed effects)     | 0.669 | 0.0117| 0.646 to 0.692  | 100.00            |

| Studies including S.T.O.N.E. | AUC   | SE    | 95% CI          | Weight (% fixed)  |
|------------------------------|-------|-------|-----------------|-------------------|
| Selmi et al.                 | 0.725 | 0.0500| 0.627 to 0.823  | 22.71             |
| Molina et al.                | 0.764 | 0.0320| 0.701 to 0.827  | 55.45             |
| Karsiya et al.               | 0.837 | 0.0510| 0.737 to 0.937  | 21.83             |
| Total (fixed effects)        | 0.771 | 0.0238| 0.724 to 0.818  | 100.00            |

| Studies including Ito’s score | AUC   | SE    | 95% CI          | Weight (% random) |
|-------------------------------|-------|-------|-----------------|-------------------|
| Richard et al.                | 0.735 | 0.0220| 0.692 to 0.778  | 33.42             |
| Bozkurt et al.                | 0.303 | 0.0200| 0.264 to 0.342  | 33.47             |
| Ito et al.                    | 0.870 | 0.0320| 0.807 to 0.933  | 33.11             |

RUS – Resorlu-Unsal score; SReSC – Seoul National University Renal Stone Complexity; AUC – area under curve; SE – standard error; CI – confidence interval
Ito’s nomogram. Three studies reported AUCs for Ito’s nomogram [4, 19, 22]. Meta-analysis yielded a pooled AUC of 0.635 (95% CI 0.361 to 0.909). Heterogeneity was very high ($I^2 = 99\%$). Statistically significant and very high heterogeneity could not be resolved by exclusion of any study.

**DISCUSSION**

There were five scoring systems which had been validated externally and well-studied; S-ReSC, R.I.R.S., S.T.O.N.E. scores, RUS and Ito’s nomogram.

**S-ReSC score.** This scoring system was introduced firstly for the prediction of SFR after percutaneous nephrolithotomy [14]. Then it was modified for outcomes of RIRS by Jung and colleagues [15]. The modified S-ReSc score is calculated sum of 1 points attended for each of following stone locations: the renal pelvis (#1), superior and inferior major calyceal groups (#2-3), and anterior and posterior minor calyceal groups of the superior (#4-5), middle (#6-7), and inferior calyx (#8-9). If the stone is in the inferior sites (#3, #8-9), one additional point per site is added.

**R.I.R.S. score.** Xiao and colleagues developed R.I.R.S. score in 2017 [16]. This score is assigned according to following criteria; (R)enal stone density ≤1000 Hounsfield Unit (HU) (1 point) or >1000 HU (2 points), the renal infundibulopelvic angle (RIPA, defined as the inner angle of the intersection of ureteropelvic axis and the axis of the lower renal calyx) of the (I)nferior pole stone (scored from 1 to 3 points as determined by a non-inferior pole stone or inferior pole stone with RIPA $>30^\circ$ or $\leq30^\circ$), (R)enal infundibular length (RIL, the distance from most distal point at bottom stone-containing calix to midpoint of lip of renal pelvis) $>25$ mm (2 points) or $\leq25$ mm (1 point), (S)tone burden (1 to 3 points for stone burden according to cumulative stone diameter ≤10 mm, >10 mm and ≤20 mm, and >20 mm, respectively).

**RUS score.** The first developed and validated scoring system for RIRS was introduced by Resorlu and colleagues in 2012 [17]. There are four criteria have equal weight (1 point for each); stone size >20 mm, lower pole stone location and RIPA <45°, stone number in different calyces >1, abnormal renal anatomy (horseshoe kidney or pelvic kidney).

**S.T.O.N.E. score.** This scoring system was derived from pre-operative radiological features of stones [18]. Name of the scoring system is an acronym of included parameters. (S) represents stone size; 1, 2 and 3 points for stones <5 mm, 5–10 mm, and ≥10 mm, respectively. (T)opography of stone is classified as distal to mid-ureter (1 point), proximal ureter, upper and middle pole of kidney (2 points), lower pole (3 points). (O)bsturation is scored with the respect of hydronephrosis degree. (N)umber of stone is scored as 1 stone = 1 point, 2 stones = 2 points, >2 stones = 3 points. Finally, (E)valuation of HU is included in the scoring system as follows; $<750$ HU = 1 point, 750–1000 HU = 2 points and $>1000$ HU = 3 points.

Ito’s nomogram. The only nomogram which was included in the systematic review was Ito’s. It was introduced by Ito and colleagues in 2014 [19]. Stone volume $\leq500$, 500$<x<1000$, 1000$<x<2000$, $>2000$ mm$^3$), lower pole calculi, operator experience ($<50$, $\geq50$), hydronephrosis, and number of stones are the parameters of nomogram. A total score is calculated according to these parameters (total score 0–25).

This systematic review and meta-analysis has not revealed superiority of any scoring system which aimed to predict surgical outcomes of RIRS. Statistically significant heterogeneity prevented results of analysis from interpreting predictive/discriminative ability of all recent scoring systems. Although heterogeneity was resolved to certain extent by the exclusion of some studies, following evident clinical obstacles will continue to be a source of heterogeneity for further studies; the inter-/intra-observer variability of scores which were calculated by many different scorers, retrospective calculation bias, discordance between studies for the exact diameter of significant residual fragments and surgical outcome assessment methods. Furthermore, some controversial parameters of scoring systems could have contributed the heterogeneity. Stone burden was the most included parameter. Its high predictive performance is still valid but there are controversies about the calculation method of it [27]. Similar to literature, stone burden unit (mm, mm$^2$, cm$^3$) showed variety among included studies. Also, heterogeneity aside, any scoring system did not demonstrate good predictive/discriminative performance (pooled AUC $\geq$0.800).

Ozbek’s study was a strong source of heterogeneity for three scoring systems; S-ReSC, R.I.R.S. and RUS [20]. AUC values for these scoring systems were reported higher in Ozbek’s study than other studies. Xiao’s study which was another cause of heterogeneity for R.I.R.S. and RUS scores, was the only article reporting higher AUC values than Ozbek’s study [16]. This two studies were responsible a high amount of heterogeneity.

After the exclusion of studies which were source of heterogeneity for each scoring system, S.T.O.N.E. score had higher pooled AUC value (0.771) than S-ReSc, R.I.R.S., and RUS scores (0.709, 0.704 and 0.669, respectively). High heterogeneity could not be resolved by exclusion of any study for Ito’s nomogram. Thus, it was not involved further interpretation.
Pooled AUC of S.T.O.N.E. score increased and heterogeneity was resolved completely after exclusion of Richard’s study [22]. However, an uncertainty arising from Karsiyakali’s study compromised the reliability of pooled AUC value. The authors of this study reported high predictive performance in favor of S.T.O.N.E score as a finding of comparative analysis involving four different scoring systems [23]. Although AUC value of S.T.O.N.E score was reported higher than other scores, pairwise analysis did not reveal any superiority for this scoring system in same study. Also the citation which addressed the developers of scoring system was not the developmental study of Molina’s. The authors have cited another S.T.O.N.E. score which developed by Okhunov and colleagues for prediction of PNL’s surgical outcomes [28]. It was developed at the same time with Molina’s system and has very similar design. But Okhunov’s system involves some PNL specific parameters such as trach length.

Except R.I.R.S. score, developmental studies of all other scoring systems did not report any comparative findings between their model and recent scoring systems [15–19]. All independent comparative studies emphasized the superiority of a different scoring system via indirect statistical analyses [4, 5, 20, 22]. This is probably due to the difficulties faced in direct comparison analyses. Same difficulty was faced during the further comparative analysis stage of this meta-analysis. At the protocol stage, we decided to perform second stage meta-analyses after obtaining of pooled AUC values for all scoring systems. But these values were derived from mixed patient cohorts (same patients from matched studies and different patients from other studies). Then we tried to have new pooled AUC values derived from same patients for scoring system couples. Also there was another requirement before pairwise comparisons; the independence of scoring systems. All of the scoring systems showed dependency because of the mutual parameters. Only matching of S-ReSC and R.I.R.S. scores provided statistically independency. But further comparison did not require due to quite similar pooled AUC values of these scores. Minimal dependency between Ito’s nomogram and S-ReSC may allow direct comparison of them [29].

Retrograde intrarenal surgery may require usage of some instruments during, before or after the procedure. Ureteral access sheath (UAS), stone basket usage or pre/postoperative double j stent placement are the common instrumentalations. Their use may influence the outcomes of RIRS [30, 31]. But surgeon attitudes towards this utilisations complicate the surgical technique standardization of RIRS. Most selected studies had reported the UAS/basket usage. But there was no enough metadata about pre/postoperative double j stent insertion rates. This was another cause for heterogeneity and poor quality beyond all the statistical and reporting problems discussed above.

Strengths and limitations

This is the first systematic review and meta-analysis on comparison of recent scoring systems predicting outcomes of RIRS. We only used AUC metadata from studies to provide inclusive findings. More reliable metadata (for example multivariable analysis results) were not exist in most of the studies. Other limitations were poor quality, the heterogeneity between the studies, cohorts differencies and dependencies between scoring systems preventing further comparisons. All of these limitations affected the generalizability of the findings.

Future research

Despite abundance of scoring system and nomogram studies, ideal system to predict surgical outcomes of RIRS is still an unmet need. A scorer friendly system showing low inter-/intraobserver variability, high discriminative/predictive ability may gain popularity and may be used widely.

CONCLUSIONS

There was no superiority of any current scoring system which aimed to predict surgical outcomes of RIRS. Although S.T.O.N.E. score showed highest AUC value, high heterogeneity between studies and dependencies between scoring systems make difficult to design a comparative statistical model to generalize these findings. Also, limitations aside, neither scoring system has demonstrated good predictive/discriminative performance.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

Research conception and design: OÖ, BÖ, HA. Data acquisition: OÖ, CB, OÇ, HMA. Statistical analysis: OÖ, CB. Data analysis and interpretation: HÇ, OÖ, HMA. Drafting of the manuscript: OÖ, EBS. Critical revision of the manuscript: BÖ, HÖ, HÇ, CMY. Obtaining funding: HA, BÖ. Administrative, technical, or material support: HMA, OÇ. Supervision: CMY, EBS. Approval of the final manuscript: BÖ, HA.
References

1. Alkan E, Saribacak A, Ozkanli AO, Basar MM, Acar O, Balbay MD. Retrograde Intrarenal Surgery in Patients Who Previously Underwent Open Renal Stone Surgery. Minim Invasive Surg. 2015; 2015: 198765.

2. Türk C, Neisius A, Petik A, et al. Guidelines on urolithiasis panel. Edn. presented at the EAU Annual Congress Milan 2021. ISBN 978-94-92671-14-1

3. Jones P, Pietropaolo A, Chew BH, Somani BK. Atlas of Scoring Systems, Grading Tools, and Nomograms in Endourology: A Comprehensive Overview from the TOWER Endourological Society Research Group. J Endourol. 2021; 35: 1863-1882.

4. Bozkurt HI, Karakoyunlu AN, Koras O, et al. External validation and comparison of current scoring systems in intracorporeal intrarenal surgery: Multi-institutional study with 949 patients. Int J Clin Pract. 2021; 75: e14097.

5. Selmi V, Sari S, Oztekin U, Canikiloglu M, Isiklay L. External Validation and Comparison of Nephrolithometric Scoring Systems Predicting Outcomes of Retrograde Intrarenal Surgery. J Endourol. 2021; 35: 781-788.

6. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021; 372: n71.

7. Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. Ann Intern Med. 2013; 158: 280-286.

8. Glasmacher SA, Stones W. Anion gap as a prognostic tool for risk stratification in critically ill patients - a systematic review and meta-analysis. BMC Anesthesiol. 2016; 16: 68.

9. Hori S, Otsuki H, Fujio K, et al. Novel prediction scoring system for simple assessment of stone-free status after flexible ureteroscopy lithotripsy: T.O.HO. score. Int J Urol. 2020; 27: 742-747.

10. Xun Y, Chen M, Liang P, et al. A Novel Clinical-Radiomics Model Pre-operatively Predicted the Stone-Free Rate of Flexible Ureteroscopy Strategy in Kidney Stone Patients. Front Med (Lausanne). 2020; 7: 576925.

11. Micali S, Sighinolfi MC, Iseppi A, et al. Initial Experience and Evaluation of a Nomogram for Outcome Prediction in Management of Medium-sized (1-2 cm) Kidney Stones. Eur Urol Focus. 2021; S2405-4569(20)30314-X.

12. Zhang Y, Li J, Zhang D, Jiao JW, Tian Y. Nomograms predicting the outcomes of endoscopic treatments for pediatric upper urinary tract calculi. Int J Urol. 2021; 28: 295-301.

13. Komey M, Odaka H, Asano j, et al. Development and internal validation of a nomogram to predict perioperative complications after flexible ureteroscopy for renal stones in overnight ureteral catheterization cases. World J Urol. 2020; 38: 2307-2312.

14. Jeong CW, Jung JW, Cha WH, et al. Seoul National University Renal Stone Complexity Score for predicting stone-free rate after percutaneous nephrolithotomy. PLoS One. 2013; 8: e65888.

15. Jung JW, Lee BK, Park YH, et al. Modified Seoul National University Renal Stone Complexity score for retrograde intrarenal surgery. Urolithiasis. 2014; 42: 335-340.

16. Xiao Y, Li D, Chen L, et al. The R.I.R.S. scoring system: An innovative scoring system for predicting stone-free rate following retrograde intrarenal surgery. BMC Urol. 2017; 17: 105.

17. Resorlu B, Unsal A, Guler H, Oztuna D. A new scoring system for predicting stone-free rate after retrograde intrarenal surgery: the ‘resorlu-unsal stone score’. Urology. 2012; 80: 512-518.

18. Molina WR, Kim FJ, Spendlove J, Pompeo AS, Sillau S, Sehrt DE. The S.T.O.N.E. E: A new assessment tool to predict stone free rates in ureteroscopy from pre-operative radiological features. Int Braz J Urol. 2014; 40: 23-29.

19. Ito H, Sakamaki K, Kawahara T, et al. Development and internal validation of a nomogram for predicting stone-free status after flexible ureteroscopy for renal stones. BJU Int. 2015; 115: 446-451.

20. Ozbek R, Senocak C, Haberal HB, Damar E, Sadioglu FE, Bozkurt OF. Comparison of scoring systems for predicting stone-free status and complications after retrograde intrarenal surgery. World J Urol. 2021; 39: 2741-2746.

21. Erbin A, Tepeler A, Buldu I, Ozdemir H, Tosun M, Binbay M. External Comparison of Recent Predictive Nomograms for Stone-Free Rate Using Retrograde Flexible Ureteroscopy with Laser Lithotripsy. J Endourol. 2016; 30: 1180-1184.

22. Richard F, Marguin J, Frontczak A, et al. Evaluation and comparison of scoring systems for predicting stone-free status after flexible ureteroscopy for renal and ureteral stones. PLoS One. 2020; 15: e0237068.

23. Karsiyanakil N, Karabay E, Erkan E, Kadihasanoglu M. Evaluation of Nephrolithometric Scoring Systems to Predict Outcomes of Retrograde Intrarenal Surgery. Urol J. 2020; 17: 352-357.

24. Park J, Kang M, Jeong CW, et al. External Validation and Evaluation of Reliability and Validity of the Modified Seoul National University Renal Stone Complexity Scoring System to Predict Stone-Free Status After Retrograde Intrarenal Surgery. J Endourol. 2015; 29: 888-893.

25. Wang C, Wang ST, Wang X, Lu J. External validation of the R.I.R.S. scoring system to predict stone-free rate after retrograde intrarenal surgery. BMC Urol. 2021; 21: 33.

26. Sfongaristos S, Gofrit ON, Mykonias I, et al. External validation of Resorlu-Unsal stone score as predictor of outcomes after retrograde intrarenal surgery. Int Urol Nephrol. 2016; 48: 1247-1252.

27. De Coninck V, Traxer O. The time has come to report stone burden in terms of volume instead of largest diameter. J Endourol. 2018; 32: 265-266.

28. Okhunov Z, Friedlander JI, George AK, et al. S.T.O.N.E. nephrolithometry: novel surgical classification system for kidney calculi. Urology. 2013; 81: 1154-1159.

29. Özcan O. About ‘Comparison of scoring systems for predicting stone free status and complications after retrograde intrarenal surgery’. World J Urol. 2021. Epub ahead of print.

30. Özcan O, Başaçağ, C, Akgül M, et al. Prediction of Possible Factors that Affect Stone Free Rate of Retrograde Intrarenal Surgery; A Multicenter Study. Namık Kemal Tip Dergisi. 2020; 8: 398-403.

31. Özcan O, Akgül HM, Başaçağ C, et al. Multi-aspect analysis of ureteral access sheath usage in retrograde intrarenal surgery: A RIRSearch group study. Asian J Urol. 2021 [Article in press].