available at www.sciencedirect.com
journal homepage: www.eu-openscience.europeanurology.com

European Association of Urology

Review – Stone Disease

Which Measure of Stone Burden is the Best Predictor of Interventional Outcomes in Urolithiasis: A Systematic Review and Meta-analysis by the YAU Urolithiasis Working Group and EAU Urolithiasis Guidelines Panel

Robert Geraghty [a,b], Amelia Pietropaolo [c,d], Lazaros Tzelves [b,d,e], Riccardo Lombardo [b,f], Helene Jung [b,g], Andreas Neisius [b,h], Ales Petrik [b,i], Bhaskar K. Somani [b,c], Niall F. Davis [b,j,k], Giovanni Gambaro [b,l], Romain Boissier [d,m], Andreas Skolarikos [b,e,], Thomas Tailly [b,d,n]*

[a] *Department of Urology, Freeman Hospital, Newcastle-upon-Tyne, UK;* [b] *Urolithiasis Guidelines Panel, European Association of Urology, Arnhem, The Netherlands;* [c] *Department of Urology, University Hospital Southampton NHS Foundation Trust, Southampton, UK;* [d] *Young Academic Urologists Urolithiasis Working Group, European Association of Urology, Arnhem, The Netherlands;* [e] *Department of Urology, National and Kapodistrian University of Athens, Sismanogleio Hospital, Athens, Greece;* [f] *Sant'Andrea Hospital, Sapienza University, Rome, Italy;* [g] *Department of Urology, University of Southern Denmark, Odense, Denmark;* [h] *Department of Urology, Bruederkrankenhaus Trier, Johannes Gutenberg University Mainz, Trier, Germany;* [i] *Department of Urology, First Faculty of Medicine, Charles University, Prague, Czechia;* [j] *Department of Urology, Beaumont Hospital, Dublin, Ireland;* [k] *Department of Surgery, Royal College of Surgeons in Ireland, Dublin, Ireland;* [l] *Division of Nephrology and Dialysis, Department of Medicine, University of Verona, Verona, Italy;* [m] *Department of Urology and Renal Transplantation, Aix-Marseille University, Marseille, France;* [n] *Department of Urology, University Hospital of Ghent, Ghent, Belgium*

Article info

Article history:
Accepted October 27, 2024

Associate Editor:
Silvia Proietti

Keywords:
Urolithiasis
Stone burden
Single linear measurement
Multidimensional measures
Stone-free rate
Predictor
Guidelines

Abstract

Background and objective: Stone size has traditionally been measured in one dimension. This is reflected in most of the literature and in the EAU guidelines. However, recent studies have shown that multidimensional measures provide better prediction of outcomes.
Methods: We performed a systematic review and meta-analysis of the prognostic accuracy of measures of stone size (PROSPERO reference CRD42022346967). We considered all studies reporting prognostic accuracy statistics on any intervention for kidney stones (extracorporeal shockwave lithotripsy [ESWL], ureterorenoscopy [URS], or percutaneous nephrolithotomy [PCNL]; Population) using multiplane measurements of stone burden (area in mm^2 or volume in mm^3; Intervention) in comparison to single-plane measurements of stone burden (size in mm; Intervention) for the study-defined stone-free rate (Outcome) in a PICO-framed question. We also assessed complication rates (overall and by Clavien-Dindo grade) and the operative time as secondary outcomes. Searches were made between 1970 and August 2023. We used the DeLong method to compare receiver operating characteristic (ROC) curves.
Key findings and limitations: Of 24 studies included in the review, 12 were eligible for comparative analysis with the DeLong test following meta-analysis of

* Corresponding author. Department of Urology, National and Kapodistrian University of Athens, Sismanogleio Hospital, Athens, Greece.
E-mail address: andskol@yahoo.com (A. Skolarikos).

https://doi.org/10.1016/j.euros.2024.10.024

prognostic accuracy. For prediction of stone-free status, the area under the ROC curve (AUC) was significantly higher for stone volume than for stone size (0.71 vs 0.67; $p < 0.001$). Subanalyses confirmed this for ESWL and URS, but not for PCNL. For URS, the AUC was also significantly higher for stone area than for stone size (0.79 vs 0.77; $p < 0.001$). Throughout all analyses, there was no difference in AUC between stone area and stone volume. There was high risk of bias for all analyses apart from the URS subanalyses.

Conclusions and clinical implications: According to the limited data currently available, stone-free rates are predicted with significantly higher accuracy using multidimensional measures of stone burden in comparison to a single linear measurement.

Patient summary: We reviewed different ways of measuring the size of stones in the kidney or urinary tract and compared their accuracy in predicting stone-free rates after treatment. We found that measurement of the stone area (2 dimensions) or stone volume (3 dimensions) is better than stone diameter (1 dimension) in predicting stone-free status after treatment.

1. Introduction

Urolithiasis is a relatively common condition for which a gradual increase in prevalence up to ~10% has been observed in the past few decades [1]. In the management of urinary stone disease, decisions on observation versus treatment and the preferred treatment modality are based in part on the size of the stone. The European Association of Urology (EAU) and other guidelines provide a decision tree and recommendations that are based on the maximum linear stone size [2,3]. The use of computed tomography (CT), which has become the gold standard for diagnosis and treatment planning, has ushered in the possibility of measuring stone burden in multiple dimensions. As a result, investigations on the value of multiplanar measurements to predict interventional outcomes have increased [4]. Intuitively, multiplanar stone measurements should represent stone burden more accurately and may therefore be a better predictor of the stone-free rate after any intervention. However, results from studies evaluating the predictive value of stone burden have not been unanimously in favor of multiplanar burden assessment in comparison to size, most often reported as cumulative stone diameter (CSD). Despite the breadth of literature available on this subject and the fact that we are guided by a measure of stone burden in everyday practice, a solid answer to the question of which measure we should be using is still lacking.

We therefore performed a systematic review and meta-analysis of studies investigating the prognostic accuracy of different stone measurements in predicting stone-free status after an intervention for urolithiasis.

2. Methods

A systematic review and meta-analysis was conducted under the auspices of the EAU according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [5] and was guided by the Methods Committee of the EAU Guidelines Office [2,6]. The PRISMA statement and review protocol are provided in the Supplementary material.

2.1. *Literature search*

A literature search was performed by a professional librarian using the Medline/PubMed, EMBASE, and Cochrane Library databases. In addition, registered randomized controlled trial protocols were screened on ClinicalTrials.gov from 1970 to May 2022. An updated search was performed in August 2023. The reference lists in all manuscripts reviewed in full-text form were also screened for eligible studies. Five independent authors (R.L., L.T., R.G., A.P., T.T.), screened the databases. Disagreements were resolved via consensus with another senior author (R.B.). Supplementary Figure 1 shows the PRISMA flow diagram.

2.2. *Eligibility criteria*

The protocol was approved by the EAU Guidelines Office and then registered in the PROSPERO database (CRD42022346967). We included randomized clinical trials and both retrospective and prospective comparative nonrandomized studies.

The inclusion criteria according to the Population, Intervention, Comparator, Outcome (PICO) framework were as follows: patients aged ≥18 yr of either sex with primary or recurrent renal tract stones of any type or composition who underwent an intervention with ureterorenoscopy (URS), percutaneous nephrolithotomy (PCNL), or extracorporeal shockwave lithotripsy (ESWL) and had at least two different measures of stone burden according to preoperative imaging. We excluded manuscripts that included only patients with neurological disorders, urogenital abnormalities (eg, horseshoe kidneys), or urinary diversions, as well as studies with fewer than ten patients. The primary outcome was stone-free status. Operative time and complication rates were secondary outcomes.

2.3. Data collection

Five authors (R.G., R.L., L.T., A.P., T.T.) independently extracted data from eligible studies, including study characteristics (author, country/center, period, study design, inclusion and exclusion criteria), patient baseline characteristics (type of treatment, definition of the stone-free rate, imaging modality and follow-up interval, number of patients, sex distribution, body mass index), and stone measurement characteristics in terms of the cumulative stone diameter (CSD) and the surface area and/or volume as available. Outcome data including numbers, proportions, cutoff values, odds ratios (ORs), area under the receiver operating characteristic (ROC) curve (AUC), the number of true and false positive and negative results, sensitivity, and specificity were extracted for: the stone-free rate, complication rates, and the operative time. If an ROC curve was presented with no sensitivity or specificity results, then these were estimated from the graph [7].

2.4. Risk-of-bias assessment

Four authors (R.B., L.T., T.T., A.P.) assessed the risk of bias (RoB) of individual studies independently using the Cochrane Collaboration Risk of Bias Tool for Randomized Controlled Trials [8] and the QUIPS tool for nonrandomized observational studies [9]. The following fields were assessed: source of the target population, method used to identify the population, recruitment period, place of recruitment, inclusion/exclusion criteria, adequate participation, baseline characteristics, proportion of the baseline sample available for analysis, dropout rate, loss to follow-up, prognostic factor definition, measurement of prognostic factor, proportion of data available for analysis, method used for missing data, outcome definition, measurement of outcome, confounder measurement, definition of confounding factors, analysis presentation, reporting of results, and overall risk of bias. The overall RoB was considered low if all domains were ranked as low, and high if at least one of the domains was ranked as high. Overall RoB is reported with the main results for context and in Section 3.7.

2.5. Statistical analysis

Each outcome was stratified by measurement type for meta-analysis, with subanalyses for the three intervention types. We used a random-effects model when heterogeneity was >50%, and a fixed-effect model when heterogeneity was <50%. Heterogeneity was assessed using I^2, τ^2, and Cochran's Q statistics. All statistical analyses were performed using R (R Foundation for Statistical Computing, Vienna, Austria) using the *meta* [10], *mada* [11], and *nsROC* [12] packages. We present odds ratios (ORs) for a random-effects or fixed-effects model as appropriate. Publication bias was assessed via visual inspection of funnel plots. For analyses that included more than two studies, we performed trim-and-fill analyses to statistically assess publication bias. Adjusted values for the trim-and-fill analysis are presented along with the calculated number of studies missing. We present forest and funnel plots, along with heterogeneity statistics (I^2, Cochran's Q, and τ^2) if the number of studies included was more than two.

Data were examined for their suitability for meta-analysis of prognostic accuracy. Studies with no data on the number of true and false positive and negative results were excluded. If data were available for the overall numbers (total number in the study and for specified outcomes, eg, stone-free number) and the sensitivity and specificity, then we calculated the numbers of true and false positive and negative results. We present diagnostic ORs (DORs) (reported as for OR, as described above), ROC curves, and AUC values. We compared AUC values between measures using DeLong's method for comparative studies only [13]. The statistical code is available in the Supplementary material.

3. Results

3.1. Study demographics

A total of 24 studies were included in the review [14–37], including 13 on ESWL, eight on URS, two on PCNL and one on both URS and PCNL. The studies were mainly retrospective comparative studies and were distributed around the world (Supplementary Table 1).

These studies involved a total of 4791 patients (63% men, 37% women), of whom 2390 underwent ESWL, 1780 underwent URS, and 621 underwent PCNL for stones in different locations in the upper urinary tract (Table 1).

3.2. Meta-analysis of models for the stone-free rate by size metric

Overall, 12 studies had ORs available for linear size (six ESWL, four URS, and two PCNL) [17,19–24,27,30,33,34,37], seven had ORs available for surface area (two ESWL, four URS, and one PCNL) [17,18,20,22,35–37], and 16 had ORs available for volume (seven ESWL, seven URS, and two PCNL) for meta-analysis.

After meta-analysis of study ORs from multivariable logistic regression analyses, the overall ORs were 1.12 (95% CI 0.97–1.28; $p < 0.001$) for linear size, 0.98 (95% CI 0.95–1.007; $p < 0.001$) for surface area, and 1.00 (95% CI 0.99–1.003; $p < 0.001$) for volume (Fig. 1A).

Heterogeneity statistics for linear size were $I^2 = 98.7\%$, $\tau^2 = 0.024$, and Cochran's $Q = 51.30$ ($p < 0.001$), with two missing studies on trim-and-fill analysis (minimal change in OR). For surface area the statistics were $I^2 = 16.21\%$, $\tau^2 = 0.0003$, and Cochran's $Q = 13.15$ ($p = 0.07$), with four missing studies on trim-and-fill analysis (minimal change in OR). For volume the statistics were $I^2 = 38.5\%$, $\tau^2 = 0.00$, and Cochran's $Q = 40.1$ ($p = 0.0002$), with eight missing studies on trim-and-fill analysis (minimal change in OR). The full analysis is reported in Supplementary Section 5.1.

3.3. Meta-analysis of prognostic accuracy for the stone-free rate

Studies reporting on prognostic accuracy included 12 on linear size [17,19–24,27,30,33,34,37], seven on surface area [17,18,20,22,35–37], and 16 on volume [14,17,18,20–24,26,27,30,32–35,37] that were eligible for meta-analysis.

Table 1 – Stone characteristics in the studies included in the review

Study	IVN	*N*	Stone size[a]			Density (HU)	Stones	Stone location
			Linear (mm)	Area (mm^2)	Volume (mm^3)		(*n*)	(*n*)
Akkas 2021 [14]	ESWL	346	8.6 ± 2.8	N/A	286±70	1079 ± 236	1	PU 173, MU 84, DU 90
Bandi 2009 [15]	ESWL	94	11.2 (range 8–15)	N/A	273.8 (range 70–770)	475 (range 177–812)	1	RP 15, LP 32, MP 7, UP 12, PU 28
Cui 2021 [16]	ESWL	72	6.92 (range 5.7–8.9)	N/A	113 (range 62–276)	N/A	1	RP 29, LP 33, MP 7, UP 2
Diamand 2017 [17]	RIRS	67	14.39 ± 9.57	153.56 ± 312.75	1770.61 ± 3841.71	564.25 ± 183.09	1	LP 36, multiple calyces 31, staghorn 12
Ergani 2021 [18]	RIRS	184	N/A	150 ± 100	2460 ± 2820	1043.7 ± 345.65	1.23 ± 0.54	RP 85, LP 55, MP 10, UP 8, PU 14
Geng 2014 [19]	ESWL	328	9.05 ± 3.1	44 ± 35.5	N/A	576 ± 256	1	N/A
Guner 2021 [34]	RIRS	170	11.8 ± 2.95	N/A	400 ± 272	1036 ± 366	1	PU 54, MU 51, DU 65
Inoue 2023 [32] [b]	RIRS	305	SF: 14.0 (10.0–21.0) nSF: 24.5 (15.0–32.0)	N/A	SF 325.5 (156.6–636.2) nSF 767.5 (439.7–1311.5)	SF: 1045 (704.0–1385.0) nSF: 1259 (972.0–1480.0)	1	LP 305
Ito 2012 [37]	RIRS	238	26.67±14.36	206.345 ± 150.37	2092.75 ± 2733.905	N/A	2.55 ± 1.95	LP 155, non-LP 83
Ito 2013 [35]	RIRS	187	N/A	209.9 ± 216.2	1394.6 ± 1762.335	N/A	3.235 ± 2.39	LP 68, MP 15, U 104
Ito 2013 [30]	RIRS	243	<20 mm 12.26 ± 4.53 ≥20 mm 37.66 ± 16.23	N/A	<20 mm 563.62 ± 453 ≥20 mm 3080.56 ± 3837.85	N/A	≥1	LP 158, non-LP 85
Kobayashi 2022 [27]	ESWL	193	8.4 ± 2.8	N/A	170 ± 160	672.8 ± 213.1	1	Renal 40, PU 111, MU 5, DU 37
Langenauer 2018 [20]	ESWL	312	12.75 (range 5.85-50.9)	348.65 (range 87–3026.8)	269.8 (range 40.75–5958.6)	506.75 (range 195.4–936.35)	1	LP 99, MP 57, UP 34, PU 116, DU 6
Lee 2015 [36]	ESWL	145	8.85 ± 2.45	41.5 ± 25.5	N/A	509.6±234	1	LP 48, non-LP 97
Merigot de Treigny 2015 [28]	RIRS	142	27 ± ± 29	N/A	599 ± 980	N/A	≥1	UP 36, MP 38, LP 95, RP 33
Oktay 2022 [33]	ESWL	220	N/A (median: success 9.5, failure 12.3)	N/A	502 ± 788 (median: success 181, failure 390)	332 ± 96 Success 317 ± 107.5 Failure 341.8 ± 87	1	Calyces 35, RP 79, PU 93, MU 13
Park 2016 [21]	ESWL	223	8.5 ± 5.2	N/A	337.2 ± 835.35	834.55 ± 378.25	1	PU 223
Tailly 2020 [22]	PCNL	313	N/A	554.3 ± 407.6	9470 ± 8950	877.5 ± 358.8	1	Renal 306, staghorn 125, RP 221, LP 62, MP 6, UP 17, PU 7
Umemoto 2021 [23]	PCNL	37	24	N/A	10 700 (range 1100–50 200)	1069 (range 405–1486)	1	Renal 37, staghorn 16
Vuruskan 2022 [31]	RIRS/ mPCNL	RIRS 244 mPCNL 271	RIRS: 16.7 ± 2.5 mPCNL: 17.1 ± 2.3	N/A	RIRS: 1192.8 ± 679±2 mPCNL: 1407.5 ± 712.8	N/A	≥1	Renal 515
Wang 2005 [24]	ESWL	80	N/A	N/A	N/A	N/A	N/A	LP 57, non-LP 23
Waqas 2017 [25]	ESWL	74	9.6 ± 3.42	39.41 ± 27.3	238.81 ± 222.34	625.97 ± 275.43	1	Renal 74, PU 59, MU 7, DU 8
Waqas 2018 [29]	ESWL	203	11.77 ± 4.58	69.44 ± 50.96	755.99 ± 1015.12	796.37 ± 328.77	1	UP 27, MP 45, LP 105, RP 26
Xun 2018 [26]	ESWL	100	8	N/A	160	749 (range 560–940)	1	PU 100

IVN = intervention; ESWL = extracorporeal shockwave lithotripsy, PCNL = percutaneous nephrolithotomy, mPCNL = mini PCNL; RIRS = retrograde intrarenal surgery, N/A = not available; SF = stone-free; nSF = not stone-free; LP = lower pole; MP = mid-pole; UP = upper pole; U = ureteric; DU = distal ureter; MU = mid-ureter; PU = proximal ureter.

[a] Results are reported as mean ± standard deviation unless otherwise indicated.

[b] Results are reported as median (interquartile range).

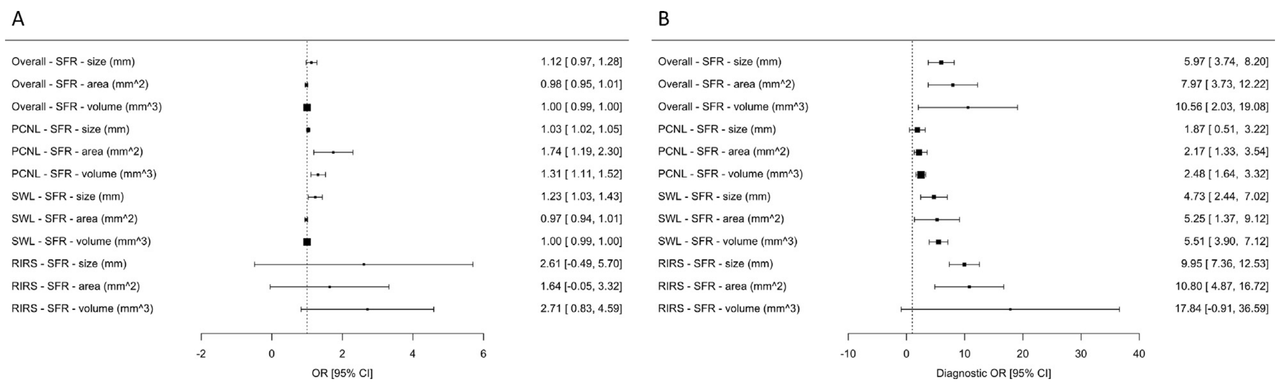

Fig. 1 – Summary forest plots of meta-analysis results for models based on measurement metrics for the stone-free rate (SFR). (A) ORs generated via meta-analysis of adjusted ORs from studies reporting multivariable logistic regression results. (B) Diagnostic ORs generated via meta-analysis of prognostic accuracy. Reference lines denote OR = 1 (ie, no difference). OR = odds ratio; CI = confidence interval; PCNL = percutaneous nephrolithotomy; SWL = shockwave lithotripsy; RIRS = retrograde intrarenal surgery.

Examination of initial ROC curves from the meta-analysis revealed that the study by Vuruskan et al [31] was clearly an outlier for volume (specificity 0.27, sensitivity 0.57). On examination of the original paper, the ROC curve appeared to be inverted and therefore could not be included in this analysis.

Overall DORs for stone-free status following any intervention were 5.97 (95% CI 3.74–8.20; $p < 0.001$; overall RoB high) for linear size, 7.97 (95% CI 3.73–12.22; $p < 0.001$; overall RoB high) for surface area, and 10.56 (95% CI 2.03–19.08; $p = 0.015$; overall RoB high) for volume (Fig. 1B). Summary AUC and sensitivity and specificity values are reported in Supplementary Table 2. Meta-analysis ROC curves are shown in Figure 2.

3.4. *Comparison of prognostic accuracy*

Studies that reported on the prognostic accuracy of two or more metrics were included in a statistical comparison of AUC values (Table 2). Four studies reported on linear size versus surface area [17,20,22,37], ten studies on linear size versus volume [17,20–23,27,30,33,34,37], and six studies on surface area versus volume [17,18,20,22,35,37].

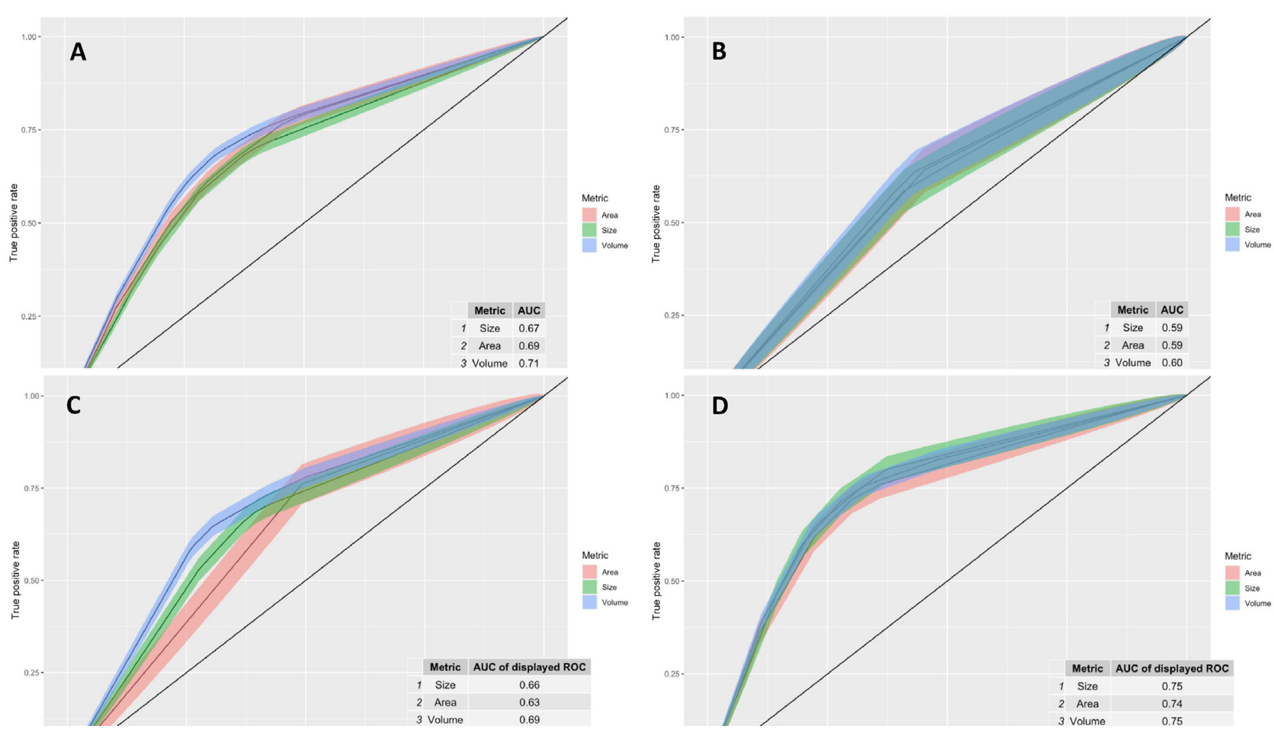

Fig. 2 – Receiver operating characteristic (ROC) curves with confidence intervals for prognostic accuracy. (A) Overall, (B) percutaneous nephrolithotomy, (C) extracorporeal shockwave lithotripsy, and (D) ureteroscopy. The reference line denotes AUC = 0.5 (ie, no difference). AUC = area under the ROC curve.

Table 2 – Comparison of AUC values according to the DeLong test

Modality	Comparison	Studies	AUC	DeLongp value
Overall	Linear size vs area	4	0.68 vs 0.69	0.24
	Linear size vs volume	10	0.67 vs 0.71	**<0.001**
	Area vs volume	6	0.69 vs 0.69	0.95
Percutaneous nephrolithotomy	Linear size vs area	1	0.59 vs 0.59	0.66
	Linear size vs volume	2	0.59 vs 0.60	0.27
	Area vs volume	1	0.59 vs 0.60	0.53
Extracorporeal shockwave lithotripsy	Linear size vs area	1	0.62 vs 0.63	0.07
	Linear size vs volume	4	0.66 vs 0.69	**0.05**
	Area vs volume	1	0.63 vs 0.64	0.78
Retrograde intrarenal surgery	Linear size vs area	2	0.77 vs 0.79	**<0.001**
	Linear size vs volume	4	0.75 vs 0.79	**<0.001**
	Area vs volume	4	0.74 vs 0.73	0.97

AUC = area under the receiver operating characteristic curve.

In the overall analysis, volume had a significantly higher AUC than linear size ($p < 0.001$), but there were no differences between other metrics.

3.5. Subanalyses by treatment modality

3.5.1. PCNL

Overall ORs for stone-free status following PCNL were 1.03 (95% CI 1.02–1.05; $p < 0.001$) for linear size, 1.74 (95% CI 1.19–2.30; $p < 0.001$) for surface area, and 1.31 (95% CI 1.11–1.51; $p < 0.001$) for volume (Fig. 1A). Heterogeneity statistics for linear size, surface area, and volume were $I^2 = 0\%$, $\tau^2 = 0$, and Cochran's $Q = 0$ ($p = 1.0$), as there was only one study.

Overall DORs were 1.87 (95% CI 0.51–3.22; $p = 0.007$; overall RoB high) for linear size, 2.17 (95% CI 1.68–2.66; $p < 0.001$; overall RoB high) for surface area, and 2.48 (95% CI 1.64–3.32; $p < 0.001$; overall RoB high) for volume (Fig. 1B). Heterogeneity statistics are detailed in Supplementary Section 6.1.1.2.

Comparison revealed that no one metric had better prognostic accuracy than another (Table 2). The full analysis is reported in Supplementary Section 6.1.

Only one study defined a cutoff for stone-free status following PCNL according to ROC analysis, which was 15 000 mm^3 [23].

3.5.2. ESWL

Overall ORs for stone-free status following ESWL were 1.23 (95% CI 1.03–1.43; $p < 0.001$) for linear size, 0.97 (95% CI 0.94–1.006; $p < 0.001$) for surface area, and 1.00 (95% CI 0.99–1.002; $p < 0.001$) for volume (Fig. 1A). Heterogeneity statistics are provided in Supplementary Section 6.2.1.3.

Overall DORs for stone-free status following ESWL were 4.73 (95% CI 2.44–7.02; $p < 0.001$; overall RoB high) for linear size, 5.25 (95% CI 1.37–9.12; $p = 0.08$; overall RoB high) for surface area, and 5.51 (95% CI 3.90–7.12; $p < 0.001$, overall RoB high) for volume (Fig. 1B). Heterogeneity statistics are provided in Supplementary Section 6.2.1.2.

The AUC was significantly higher for volume than for linear size ($p = 0.05$), but there were no other significant differences between the size metrics (Table 2). The full analysis is reported in Supplementary Section 6.2.

There were multiple study-defined cutoff values for the stone size metrics (Table 3).

Table 3 – Study-defined size cutoffs for extracorporeal shockwave lithotripsy and ureterorenoscopy by measure

Study	Linear (mm)	Area (mm^2)	Volume (mm^3)
Extracorporeal shockwave lithotripsy			
Akkas 2021 [14]	NA	NA	311
Kobayashi 2022 [27]	8.5	NA	290
Langenauer 2018 [20]	8	41.7	354.8
Lee 2015 [36]	NA	48	NA
Oktay 2022 [33]	10.8	NA	293
Park 2016 [21]	7	NA	NA
Wang 2005 [24]	12	NA	700
Xun 2018 [26]	10	NA	158.5
Ureterorenoscopy			
Diamand 2017 [17]	11.8	64.3	532
Ergani 2021 [18]	NA	148	1540
Guner 2021 [34]	11.5	NA	257
Ito 2012 [37]	23	150	1120
Ito 2013 [35]	NA	125	840
Ito 2013 [30]	20	NA	NA
Merigot de Treigny 2015 [28]	20	NA	1131
Vuruskan 2022 [31]	NA	NA	1416

NA = not applicable.

3.5.3. URS

Overall ORs for stone-free status following URS were 2.61 (95% CI −0.49 to 5.70; $p = 0.10$) for linear size, 1.64 (95% CI −0.05 to 3.32; $p = 0.06$) for surface area, and 2.71 (95% CI 0.83–4.59; $p = 0.005$) for volume (Fig. 1A). Heterogeneity statistics are detailed in Supplementary Section 6.3.1.3.

Overall DORs for stone-free status following URS were 9.95 (95% CI 7.36–12.53; $p < 0.001$; overall RoB low) for linear size, 10.80 (95% CI 4.87–16.72; $p = 0.0004$; overall RoB high) for surface area, and 17.84 (95% CI −0.91 to 36.59; $p = 0.06$; overall RoB low) for volume (Fig. 1B). Heterogeneity statistics are detailed in Supplementary Section 6.3.1.2.

The AUCs were significantly higher for surface area and volume than for linear size (both $p < 0.001$; Table 2). The full analysis is presented in Supplementary Section 6.3.

There were multiple study-defined cutoff values for the stone size metrics (Table 3).

3.6. Complication rates

None of the studies identified in the systematic review reported on the influence of the stone burden on complication rates.

3.7. Operative time

Two studies reported on the influence of stone burden on operative time [22,34]. Guner et al [34] found that both stone size and stone volume were significantly correlated to operative time, with Spearman's ρ values of 0.510 and 0.540, respectively. A regression analysis by Tailly et al [22] showed that all measures for stone burden were predictive of operative time. In multivariate models, surface area and volume were independent predictors of operative time, whereas Hounsfield units and stone complexity were also significant in the CSD model for prediction of operative time.

3.8. RoB analysis

RoB assessment results according to the QUIPS tool are shown in Supplementary Table 3. In summary, most studies were at low RoB for domain 1 (study participation), domain 4 (outcome measurement), and domain 6 (statistical analysis and reporting). Several studies were ranked as having moderate to high RoB for domain 2 (study attrition), domain 3 (prognostic factor measurement), and domain 5 (study confounding).

4. Discussion

Stone burden is arguably the most important factor guiding endourologists in their choice of procedure for treatment of urolithiasis. Stone burden has been historically described in terms of the maximum diameter or CSD, which is also used in urolithiasis guidelines globally [3]. As stones may vary in shape, it makes sense that this linear measurement may not provide sufficient information about the true stone burden, while stone volume is unquestionably the most accurate representation of the amount of stone to be treated. However, it is still unclear if a more accurate measurement of stone burden would allow more accurate prediction of outcomes after urolithiasis interventions.

To the best of our knowledge, this is the first systematic review and meta-analysis to evaluate whether any of the measures of stone burden is superior as a predictor of the stone-free rate, operative time, or complication rates after any type of stone treatment.

Although we were able to include 24 studies in the systematic review, only a few studies per procedure were available for evaluating if any measurement method is superior to the other methods in predicting stone-free status after an intervention. Only two studies were available for PCNL, and four studies each for URS and ESWL.

While OR and DOR values cannot provide any information on whether any measurement method is superior, a higher DOR for a specific measure of the stone burden does indicate a better-performing test. A reason for the low DOR for PCNL may be the complexity of the procedure and the fact that multiple other factors such as stone complexity, location, and plurality, which are not captured in this meta-analysis, may influence the result more than for URS and ESWL.

However, our statistical analysis of AUC values does support the intuitive hypothesis that stone volume may be a better measure to consider when treating stones. Overall and for the ESWL and URS modalities, volume performed better than linear stone size in predicting stone-free status. While our overall AUC values of 0.67 for linear size, 0.69 for surface area, and 0.71 for volume are very close, from a statistical perspective the AUCs for surface area and volume are significantly better than the AUC for linear size. However, whether or not this translates into a clinically relevant difference or whether a different measure of stone burden would change the choice of treatment or treatment planning remains to be evaluated.

Considering the limited number of studies available for comparison, the RoB, the relatively small sample size in each of the studies, and the close range for the AUC values, the superiority of volume over linear size should be interpreted with caution. Additionally, it should be highlighted that the studies included in this meta-analysis have treated stones of considerably different sizes, in different locations, and used different definitions of success as is clear from Supplemental Table 1 and Table 1. Indeed, for the procedure of PCNL, no superiority could be identified, while these are without question, the largest and most complexly shaped stones treated. This may in part be due to the fact that more complex stones will have a lower volume compared to ellipsoid-shaped stones of the same maximum diameter, while these patients have a higher risk of residual fragments after the procedure. For the procedures of ESWL and URS, with which usually smaller stone burdens are treated, volume then again was a significantly more accurate predictor of stone-free status than linear size.

Interestingly, and in contrast to this finding, both Ito et al [30] and Merigot de Treigny et al [28] found that stone volume was not a more accurate predictor of the outcome in question for stones <2 cm in size, whereas volume was clearly the more predictive factor for stone-free status after URS for larger stones. Yamashita et al [38] similarly demonstrated that stone volume was not an independent predictor of the stone-free rate for ureteral stones, while it was for kidney stones, hypothesizing this could be because of the smaller size of ureteral stones. These reports indicate that stone volume may not be the best outcome predictor for stones of all sizes and in all locations.

When evaluating the literature on predictors of success after stone interventions, it is clear from the nomograms and scoring systems available that stone burden, regardless of measure, is not the only variable to consider [39]. Interestingly, most of the scoring systems use CSD as the measure of stone burden, with surface area or volume rarely used as a parameter. All the scoring systems and nomograms were developed in the CT imaging era, which means that, in theory, these data could have been available. However, it is possible that acquisition of these parameters may have been more difficult in the past, while multiple software packages are now available for automatic or at least easy acquisition of the relevant information from multiplanar CT images [4].

Although we did not have sufficient data available for meta-analysis, intuitively, surface area and volume would seem to be more accurate predictors of the operative time needed in comparison to CSD. Once again, many other variables should be taken into account, such as stone complexity, number of stones to be treated, stone density, and pelvicalyceal anatomy, among others.

The most important limitation of our analysis is the large heterogeneity between the studies that cannot be overlooked. The definitions of stone-free status and timing of this assessment varied widely between studies. The definitions of maximum diameter, surface area, and stone volume and the methods for determining these parameters, whether way formula, tracing, measurement, or software, also differed among the studies included in the meta-analysis. This makes pooling and comparison of results difficult. In addition, treatment strategies and equipment may have varied significantly between studies, influencing the outcomes. It should also be mentioned that all the OR and AUC values were developed using just a training set and no validation data set was used to strengthen the outcomes, which carries a risk of model overfitting to the training set. Lastly, the statistical significance of the model does not necessarily indicate good prognostic accuracy, as the AUC values are mainly <0.7.

Volume measurements from noncontrast low-dose CT scans are easily achieved via scanner-specific software or even free open-source software that can easily be downloaded on any computer. Although it is still unclear whether knowledge of stone volume will prompt clinicians to change their practice in comparison to CSD, this at least has lowered the threshold to assess and report stone volume in patients with urolithiasis who may need to undergo surgical intervention. We would therefore suggest that researchers should report multiple measures of the stone burden when possible.

While cutoffs have been based on CSD for a few decades, reference ranges for surface area and volume to help guide surgeons to a certain procedure are largely unknown. The cutoffs that have been reported in the literature included in this analysis may serve as preliminary information towards a better understanding of these cutoffs in future literature. These cutoff values however are based on small series and cannot be pooled due to the heterogeneity of stone locations, treatment modalities, and strategies as well as many other confounding factors.

Further research is needed to strengthen the data and effectively compare different stone burden measures and their predictive accuracy for stone-free status, operative time, and complication rates. To this end, a prospective study (PRAISE; Predictive Accuracy of Initial Stone burden Evaluation; ClinicalTrials.gov NCT04746378) of different measures of stone size for all interventions may prove useful.

5. Conclusions

Our systematic review and meta-analysis revealed some evidence that the accuracy of predictive models based on multiplanar measurements of stone volume and surface area are statistically superior to models based on single-plane measurements alone for prediction of stone-free status after intervention. On subanalysis, there seemed to be no significant differences between metrics for stone-free status following PCNL. However, volume is a significantly better predictor of stone-free status in comparison to linear size following both ESWL and URS. There was insufficient evidence for meta-analysis of complication rates and operative time. Further studies are needed to strengthen these findings before guidelines can be changed to reflect new measurement practices. We call on researchers to report multiple measures of stone burden whenever possible.

Author contributions: Andreas Skolarikos had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Geraghty, Pietropaolo, Tzelves, Skolarikos, Tailly.
Acquisition of data: Geraghty, Pietropaolo, Tzelves, Skolarikos, Tailly.
Analysis and interpretation of data: Geraghty, Pietropaolo, Tzelves, Skolarikos, Tailly.
Drafting of the manuscript: Geraghty, Pietropaolo, Tzelves, Skolarikos, Tailly.
Critical revision of the manuscript for important intellectual content: Geraghty, Pietropaolo, Tzelves, Lombardo, Jung, Neisius, Petrik, Somani, Davis, Gambaro, Boissier, Skolarikos, Tailly.
Statistical analysis: Geraghty, Pietropaolo, Tzelves, Skolarikos, Tailly.
Obtaining funding: Skolarikos, Tailly.
Administrative, technical, or material support: Geraghty, Pietropaolo, Tzelves, Skolarikos, Tailly.
Supervision: Skolarikos, Tailly.
Other: None.

Financial disclosures: Andreas Skolarikos certifies that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (eg, employment/affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), are the following: None.

Funding/Support and role of the sponsor: This work was supported by the European Association of Urology. The sponsor played a role in collection and management of the data and in preparation and review of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.euros.2024.10.024.

References

[1] Chewcharat A, Curhan G. Trends in the prevalence of kidney stones in the United States from 2007 to 2016. Urolithiasis 2020;3:278–1213. https://doi.org/10.1007/s00240-020-01210-w.
[2] Skolarikos A, Jung G, Neisius A, et al. EAU guidelines on urolithiasis. Arnhem, The Netherlands: European Association of Urology; 2024.
[3] Jiang P, Xie L, Arada R, Patel RM, Landman J, Clayman RV. Qualitative review of clinical guidelines for medical and surgical management of urolithiasis: consensus and controversy 2020. J Urol 2021;205:999–1008. https://doi.org/10.1097/ju.0000000000001478.
[4] Panthier F, Kutchukian S, Ducousso H, et al. How to estimate stone volume and its use in stone surgery: a comprehensive review. Actas

Urol Esp 2024;48:71–8. https://doi.org/10.1016/j.acuroe.2023.08.009.

[5] Moher D, Liberati A, Tetzlaff J, Altman DG. PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009;6:e1000097.

[6] Knoll T, Omar MI, MacLennan S, et al. Key steps in conducting systematic reviews for underpinning clinical practice guidelines: methodology of the European Association of Urology. Eur Urol 2018;73:290–300. https://doi.org/10.1016/j.eururo.2017.08.016.

[7] Gheibi S, Mahmoodzadeh A, Kashfi K, Jeddi S, Ghasemi A. Data extraction from graphs using Adobe Photoshop: applications for meta-analyses. Int J Endocrinol Metab 2019;17:e95216.

[8] Higgins JPT, Altman DG, Gotzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928. https://doi.org/10.1136/bmj.d5928.

[9] Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. Ann Intern Med 2013;158:280–6. https://doi.org/10.7326/0003-4819-158-4-201302190-00009.

[10] Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. BMJ Mental Health 2019;22:153. https://doi.org/10.1136/ebmental-2019-300117.

[11] Doebler P, Holling H, Sousa-Pinto B. Meta-analysis of diagnostic accuracy with mada. The Comprehensive R Archive Network; 2015. https://cran.r-project.org/web/packages/mada/vignettes/mada.pdf.

[12] Pérez-Fernández S, Martínez-Camblor P, Filzmoser P, Corral N. nsROC: an R package for non-standard ROC curve analysis. R J 2019;10:55. https://doi.org/10.32614/rj-2018-043.

[13] DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837–45.

[14] Akkas F, Culha MG, Ayten A, et al. A novel model using computed tomography parameters to predict shock wave lithotripsy success in ureteral stones at different locations. Actas Urol Esp 2022;46:114–21. https://doi.org/10.1016/j.acuroe.2021.01.011.

[15] Bandi G, Meiners RJ, Pickhardt PJ, Nakada SY. Stone measurement by volumetric three-dimensional computed tomography for predicting the outcome after extracorporeal shock wave lithotripsy. BJU Int 2009;103:524–8. https://doi.org/10.1111/j.1464-410x.2008.08069.x.

[16] Cui HW, Tan TK, Christiansen FE, Osther PJS, Turney BW. The utility of automated volume analysis of renal stones before and after shockwave lithotripsy treatment. Urolithiasis 2021;49:219–26. https://doi.org/10.1007/s00240-020-01212-8.

[17] Diamand R, Idrissi-Kaitouni M, Coppens E, Roumeguère T, Legrand F. Évaluation de la taille des calculs urinaires avant urétéroscopie souple: quelle mesure choisir? Prog Urol 2018;28:62–70. https://doi.org/10.1016/j.purol.2017.09.014.

[18] Ergani B, Kozacioglu Z. Effects of three-dimensional measurement of the urinary stone size on the surgical outcomes of retrograde intrarenal stone surgery. J Urol Surg 2021;8:111–7. https://doi.org/10.4274/jus.galenos.2021.0032.

[19] Geng JH, Tu HP, Shih PMC, et al. Noncontrast computed tomography can predict the outcome of shockwave lithotripsy via accurate stone measurement and abdominal fat distribution determination. Kaohsiung J Med Sci 2015;31:34–41. https://doi.org/10.1016/j.kjms.2014.10.001.

[20] Langenauer J, Betschart P, Hechelhammer L, et al. Advanced non-contrasted computed tomography post-processing by CT-calculometry (CT-CM) outperforms established predictors for the outcome of shock wave lithotripsy. World J Urol 2018;36:2073–80. https://doi.org/10.1007/s00345-018-2348-x.

[21] Park HS, Gong MK, Yoon CY, Moon DG, Cheon J, Choi YD. Computed tomography-based novel prediction model for the outcome of shockwave lithotripsy in proximal ureteral stones. J Endourol 2016;30:810–6. https://doi.org/10.1089/end.2016.0056.

[22] Tailly T, Nadeau BR, Violette PD, et al. Stone burden measurement by 3D reconstruction on noncontrast computed tomography is not a more accurate predictor of stone-free rate after percutaneous nephrolithotomy than 2D stone burden measurements. J Endourol 2020;34:550–7. https://doi.org/10.1089/end.2019.0718.

[23] Umemoto T, Hasegawa M, Kawakami M, et al. Impact of stone volume on treatment outcomes of percutaneous nephrolithotripsy. Asian J Endosc Surg 2022;15:599–607. https://doi.org/10.1111/ases.13059.

[24] Wang LJ, Wong YC, Chuang CK, et al. Predictions of outcomes of renal stones after extracorporeal shock wave lithotripsy from stone characteristics determined by unenhanced helical computed tomography: a multivariate analysis. Eur Radiol 2005;15:2238–43. https://doi.org/10.1007/s00330-005-2742-9.

[25] Waqas M, Khan MA, Iqbal MW, Akbar MK, Saqib IUD, Akhter S. Non-contrast computed tomography scan based parameters of ureteric stones affecting the outcome of extracorporeal shock wave lithotripsy. Cureus 2017;9:e1227. https://doi.org/10.7759/cureus.1227.

[26] Xun Y, Li J, Geng Y, et al. Single extracorporeal shock-wave lithotripsy for proximal ureter stones: can CT texture analysis technique help predict the therapeutic effect? Eur J Radiol 2018;107:84–9. https://doi.org/10.1016/j.ejrad.2018.08.018.

[27] Kobayashi M, Waseda Y, Fuse H, Takazawa R. Variables measured on three-dimensional computed tomography are preferred for predicting the outcomes of shock wave lithotripsy. World J Urol 2022;40:569–75. https://doi.org/10.1007/s00345-021-03861-9.

[28] Merigot de Treigny O, Nasr EB, Almont T, et al. The cumulated stone diameter: a limited tool for stone burden estimation. Urology 2015;86:477–81. https://doi.org/10.1016/j.urology.2015.06.018.

[29] Waqas M, Saqib I, Jamil MI, Khan MA, Akhter S. Evaluating the importance of different computed tomography scan-based factors in predicting the outcome of extracorporeal shock wave lithotripsy for renal stones. Investig Clin Urol 2018;59:25–31. https://doi.org/10.4111/icu.2018.59.1.25.

[30] Ito H, Kawahara T, Terao H, et al. Utility and limitation of cumulative stone diameter in predicting urinary stone burden at flexible ureteroscopy with holmium laser lithotripsy: a single-center experience. PLoS One 2013;8:e65060. https://doi.org/10.1371/journal.pone.0065060.

[31] Vuruskan E, Dilek O, Karkin K, Unal U, Ayhan L, Sener NC. Volume should be used instead of diameter for kidney stones between 10 and 20 mm to determine the type of surgery and increase success. Urolithiasis 2022;50:215–21. https://doi.org/10.1007/s00240-022-01305-6.

[32] Inoue T, Hamamoto S, Okada S, et al. Single-session impact of high-power laser with Moses technology for lower pole stones in retrograde intrarenal surgery: retrospective study. J Clin Med 2022;12:301. https://doi.org/10.3390/jcm12010301.

[33] Oktay C, Çoraplı M, Tutuş A. The usefulness of the Hounsfield unit and stone heterogeneity variation in predicting the shockwave lithotripsy outcome. Diagn Interv Radiol 2022;28:187–92. https://doi.org/10.5152/dir.2022.20945.

[34] Guner E, Danacioglu YO, Akkas F, et al. Factors predicting duration and success of semirigid ureteroscopy for ureteral stones in different localizations. Arch Esp Urol 2021;74:335–42.

[35] Ito H, Kawahara T, Terao H, et al. Evaluation of preoperative measurement of stone surface area as a predictor of stone-free status after combined ureteroscopy with holmium laser lithotripsy: a single-center experience. J Endourol 2013;27:715–21. https://doi.org/10.1089/end.2012.0548.

[36] Lee HY, Yang YH, Lee YL, et al. Noncontrast computed tomography factors that predict the renal stone outcome after shock wave lithotripsy. Clin Imaging 2015;39:845–50. https://doi.org/10.1016/j.clinimag.2015.04.010.

[37] Ito H, Kawahara T, Terao H, et al. The most reliable preoperative assessment of renal stone burden as a predictor of stone-free status after flexible ureteroscopy with holmium laser lithotripsy: a single-center experience. Urology 2012;80:524–8. https://doi.org/10.1016/j.urology.2012.04.001.

[38] Yamashita S, Iwahashi Y, Deguchi R, Kikkawa K, Kohjimoto Y, Hara I. Three-dimensional mean stone density on non-contrast computed tomography can predict ureteroscopic lithotripsy outcome in ureteral stone cases. Urolithiasis 2020;48:547–52. https://doi.org/10.1007/s00240-020-01178-7.

[39] 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;35:1863–82. https://doi.org/10.1089/end.2021.0124.