

Comparing Performance of Automated and Semi-automated Methods for Measuring Kidney Stone Volumes: Can Machine Replace Man?

Authors: *Jackson J.S. Cabo,¹ Andrew Amenyogbe,¹ Karen L. Stern¹

1. Department of Urology, Mayo Clinic Arizona, Phoenix, USA

*Correspondence to cabo.jackson@mayo.edu

Disclosure: Cabo and Stern have received an educational grant from Calyxo Inc. for this study, with payments to the institution. Amenyogbe has declared no conflicts of interest.

Acknowledgements: The authors would like to thank Victoria Edmonds, Mayo Clinic Arizona, Phoenix, USA; Christopher Ballantyne, Bon Secours, Greenville, South Carolina, USA; and the Mayo Clinic AI Core for collaboration on this work.

Keywords: Kidney stone, stone volume, suction ureteroscopy.

Citation: EMJ Urol. 2026;14[1]:52-53.
<https://doi.org/10.33590/emjurol/77C92T7P>

BACKGROUND AND AIMS

With recent emphasis placed on using stone volumes in surgical outcomes research,¹⁻⁶ there is a need to assess the accuracy and utility of available tools. The authors sought to assess the accuracy of AI-calculated volumes compared to semi-automated methods^{5,6} and evaluate whether cumulative stone diameter, semi-automated (SA) stone volume calculation, or a fully automated AI method was a better predictor of stone-free status following suction-augmented ureteroscopy (Figure 1).

METHODS

A total of 171 CT scans were included (96 pre-operation and 75 post-operation). Cumulative stone diameter was measured manually. Stone volumes were assessed using two semi-automated segmentation applications (QSAS [Mayo Clinic, Rochester, Minnesota, USA], 3D-Slicer [Chitubox, Schenzhen, China]), which require

investigator annotation of the region of interest. A Mayo Clinic-developed AI programme calculated stone volumes in a fully automated fashion. Pearson correlation assessed the association between AI-estimated and semi-automated volumes. Sensitivity and specificity of the AI model were assessed for absolute stone-free rate (SFR-Grade A) on post-operative scans. Receiver operating characteristic analysis evaluated accuracy of pre-operative stone burden metrics in predicting stone-free status using SFR-Grade A (no residual fragments) and SFR-Grade C, stone-free rate with residual fragments between 2.1-4.0 mm in size (no fragments >4 mm) criteria.

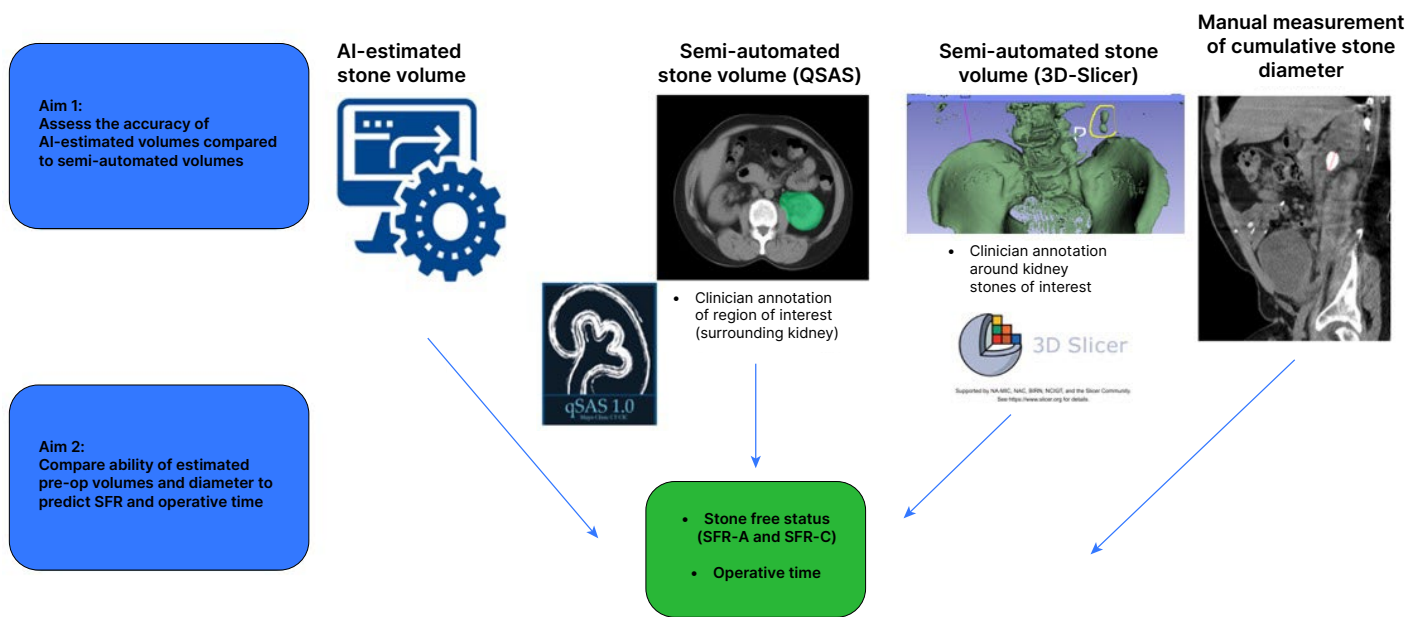
RESULTS

AI-estimated stone volumes showed strong linear correlations with both 3D-Slicer (R=0.95; p<0.001; mean difference: -0.31 mm³; interquartile range: -13.06-25.81 mm³) and QSAS (R=0.95; p<0.001; mean difference: 0 mm³; interquartile range: -12.0-8.0 mm³) calculated volumes. Among post-operative scans, strong correlations persisted with QSAS (R=0.88; p<0.001) and 3D-Slicer (R=0.86; p<0.001).

Among 59 patients with eligible pre- and post-operative imaging, the AI model demonstrated a sensitivity of 85.7% and specificity of 88.9% for SFR-Grade A. In cases of incorrect stone-free determination, parenchymal stones were identified in both false-negative cases, and the largest false-positive residual burden (16 mm³) occurred in the right moiety of a horseshoe kidney.

Among 69 patients with a single ureteroscopy and complete follow-up imaging within 3 months, no significant difference was found between preoperative volumetric and diameter measurements in predicting SFR-Grade A. Cumulative pre-operative diameter outperformed

Figure 1: Study schema and primary aims.



SFR: Stone-Free Rate; SFR-A: Stone-Free Rate-Grade A; SFR-C: Stone-Free Rate-Grade C.

QSAS-calculated volume in predicting SFR-Grade C (area under the curve: 0.78 versus 0.62; DeLong Test $p=0.037$). Sub-analysis of flexible and navigable ureteric access sheath cases revealed no difference between any measurement in predicting SFR-Grade A or -Grade C.

All pre-operative stone measurements correlated significantly with operative time; however, semi-automated volumes from 3D-Slicer ($R^2=0.41$) and QSAS ($R^2=0.37$) explained more variation in operative time than cumulative diameter ($R^2=0.25$) or AI-estimated volume ($R^2=0.24$).

CONCLUSION

A fully automated, AI-driven method for stone volume determination was highly accurate, offering an efficient option for estimating pre-operative or residual stone burden.

The AI model was 85.7% sensitive and 88.9% specific for determining SFR-Grade A without clinician annotation, with errors concentrated in small low-

attenuation stones and anatomic variants. Pre-operative volumetric measurements did not outperform cumulative diameter in predicting stone-free status.

References

1. Cabo J et al. Comparing performance of automated and semi-automated methods for measuring kidney stone volumes: can machine replace man? Abstract A0417. EAU Congress, 13-16 March, 2026.
2. Cumanas AD et al. Efficient and accurate computed tomography-based stone volume determination: development of an automated artificial intelligence algorithm. *J Urol.* 2024;211(2):256-65.
3. Geraghty R et al. 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. *Eur Urol Open Sci.* 2024;71:22-30.
4. Matlaga BR et al. Residual stone volume predicts health care consumption and stone events: analysis of two-year results of the ASPIRE study. *J Endourol.* 2026;40(3):328-34.
5. Slots C et al. Stone metrics: is stone volume the new king? *Curr Opin Urol.* 2026; DOI:10.1097/MOU.0000000000001376.
6. Sajdak G et al. Rock solid measurements: a comparison of three methods of kidney stone volume assessment. *J Endourol.* 2025;39(8):856-61.