The Canadian Anatomic Kidney Score: Assessment of reproducibility among surgeons and trainees across Canada

William Luke[1,2], Danny Matti[1,2], Juliano Offerni[1,2,3], Victoria Lebedeva[4], Zoe Myers[5], Haider Abed[1,2], Pavel Roshanov[1,6], Carol Wang[1,6], Alp Sener[1,2], Patrick P. Luke[1,2]
[1]Schulich School of Medicine & Dentistry, Western University, London, ON, Canada; [2]Division of Urology, Department of Surgery, Western University, London, ON, Canada; [3]Maxy Rady College of Medicine, University of Manitoba, Winnipeg, MB, Canada; [4]London Health Sciences Centre, London, ON, Canada; [5]Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada; [6]Division of Nephrology, Department of Medicine, Western University, London, ON, Canada

Cite as: Luke W, Matti D, Offerni J, et al. The Canadian Anatomic Kidney Score: Assessment of reproducibility among surgeons and trainees across Canada. *Can Urol Assoc J* 2026 February 13; Epub ahead of print. http://dx.doi.org/10.5489/cuaj.9484

Published online February 13, 2026

Corresponding author: Dr. Patrick P. Luke, Schulich School of Medicine & Dentistry, Western University, London, ON, Canada; Patrick.Luke@lhsc.on.ca

***

KEY MESSAGES

- Current renal transplantation cost and availability is limited by pre-implantation graft assessment techniques.
- The Canadian Anatomic Kidney Score (CAKS) is a rapid, reproducible renal graft assessment system.
- Further evaluation is required to determine its role in high KDPI donors and its combinatorial value with other clinically available data.

ABSTRACT

Introduction: The Canadian Anatomic Kidney Score (CAKS) is a quantitative score used to assess donor kidneys, which has been shown to predict transplant outcomes. We hypothesized that the quantification of this score is highly reproducible across surgeons with varying levels of experience.

Methods: An anonymous survey was distributed to surgical attendings, fellows, and residents at teaching centers across Canada. The survey included photographs of six distinct kidneys (two images for each) taken during intraoperative back-table dissection. Participants evaluated the kidneys based on vascular features, anatomy, and sticky fat. Scores ranged from 0–2 points for each category. Comparisons of mean scores among residents, fellows, and staff were made using one-way ANOVA, and the intraclass correlation coefficient (ICC) was calculated for each scoring factor and overall.

Results: Among the 35 respondents, 570 sets of evaluations were performed, with a mean evaluation time of one minute 37 seconds per kidney. Across the 18 survey questions, there was

no significant difference in scoring among fellows, residents, and attendings, except for two of the six "sticky fat" image sets. The ICC for the CAKS overall was 0.78 when evaluated against a gold standard, and 0.80 when compared between raters, indicating excellent agreement between groups.
Conclusions: CAKS is reproducible by surgeons across different levels of training and various centers in Canada. This scoring system provides a reliable means to convey quantitative anatomic information between transplant professionals.

INTRODUCTION

Over the past decades, the gap between donor organ availability and recipient needs continues to widen.[1] To prevent the discard of viable grafts, optimal, multifactorial assessments are essential. Currently, donor suitability assessments include scores predicting transplant outcomes. One such score is the Kidney Donor Profile Index (KDPI), a quantitative measure designed to provide a cumulative risk of graft failure, validated within the North American population.[2–4] The KDPI utilizes 10 donor history criteria associated with transplant outcomes, but has limited predictive ability.[4,5] Notably, the KDPI does not account for physical and microscopic evaluations of the graft. Pre-implantation renal biopsy is another common assessment practice. But the various available scoring systems are limited by lack of accurate correlation with graft function and survival.[6–9] Renal biopsies also introduce delays of 2–8 hours in cold ischemic time and incur personnel costs for processing samples at various stages.

Intraoperative organ assessment, often used as a subjective adjunctive tool, has also been linked to organ discard rates despite the absence of standardized evaluation methods or strong supporting evidence.[10] To address these challenges, our group developed the Canadian Anatomic Kidney Score (CAKS)—a quantitative, multifactorial system shown to correlate with graft outcomes independent of KDPI or pathological scoring systems.[11] The CAKS evaluates macroscopic features such as adherent fat, atherosclerosis, and cysts/scars. Each of these factors independently has previously been shown to be associated with kidney disease. Adherent fat has been associated with chronic kidney disease progression in diabetic patients and operative difficulty during partial nephrectomy.[12,13] Macroscopic renal artery atherosclerosis is associated with graft discard and primary non-function.[10,14] Renal cysts/scars may be associated with glomerular sclerosis and worsened long-term function in grafts from living donors.[15] Our previous study demonstrated association between the CAKS and 1-year eGFR without adjustment for KDPI or Remuzzi Score. This association was further augmented when adjusting for KDPI and Remuzzi Score.[11]

As the CAKS represents an expense-free, validated assessment tool to accompany KDPI assessment, we further assess the reproducibility of the test among surgeons. We hypothesize

that this system is reproducible across surgeons with variable levels of experience, facilitating its implementation across institutions. If so, this system could improve the consistency and granularity of graft evaluations and enable improved communication between donor and recipient teams.

METHODS

This study received ethical approval from the institutional review board (REB-120254). An anonymous survey and scoring catalogue were distributed to attending kidney transplant surgeons (identified through the Canadian Society of Transplantation), surgical kidney transplant fellows, and a group of urology residents at teaching centres across Canada. Survey responses were collected using the Survey MonkeyÒ platform over a six-month period. Data was collected regarding participant surgical specialty, demographics, centre transplant volume, and perceived CAKS for each renal unit. Time per evaluation was calculated by comparing survey start and end timestamps. The surveys included a description of the visual component of the CAKS (Table 1) and six separate kidneys (two photos each) taken at the time of intra-operative back table dissection within the past year. An example is provided in Figure 1. The six donor kidneys were selected to provide a range of CAKS values to assess the reproducibility of the system. Gold standard scores were collected based on an agreement between the developers (PL, attending; JO, fellow) before the distribution of the surveys. The developers' assessments for each kidney are described in Table 2.

The difference between these values was evaluated in subgroups of residents, fellows and staff using a Kruskal-Wallis non-parametric one-way ANOVA of ranks for each question and for each CAKS factor.

We performed inter-rater reliability analyses to investigate the degree of consistency between the scores assigned by the individual rates and a predefined gold standard. To evaluate the agreement between raters and the gold standard, we calculated the Intraclass Correlation Coefficient (ICC) using the “psych” package in R version 4.3.2.[16] Reliability is categorized as poor, moderate, good, and excellent corresponding to values of <0.25, 0.25-0.50, 0.50-0.75, and >0.75, respectively.[17] We fit a two-way mixed-effects model tailored for a single measurement aimed at capturing consistency. This process was repeated for each rater. We calculated the mean ICC and its 95% confidence interval (CI). Raters with missing data were excluded when computing the ICC.

We then examined the level of agreement between raters. We calculated ICCs using a two-way mixed-effects model tailored for a single measurement aimed at capturing absolute agreement. We excluded from this analysis raters with missing responses and those who rated fewer than 3 images.

RESULTS

Out of 72 surgeons and surgical trainees invited to participate in the survey, 35 (49%) completed at least the demographic portion of the survey (Table 3). The median time required to complete

the survey was 9 minutes and 30 seconds, resulting in a mean of 1 minute 37 seconds per kidney evaluated. Of the participants, 23 had training in urology, 10 in general surgery, and two in vascular surgery. Sixty percent were attending physicians, 14% were fellows, and 26% were residents. Nearly half of centres perform 100-149 transplants per year, but there was a wide range of reported transplant volumes.

Overall, 570/620 (92%) of evaluations were completed by the respondents. Across 16/18 sets of images, the Kruskal-Wallis test revealed no significant differences in assessments between developers and respondents. However, in two sets of images (sticky fat evaluations) significant differences were observed. In both question sets, the less experienced trainees rated renal units less favourably than developers (median resident:fellow:staff:developer score; 0.5:0:0:0, $p = 0.011$ – 1.0:0.5:0.5:0, $p = 0.020$).

The ICC analyses (assessing the degree of consistency compared to the gold standard) revealed excellent consistency across the CAKS evaluation overall, with individual components demonstrating good reliability (Table 4). The ICC analyses evaluating the degree of absolute agreement among raters across the CAKS overall, also revealed excellent correlation (Table 5).

DISCUSSION

This study demonstrates that the CAKS is a reproducible scoring system, applicable across various levels of expertise through visual inspection of photographs.[18] The Banff working group on pre-implantation biopsy evaluated the ICC of 12 (median ICC = 0.46 - moderate) commonly reported histopathologic features. They found eight demonstrated moderate or worse reliability, with only the total number of glomeruli demonstrating excellent reliability.[17] The CAKS intraclass correlation coefficients demonstrated excellent reliability overall, both between raters and in comparison to the gold standard score (ICC=0.783, ICC=0.800 respectively). Our data lends credence to its use as a standardized tool for macroscopic donor kidney assessment, notably histologic assessment by transplant pathologists demonstrates poor agreement when compared with CAKS trainee-experts, thus CAKS likely has superior reproducibility to current pre-implantation biopsy criteria.

The low percentage of unanswered questions highlights this tool's ease of use and consistency, even among first time users. Despite relying solely on visual representation without tactile cues, raters scored in consistent patterns. Residents had more difficulty accurately assessing the degree of sticky fat compared to the gold standard assessors, likely due to the absence of tactile feedback. In clinical settings, where tactile evaluation is possible, junior surgeons may be able to apply CAKS even more reliably as indicated by this study.

The mean survey time of 9 minutes and 30 seconds minutes (1 minute and 37 seconds per renal unit) likely overestimates the time required for CAKS assessment in a real-world clinical setting. In clinical practice, scoring is typically performed during initial procurement or the back-table procedure, where experienced surgeons can complete the evaluation in seconds. By contrast, histologic assessment requires tissue sampling, pathologist review, and introduces delays of several hours, factors that can contribute to increased cold ischemic time and resource

use.[19] CAKS thus offers a rapid, cost-free alternative that can be integrated seamlessly into routine workflow.

This study complements previous work demonstrating that CAKS is independently associated with graft function, including 1-year estimated GFR, even after adjustment for KDPI and biopsy findings.[16] Taken together, these findings support the use of CAKS not only as a reproducible assessment tool, but also as a means of improving communication between members of the transplant team. Ideally, CAKS would be determined at the time of procurement and related alongside other donor information such as number of vessels, ureteral conditions (number and injury), perfusion pump parameters, and flush conditions. Whether CAKS alone or in combination with KDPI and other indicators can help optimize use of marginal kidneys remains an important area for future research.

This study has several limitations. First, exclusion of participants with incomplete survey data may have biased the ICC estimates towards a more favorable outcome, as less confident raters were excluded from the analysis. Second, this study may be limited by exposure bias. Respondents were shown standardized photographs that captured only a portion of the renal surface, which is not representative of real-world practice. In clinical settings, surgeons have full visual and tactile access to the graft, which may improve accuracy, particularly for evaluating sticky fat and vascular calcifications.

CONCLUSIONS

CAKS is a quantitative scoring system for donor kidneys designed to improve communication between surgeons and physicians in a cost- and time-efficient manner. This study demonstrates that the visual component of CAKS is reproducible across levels of surgical training and provides consistent assessments of donor kidney anatomy.

Future work will evaluate the utility of CAKS in predicting early graft failure among high KDPI kidneys and explore the performance of a composite score that integrates CAKS with KDPI, perfusion pump parameters, and biopsy findings to better inform transplant decision-making.

REFERENCES

1. Summary statistics on organ transplants, wait-lists and donors | CIHI. https://www.cihi.ca/en/summary-statistics-on-organ-transplants-wait-lists-and-donors. Accessed October 30, 2022.
2. Dahmen M, Becker F, Pavenstädt H, et al. Validation of the kidney donor profile index (KDPI) to assess a deceased donor's kidneys' outcome in a European cohort. *Sci Rep* 2019;9:11234. https://doi.org/10.1038/s41598-019-47772-7
3. Shantier M, Li Y, Ashwin M, et al. Use of the living kidney donor profile index in the Canadian kidney transplant recipient population: A validation study. *Can J Kidney Health Dis* 2020:7:2054358120906976. https://doi.org/10.1177/2054358120906976
4. Prado NP, Silva CK da, Meinerz G, et al. Usefulness of kidney donor profile index (KDPI) to predict graft survival in a South Brazilian Cohort. *J Bras Nefrol Orgao Of Soc Bras E Lat-Am Nefrol* 2020;42:211-8. https://doi.org/10.1590/2175-8239-JBN-2018-0263
5. Bachmann Q, Haberfellner F, Büttner-Herold M, et al. The kidney donor profile index (KDPI) correlates with histopathologic findings in post-reperfusion baseline biopsies and predicts kidney transplant outcome. *Front Med* 2022;9:875206. https://doi.org/10.3389/fmed.2022.875206.
6. Lentine KL, Kasiske B, Axelrod DA. Procurement biopsies in kidney transplantation: More information may not lead to better decisions. *J Am Soc Nephrol* 2022;32:1835-7. https://doi.org/10.1681/ASN.2021030403
7. Chen K, Guntur G, Stalam T, et al. Deceased-donor kidney biopsy scoring systems for predicting future graft function: A comparative study. *Transplant Proc* 2021;53:906-12. https://doi.org/10.1016/j.transproceed.2020.09.002
8. Amenábar JJ, Camacho JA, Gómez-Larrambe N, et al. Prognostic utility of preimplantation kidney biopsy from deceased older donors in first year post-transplant renal function. *Nefrol Publicacion Of Soc Espanola Nefrol* 2016;36:33-41. https://doi.org/10.1016/j.nefro.2015.10.009
9. Girolami I, Gambaro G, Ghimenton C, et al. Pre-implantation kidney biopsy: Value of the expertise in determining histological score and comparison with the whole organ on a series of discarded kidneys. *J Nephrol* 2020;33:167-76. https://doi.org/10.1007/s40620-019-00638-7
10. Keijbeck A, Veenstra R, Pol RA, et al. The association between macroscopic arteriosclerosis of the renal artery, microscopic ateriosclerosis, organ discard, and kidney transplant outcome. *Transplantation* 2020;104:2567-74. https://doi.org/10.1097/TP.0000000000003189
11. Matti D, Offerni J, Roshanov PS, et al. Canadian anatomic kidney score: Quantitative macroscopic assessment of donor kidney quality for transplantation. *Transplant Direct* 2024;10:e1604. https://doi.org/10.1097/TXD.0000000000001604
12. Fang L, Li H, Zhang T, et al. Analysis of predictors of adherent perinephric fat and its impact on perioperative outcomes in laparoscopic partial nephrectomy: A retrospective case-control study. *World J Surg Oncol* 2021;19:319. https://doi.org/10.1186/s12957-021-02429-6
13. Roever L, Tse G, Biondi-Zoccai G. Perirenal fat and chronic kidney disease in patients with diabetes. *Diabetes* 2021;70:2190-1. https://doi.org/10.2337/dbi21-0032

14. Miret Alomar E, Moreso F, Toapanta N, et al. Surgeon preimplantation macroscopic graft appraisal improves risk stratification of deceased kidney donors: a prospective study. *Minerva Urol Nephrol* 2022;74:615-24. https://doi.org/10.23736/S2724-6051.21.04345-7
15. Qiu W, Jiang Y, Wu J, et al. Simple cysts in donor kidney contribute to reduced allograft function. *Am J Nephrol* 2017;45:82-8. https://doi.org/10.1159/000453078
16. Revelle W. psych: Procedures for psychological, psychometric, and personality research. 2025;2.5.6. https://CRAN.R-project.org/package=psych. Accessed September 20, 2025.
17. Liapis H, Gaut JP, Klein C, et al. Banff histopathological consensus criteria for preimplantation kidney biopsies. *Am J Transplant* 2017;17:140-50. https://doi.org/10.1111/ajt.13929
18. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;15:155-63. https://doi.org/10.1016/j.jcm.2016.02.012
19. Dare AJ, Pettigrew GJ, Saeb-Parsy K. Preoperative assessment of the deceased-donor kidney: From macroscopic appearance to molecular biomarkers. *Transplantation* 2014;97:797-807. https://doi.org/10.1097/01.TP.0000441361.34103.53

FIGURES AND TABLES

Figure 1. Example of back table images presented for respondent scoring.

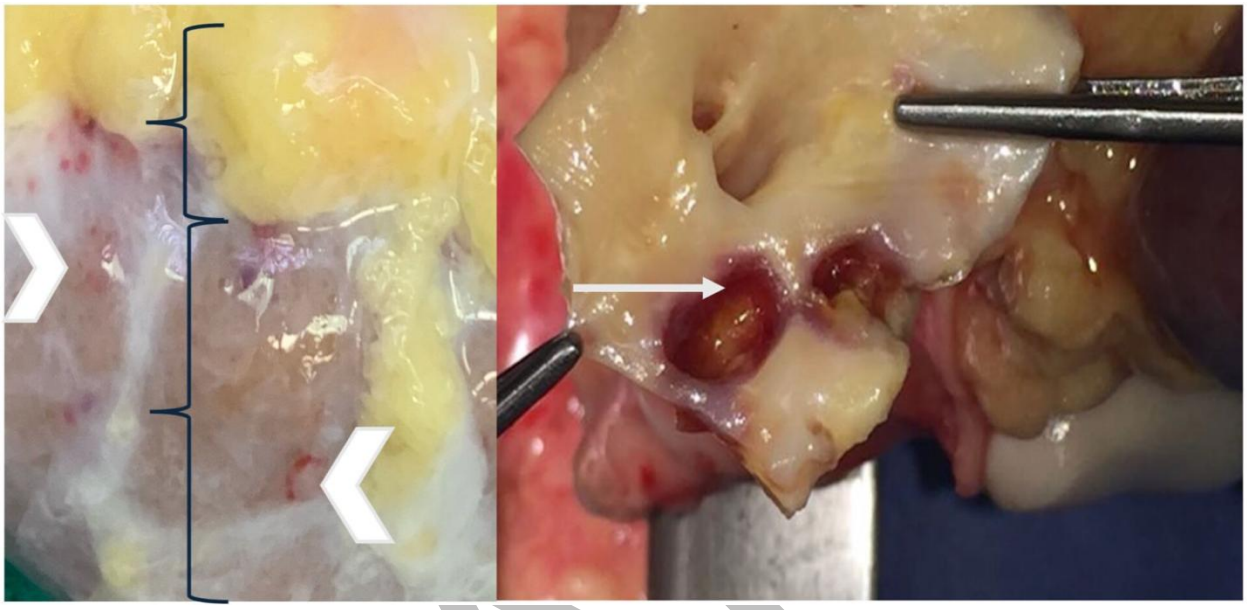

Table 1. The CAKS component definitions and scoring criteria

Score	Vasculature†	Sticky fat	Anatomy
Definition	Arteriopathy	Adipose tissue adherent to renal capsule that cannot be removed without damage	Cysts or other significant anomaly reported
0	No calcifications in the renal artery or aorta	No sticky fat remains after dissection	No cysts
0.5	Few calcified plaques in the aorta	Almost all sticky fat removed	<3 cysts <0.3 cm
1.0	Significant calcified or ulcerated plaques in the aorta, but the renal artery is clear	Sticky fat removed, but there remain areas of intense fibrosis	3-6 cysts <0.3 cm OR 1 cyst 0.3–1 cm
1.5	Calcified plaque on the ostium of the renal artery	Majority of the fat was intensively adhered and cannot be removed without damage	Between 7–9 cysts OR 1 cyst 1–2 cm

2.0	Calcified plaque in the renal artery or any branch	Intense adherent fat, which is impossible to remove	>9 cysts OR 1 cyst >2 cm

The Canadian Anatomic Kidney Score (CAKS) is scored out of 6 points and is defined by the vasculature, anatomy, and sticky fat content of each kidney. Scores were assigned during the back-table dissection prior to implantation. †Angiodysplasia: The presence of angiodysplasia or aneurysm should add a score of 0.5 in addition to your vasculature score (with a max score of 2.0).

TABLE 2. Gold-standard developer scores of the six image sets distributed to survey respondents

Survey question	CAKS Total	Vascular	Sticky fat	Anatomic
1	3.0	1.5	1.0	0.5
2	2.0	2.0	0	0
3	0	0	0	0
4	0.5	0	0	0.5
5	0	0	0	0
6	3.0	1.5	1.0	0.5

Table 3. Characteristics of survey respondents	
Baseline characteristics	**Summary**
Total participants	N=35
Median age (IQR), years	40.0 (32.25, 49.0)
Median years of practice (IQR), years*	4.5 (1.0, 13.75)
Focus of surgical practice (n, %)	
Urology	23 (65.7)
General surgery	10 (28.6)
Vascular surgery	2 (5.7)
Level of training (n, %)	
Staff	21 (60.0)
Transplant fellow	5 (14.3)
Resident	9 (25.7)
Number of kidney transplants at local center per year (%)	
30–50	4 (11.4)
50–100	12 (34.3)
100–150	15 (42.9)
>150	4 (11.4)

IQR: interquartile range.

Table 4. Inter-rater reliability analyses: degree of consistency between respondent and gold-standard score

Factor (No. of raters)	ICC	95% CI
CAKS overall (34)	0.783	0.728–0.838
Sticky fat (35)	0.656	0.577–0.735
Vasculature (34)	0.727	0.660–0.793
Anatomy (35)	0.552	0.470–0.633

Reliability is categorized as poor, moderate, good, and excellent, corresponding to values of <0.25, 0.25–0.5, 0.5–0.75, and >0.75, respectively. CAKS: Canadian Anatomic Kidney Score; CI: confidence interval; ICC: intraclass correlation coefficient.

Table 5. Inter-rater reliability analyses: Degree of absolute agreement between individual respondents on a per-question basis

Factor (No. of raters)	ICC	95% CI
CAKS overall (23)	0.800	0.599–0.961
Sticky fat (33)	0.669	0.427–0.925
Vasculature (25)	0.656	0.413–0.921
Anatomy (28)	0.584	0.339–0.896

Footnote: Reliability is categorized as poor, moderate, good, and excellent, corresponding to values of <0.25, 0.25–0.5, 0.5–0.75, and >0.75, respectively. CAKS: Canadian Anatomic Kidney Score; CI: confidence interval; ICC: intraclass correlation coefficient.