Original Article

Integrating artificial intelligence in renal cell carcinoma: evaluating ChatGPT's performance in educating patients and trainees

J. Patrick Mershon[1], Tasha Posid[1], Keyan Salari[2], Richard S. Matulewicz[3], Eric A. Singer[1], Shawn Dason[1]

[1]Division of Urologic Oncology, The Ohio State University Comprehensive Cancer Center, Columbus, OH, USA; [2]Department of Urology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA; [3]Urology Service, Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Contributions: (I) Conception and design: S Dason, JP Mershon; (II) Administrative support: EA Singer, S Dason, T Posid; (III) Provision of study materials or patients: EA Singer, S Dason, K Salari, RS Matulewicz; (IV) Collection and assembly of data: JP Mershon, T Posid; (V) Data analysis and interpretation: JP Mershon, T Posid; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Shawn Dason, MD. Division of Urologic Oncology, The Ohio State University Comprehensive Cancer Center, 2121 Kenny Road, Columbus, OH 43210, USA. Email: shawn.dason@osumc.edu.

Background: OpenAI's ChatGPT is a large language model-based artificial intelligence (AI) chatbot that can be used to answer unique, user-generated questions without direct training on specific content. Large language models have significant potential in urologic education. We reviewed the primary data surrounding the use of large language models in urology. We also reported findings of our primary study assessing the performance of ChatGPT in renal cell carcinoma (RCC) education.

Methods: For our primary study, we utilized three professional society guidelines addressing RCC to generate fifteen content questions. These questions were inputted into ChatGPT 3.5. ChatGPT responses along with pre- and post-content assessment questions regarding ChatGPT were then presented to evaluators. Evaluators consisted of four urologic oncologists and four non-clinical staff members. Medline was reviewed for additional studies pertaining to the use of ChatGPT in urologic education.

Results: We found that all assessors rated ChatGPT highly on the accuracy and usefulness of information provided with overall mean scores of 3.64 [±0.62 standard deviation (SD)] and 3.58 (±0.75) out of 5, respectively. Clinicians and non-clinicians did not differ in their scoring of responses (P=0.37). Completing content assessment improved confidence in the accuracy of ChatGPT's information (P=0.01) and increased agreement that it should be used for medical education (P=0.007). Attitudes towards use for patient education did not change (P=0.30). We also review the current state of the literature regarding ChatGPT use for patient and trainee education and discuss future steps towards optimization.

Conclusions: ChatGPT has significant potential utility in medical education if it can continue to provide accurate and useful information. We have found it to be a useful adjunct to expert human guidance both for medical trainee and, less so, for patient education. Further work is needed to validate ChatGPT before widespread adoption.

Keywords: Artificial intelligence (AI); renal neoplasm; renal cell carcinoma (RCC); health education

Submitted Dec 05, 2023. Accepted for publication Apr 10, 2024. Published online May 21, 2024.
doi: 10.21037/tcr-23-2234
View this article at: https://dx.doi.org/10.21037/tcr-23-2234

Introduction

In November 2022 OpenAI announced public access to a new online program, ChatGPT, that has rapidly become a cultural sensation. ChatGPT is an example of a "limited", or "narrow" artificial intelligence (AI) that can be used to answer unique and novel questions across a variety of subject matters without direct training. Limited or narrow AI are programs with a specific set of constraints and output types and can only handle tasks within their programming or training parameters. General AI has not yet been achieved but would be capable of learning and solving problems of any kind or format, on the level of a human mind and without concrete limits to its abilities. Concerns exist about its long-term applications in a variety of fields, including undergraduate education, scientific and creative writing, and in medical contexts (1,2). Many of the alarms being raised about ChatGPT concern its ability to mislead readers about the source of submitted material and the possibility of dangerous misinformation reaching a vulnerable audience (3-5). One particularly alarming trend has been ChatGPT's tendency to generate entirely fake literature citations to support its claims, which it will present as fact (6). These concerns have prompted an explosion of ethical and philosophical debates about the role of AI within medicine and for patient-facing applications (7-9).

Despite the risks of using AI in medicine, its potential benefits cannot be ignored and will force the field to wrestle with how and when (not if) to appropriately develop and employ this tool (7,10). A clear role for AI exists within medical education, a rapidly changing field that needs to respond quickly to new data to keep up with technological and scientific progress (11-13). Early perspectives of the use of ChatGPT for medical education have overall been cautiously optimistic, with focus on its limited ability to return consistently accurate technical information but with a high level of personalization and adaptability that could make it an excellent support tool for educators (14,15).

Another key area of study is AI use in patient education (10). AI-based decision aids for patients have been in existence for over 20 years, typically in more primitive forms than the now extremely user-friendly ChatGPT. ChatGPT is still in its infancy and much of the literature surrounding its value to patients is opinion-based and limited, but there has been a recent explosion in research attempting to evaluate its capabilities and limitations within the medical field. In the past year alone there have been assessments of its ability to answer questions regarding cirrhosis and hepatocellular carcinoma (16), total hip arthroplasty (17), obstructive sleep apnea (18), diabetes (19), and many other topics. Specific to urology, men's health (20), pediatric urology (21), and prostate cancer (22,23) have been investigated, though notably no work has yet been done within kidney cancer specifically.

In this pilot study we attempt to establish in general, primarily qualitative terms the guideline-concordance and subjective utility of ChatGPT's response to typical questions regarding the management of renal masses and renal cell carcinoma (RCC). We attempt to assess the ideal audience that would most benefit from interacting with ChatGPT to learn about renal malignancies and other urologic disease states. In conclusion, we review the current state of the literature as regards patient and trainee education using ChatGPT and other AI tools.

Highlight box

Key findings

- Clinicians and laypeople both rated ChatGPT highly on accuracy and usefulness based on responses to basic questions about small renal masses and renal cell carcinoma.

What is known and what is new?

- ChatGPT and other artificial intelligence (AI) tools are increasingly used by patients seeking information and decision-making help with new medical diagnoses. The accuracy and utility of this information is unknown and is likely subject-specific.
- Reviewing ChatGPT responses improved confidence in clinical information provided by the program and increased agreement that it should be used for medical education.

What is the implication, and what should change now?

- Clinician awareness of AI capabilities is critical to guide patients and medical trainees in its safe and effective use in urology. Ongoing assessments of safety and accuracy are needed to justify its use in urology and in medical applications in general.

Methods

Assessment development

Our first task was to develop a standardized set of questions regarding the diagnosis and management of renal masses and RCC that could reasonably be answered by ChatGPT. We reviewed three major professional society guidelines published prior to the AI learning end date regarding renal masses and RCC diagnosis and management. These

guidelines were chosen to fairly assess the ability of ChatGPT to process and synthesize the data available to it rather than testing its ability to extrapolate beyond its accessible knowledge. To this end, the American Society of Clinical Oncology (ASCO) 2022 guideline was initially reviewed but not included due to its publication after the AI learning end date. Using the guidelines below, we generated topic points with relative consensus among professional organizations to have an objective basis for assessing the accuracy of ChatGPT responses.

(I) Management of Small Renal Masses: American Society of Clinical Oncology Clinical Practice Guideline (24);
(II) NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®) - Kidney Cancer (25);
(III) Renal Mass and Localized Renal Cancer: Evaluation, Management, and Follow-up: AUA Guideline (26).

The primary author (J.P.M.) generated an initial list of questions which was then reviewed by two urologic oncologists (S.D., E.A.S.) to reach a consensus list of fifteen content questions to submit to ChatGPT 3.5 in February of 2023, with the goal being a mix of patient- and medical trainee-level questioning. J.P.M., S.D. and E.A.S. additionally reviewed the Canadian Urological Association guideline to ensure questions appropriately addressed international variations in small renal mass (SRM) management recommendations (27). The final list was input sequentially into the program and the responses recorded in a single document. To assess any potential bias of reviewers and to record changes in attitude after assessment, we obtained pre- and post-content assessment information regarding familiarity with and attitudes towards ChatGPT or other AI tools. The entire assessment tool is included in Appendix 1.

Assessment

We created a survey through Qualtrics (Qualtrics, Provo, UT, USA) with pre- and post-assessment questions as well as the fifteen content questions and answers (please see Appendix 1). The pre- and post-assessment questions asked about familiarity with ChatGPT and then asked assessors to select learner groups that would benefit most from its use, with choices of patients, medical students, residents, and attending physicians. Four fellowship-trained urologic oncologists (S.D., E.A.S., R.S.M., K.S.) utilized the assessment tool independently. After this initial review we extended the survey to four non-clinical reviewers on staff to obtain more qualitative data regarding a layperson's impression of the information. Each of these staff members have a Bachelor of Arts (BA) or Bachelor of Science (BS) but none have any specific medical training, do not educate residents or patients, and were chosen at random from the available research staff at the Ohio State University.

For the assessment portion, each of the fifteen questions was included alongside ChatGPT's response to that specific input, copied verbatim with no editing. For each question, the assessor was asked to rate the accuracy of the response on a Likert scale of 1 to 5 (1 being entirely inaccurate and 5 being entirely accurate). Reviewers compared the responses to the three guideline statements issued above as an objective benchmark. There was also an option to indicate dangerous or nonsensical information that could represent harm to patients or trainees. Assessors were then asked how useful the response was for an audience with a novice-level understanding equivalent to a typical patient with a new diagnosis of an SRM or RCC. Responses to this question were on a Likert scale of 1 to 5 (1 being entirely not at all useful and 5 being extremely useful). Without an objective benchmark all reviewers were asked to use their best judgment given their experience with the reading and comprehension levels of typical patients and medical trainees. Pre- and post-assessment questions were asked regarding overall subjective impressions of ChatGPT accuracy and usefulness in medical and patient education, with the addition of free text boxes to allow for qualitative discussions of the content responses and for assessors to describe their overall impression of the tool.

Statistical analysis

Data were collected from the Qualtrics database in raw format. Mean scores and standard deviation for each question score were calculated. Differences between pre- and post-assessment questions were then analyzed using paired-sample *t*-tests with a P value cutoff of 0.05 assigned for significance. Clinician and non-clinician scores were compared using a simple *t*-test. Selected quotations from reviewer's qualitative impressions were included in the final manuscript at the discretion of the primary author to best represent the reviewer's impressions of the tool.

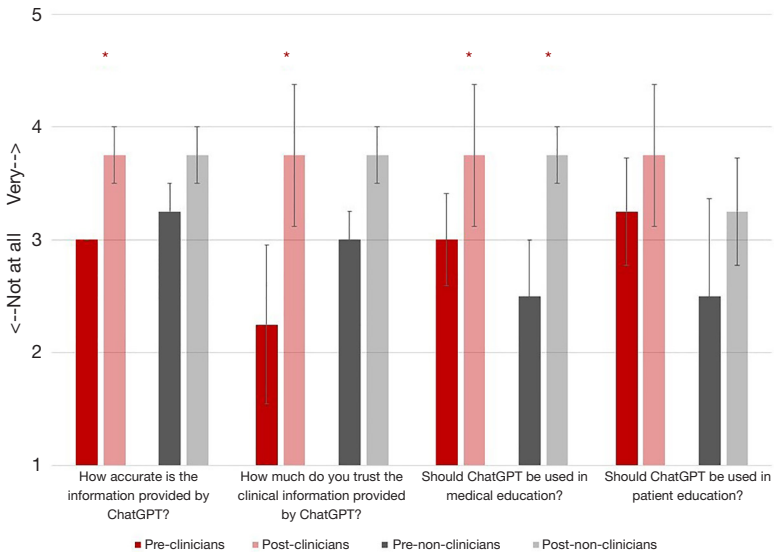

Figure 1 Change in responses to qualitative questions regarding ChatGPT before and after completing the survey tool and reviewing ChatGPT responses to prompts. Assessors were asked about the accuracy and trustworthiness of responses and then asked to rate its utility in both medical and patient education. Error bars represent +/– SEM. Asterisks denote statistically significant differences. SEM, standard error of the mean.

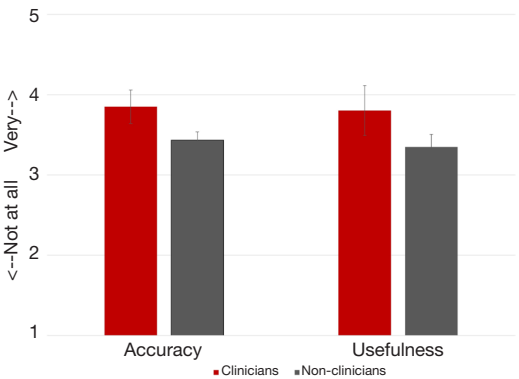

Figure 2 Perceptions of ChatGPT responses to the prompt questions between clinicians and non-clinicians. Scores of 1–5 were assigned on a Likert scare rating the accuracy (compared to guideline statements) and the usefulness of each prompt answer with a score of 5 representing "very accurate" or "very useful". Error bars represent +/– SEM. SEM, standard error of the mean.

Results

Pre-assessment questions: before the assessment, familiarity with ChatGPT did not differ between clinicians (mean =2.75/5) and non-clinicians (mean =2.5/5). Of those who had heard of it, 5/8 respondents said they had heard about it from social media and 3/8 said they had heard about it from colleagues in healthcare. Before completing the assessment questions, clinicians and non-clinicians had overall neutral assessments of the accuracy of the information provided by ChatGPT (mean score 3±0 for clinicians, 3.25±0.5 for non-clinicians), with a lower opinion of the accuracy of clinical information (mean score 2.625±1.06, *Figure 1*). Similarly, clinicians were neutral regarding the use of ChatGPT for medical (mean 3±0.81) and patient (mean 3.25±0.95) education.

We found that all assessors rated ChatGPT highly on the accuracy and usefulness of information provided in response to the generated questions, with scores differing slightly by question (Figures S1,S2). Clinicians rated content answers with mean scores of 3.85 [±0.42 standard deviation (SD)] out of 5 for accuracy and 3.8 (±0.62) out of 5 for usefulness (*Figure 2*). Non-clinicians gave slightly lower scores with mean scores of 3.43 (±0.77) for accuracy and 3.35 (±0.90) for usefulness but these groups were not significantly different in their assessments (P=0.37 for accuracy, P=0.21 for usefulness).

Completing the content assessment improved clinician confidence in the accuracy of information generated by ChatGPT (mean improvement of 0.75 on a 5-point Likert scale, P=0.01) and in the accuracy of clinical information provided (mean improvement of 1.5, P=0.01, *Figure 1*).

Both clinicians and non-clinicians were more likely to agree that ChatGPT or another AI tool could be useful in medical education after reviewing answers (mean improvement of 0.75 and 1.25 for clinicians and non-clinicians, respectively, P=0.007). There was no significant change in attitude towards ChatGPT use in patient education (mean improvement 0.5, P=0.30).

All assessors were skeptical of the utility of ChatGPT as an education tool before completing content assessment questions. One relevant quote stated, "*Patients should always seek a provider and ask questions to their provider about what the AI said*." After content assessment, reviewers were overall positive with their impressions of the tool. "*This technology feels very beneficial to audiences without vast medical knowledge (i.e., medical students and patients). However, I do feel that for certain audiences the medical jargon and/or quantity of information may confuse or worry them. I think it is a great supplemental tool when one doesn't have immediate access to a physician to gain an individual perspective*." Another commenter noted: "*It was much more accurate and useful than I anticipated*." However, one concern brought up was the source of content/data retrieved, with multiple reviewers wishing it could provide citations, "*Question 9 cited studies with no actual references, authors, or links*."

Discussion

Most of our content assessors were aware of ChatGPT through social media but had limited personal experience with the AI tool. Initially, assessor opinions were either skeptical or neutral regarding its information accuracy and its utility in medical and patient education. This is consistent with a large global survey of over 450 urologic clinicians; overall providers use ChatGPT for research and academic pursuits but do not use it for patient care (5). However, during content assessment, both clinicians and non-clinicians rated answer quality to be quite high both in terms of accuracy and usefulness to an audience with limited clinical knowledge equivalent to a typical patient (*Figure 1*). Reviewing the answers improved perceptions of the tool and its capabilities as well as its utility in educating novices (*Figure 2*). This study provides critical data assessing ChatGPT's accuracy and utility in directly answering medical questions at the level of a trainee or patient regarding RCC. However, this is only a pilot study, and much work remains to be done in this critical area.

AI-supported healthcare tools have been in use for decades with varying degrees of efficacy and sophistication; recently, Jayakumar *et al.* developed an AI-enabled patient decision tool for total knee replacement that improved decision quality, patient perception of shared decision making and treatment satisfaction without changing surgical intervention rates (28). Multiple AI chatbots have shown a remarkable ability to formulate patient-friendly responses that can minimize the likelihood of confusion and alarm often experienced by patients searching online about their conditions (29).

The literature surrounding ChatGPT itself and its value to patients is limited, but there has been a recent push to evaluate its capabilities and limitations. As a whole, these studies have lauded the capabilities of the new technology. Gabriel *et al.* [2023] directly compared ChatGPT responses to a human-generated patient information handout about radical prostatectomy and found 79% to be concordant and comparable, with 93% containing pertinent and accurate information (23). Durairaj *et al.* [2023] asked expert rhinoplasty surgeons to rate answers to typical patient questions from both a fellow surgeon and from ChatGPT 3.5 and found that the AI tool outperformed the surgeons with higher ratings in accuracy, completeness, and overall quality (30). In a head-to-head comparison, the AI-generated answer was the preferred response over 80% of the time. Similar results were shown with general medical questions posed by patients online, with licensed healthcare professionals preferring ChatGPT responses to physician responses 79% of the time, with nearly 10-fold higher rates of empathetic responses (31).

ChatGPT, however, did not perform as well on all assessments; Musheyev *et al.* [2023] found that ChatGPT 3.5 had moderately high information quality when responding to patient questions about multiple urologic malignancies, but only moderate understandability and low actionability and actually underperformed compared to other AI Chatbots (32). Similar results were seen in a broad survey of the most common urologic conditions with ChatGPT responses rated moderate on validated tools by two urologists (33). Coskun *et al.* [2023] evaluated prostate cancer patient information and found that ChatGPT performed only moderately well, with a mean score of 3.62 on a 5-point Likert scale of general quality (22). The authors concluded that the current version of the technology should be viewed with caution and could be further optimized before deployment as a patient education tool.

Education represents another key area of development within the AI sphere. Many articles discuss ChatGPT and

other AI tools as a double-edged sword within education as a whole, with rapid information delivery balanced by issues of possible plagiarism and offloading important intellectual work (10,13,34). Many others focus on its flaws; making up scientific papers and other sources that do not exist and occasionally answering confidently with false information (4,6,12,29). Notably we did not see this occur with our limited number of content-based questions. Its current failure to cite sources and to generate occasional misinformation is an extremely valid concern but seems to be primarily a technical challenge that could ideally be addressed with ongoing programming improvements (6). Future iterations of this technology should be even more powerful, with the ability to rapidly trawl the internet for all currently available data on a specific topic and provide more up to date information than a human could possibly generate (35). The currently available version of ChatGPT 3.5 does not have access to an up-to-date version of the internet and is time-locked at early 2022, limiting its real-time accuracy.

In the realm of medical education specifically, AI chatbots can (almost always) give accurate technical information but with a high level of personalization and adaptability (14,36). One notable example of its extreme power is its ability to generate and execute interactive medical simulations in text form for learners; Scherr *et al.* [2023] used ChatGPT 3.5 to create Advanced Cardiac Life Support (ACLS) and intensive care unit (ICU) scenarios with opportunities for improvisation and real-time feedback, with the goal to enhance medical student readiness for clinical clerkships (15). ChatGPT's achievement of a passing score on the United States Medical Licensing Examination (USMLE) Step 1 licensing exam generated headlines and its ability to logically justify its answers prompted many to argue for its immense utility as a study and training tool for medical students learning how to tackle both clinical and exam problems (36,37). Most literature on this topic highlights the need for caution, but focus on our opportunity to shape the implementation of this technology (12,15). Specific recommendations include AI literacy training and increased focus on source evaluation and evidence-based medicine, with heightened effort to teach empathy and good communication skills as the technology evolves (13).

The recent focus on AI is in many ways just one facet of rapid technologic advancement impacting medicine, with telehealth being the iteration prior. Particularly during the coronavirus disease 2019 (COVID-19) pandemic, this technology allowed physicians to continue to achieve patient care goals safely and effectively. Patients on the whole welcomed the change, with a majority satisfied with its use for their care and not feeling depersonalization from its use (38). Younger and more tech-savvy patients are even more likely to embrace these technological changes in medicine, and physicians should take note (30). While adoption of telehealth was difficult for some, its utility was immense and its introduction into healthcare has expanded our options for interacting with and treating patients in ways that meet their needs. AI will be a similarly impactful tool, but the burden is on physicians to use it safely and well.

Our pilot study is limited primarily by its size, with only four urologic oncologists assessing the tool's accuracy against an objective benchmark of relevant guidelines. Additionally, our assessment tool was not validated (as no validated tools for assessing AI responses currently exist). The possibility of bias given reviewers' knowledge of the information source is real but was unavoidable in this version of our study. Future work could consider blinding reviewers to the source of information to generate a more objective assessment of information quality, as has been done in several small early studies (30). This work was intended primarily to generate initial, more qualitative impressions of the tool from the perspectives of providers and laypeople to further hypothesis generation and study. We intend to expand upon this work in the future within our department to include more reviewers, more sophisticated assessment metrics, and to query other common topics within urology. It should also be noted that improvement of the AI underlying ChatGPT may date our results and repeating the analysis frequently with the most up-to-date version of the software will be extremely important. Since this work was completed, ChatGPT 4.0 and countless other AI tools have been released and will need similar assessments prior to widespread adoption in a medical context.

Conclusions

We found that clinicians and lay assessors consistently rated ChatGPT highly on the accuracy and usefulness of information provided in response to questions regarding the management of SRMs and RCC. Completing content assessment improved confidence in the accuracy of ChatGPT's information and increased agreement that it should be used for medical education. These results are an

early, informal evaluation of the capabilities of evolving AI tools but show great promise for this new technology.

Understanding how to leverage ChatGPT and other AI tools effectively and safely will be critical in the coming years in medicine as in many other fields (10). This is an extremely difficult area to study but the involvement of physicians and sub-specialty trained surgeons is critical to help shape AI into a positive force for patients and trainees alike (5,31,39). Just as clinicians have had to adjust their education and counseling strategies with the advent of "Dr. Google", AI proliferation will fundamentally shift how patients and trainees interact with medical information and we need to prepare ourselves for a new era (13). This pilot study is a first step towards understanding the power and pitfalls of this new tool and will facilitate ongoing study of this critical topic.

Acknowledgments

Funding: This work was supported by grants from the National Cancer Institute (No. 2P30CA016058, No. K08 CA259452) and from a Memorial Sloan Kettering Cancer Center Grant (No. P30 CA008748) to R.S.M.

Footnote

Data Sharing Statement: Available at https://tcr.amegroups.com/article/view/10.21037/tcr-23-2234/dss

Peer Review File: Available at https://tcr.amegroups.com/article/view/10.21037/tcr-23-2234/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://tcr.amegroups.com/article/view/10.21037/tcr-23-2234/coif). E.A.S. serves as an unpaid editorial board member of *Translational Cancer Research* from January 2023 to December 2024. R.S.M. reports receiving grants support from National Cancer Institute (No. 2P30CA016058, No. K08 CA259452) and from Memorial Sloan Kettering (No. P30 CA008748). S.D. reports that receiving consulting fees from Bristol Myers Squibb and Educational funding from Intuitive Surgical. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Stokel-Walker C. AI bot ChatGPT writes smart essays - should professors worry? Nature 2022. [Epub ahead of print]. doi: 10.1038/d41586-022-04397-7.
2. Biswas S. ChatGPT and the Future of Medical Writing. Radiology 2023;307:e223312.
3. Gao CA, Howard FM, Markov NS, et al. Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers. bioRxiv 2022;2022.12.23.521610.
4. Brewster J, Arvanitis L, Sadeghi M. The next great misinformation superspreader: how ChatGPT could spread toxic misinformation at unprecedented scale. Newsweek LLC; NewsGuard; 2023.
5. Eppler M, Ganjavi C, Ramacciotti LS, et al. Awareness and Use of ChatGPT and Large Language Models: A Prospective Cross-sectional Global Survey in Urology. Eur Urol 2024;85:146-53.
6. Weiser B. Here's What Happens When Your Lawyer Uses ChatGPT. The New York Times; 2023.
7. Shen Y, Heacock L, Elias J, et al. ChatGPT and Other Large Language Models Are Double-edged Swords. Radiology 2023;307:e230163.
8. Xu L, Sanders L, Li K, et al. Chatbot for Health Care and Oncology Applications Using Artificial Intelligence and Machine Learning: Systematic Review. JMIR Cancer 2021;7:e27850.
9. Cacciamani GE, Collins GS, Gill IS. ChatGPT: standard reporting guidelines for responsible use. Nature 2023;618:238.
10. Gabrielson AT, Odisho AY, Canes D. Harnessing Generative Artificial Intelligence to Improve Efficiency Among Urologists: Welcome ChatGPT. J Urol 2023;209:827-9.
11. Suárez A, Adanero A, Díaz-Flores García V, et al. Using a Virtual Patient via an Artificial Intelligence Chatbot to

Develop Dental Students' Diagnostic Skills. Int J Environ Res Public Health 2022;19:8735.
12. Boscardin CK, Gin B, Golde PB, et al. ChatGPT and Generative Artificial Intelligence for Medical Education: Potential Impact and Opportunity. Acad Med 2024;99:22-7.
13. Jamal A, Solaiman M, Alhasan K, et al. Integrating ChatGPT in Medical Education: Adapting Curricula to Cultivate Competent Physicians for the AI Era. Cureus 2023;15:e43036.
14. Mogali SR. Initial impressions of ChatGPT for anatomy education. Anat Sci Educ 2024;17:444-7.
15. Scherr R, Halaseh FF, Spina A, et al. ChatGPT Interactive Medical Simulations for Early Clinical Education: Case Study. JMIR Med Educ 2023;9:e49877.
16. Yeo YH, Samaan JS, Ng WH, et al. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. Clin Mol Hepatol 2023;29:721-32.
17. Mika AP, Martin JR, Engstrom SM, et al. Assessing ChatGPT Responses to Common Patient Questions Regarding Total Hip Arthroplasty. J Bone Joint Surg Am 2023;105:1519-26.
18. Campbell DJ, Estephan LE, Mastrolonardo EV, et al. Evaluating ChatGPT responses on obstructive sleep apnea for patient education. J Clin Sleep Med 2023;19:1989-95.
19. Sharma S, Pajai S, Prasad R, et al. A Critical Review of ChatGPT as a Potential Substitute for Diabetes Educators. Cureus 2023;15:e38380.
20. Shah YB, Ghosh A, Hochberg AR, et al. Comparison of ChatGPT and Traditional Patient Education Materials for Men's Health. Urol Pract 2024;11:87-94.
21. Caglar U, Yildiz O, Meric A, et al. Evaluating the performance of ChatGPT in answering questions related to pediatric urology. J Pediatr Urol 2024;20:26.e1-5.
22. Coskun B, Ocakoglu G, Yetemen M, et al. Can ChatGPT, an Artificial Intelligence Language Model, Provide Accurate and High-quality Patient Information on Prostate Cancer? Urology 2023;180:35-58.
23. Gabriel J, Shafik L, Alanbuki A, et al. The utility of the ChatGPT artificial intelligence tool for patient education and enquiry in robotic radical prostatectomy. Int Urol Nephrol 2023;55:2717-32.
24. Finelli A, Ismaila N, Bro B, et al. Management of Small Renal Masses: American Society of Clinical Oncology Clinical Practice Guideline. J Clin Oncol 2017;35:668-80.
25. Motzer RJ, Jonasch E, Agarwal N, et al. Kidney Cancer, Version 3.2022, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw 2022;20:71-90.
26. Campbell SC, Clark PE, Chang SS, et al. Renal Mass and Localized Renal Cancer: Evaluation, Management, and Follow-Up: AUA Guideline: Part I. J Urol 2021;206:199-208.
27. Richard PO, Violette PD, Bhindi B, et al. Canadian Urological Association guideline: Management of small renal masses - Full-text. Can Urol Assoc J 2022;16:E61-75.
28. Jayakumar P, Moore MG, Furlough KA, et al. Comparison of an Artificial Intelligence-Enabled Patient Decision Aid vs Educational Material on Decision Quality, Shared Decision-Making, Patient Experience, and Functional Outcomes in Adults With Knee Osteoarthritis: A Randomized Clinical Trial. JAMA Netw Open 2021;4:e2037107.
29. Hopkins AM, Logan JM, Kichenadasse G, et al. Artificial intelligence chatbots will revolutionize how cancer patients access information: ChatGPT represents a paradigm-shift. JNCI Cancer Spectr 2023;7:pkad010.
30. Durairaj KK, Baker O, Bertossi D, et al. Artificial Intelligence Versus Expert Plastic Surgeon: Comparative Study Shows ChatGPT "Wins" Rhinoplasty Consultations: Should We Be Worried? Facial Plast Surg Aesthet Med 2023. [Epub ahead of print]. doi: 10.1089/fpsam.2023.0224.
31. Ayers JW, Poliak A, Dredze M, et al. Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. JAMA Intern Med 2023;183:589-96.
32. Musheyev D, Pan A, Loeb S, et al. How Well Do Artificial Intelligence Chatbots Respond to the Top Search Queries About Urological Malignancies? Eur Urol 2024;85:13-6.
33. Szczesniewski JJ, Tellez Fouz C, Ramos Alba A, et al. ChatGPT and most frequent urological diseases: analysing the quality of information and potential risks for patients. World J Urol 2023;41:3149-53.
34. O'Connor S. Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse? Nurse Educ Pract 2023;66:103537.
35. Adams D, Chuah KM. Artificial Intelligence-Based Tools in Research Writing: Current Trends and Future Potentials. Artificial Intelligence in Higher Education 2022:169-84.
36. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS Digit Health 2023;2:e0000198.
37. Gilson A, Safranek CW, Huang T, et al. How Does ChatGPT Perform on the United States Medical

Licensing Examination (USMLE)? The Implications of Large Language Models for Medical Education and Knowledge Assessment. JMIR Med Educ 2023;9:e45312. Erratum in: JMIR Med Educ 2024;10:e57594.
38. Amparore D, Campi R, Checcucci E, et al. Patients' perspective on the use of telemedicine for outpatient urological visits: Learning from the COVID-19 outbreak. Actas Urol Esp (Engl Ed) 2020;44:637-8.
39. Javid M, Reddiboina M, Bhandari M. Emergence of artificial generative intelligence and its potential impact on urology. Can J Urol 2023;30:11588-98.

Cite this article as: Mershon JP, Posid T, Salari K, Matulewicz RS, Singer EA, Dason S. Integrating artificial intelligence in renal cell carcinoma: evaluating ChatGPT's performance in educating patients and trainees. Transl Cancer Res 2024. doi: 10.21037/tcr-23-2234

Appendix 1 Assessment Tool with ChatGPT Responses

Thank you again for agreeing to participate in our investigation! You will complete this assessment in Qualtrics. When you're ready, please click this link to begin, you will see all prompts and be able to enter all answers via this link:

[Qualtrics Link Here]

Please note, you will need to complete all questions and answers in a single sitting. This will take approximately 45-50 minutes.

You will begin by answering several questions (1-9) regarding your current knowledge and perspectives on the OpenAI ChatGPT Large Language Model. Please answer these questions as instructed by the question prompt using the buttons to select the option that best matches your perspective/experience.

For the following 15 questions (10-24), please refer to the below document. Using recent guidelines available regarding the management of kidney cancer, we generated the below 15 questions. The guidelines used to generate these questions are as below:

1. Management of Small Renal Masses: American Society of Clinical Oncology Clinical Practice Guideline[1]
2. NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®) - Kidney Cancer[2]
3. Renal Mass and Localized Renal Cancer: Evaluation, Management, and Follow-up: AUA Guideline[3]

We then input these questions verbatim into the ChatGPT interface and copied and pasted the responses exactly as you'll see them in Qualtrics. Please read the question we entered and the ChatGPT response, then rate the ChatGPT response across two parameters.

1) Rate the accuracy of the response rated on a 5-point scale (5 = entirely accurate, 1 = entirely inaccurate, and 0 = dangerous or nonsensical information or representing potential hazard to patients or medical trainees).

2) Rate each response on its usefulness or appropriateness in regard to the question, where 5 = extremely useful and 1 = not useful at all. For example, a technically accurate but incomplete answer could receive a 5 in accuracy but a 1 in usefulness. **of note, the ChatGPT model was trained on data that only extended to the end of 2021, and its accuracy and usefulness should be assessed in that context. We expect complete accuracy to be interpreted as accurate given the information available to the model in late 2021.*

For questions 25-30, please answer according to the Qualtrics survey instructions regarding your perspective on the ChatGPT interface after having reviewed its responses. In the final question, please include any open-text thoughts or reactions you may have regarding this technology, or please feel free to comment on your impressions of its implications for medical trainee or patient education. Additionally, please note any particularly impressive or troubling answers and comment on those here by referencing the question number for additional review by our team.

Pre-assessment questions for investigators

1. How familiar are you with ChatGPT?
   Not at all familiar with it
   Not very familiar with it
   Somewhat familiar with it
   Very familiar with it
   Extremely familiar with it

https://dx.doi.org/10.21037/tcr-23-2234

2. If you have heard of ChatGPT, where have you heard about or learned about it? (Check all that apply)
   _____ Heard about it from friends or family
   _____ Heard about it from colleagues in the healthcare field
   _____ Heard about it from patients
   _____ Heard about it on social media or the internet
   _____ Heard about it from television or the news
   _____ Have used it personally
   If they select this one: How many times have you used it?
   1-2 times
   3-5 times
   5-10 times
   10+ times

3. How accurate do you think the information provided by ChatGPT is?
   Inaccurate all of the time
   Inaccurate most of the time
   Inaccurate at least half of the time
   Accurate at least half of the time
   Accurate most of the time
   Accurate all of the time

4. How much do you trust the non-clinical information provided by ChatGPT?
   Do not trust at all
   Somewhat distrust
   Neutral / neither trust or don't trust
   Somewhat trust
   Trust very much

5. How much do you trust the clinical information provided by ChatGPT?
   Do not trust at all
   Somewhat distrust
   Neutral / neither trust or don't trust
   Somewhat trust
   Trust very much

6. For whom are the type of answers provided by ChatGPT most appropriate? (Rank order if easy, otherwise just choose most appropriate).
   Patients
   Medical students
   Residents
   Attending physicians
7. Do you agree that Chatbots or other automated tools should be used in medical education? (i.e. for medical students or residents)
   Strongly disagree
   Somewhat disagree
   Neutral/neither agree or disagree
   Somewhat agree
   Strongly agree

 https://dx.doi.org/10.21037/tcr-23-2234

8. Do you agree that Chatbots or other automated tools should be used in patient education? (i.e. for newly diagnosed patients or those considering certain treatment options)
   Strongly disagree
   Somewhat disagree
   Neutral/neither agree or disagree
   Somewhat agree
   Strongly agree

9. What are your thoughts regarding AI and ChatGPT in particular being used for medical education? For patient education?
   a. Free text box

ChatGPT Questions and Responses

10. Input: **Tell me what every patient should know about kidney cancer**

ChatGPT Response: Here are a few key things that every patient should know about kidney cancer:

- Kidney cancer is a type of cancer that starts in the kidneys. It can occur in people of any age, but it is most commonly diagnosed in people over the age of 60.
- The most common type of kidney cancer is called renal cell carcinoma (RCC). Other types of kidney cancer include transitional cell carcinoma, Wilms tumor, and renal sarcoma.
- The most common symptom of kidney cancer is blood in the urine, but other symptoms can include pain in the side or lower back, a lump in the abdomen, fatigue, unexplained weight loss, and fever.
- Kidney cancer is typically diagnosed using imaging tests such as CT scans, MRIs, or ultrasounds. A biopsy may also be done to confirm the diagnosis.
- Treatment options for kidney cancer depend on the stage and location of the tumor, as well as the patient's overall health. Surgery is often the main treatment, and may involve partial or total removal of the kidney. Other treatments may include thermal ablation, radiation therapy, chemotherapy, or targeted therapy.
- Kidney cancer can sometimes come back after treatment, so regular follow-up care is important. This may include imaging tests and blood tests to monitor for any signs of recurrence.
- Certain risk factors can increase a person's chance of developing kidney cancer, including smoking, obesity, high blood pressure, and a family history of the disease. Avoiding these risk factors and maintaining a healthy lifestyle can help reduce the risk of developing kidney cancer.

It's important to work with a team of healthcare professionals, including a urologist, oncologist, and other specialists as needed, to develop a personalized treatment plan that takes into account the individual patient's unique needs and goals. Support groups and other resources can also be helpful for patients and their families.

11. Input: **Tell me about the risk factors for kidney cancer**.

ChatGPT Response: Kidney cancer, also known as renal cell carcinoma, can have a variety of risk factors, including:

- Age: The risk of kidney cancer increases with age, especially for those over the age of 60.

- Smoking: Cigarette smoking is a significant risk factor for kidney cancer, as tobacco smoke contains numerous harmful chemicals that can damage the kidneys.
- Obesity: People who are overweight or obese have an increased risk of developing kidney cancer, as excess body fat can increase inflammation and hormone levels that may promote the growth of cancer cells.
- High blood pressure: Chronic high blood pressure can cause damage to the small blood vessels in the kidneys, which may increase the risk of developing kidney cancer.

 https://dx.doi.org/10.21037/tcr-23-2234

- Exposure to certain chemicals: Exposure to certain chemicals such as cadmium, asbestos, and organic solvents has been linked to an increased risk of kidney cancer.
- Family history: Individuals with a family history of kidney cancer have a higher risk of developing the disease.
- Certain genetic conditions: Inherited genetic conditions such as von Hippel-Lindau disease, Birt-Hogg-Dube syndrome, and hereditary papillary renal cell carcinoma can increase the risk of developing kidney cancer.

It's important to note that having one or more of these risk factors does not necessarily mean that an individual will develop kidney cancer, and some people who develop kidney cancer may not have any known risk factors. Regular medical check-ups and screening tests can help identify kidney cancer early, when it is more treatable.

12. Input: **Why does kidney cancer respond better to immunotherapy than chemotherapy?**

ChatGPT Response: Kidney cancer is known to be relatively resistant to chemotherapy, meaning that it does not respond as well to chemotherapy drugs as some other types of cancer. However, kidney cancer has shown promise in responding to immunotherapy treatments.

- Immunotherapy works by helping the body's immune system recognize and attack cancer cells. In the case of kidney cancer, immunotherapy drugs called immune checkpoint inhibitors have been shown to be effective. These drugs block certain proteins that can prevent the immune system from recognizing and attacking cancer cells. By blocking these proteins, the immune system is able to more effectively target and attack the cancer cells.
- In contrast, chemotherapy drugs work by directly killing rapidly dividing cells, including cancer cells. However, kidney cancer cells tend to divide more slowly than some other types of cancer cells, which can make them less responsive to chemotherapy.
- Additionally, because chemotherapy drugs do not differentiate between cancer cells and healthy cells, they can cause significant side effects such as nausea, hair loss, and damage to healthy cells in the body. Immunotherapy, on the other hand, can be more targeted and may have fewer side effects.

Overall, while every case of cancer is different and the best treatment approach will depend on individual factors, immunotherapy has shown promise as a treatment for kidney cancer and may be a viable option for some patients.

13. Input: **Under what circumstances should a biopsy be obtained for a small renal mass?**

ChatGPT Response: A biopsy may be recommended for a small renal mass in certain circumstances, depending on various factors such as the size and characteristics of the mass, the patient's medical history, and other diagnostic test results. In general, a biopsy may be recommended for a small renal mass if:

- The mass is suspicious for cancer: If imaging tests such as an ultrasound, CT scan, or MRI indicate that the renal mass has characteristics that are suggestive of cancer, a biopsy may be recommended to confirm the diagnosis.
- The patient is at high risk of complications: For some patients, such as those with significant medical comorbidities or other factors that increase the risk of complications from surgery, a biopsy may be recommended to help determine whether the renal mass is cancerous and whether surgical intervention is necessary.
- The patient is a candidate for active surveillance: In some cases, if the renal mass is small and low-risk, a patient may be a candidate for active surveillance rather than immediate surgical removal. In this case, a biopsy may be recommended to confirm the diagnosis and help guide ongoing surveillance and treatment.
- It's important to note that in some cases, a biopsy may not be recommended for a small renal mass, particularly if the mass is small and low-risk or if the patient is not a good candidate for biopsy due to medical comorbidities or other factors. Ultimately, the decision to obtain a biopsy for a small renal mass should be made on an individual basis, taking into account the patient's unique circumstances and medical history.

 https://dx.doi.org/10.21037/tcr-23-2234

14. **Input**: **What are the risks of renal mass biopsy? Are they always accurate?**

ChatGPT Response: While renal mass biopsy is generally considered a safe procedure, as with any medical procedure, there are potential risks and complications. Some of the possible risks associated with renal mass biopsy include:

- Bleeding: The biopsy needle can cause bleeding, which can be mild or, in rare cases, severe.
- Infection: There is a risk of infection at the biopsy site, although this is rare.
- Damage to other organs: In rare cases, the biopsy needle can puncture other nearby organs or blood vessels, which can cause damage and potentially require additional medical interventions.
- Diagnostic error: Although renal mass biopsies are generally accurate, there is a small risk of diagnostic error or sampling error, which means that the biopsy results may not be completely accurate.

Regarding the accuracy of renal mass biopsies, studies have shown that they are generally reliable and accurate, particularly for larger renal masses. However, the accuracy of the biopsy can depend on a number of factors, including the size and location of the mass, the type of biopsy performed, and the skill and experience of the healthcare provider performing the biopsy.

It's important to note that a biopsy is just one tool in the diagnostic process, and other diagnostic tests may be necessary to confirm a diagnosis or rule out other conditions. Patients should discuss the potential risks and benefits of renal mass biopsy with their healthcare provider to determine whether the procedure is appropriate for them.

15. **Input**: **What type of patients are better candidates for thermal ablation than partial nephrectomy for a renal tumor?**

ChatGPT Response: Thermal ablation and partial nephrectomy are two different treatment options for renal tumors, and the choice between them will depend on various factors, including the size and location of the tumor, the patient's overall health, and other medical considerations. In general, patients who may be better candidates for thermal ablation than partial nephrectomy include those who:

- Have small tumors: Thermal ablation may be a good option for patients with small renal tumors (less than 4 cm in size) that are limited to one area of the kidney.
- Have medical comorbidities: For some patients, such as those with significant medical comorbidities, thermal ablation may be a better option than partial nephrectomy because it is generally less invasive and has a lower risk of complications.
- Have multiple tumors: Patients who have multiple renal tumors may be better candidates for thermal ablation, as it is less invasive than partial nephrectomy and can be used to treat multiple tumors in a single session.
- Are not good candidates for surgery: For patients who are not good candidates for surgery due to underlying medical conditions or other factors, thermal ablation may be a viable treatment option.
- It's important to note that the decision to perform thermal ablation or partial nephrectomy for a renal tumor should be made on an individual basis, taking into account the patient's unique circumstances and medical history. Patients should discuss the potential risks and benefits of each treatment option with their healthcare provider to determine which option is best for them

16. **Input**: **What are the advantages and disadvantages of thermal ablation compared to surgery for a renal tumor?**

ChatGPT Response: Thermal ablation and surgery (such as partial nephrectomy) are two different treatment options for renal tumors, and each has its own advantages and disadvantages. Some of the advantages and disadvantages of thermal ablation compared to surgery are:

 https://dx.doi.org/10.21037/tcr-23-2234

Advantages of thermal ablation:
- Less invasive: Thermal ablation is less invasive than surgery, as it involves using a needle or probe to deliver heat or cold to the tumor, rather than making an incision in the body.
- Shorter recovery time: Because thermal ablation is less invasive than surgery, patients typically have a shorter recovery time and can often return to normal activities more quickly.
- Less blood loss: Thermal ablation generally involves less blood loss than surgery, which may be an advantage for patients who are at risk for bleeding or who have a history of bleeding disorders.
- Can be repeated: If the tumor is not completely eliminated with thermal ablation, the procedure can often be repeated.

Disadvantages of thermal ablation:
- Limited to smaller tumors: Thermal ablation is generally only suitable for smaller tumors (less than 4 cm in size) that are located in a specific area of the kidney.
- Higher risk of incomplete treatment: There is a risk that thermal ablation may not completely eliminate the tumor, which can lead to the need for additional treatments.
- Limited information about long-term outcomes: Because thermal ablation is a relatively new treatment for renal tumors, there is limited information available about its long-term outcomes.

Advantages of surgery:
- Complete removal of the tumor: Surgery allows for complete removal of the tumor, which can help ensure that the cancer is completely eliminated.
- Can be used for larger tumors: Surgery can be used to treat larger tumors that may not be suitable for thermal ablation.
- Provides a definitive diagnosis: Surgical removal of the tumor allows for a definitive diagnosis, which can help guide further treatment.

Disadvantages of surgery:
- More invasive: Surgery is more invasive than thermal ablation, as it involves making an incision in the body.
- Longer recovery time: Patients typically have a longer recovery time after surgery than after thermal ablation.
- Higher risk of complications: Surgery carries a higher risk of complications than thermal ablation, including bleeding, infection, and damage to other organs.

It's important to note that the choice between thermal ablation and surgery will depend on various factors, including the size and location of the tumor, the patient's overall health, and other medical considerations. Patients should discuss the potential risks and benefits of each treatment option with their healthcare provider to determine which option is best for them.

17. Input: **Compare and contrast radical v. partial nephrectomy for a renal mass**.

ChatGPT Response: Radical nephrectomy and partial nephrectomy are two surgical treatment options for a renal mass. While both procedures involve removing a portion of the kidney, there are some key differences between them.

- Radical nephrectomy involves removing the entire kidney, along with the surrounding tissue and lymph nodes. This procedure is typically reserved for larger or more aggressive tumors that are not confined to a single area of the kidney, and it may be recommended if the tumor is located near the renal vein or other critical structures.
- Partial nephrectomy, on the other hand, involves removing only the tumor and a small margin of healthy tissue surrounding it, while preserving the rest of the kidney. This procedure is generally recommended for smaller tumors that are confined to a specific area of the kidney, and it may be preferred for patients who have a higher risk of kidney failure or who only have one kidney.

 https://dx.doi.org/10.21037/tcr-23-2234

Here are some of the key differences between radical and partial nephrectomy:

- Kidney preservation: Partial nephrectomy preserves more of the kidney tissue than radical nephrectomy, which removes the entire kidney. This can help maintain kidney function and reduce the risk of developing chronic kidney disease or needing dialysis after the surgery.
- Surgical complexity: Partial nephrectomy is a more complex surgical procedure than radical nephrectomy, as it involves removing only a portion of the kidney while preserving the surrounding tissue and structures. This can make the procedure more challenging and may require a more experienced surgeon.
- Risks of complications: Radical nephrectomy carries a higher risk of complications than partial nephrectomy, as it involves removing the entire kidney and may require more extensive surgery. In addition, patients who undergo radical nephrectomy may be at increased risk of developing chronic kidney disease or needing dialysis after the surgery.
- Cancer recurrence: Some studies suggest that partial nephrectomy may be associated with a lower risk of cancer recurrence than radical nephrectomy, particularly for smaller tumors.
- Overall survival: The choice between radical and partial nephrectomy may depend on the patient's overall health and medical history. In some cases, radical nephrectomy may be necessary to ensure complete removal of the tumor, while in other cases, partial nephrectomy may be a better option to preserve kidney function and improve overall survival.

Overall, the choice between radical and partial nephrectomy will depend on various factors, including the size and location of the tumor, the patient's overall health, and other medical considerations. Patients should discuss the potential risks and benefits of each treatment option with their healthcare provider to determine which option is best for them.

18. Input: **What tumor characteristics argue in favor of active surveillance for a renal mass?**

ChatGPT Response: Active surveillance, also known as watchful waiting, may be an appropriate treatment option for patients with small renal masses (less than 4 cm) or those who are at high risk for surgical complications due to their age or medical history. The decision to pursue active surveillance depends on several factors, including the tumor characteristics.

Here are some of the tumor characteristics that may favor active surveillance for a renal mass:

- Small size: Tumors that are small (less than 4 cm) are more likely to be benign or slow-growing and may not require immediate treatment.
- Low growth rate: If the tumor is growing slowly or not at all, it may not require immediate treatment.
- Low complexity: If the tumor has a low complexity score on imaging studies, meaning it has a well-defined margin and low vascularity, it may be less aggressive and may not require immediate treatment.
- Low risk of metastasis: Tumors that are at low risk of spreading to other organs or tissues may be less aggressive and may not require immediate treatment.
- Co-existing medical conditions: Patients who have other medical conditions that may increase the risk of complications from surgery, such as heart or lung disease, may be better candidates for active surveillance.

It is important to note that active surveillance is not appropriate for all patients with renal masses, as some tumors may be more aggressive and require immediate treatment. Patients should discuss the potential risks and benefits of active surveillance with their healthcare provider to determine if it is a suitable treatment option for their specific case.

19. Input: **Does active surveillance increase the risk of metastasis in renal masses compared to surgery?**

ChatGPT Response: Active surveillance for renal masses is generally considered to be a safe and effective treatment option for select patients with small renal masses or those who are at high risk for surgical complications. However, there is a potential risk of disease progression and metastasis associated with active surveillance.

 https://dx.doi.org/10.21037/tcr-23-2234

Studies have shown that the rate of disease progression and metastasis in patients undergoing active surveillance for small renal masses is low. In fact, one study found that the risk of metastasis was less than 1% per year for tumors less than 4 cm in size that were being actively monitored.

However, there is some evidence to suggest that active surveillance may be associated with a higher risk of disease progression and metastasis compared to surgery. A systematic review of 15 studies found that the risk of disease progression was higher in patients undergoing active surveillance compared to those undergoing surgery.

The decision to pursue active surveillance or surgery for a renal mass should be based on a careful assessment of the patient's individual risk factors and preferences. Patients who are at high risk for surgical complications or who have small, low-risk tumors may be good candidates for active surveillance. However, patients with larger or more aggressive tumors may require surgery to ensure complete removal of the tumor and to reduce the risk of disease progression and metastasis. It is important for patients to discuss the potential risks and benefits of both treatment options with their healthcare provider to make an informed decision about their care.

20. Input: **What are the indications for genetic testing for a patient with a small renal mass?**

ChatGPT Response: Genetic testing may be considered for patients with a small renal mass to help identify any underlying genetic syndromes that may increase their risk of developing kidney cancer. The following are some of the indications for genetic testing for a patient with a small renal mass:

- Personal or family history of kidney cancer: Patients who have a personal or family history of kidney cancer may be at increased risk for hereditary forms of kidney cancer and may benefit from genetic testing.
- Young age at diagnosis: Patients who are diagnosed with kidney cancer at a younger age, particularly under the age of 46, may be at increased risk for hereditary forms of kidney cancer.
- Bilateral or multifocal kidney tumors: Patients who have multiple kidney tumors, either in the same kidney or in both kidneys, may be at increased risk for hereditary forms of kidney cancer.
- Specific clinical features: Patients who have specific clinical features, such as skin lesions, eye abnormalities, or tumors in other organs, may be at increased risk for hereditary forms of kidney cancer.

The most common hereditary forms of kidney cancer are von Hippel-Lindau (VHL) syndrome, hereditary leiomyomatosis and renal cell cancer (HLRCC), and Birt-Hogg-Dubé (BHD) syndrome. These syndromes are associated with specific genetic mutations and are characterized by an increased risk of developing multiple kidney tumors and other types of cancer.

Patients who are found to have a hereditary form of kidney cancer may require more frequent surveillance and may be candidates for earlier or more aggressive treatment. Genetic counseling can help patients understand the potential risks and benefits of genetic testing and can assist in the development of a personalized screening and management plan.

21. Input: **What are preferred first-line systemic therapies for advanced clear cell renal cell carcinoma?**

ChatGPT Response: Clear cell renal cell carcinoma (ccRCC) is the most common type of kidney cancer, and systemic therapy is the standard of care for patients with advanced or metastatic disease. The choice of first-line systemic therapy will depend on a variety of factors, including the patient's overall health, the extent of the disease, and the potential side effects of each therapy. The following are some of the preferred first-line systemic therapies for advanced ccRCC:

- Tyrosine kinase inhibitors (TKIs): TKIs are oral medications that target specific proteins involved in the growth and spread of cancer cells. Sunitinib and pazopanib are two TKIs that are commonly used as first-line therapy for advanced

 https://dx.doi.org/10.21037/tcr-23-2234

ccRCC.
- Immune checkpoint inhibitors: Immune checkpoint inhibitors are medications that help to stimulate the immune system to attack cancer cells. Nivolumab and ipilimumab are two immune checkpoint inhibitors that may be used as first-line therapy for patients with intermediate or poor-risk advanced ccRCC.
- Combination therapy: Combination therapy, using both a TKI and an immune checkpoint inhibitor, has been shown to be effective in patients with advanced ccRCC. The combination of pembrolizumab and axitinib is one example of a TKI/immune checkpoint inhibitor combination that is commonly used as first-line therapy.
- mTOR inhibitors: mTOR inhibitors are another class of medications that may be used as first-line therapy for patients with advanced ccRCC. Everolimus is an example of an mTOR inhibitor that has been approved for the treatment of advanced ccRCC.

The choice of first-line therapy will depend on a variety of factors, including the patient's overall health, the extent of the disease, and the potential side effects of each therapy. Your healthcare provider can help you determine which therapy is most appropriate for your specific case.

22. Input: **Is there a survival or other benefit to lymph node dissection at the time of nephrectomy for clear cell renal cell carcinoma?**

ChatGPT Response: The role of lymph node dissection (LND) in the management of clear cell renal cell carcinoma (ccRCC) is a topic of ongoing debate. Lymph node involvement is a known predictor of poor prognosis in patients with ccRCC, and LND is commonly performed at the time of nephrectomy in an attempt to remove any potentially involved lymph nodes.

However, there is no clear consensus on whether LND provides a survival or other benefit for patients with ccRCC. Some studies have suggested that LND may improve overall survival and reduce the risk of cancer recurrence, particularly in patients with larger tumors or a higher likelihood of lymph node involvement.

Other studies have found no significant survival benefit to LND and have raised concerns about the potential risks and side effects of the procedure, such as bleeding, infection, and lymphedema. Furthermore, LND may increase the risk of complications during surgery and may prolong the operative time and hospital stay.

As a result, the decision to perform LND in patients with ccRCC should be individualized and based on a careful consideration of the potential risks and benefits. Your healthcare provider can help you understand the potential risks and benefits of LND and can assist in the development of a personalized treatment plan based on your specific case.

23. Input: **What are the indications for adjuvant systemic therapy after nephrectomy for clear cell renal cell carcinoma?**

ChatGPT Response: Adjuvant systemic therapy is a form of treatment given after surgery with the aim of reducing the risk of cancer recurrence. In the case of clear cell renal cell carcinoma (ccRCC), adjuvant systemic therapy may be considered in patients with a high risk of cancer recurrence following nephrectomy.

The following are some of the indications for adjuvant systemic therapy after nephrectomy for ccRCC:
- High-risk features: Patients with high-risk features, such as large tumor size, tumor invasion into surrounding tissues, or lymph node involvement, may be considered for adjuvant systemic therapy. The presence of certain genetic mutations may also increase the risk of cancer recurrence and may be an indication for adjuvant systemic therapy.
- Incomplete resection: Patients who undergo nephrectomy but have evidence of residual disease, such as positive surgical margins, may benefit from adjuvant systemic therapy.

 https://dx.doi.org/10.21037/tcr-23-2234

- Metastatic disease: In some cases, adjuvant systemic therapy may be given to patients with metastatic ccRCC who have undergone nephrectomy. This may be done in an attempt to improve overall survival and reduce the risk of cancer recurrence.

The choice of adjuvant systemic therapy will depend on a variety of factors, including the patient's overall health, the extent of the disease, and the potential side effects of each therapy. Some of the options for adjuvant systemic therapy for ccRCC include targeted therapy with tyrosine kinase inhibitors, immunotherapy with checkpoint inhibitors, or a combination of both.

The decision to recommend adjuvant systemic therapy should be individualized and based on a careful consideration of the potential risks and benefits. Your healthcare provider can help you understand the potential risks and benefits of adjuvant systemic therapy and can assist in the development of a personalized treatment plan based on your specific case.

24. **Input**: **What patient factors increase the risk of relapse or metastasis after treatment for clear cell renal cell carcinoma?**

    ChatGPT Response: Clear cell renal cell carcinoma (ccRCC) can be an aggressive cancer that has the potential to relapse or metastasize after treatment. The risk of relapse or metastasis depends on several patient factors. Some of the patient factors that can increase the risk of relapse or metastasis after treatment for ccRCC include:

    - Tumor size: Larger tumors are associated with a higher risk of relapse or metastasis. Tumors greater than 7 cm in size are considered to be high-risk.
    - Tumor grade: The grade of the tumor is a measure of how abnormal the cells appear under a microscope. High-grade tumors are associated with a higher risk of relapse or metastasis.
    - Stage of the tumor: The stage of the tumor is a measure of how advanced the cancer is at the time of diagnosis. Patients with advanced stage tumors are at a higher risk of relapse or metastasis.
    - Lymph node involvement: The presence of cancer cells in the lymph nodes indicates that the cancer has spread beyond the kidney and is associated with a higher risk of relapse or metastasis.
    - Vascular invasion: The presence of cancer cells in blood vessels or the renal vein indicates a higher risk of relapse or metastasis.
    - Genetic mutations: Certain genetic mutations, such as mutations in the von Hippel-Lindau (VHL) gene, are associated with a higher risk of relapse or metastasis.
    - Prior treatment: Patients who have received prior treatment for ccRCC may be at a higher risk of relapse or metastasis, especially if the treatment was not successful in removing all of the cancer.

    It is important to note that these factors are not absolute predictors of relapse or metastasis, and each patient's individual case may vary. Your healthcare provider can help you understand your individual risk factors and can assist in the development of a personalized treatment plan.

Post-assessment questions for investigators

25. How accurate do you think the information provided by ChatGPT is?
    Inaccurate all of the time
    Inaccurate most of the time
    Inaccurate at least half of the time
    Accurate at least half of the time
    Accurate most of the time
    Accurate all of the time

 https://dx.doi.org/10.21037/tcr-23-2234

26. How much do you trust clinical information provided by ChatGPT?
    Do not trust at all
    Somewhat distrust
    Neutral / neither trust or don't trust
    Somewhat trust
    Trust very much

27. For whom are the type of answers provided by ChatGPT most appropriate? (Rank order if easy, otherwise just choose most appropriate).
    Patients
    Medical students
    Residents
    Attending physicians
28. Do you agree that Chatbots or other automated tools should be used in medical education? (i.e. for medical students or residents)
    Strongly disagree
    Somewhat disagree
    Neutral/neither agree or disagree
    Somewhat agree
    Strongly agree

29. Do you agree that Chatbots or other automated tools should be used in patient education? (i.e. for newly diagnosed patients or those considering certain treatment options)
    Strongly disagree
    Somewhat disagree
    Neutral/neither agree or disagree
    Somewhat agree
    Strongly agree

30. Overall perceptions of ChatGPT and/or answers provided by ChatGPT: _____ (free text box).

 https://dx.doi.org/10.21037/tcr-23-2234

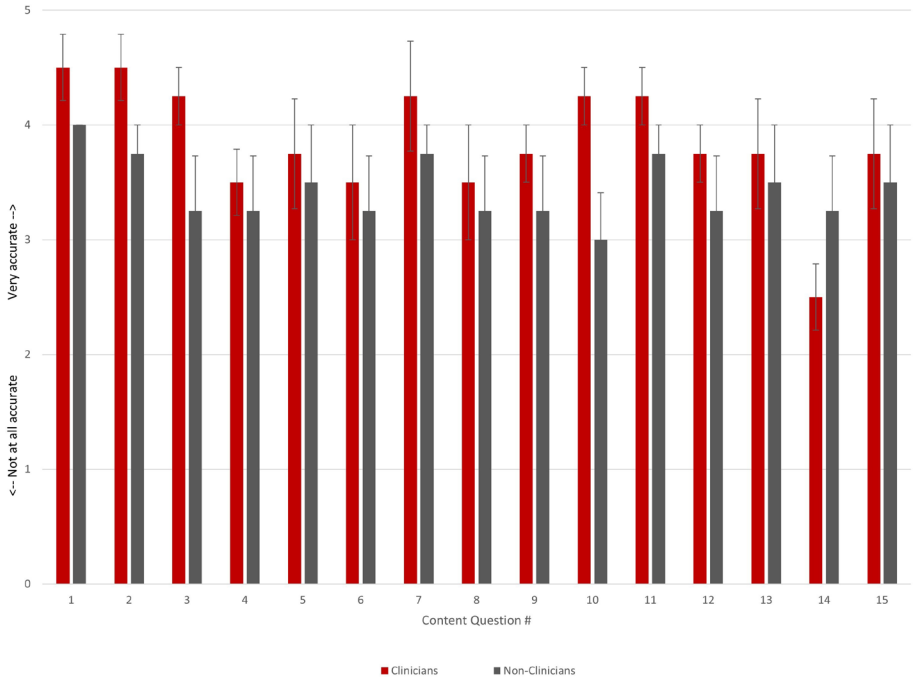

Figure S1 Perceptions of the accuracy of ChatGPT responses to each prompt question comparing the answer to appropriate guideline statements from clinical and non-clinical staff. Question number corresponds to each content question and response in Appendix 1: "assessment tool". Scores of 1–5 were assigned on a Likert score with a score of 5 representing "very accurate". Error bars represent +/– SEM.

https://dx.doi.org/10.21037/tcr-23-2234

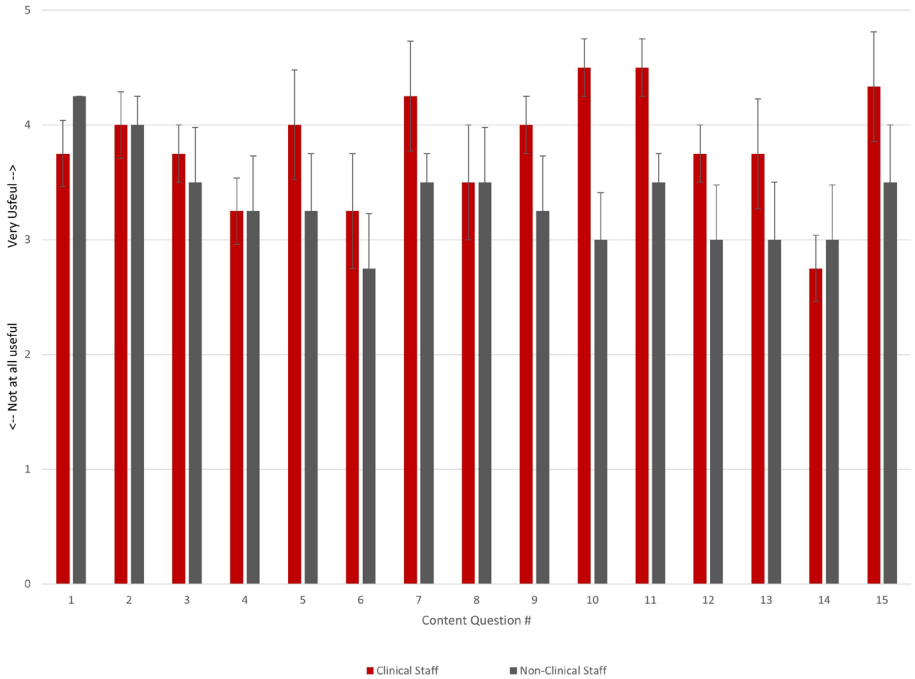

Figure S2 Perceptions of the usefulness of ChatGPT responses to each prompt question, judged subjectively against the comprehension ability of a typical patient without medical training. Question number corresponds to each content question and response in Appendix 1: "assessment tool". Scores of 1–5 were assigned on a Likert score with a score of 5 representing "very useful". Error bars represent +/– SEM.

https://dx.doi.org/10.21037/tcr-23-2234