J Chin Med Assoc

BRIEF COMMUNICATION

Limitations of GPT-4 as a geriatrician in geri-oncology case conference: A case series

Ling-Yuk Kuk[a,*], Dora Lai-Wan Kwong[b], Wing-Lok Wendy Chan[b], Yat-Fung Shea[a]

[a]Geriatrics Division, Department of Medicine, Queen Mary Hospital, Hong Kong, China; [b]Department of Clinical Oncology, Center of Cancer Medicine, Queen Mary Hospital University of Hong Kong, Hong Kong, China

Abstract

Generative pre-trained transformer 4 (GPT-4) is an artificial intelligence (AI) system with a chat interface. The number of studies testing GPT-4 in clinical applications has been increasing. We hypothesized that GPT-4 would be able to suggest management strategies for medical issues in elderly oncology patients, similar to those provided by geriatricians. We compared the responses of GPT-4 to those of a geriatrician for four oncological patients. After these case conferences, none of the patients required admission for medical consultation. In three out of four scenarios, GPT-4 was able to offer a multidisciplinary approach in the first prompt. In all three scenarios, GPT-4 identified medication-related side effects and suggested appropriate medications in the first prompt. However, GPT-4 was unable to suggest initial dosages of medications to be used in the first prompt and was unable to suggest a more humanistic and non-pharmacological approach to anorexia, even with a follow-up prompt. In conclusion, GPT-4 may be used as a screening tool to provide potential rudimentary directions for management, which can then be reviewed by medical professionals before considering a formal consultation for more tailored and refined opinions from specialists.

Keywords: Case conference; Geriatrics; GPT-4; Oncology

1. INTRODUCTION

Artificial intelligence (AI) has been increasingly studied and utilized in medical consultations, such as those involving drug interactions or medical image analyses.[1] Generative pre-trained transformer 4 (GPT-4), developed by OpenAI, is an AI system with a chat interface,[1] which has been trained on internet texts and fine-tuned on more specific datasets with the help of human reviewers. In a typical "session," a user enters a query, and GPT-4 generates a natural-language "response" within seconds. GPT-4 has been tested and has shown promising performance in medical examinations.[1] Its potential use has also been explored in geriatric patients, as it may assist elderly patients with prolonged admissions with diagnostic difficulties.[2] However, GPT-4 has limitations, such as its ability to analyze medical histories related to cognitive impairments.[3]

ChatGPT, which is an earlier version of an AI chatbot, has been suggested as a resource for providing useful personal information on personal obesity management, healthy lifestyles, and vaccination information for disease prevention.[4,5] There has been an increasing number of consultations to geriatricians from other specialties, such as clinical oncologists, regarding medical issues in elderly patients.[6] Because GPT-4 is the latest and most advanced form of AI chatbot, it may be able to provide more comprehensive suggestions for medical problems or can act as a "clinical consultant." To our knowledge, there have been no studies exploring the role of GPT-4 as a geriatrician in geri-oncology case conferences. Based on our previous work, GPT-4 was able to offer comprehensive multidisciplinary suggestions for medical problems and was particularly effective in identifying medication-related side effects. We hypothesized that GPT-4 would be able to suggest a comprehensive multidisciplinary management plan for medical issues and detect potential medication-related side effects in elderly oncology patients, similar to those provided by geriatricians.

2. METHODS

Between April 1 and May 31, 2023, monthly geri-oncology case conferences started, and four oncology patients sought advice from geriatricians regarding their medical issues. We included all four patients in our case series. These conferences involved a clinical oncologist, an oncology nurse, and a geriatrician. The clinical oncologist presented the patients' histories and specific medical issues, while the geriatrician provided advice in response. The clinical oncologist followed the management plan during subsequent outpatient consultations, usually within 1 to 2 weeks. The geriatrician (Y.-F.S.) summarized the clinical history and questions from the clinical oncologist during the case conference (as shown in the Supplementary File 1, http://links.lww.com/JCMA/A225) and entered the data into the GPT-4 chatbot in June 2023. The GPT-4 chatbot was offered by poe.com, and no additional training of the system was conducted. GPT-4's responses were extracted (Supplementary File 1, http://links.lww.com/JCMA/A225) and compared with the responses of the geriatrician (Table 1), particularly examining the comprehensiveness of suggestions involving multidisciplinary teams, the ability to identify medication-related side effects, and the appropriateness of suggested medications and dosages.

* *Address correspondence. Dr. Ling-Yuk Kuk, Department of Medicine, Queen Mary Hospital, 102 Pokfulam Road, Hong Kong, China. E-mail address: kuklingyuk@gmail.com (L.-Y. Kuk).*

Conflicts of interest: The authors declare that they have no conflicts of interest related to the subject matter or materials discussed in this article.

Journal of Chinese Medical Association. (2024) 87: 148-150.

Received August 22, 2023; accepted September 23, 2023.

doi: 10.1097/JCMA.0000000000001032

Table 1

A summary of responses from GPT-4 vs geriatricians in geriatrics-oncology conference

Patients	GPT-4's reply	Geriatrician's reply	Comparing GPT-4 (first prompt) and geriatrician
The patient was an 87-yr old man living in an old age home, non-drinker, and ex-chronic smoker, with cancer of stomach, pending operative decision.	Need to perform a comprehensive geriatric assessment to evaluate the patient's overall health and functional status. Recommend exercise and nutritional intervention before surgery or radiotherapy. Nutritional support. Review of medications. Provide guidance on pain management.	Operative risk could be assessed using the acute coronary syndrome calculator. To consider dietician support. To consider referring to outpatient physiotherapy and occupational therapy training.	GPT-4 made the same suggestions as geriatrician in offering nutritional assessment and exercise therapy as a form of prerehabilitation.
An 80-yr old woman with colon cancer and lung metastasis and was noted to have ankle swelling, tiredness, and poor appetite.	Related to ankle swelling: Watch out for side effect of medication: amlodipine. Look out for venous insufficiency, lymphedema, and infection. Look out for cancer-related complications. Related to tiredness and anorexia: Watch out and address anemia. Consult dietician. Review medications. Encourage tolerable physical activity. Watch out for emotional or psychological issues, for example, anxiety and depression.	Ankle swelling was suspected to be due to calcium channel blocker. Amlodipine could be taken off and try angiotensin converting enzyme inhibitor, for example, losartan. Look out for depression, for example, using screening tool like Geriatric Depression Scale 15. Remove all dietary restrictions if any. Try to use flavor enhancer. Use appetite stimulating agents, for example, megestrol.	GPT-4 was able to identify potential side effects of calcium channel blockers and able to consider other differential diagnoses. GPT-4 was able to suggest a multidisciplinary approach involving dietician and to look out for mood problem. GPT-4 was unable to consider humanistic non-pharmacological approach for management of anorexia.
A 67-yr old woman with carcinoma of the breast with lymph node and bone metastases was concerned by clinical oncologist to have low mood and peripheral numbness.	Mood management Consider the usage of depression screening tool, for example, Geriatric Depression Scale 15. Multiple disciplinary approaches—including psychologist and psychiatrist. Pharmacotherapy: consider the use of SSRI or SNRI. Counseling or psychotherapy, for example, cognitive behavioral therapy. Encourage the patient to participate in support groups or engaging with families or friends. Peripheral numbness management: Perform a neurological examination. Modify chemotherapy regimen. Medications: consider the usage of duloxetine or gabapentin. Refer to a physical therapist and occupational therapist.	To assess patient's mood problem especially whether there is any suicidal ideation. Consider screening for any depression using Geriatric Depression Scale 15. May consider to start antidepressant, for example, SSRI (eg, citalopram 10 mg), SNRI (eg, venlafaxine 37.5 mg), or serotonin modulators (eg, vortioxetine 5 mg). To consider referring to a clinical psychologist for assessment and counseling. For peripheral limb numbness which could be related to previous chemotherapy, we may refer the patient to neurologist for nerve conduction study and consider medications, for example, gabapentin (100-200 mg tds po) and pregabalin (eg, 50 mg BD) for control.	GPT-4 was able to suggest the correct screening tools for depression in elderly patients. GPT-4 was able to suggest a multidisciplinary approach involving psychologist and psychiatrist to handle the potential mood problem. GPT-4 was able to suggest the medications (including the class) for depression or chemotherapy-related peripheral numbness. GPT-4 was unable to suggest the initial dosages of antidepressants.
A 71-yr old woman with metastatic follicular carcinoma with cauda equina syndrome and was now on cabozantinib with proteinuria.	Regular urine test for proteinuria. To consider reducing the dosage of cabozantinib. Consider starting angiotensin converting enzyme inhibitor or angiotensin receptor blockers. If proteinuria persists, consider stopping cabozantinib. Look for the other causes of proteinuria, for example, hypertensive kidney disease.	Suggested to start angiotensin converting enzyme inhibitor, for example, perindopril with gradual titration to 8 mg. Reduced amlodipine to 5 mg daily po. To check fasting glucose, HbA1c for any DM nephropathy. If persistent proteinuria, consider refer renal team for assessment.	GPT-4 was able to suggest the class of medications for management of proteinuria and was able to suggest clinician to look out for other potential causes. GPT-4 was unable to suggest dosages of medications for proteinuria and how existing medications may need to be adjusted. GPT-4 was unable to suggest referral to appropriate specialist in case of persistence of the problem.

GPT-4 = generative pre-trained transformer 4; SNRI = selective nor-adrenaline reuptake inhibitor; SSRI = selective serotonin reuptake inhibitor.

3. RESULTS

Four patients' cases, which exhibited different underlying malignancies, were discussed in the case conferences. Multiple medical issues were discussed, including mood assessment, management of peripheral numbness, proteinuria management, adjustment of antihypertensives, and potential preparation for surgery and radiotherapy. All four scenarios involved multidisciplinary management, with three scenarios (patients 2 to 4) focusing on medication, including initial dosage suggestions, and one scenario (patient 2) involving non-pharmacological management of anorexia. After these case conferences, none of the patients required admission for medical consultation.

Regarding multidisciplinary management, GPT-4 was able to provide suggestions in three out of four scenarios (excluding patient 4 who failed to consider referral to a nephrologist in case of persistent proteinuria). GPT-4 could also suggest the use of the Geriatric Depression Scale to screen for depression.[7] In all three scenarios involving medication management, GPT-4 was able to offer suggestions similar to those of a geriatrician, including detection of amlodipine-induced lower limb swelling, suggestions for drugs that may be useful in managing chemotherapy-related peripheral numbness, and recommendations for medications to manage proteinuria, such as angiotensin converting enzyme inhibitors or angiotensin receptor blockers. However, GPT-4 was not able to suggest initial dosages of medications, such as antidepressants in the first prompt and suggest the titration of existing medications. In addition, GPT-4 was not able to offer non-pharmacological and more humanistic management of anorexia, such as the usage of flavor enhancers and the removal of dietary restrictions, in the first prompt.

Based on the limitations of initial responses, follow-up questions were asked for patients 2 to 4 (Supplementary File 1, http://links.lww.com/JCMA/A225). GPT-4 was able to provide suggestions for initial medication dosages and recommended referral to a nephrologist if proteinuria persisted for patient 4. However, GPT-4 remained incapable of offering non-pharmacological management for anorexia (patient 2).

4. DISCUSSION

Based on our study, GPT-4 demonstrated the ability to offer suggestions for managing medical issues in elderly cancer patients using a multidisciplinary approach. It was also able to detect potential side effects of medications and recommend potentially beneficial alternatives. GPT-4's advice could be consulted before formal consultation to a geriatrician. However, GPT-4 may not provide dosage recommendations initially, and follow-up questions are needed to obtain such information. Alternatively, doctors can refer to pharmacological textbooks for dosage information. Despite these findings, responses from GPT-4 have not been fully assessed by the medical community, and there may be potential medicolegal issues. Compared with GPT-4, geriatricians can emphasize the importance of monitoring blood pressure, watching out for drug allergies or anaphylaxis, and adjusting dosages according to renal function for certain medications.

GPT-4 lacks the ability to suggest humanistic and non-pharmacological approaches, such as the management of anorexia. We speculate that this limitation may be related to the way it "learns," which is largely based on pattern recognition and statistical association. The responses generated by GPT-4 exhibit a strong framework or checklist-like tone. Therefore, a more individualized management plan would still require geriatrician input.

Our study was limited by a small sample size, and therefore the findings should be interpreted as preliminary. One also needs to be cautious about the clinical information entered into the chatbot, as incorrect clinical information may lead to misinterpretation by GPT-4. The newer version of GPT-4 should be retested once it becomes available. The AI system should be tested in various clinical aspects, and familiarity should be established to avoid misuse. One proposed approach is to utilize GPT-4 as a screening tool to provide initial management directions, which can then be reviewed by medical professionals before considering a formal consultation for more tailored and refined opinions from specialists. It is hoped that this piece can inspire thoughts and trials regarding the delicate use of AI in medical contexts as AI continues to develop and evolve.

GPT-4 shows great potential in suggesting multidisciplinary management for medical issues in elderly oncology patients, particularly in identifying medication-related side effects and suggesting medications for their medical conditions. However, certain limitations, such as the absence of human touch and failure to provide advice on potential side effects of new medications, were also identified. Therefore it is advisable to continue consulting geriatricians at present.

APPENDIX A. SUPPLEMENTARY DATA

Supplementary data related to this article can be found at http://links.lww.com/JCMA/A225.

REFERENCES

1. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. *N Engl J Med* 2023;13:1233–9.
2. Shea YF, Lee CMY, Ip WCT, Luk DWA, Wong SSW. Use of GPT-4 to analyze medical records of patients with extensive investigations and delayed diagnosis. *JAMA Netw Open* 2023;6:e2325000.
3. Shea YF, Ma NC. Limitations of GPT-4 in analyzing real-life medical notes related to cognitive impairment. *Psychogeriatrics* 2023;23:885–7.
4. Arslan S. Exploring the potential of Chat GPT in personalized obesity treatment. *Ann Biomed Eng* 2023;51:1887–8.
5. Biswas SS. Role of Chat GPT in public health. *Ann Biomed Eng* 2023;51:868–9.
6. Soo WK, King MT, Pope A, Parente P, Dārziņš P, Davis ID. Integrated geriatric assessment and treatment effectiveness (INTEGERATE) in older people with cancer starting systemic anticancer treatment in Australia: a multicentre, open-label, randomised controlled trial. *Lancet Healthy Longev* 2022;3:e617–27.
7. Chan AC. Clinical validation of the geriatric depression scale (GDS): Chinese version. *J Aging Health* 1996;8:238–53.