ARTICLE IN PRESS

Surgery xxx (2023) 1-3



Contents lists available at ScienceDirect

Surgery

journal homepage: www.elsevier.com/locate/surg



How appropriate are answers of online chat-based artificial intelligence (ChatGPT) to common questions on colon cancer?

ChatGPT is a large language model created by OpenAI that has been trained on a diverse set of text data to understand human language and respond to questions and prompts naturally. ChatGPT was designed to assist with various tasks, such as answering questions, generating text, and completing prompts to help people more effectively and efficiently communicate with natural language.

With the increasing popularity and use of ChatGPT, the general public may use it to gain information on various medical conditions. It is still unclear whether or not ChatGPT can provide accurate information concerning education questions. Therefore, the present study is aimed to assess the appropriateness and consistency of ChatGPT answers to questions on different aspects of colon cancer management.

The present study followed a similar methodology to that reported in a recent research letter that assessed the appropriateness of ChatGPT responses to fundamental questions on cardiovascular disease prevention. Thirty-eight questions on the topics of prevention, diagnosis, and management of colon cancer were generated (Figure 1). These questions were based on our clinical experience and patient information handouts from the American Society of Colon and Rectal Surgeons (ASCRS).

Each question was asked twice to ChatGPT, and both answers were recorded to examine the consistency of answers. Three geographically diverse colorectal surgery experts were emailed the answers and an assessment form. The experts were 2 male and 1 female surgical consultants from Italy, Israel, and Brazil; they were selected based on their clinical expertise and publications on colon cancer. The experts were informed about the study design and purpose and were requested to rate the answers of ChatGPT as appropriate (consistent and factually accurate), inappropriate (consistent and factually inaccurate or contains incorrect information), or inconsistent. The first author reviewed the assessment forms returned by the experts, and the agreement between their ratings was assessed using inter-rater reliability. Two additional colorectal surgeons (S.E. and N.H.) assessed whether or not the answers of ChatGPT were consistent with the 2022 ASCRS practice parameters on colon

Twenty-seven answers of ChatGPT were rated as appropriate by the 3 experts, whereas 6 answers were rated as appropriate by 2 experts, and 4 were rated as appropriate by only 1 expert. All experts rated the answers to 1 question (number 15) as inconsistent. A sample of inappropriate answers of ChatGPT is shown in Figure 2.

The percentage of answers rated as appropriate varied among the experts (78.9%, 81.6%, and 97.4%). Overall, at least 2 of 3 experts rated the answers as appropriate for 86.8% (95% CI: 71.9%—95.6%) of questions. The inter-rater reliability was 79.8% (95% CI: 71.3%—86.8%). Nineteen (95%) of the 20 questions that applied to the 2022 ASCRS practice parameters for colon cancer were concordant. The answers to 1 question (number 15) showed some discordance.

Approximately 87% of the answers provided by ChatGPT to 38 questions on colon cancer management were deemed appropriate and consistent by at least 2 independent experts. This percentage is similar to the 84% reported in another study¹ that examined the appropriateness of ChatGPT answers to questions in a different medical discipline. Inappropriate ChatGPT answers involved incorrect information regarding the potential risk of hemorrhoids causing chronic colon inflammation that might increase the risk of colon cancer. It also incorrectly suggested that colorectal surgeons can deliver chemotherapy and radiotherapy and that radiation therapy may be used to shrink the tumor after surgery. Inconsistency of answers was noted by at least 1 expert for 8 questions, which involved 1 of the 2 answers missing some information related to preventive methods for colon cancer, screening methods, age to start screening, symptoms of colon cancer, multidisciplinary team management, and side effects of surgery for colon cancer.

The answers of ChatGPT were largely concordant with ASCRS practice parameters, except when answering the question on when and how often screening for colon cancer be done, as it did not include double-contrast barium enema as was stated in the ASCRS parameters. Most ChatGPT answers consistently recommended consulting health care providers and professionals. However, the guideline and practice parameters of medical associations were not implied. The role of professional surgical associations in vetting the medical information provided by the AI platform to patients and taking the lead should be considered. The main limitation of the present study is that ChatGPT is not meant for medical use. The current ChatGPT statistics in 2023 indicate that ChatGPT attained 100 million active users by January 2023 and grew by 9,900% in 2 months, rendering it the fastest-growing platform.⁴ However, there is no information on the specific use or type of questions ChatGPT is asked by its users. With growing research, future survey studies may shed light on the frequency of the use of ChatGPT by patients as a source of medical information.

S.H. Emile et al. / Surgery xxx (2023) 1-3

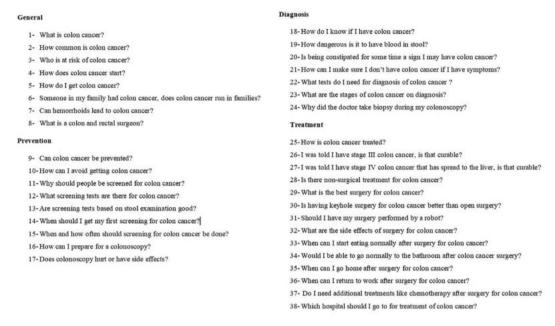


Figure 1. Thirty-eight questions on the diagnosis and management of colon cancer that were asked to ChatGPT.



Figure 2. A sample of inappropriate answers by ChatGPT.

In addition, although the study attempted to evaluate different aspects of colon cancer care, the current list of simple questions cannot fully assess this wide area. Nonetheless, the answers of ChatGPT to most questions were detailed and seemingly accurate. Therefore, ChatGPT may become a popular educational and informative tool in clinical practice.

Acknowledgments

The first paragraph of the introduction, describing ChatGPT, was generated by ChatGPT. During the preparation of this work, the authors used ChatGPT 3.5 to assess the appropriateness of its answers to questions on the management of colon cancer. After using this tool/service, the authors reviewed and edited the content as needed and took full responsibility for the publication's content.

Sameh Hany Emile, MD Colorectal Surgery Department Ellen Leifer Shulman and Steven Shulman Digestive Disease Center, Cleveland Clinic Florida Weston, FL Colorectal Surgery Unit Mansoura University Hospitals Mansoura, Egypt

Twitter: @dr_samehhany81

Nir Horesh, MD

Colorectal Surgery Department Ellen Leifer Shulman and Steven Shulman Digestive Disease Center, Cleveland Clinic Florida Weston, FL

> Department of Surgery and Transplantation Sheba Medical Center Ramat-Gan, Israel

Twitter: @nirhoresh

Michael Freund, MD

Department of General Surgery, Shaare Zedek Medical Center, Faculty of Medicine, Hebrew University of Jerusalem, Israel

Twitter: @mikifreund

Gianluca Pellino, MD Colorectal Surgery, Vall d'Hebron University Hospital, Universitat Autonoma de Barcelona UAB, Barcelona, Spain S.H. Emile et al. / Surgery xxx (2023) 1-3

Department of Advanced Medical and Surgical Sciences Università degli Studi della Campania "Luigi Vanvitelli" Naples, Italy

Twitter: @GianlucaPellino

Lucia Oliveira, MD

Director Anorectal Physiology Department of Rio de Janeiro and

CEPEMED

Rio de Janeiro, Brazil

Twitter: @luciacolorectal

Anjelli Wignakumar, MD

Department of Plastic Surgery, St. Mary's Hospital, London, UK

Twitter: @AWignakumar

Steven D. Wexner, MD*

Colorectal Surgery Department

Ellen Leifer Shulman and Steven Shulman Digestive Disease Center, Cleveland Clinic Florida

Weston, FL

* Corresponding author.

E-mail address: wexners@ccf.org (S.D. Wexner); Twitter: @CleveClinicFL, @SWexner

Accepted 18 June 2023

References

- Sarraju A, Bruemmer D, Van Iterson E, Cho L, Rodriguez F, Laffin L. Appropriateness of cardiovascular disease prevention recommendations obtained from a popular online chat-based artificial intelligence model. JAMA. 2023;329:842–844.
- Vogel JD, Felder SI, Bhama AR, et al. The American Society of Colon and Rectal Surgeons clinical practice guidelines for the management of colon cancer. Dis Colon Rectum. 2022;65:148–177.
- ASCRS. Colorectal cancer screening and surveillance: clinical guideline and rationale. https://fascrs.org/healthcare-providers/education/clinical-practice-guidelines/ colorectal-cancer-screening-and-surveillance-clini. Accessed February 27, 2023.
- Nerdy Nav. 97+ ChatGPT statistics and user numbers in May 2023. https:// nerdynav.com/chatgpt-statistics. Accessed 1 June 2023.