

Mediastinum & Esophagus: Short Report

ChatGPT vs Expert-Guided Care Pathways for Postesophagectomy Symptom Management



Mohamad K. Abou Chaar, MD,¹ Giovanna Grigsby-Rocca, BS,² Ming Huang, PhD,³ and Shanda H. Blackmon, MD, MPH⁴

ABSTRACT

BACKGROUND The objective of this study was to compare generative artificial intelligence-initiated care pathways, using ChatGPT, with expert-guided consensus-initiated care pathways from AskMayoExpert (AME) for symptom management of esophageal cancer patients after esophagectomy.

METHODS A formal protocol for development of 9 AME care pathways was followed for specific patient-identified domains after esophagectomy for esophageal cancer. Domain scores were measured and assessed through the Upper Digestive Disease tool. These care pathways were developed by experts validated by a consensus-driven methodology. ChatGPT was used to answer specific questions similar to the AME care pathway on April 9, 2023, and March 28, 2024. To compare outcomes, answers were recorded, and algorithms were compared with a survey tool composed of 5 questions.

RESULTS Both modalities were able to provide a clear definition with multidirectional management options for all 9 domains: dysphagia, generalized dumping, gastrointestinal dumping, pain, regurgitation, heartburn, nausea, physical health, and mental health. When provided with a simple prompt, ChatGPT 3.5 failed to provide a comprehensive stepwise approach for providers, any testing recommendations, or any form of triage process. However, ChatGPT 4.0 provided plans, similar to AME care pathways, when a sophisticated prompt was used.

CONCLUSIONS Generative artificial intelligence-initiated care pathways can be used by physicians as a supplementary tool to guide provider management of patients with complex symptoms after esophagectomy. This technology will continue to advance but is currently insufficient to solely guide clinical management of complex patients with severe symptoms.

(Ann Thorac Surg Short Reports 2024;2:674-679)

© 2024 The Authors. Published by Elsevier Inc. on behalf of The Society of Thoracic Surgeons.
This is an open access article under the CC BY-NC-ND license
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

The evolution of generative artificial intelligence (AI) and chatbot development have been marked by significant milestones. In the early 2000s, basic rule-based systems laid the foundation for this technology.¹ Subsequent years witnessed breakthroughs with the advent of machine learning and natural language processing. As of 2023, advanced models like GPT-3 showcase the cutting edge,

IN SHORT

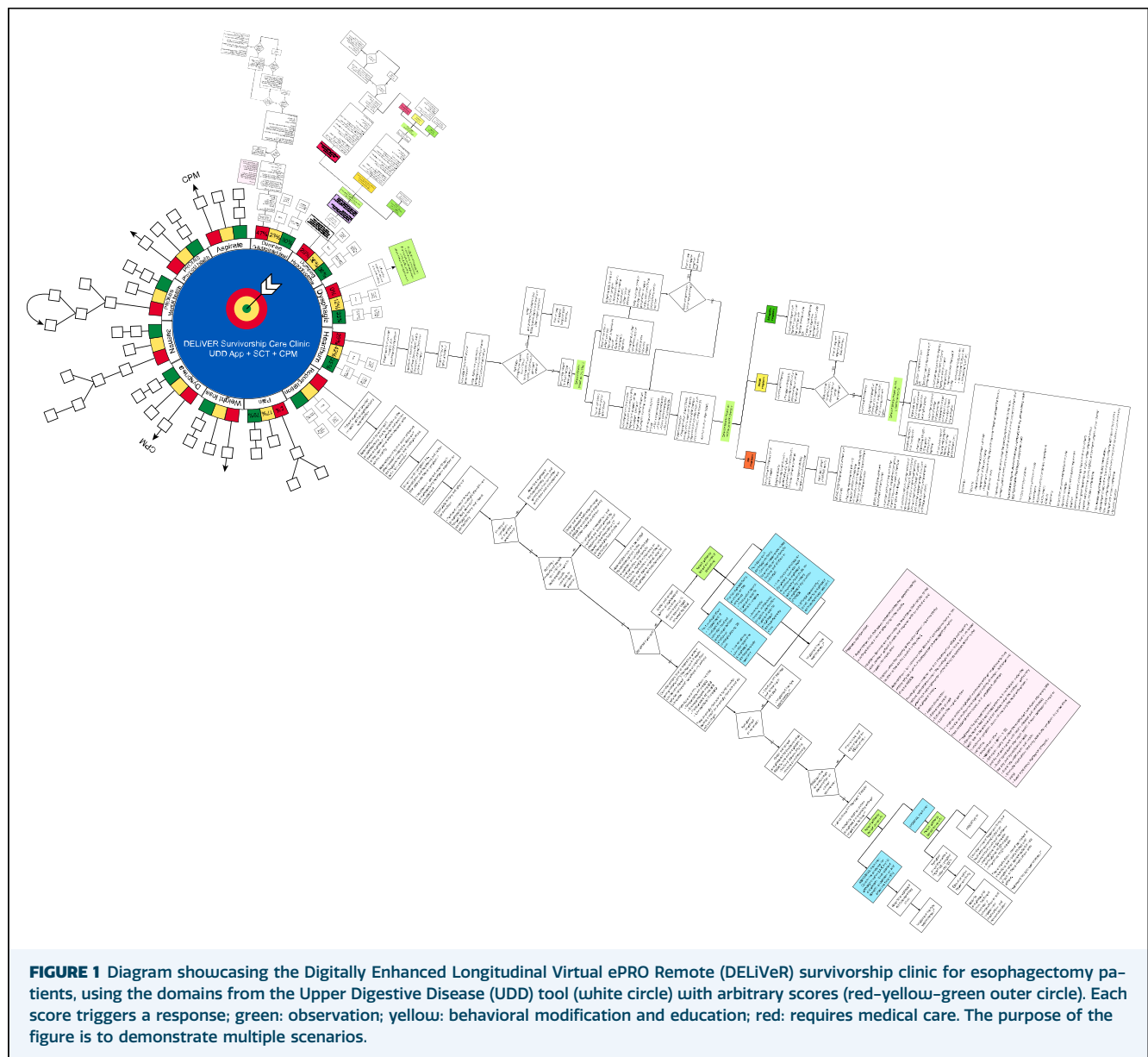
- Generative artificial intelligence-initiated care pathway models are solely insufficient to guide provider management of patients with complex symptoms after esophagectomy.
- Generative artificial intelligence might be a better tool to assist physicians with clinical care.
- This technology requires detailed instructions to provide accurate feedback.

Accepted for publication Jun 4, 2024.

Presented at the Seventieth Annual Meeting of the Southern Thoracic Surgical Association, Orlando, FL, Nov 2-5, 2023.

¹Division of Thoracic Surgery, Department of Surgery, Mayo Clinic, Rochester, Minnesota; ²Mayo Clinic Alix School of Medicine, Mayo Clinic, Rochester, Minnesota; ³Department of Health Data Science and Artificial Intelligence, University of Texas Health Science Center at Houston, Houston, Texas; and ⁴Department of Surgery, Baylor College of Medicine, Houston, Texas

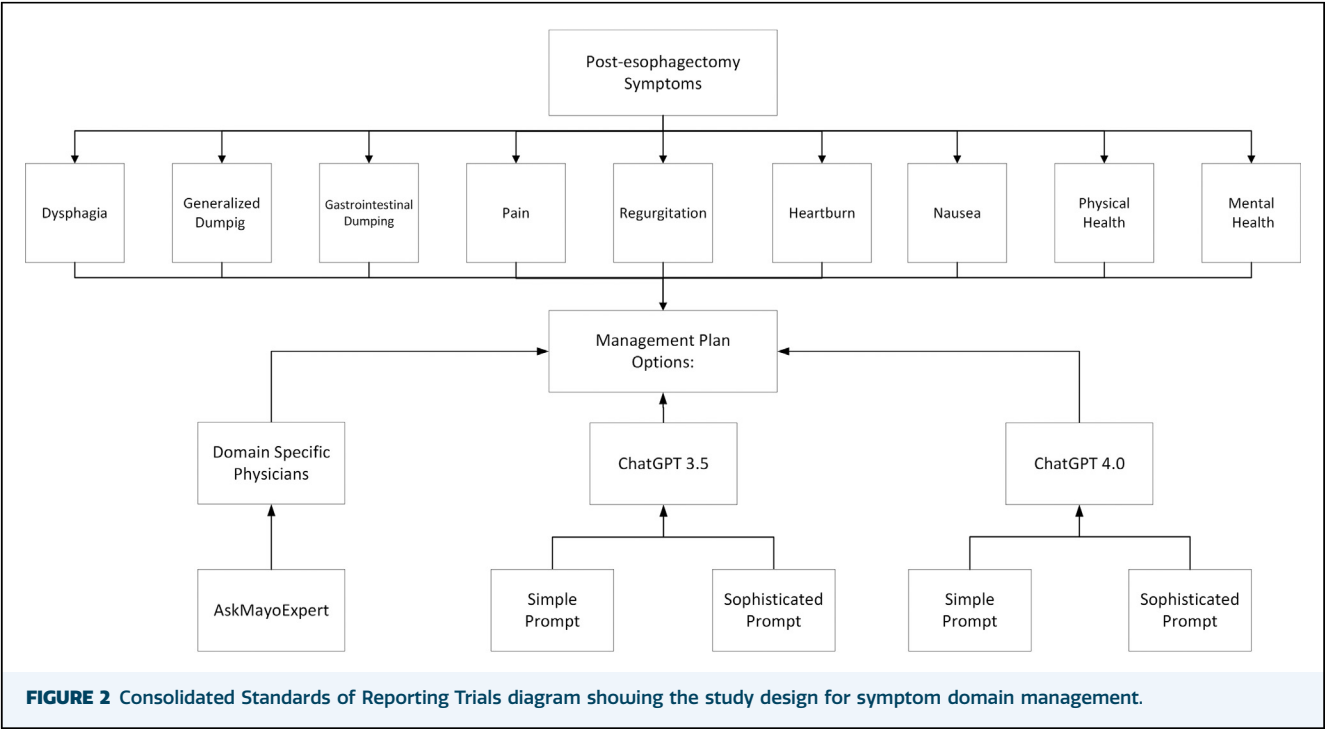
Address correspondence to Dr Blackmon, Professor of Surgery, Baylor College of Medicine, Executive Director, Baylor College of Medicine Lung Institute, 7200 Cambridge Street, Houston, TX 7703; email: shanda.blackmon@bcm.edu.



emphasizing the continual progression of chatbot technology.² As generative AI becomes more mainstream, patients and providers are likely to turn to chatbots and AI-driven tools for health care advice. Numerous projects are already attempting to use AI for clinical decision-making, patient education, and patient treatment support.³

Care pathway models (CPMs) serve as navigational frameworks, orchestrating a patient's journey from diagnosis to the achievement of targeted health outcomes. These models aim to enhance overall health while concurrently mitigating the intensity of care, contributing to a streamlined and effective approach to patient well-being.⁴

With the development of new treatment modalities for esophageal cancer and the noticed improvement in survival rates, providers are required to manage symptoms that patients continue to struggle with years after curative surgery.⁵ AI has the potential to reduce provider burden and to enhance patient care by generating or supplementing existing care management plans. To assess reliability of current AI-generated information, we aimed to compare generative AI-initiated care pathways, using ChatGPT, with expert-guided consensus-initiated care pathways from AskMayoExpert (AME)⁴ for symptom management of patients with esophageal cancer after esophagectomy.



MATERIAL AND METHODS

A formal protocol was followed for development of 9 AME CPMs that encompassed specific patient-identified symptom domains. These domains were dysphagia, generalized/systemic dumping, gastrointestinal dumping, pain, regurgitation, heartburn, nausea, physical health, and mental health, all of which are measured and scored with the Upper Digestive Disease tool.⁶ Domain-expert clinicians who are caring for patients with esophageal cancer within a single institute, Mayo Clinic, met with a certified nurse practitioner who is trained in developing CPM, along with the authors (M.K.A.C., S.H.B.), between the period of August 2022 and February 2023. Expert-derived CPM drafts were edited and approved by the study authors and domain-expert clinicians and ultimately finalized and saved (Figure 1).

On April 9, 2023, during a single session, ChatGPT 3.5 (OpenAI) was accessed to answer specific questions similar to the AME care pathway for each domain. The same question construct (simple prompt) was used for each domain:

What are the treatments recommended by physicians to manage [domain name] after an esophagectomy, and how do they assess and address the severity of the condition?

To compare outcomes from the 2 pathways, answers were recorded, and algorithms were manually compared with a survey tool composed of 5 metrics:

1. Clear definition of specific domain is provided
2. Multidirectional options are offered
3. Actionable guidance is provided to enable a provider to follow a stepwise process for management of symptom domain
4. Indications for testing are provided
5. Triage pathways are presented

On March 28, 2024, during a single session, ChatGPT 4.0 was accessed to answer a more specific question formulated from the questions of the survey tool (sophisticated prompt). The same question construct was used for each domain. ChatGPT 4.0 was prompted the same question 3 different times, on 3 separate tabs, to ensure consistency of the answers provided. Later, ChatGPT 4.0 was used to compare the 3 answers.

Please provide a clear definition of [domain name] and then provide multidirectional options for physicians to manage [domain name] after an esophagectomy.

Can you also provide actionable guidance to enable a provider to follow a stepwise process for

Domain	AskMayoExpert Care Pathway Model	ChatGPT 3.5	ChatGPT 4.0
1. General Internal Medicine	95%	85%	90%
2. Cardiology	92%	80%	88%
3. Oncology	90%	78%	85%
4. Neurology	88%	75%	82%
5. Pediatrics	93%	83%	89%
6. Geriatrics	91%	81%	87%
7. Infectious Disease	94%	84%	91%
8. Endocrinology	96%	86%	92%
9. Rheumatology	97%	87%	93%

AskMayoExpert Care Pathway Score										
Variable	Dysphagia	Generalized Dumping	Gastrointestinal Dumping	Pain	Regurgitation	Heartburn	Nausea	Physical Health	Mental Health	
Clear definition of specific domain is provided	1	1	1	1	1	1	1	1	1	
Multidirectional options are offered	1	1	1	1	1	1	1	1	1	
Actionable guidance is provided to enable a provider to follow a stepwise process for management of symptom domain	1	1	1	1	1	1	1	1	1	
Indications for testing are provided	1	1	1	1	1	1	1	1	1	
Triage pathways are presented	1	1	1	1	1	1	1	1	1	
Total	5	5	5	5	5	5	5	5	5	
ChatGPT 3.5 Score (Simple Prompt)										
Variable	Dysphagia	Generalized Dumping	Gastrointestinal Dumping	Pain	Regurgitation	Heartburn	Nausea	Physical Health	Mental Health	
Clear definition of specific domain is provided	1	1	1	1	1	1	1	1	1	
Multidirectional options are offered	1	1	1	1	1	1	1	1	1	
Actionable guidance is provided to enable a provider to follow a stepwise process for management of symptom domain	0	0	0	0	0	0	0	0	0	
Indications for testing are provided	0	0	0	0	0	0	0	0	0	
Triage pathways are presented	0	0	0	0	0	0	0	0	0	
Total	2	2	2	2	2	2	2	2	2	
ChatGPT 4.0 Score (Sophisticated Prompt)										
Variable	Dysphagia	Generalized Dumping	Gastrointestinal Dumping	Pain	Regurgitation	Heartburn	Nausea	Physical Health	Mental Health	
Clear definition of specific domain is provided	1	1	1	1	1	1	1	1	1	
Multidirectional options are offered	1	1	1	1	1	1	1	1	1	
Actionable guidance is provided to enable a provider to follow a stepwise process for management of symptom domain	1	1	1	1	1	1	1	1	1	
Indications for testing are provided	1	1	1	1	1	1	1	1	1	
Triage pathways are presented	1	1	1	1	1	1	1	1	1	
Total	5	5	5	5	5	5	5	5	5	
aResults are based on binary configuration (1 = yes, 0 = no).										

management of postesophagectomy [domain name] symptoms?

Please also provide if there is any indication for testing and/or triage.

Finally, the simple prompt was given to ChatGPT 4.0 and the sophisticated prompt was given to ChatGPT 3.5 and domain answers were compared (Figure 2).

RESULTS

ChatGPT 3.5 and AME CPMs were able to provide a definition with multidirectional management options for all 9 domains. With the simple prompt, ChatGPT failed to provide a comprehensive step-wise approach for providers to follow during management, any testing recommendations, or any form of triage process (Table). Overall score for each domain was 2/5 (40%) compared with 5/5 (100%) to the AME care pathways. The ChatGPT 3.5 answers were formatted the same for all domains. Each answer started with a short paragraph that briefly explained the domain, followed by multiple bullet points with headers that included the management options. ChatGPT 3.5 concluded each of the answers with a statement recommending working closely with the health care provider to manage the symptoms associated with each domain.

ChatGPT 3.5 did not differentiate between heartburn and regurgitation and provided the same management for each. We asked the chatbot to define each of these terms, and it provided similar answers for both with almost no differences. On the other hand, ChatGPT 3.5 was able to discriminate between hypoglycemic and gastrointestinal dumping in management and definition. Answers did not change when the questions were asked multiple times in a row.

The updated version of the generative AI tool, ChatGPT 4.0, showed improvement with responses to the sophisticated prompt in all domains. It was able to provide a detailed plan starting with multidirectional management options, actionable guidance for management, and finally indications for testing and triage. It was able to recommend accurate medications for symptomatic treatment when needed. It concluded each management plan with recommendations to closely monitor patients for signs of improvement (Table). Scores were identical for ChatGPT 4.0 and AME care pathways with a total of 5/5 (100%).

When ChatGPT 4.0 was asked to compare the 3 responses, it showed that all 3 attempts to answer

the domain-specific question had the same theme construct. It was also capable of listing differences between each run, and then it emphasized the uniqueness of each response.

ChatGPT 3.5 answers drastically improved when the sophisticated prompt was used but was not as detailed as the answers provided by ChatGPT 4.0. Also, ChatGPT 4.0 was able to provide a detailed plan when the simple prompt was used but it failed to address triage options.

COMMENT

When it comes to the management of complex patients, specifically esophageal cancer patients treated with curative esophagectomy, expert-derived CPMs remain superior to data generated by chatbots. However, when generative AI models are provided with more detailed instructions, their responses improve, although the issue of inconsistency in providing answers to the same proposed question multiple times raises concern. Even with the rapid development and astounding capability of ChatGPT 4.0 in providing medical information,⁷ it is still limited in offering a comprehensive management plan for patients. Despite the initial high costs of such technologic infrastructure, it is expected to result in long-term savings because of its positive impact on workflow, diagnosis, and treatment processes.⁸ The reliance of generative AI models on a variety of resources, some of which might not have scientifically vetted data, is a major concern for their implementation in patient care.⁹ To avoid this risk, continual clinical assessment and evaluation of generative AI models are highly recommended. Our study was limited to the use of 1 model of generative AI examples and the inability to secure external validation to the comparison that was made between the 2 symptom management approaches. Regardless of that, it seems that generative AI is still far from being specific or reliable enough to be heavily depended on when it comes to patient care and clinical decision-making.¹⁰

With the rapid advancement in machine learning and AI, we believe that generative AI models in the future will function as an adjunct to physicians by aiding them in obtaining information, verifying results, and possibly suggesting plans of care. In addition, some patients might use these tools to look up their symptoms and possible management strategies. Physicians

are expected nowadays to teach patients how to use these tools effectively to obtain accurate data.

We conclude that generative AI can be a useful supplementary tool in clinical practice but should not be used by patients or solely by providers to determine next steps in symptom management.

FUNDING SOURCES

Shanda H. Blackmon receives funding grants from STERIS Corporation and Medtronic.

DISCLOSURES

Shanda H. Blackmon reports a relationship with Medtronic that includes: speaking and lecture fees; and patent ##2018/0221026 A1 US has been issued to Shanda H. Blackmon.

REFERENCES

1. Adamopoulou E, Moussiades L. Chatbots: history, technology, and applications. *Machine Learn Appl*. 2020;2:100006.
 2. Zhang P, Kamel Boulos MN. Generative AI in medicine and healthcare: promises, opportunities and challenges. *Future Internet*. 2023;15:286.
 3. Dwivedi YK, Kshetri N, Hughes L, et al. "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *Int J Inform Manag*. 2023;71:102642.
 4. Noseworthy JH. What is ahead for Mayo Clinic? *Mayo Clin Proc*. 2014;89:440–443.
 5. Abou Chaar MK, Godin A, Saddoughi SA, et al. Patients struggle with severe symptoms even after surviving esophagectomy for esophageal cancer. *Ann Thorac Surg Short Rep*. 2023;2:98–102.
 6. Abou Chaar MK, Yost KJ, Lee MK, et al. Developing & integrating a mobile application tool into a survivorship clinic for esophageal cancer patients. *J Thorac Dis*. 2023;15:2240.
 7. Walker HL, Ghani S, Kuemmerli C, et al. Reliability of medical information provided by ChatGPT: assessment against clinical guidelines and patient information quality instrument. *J Med Internet Res*. 2023;25:e47479.
 8. Novak LL, Russell RG, Garvey K, et al. Clinical use of artificial intelligence requires AI-capable organizations. *JAMIA Open*. 2023;6:ooad028. <https://doi.org/10.1093/jamiaopen/ooad028>
 9. Khowaja SA, Khowaja P, Dev K, Wang W, Nkenyereye L. ChatGPT needs SPADE (Sustainability, PrivAcy, Digital divide, and Ethics) evaluation: a review. Preprint. Posted online March 27, 2024. arXiv:2305.03123v3 [cs.CV]. <http://doi.org/10.48550/arXiv.2305.03123>
 10. Haemmerli J, Sveikata L, Nouri A, et al. ChatGPT in glioma adjuvant therapy decision making: ready to assume the role of a doctor in the tumour board? *BMJ Health Care Inform*. 2023;30:e100775.
-