International Journal of Research in Medical Sciences
Sabuwala NM et al. Int J Res Med Sci. 2026 Sep;14(9):xxx-xxx
www.msjonline.org
pISSN 2320-6071 | eISSN 2320-6012
DOI: https://dx.doi.org/10.18203/2320-6012.ijrms20262803

Original Research Article

A cross-sectional study to assess and compare artificial intelligence-generated content for medical professionals with UpToDate on management of acute coronary syndrome

Naqeeya M. Sabuwala[1]*, Tayyaba A. Mufti[2], Hafiz Z. Iqbal[3], Ravi J. Rathod[4], Sriram Pathuri[5]

[1]Department of Cardiology, Narendra Modi Medical College, Ahmedabad, Gujarat, India
[2]Department of Emergency Medicine, University Hospital Waterford, Waterford, Ireland
[3]Department of Anesthesia, University hospital Waterford, Waterford, Waterford, Ireland
[4]Department of Internal Medicine, Pandit Deendayal Upadhyay medical College, Rajkot, Gujarat, India
[5]Department of Cardiology, Katuri Medical College, Guntur, Andhra Pradesh, India

Received: 07 April 2026
Revised: 05 August 2026
Accepted: 06 August 2026

***Correspondence:**
Dr. Naqeeya M. Sabuwala,
E-mail: sabuwalanaqeeya@gmail.com

ABSTRACT

Background: The rapid rise of artificial intelligence (AI) in medicine offers new ways to generate educational content, but its readability and accessibility for both clinicians and patients remain underexplored. This study compares AI-generated medical information to a standard clinical reference.
Methods: This cross-sectional study analyzed educational content on acute coronary syndrome (ACS) management generated by DeepSeekAI and retrieved from UpToDate. Readability was assessed using four established metrics: Flesch reading ease (FRE), Flesch-Kincaid grade level (FKGL), SMOG index, and difficult word percentage. Data were collected across six distinct topics related to ACS management, and the scores for each tool were statistically compared.
Results: Readability analysis revealed notable disparities between the two sources. UpToDate consistently produced content with significantly higher FRE scores (e.g., UpToDate FRE up to 40.2 vs. DeepSeek AI FRE down to 2.7 for topic 1) and considerably lower FKGLs, SMOG Indices, and percentages of difficult words across all examined topics. While both sources generated text with relatively high reading levels indicative of medical content, DeepSeekAI consistently demonstrated lower readability, suggesting more complex vocabulary and sentence structures.
Conclusions: UpToDate content on ACS management is more readable and accessible for clinicians than DeepSeekAI-generated content, highlighting the need to improve readability in AI-generated medical materials for effective clinical use.

Keywords: Artificial intelligence, DeepSeekAI, UpToDate, Readability, Acute coronary syndrome, Medical education, Flesch reading ease, Flesch-Kincaid grade level, SMOG index

INTRODUCTION

An acute myocardial infarction is defined as an acute myocardial injury due to insufficient blood supply to myocardium which causes significant increase of cardiac troponins (cTn). myocardial ischemia, which is defined and diagnosed by certain clinical, electrocardiographic, imaging and angiographic criteria.[1] Acute MI is a common condition encountered in clinical practice, requiring accurate and timely diagnosis and management. Timely diagnosis and medical intervention is a crucial step in acute MI management. Medical education on this condition is critical, as it ensures healthcare professionals stay informed about evolving diagnostic approaches, treatment

guidelines, and potential complications. Up-to-date knowledge enables clinicians to provide better care, reduce the risk of misdiagnosis, and improve patient outcomes. UpToDate is a subscription-based, evidence-based clinical decision support resource widely used by medical professionals. It provides access to the latest medical information, including topic reviews, drug information, and patient education materials, to help clinicians make informed decisions at the point of care.[2]

Recently, AI tools have emerged as supplementary resources in medical education. Large language models (LLMs), such as DeepSeek AI, can generate real-time content on a wide range of medical topics. Designed specifically for clinical and scientific domains, DeepSeek AI is trained on biomedical literature and aims to deliver clinically relevant and accurate information.[3,4] These tools offer advantages such as accessibility and rapid summarization, but concerns remain regarding the accuracy, source verification, and potential for over-reliance without expert review. The use of these technologies raises a number of ethical concerns, including the possibility of plagiarism and errors, as well as a possible disparity in accessibility between high- and low-income nations should the program start to charge. This is why it will soon be necessary to reach an agreement on how to control the use of chatbots in scientific writing.[5]

In this study, the authors compare the content generated by DeepSeek AI with that from UpToDate on the topic of management of acute myocardial infarction. The focus is on evaluating readability using standardized metrics such as the FKGL and reading ease scores. This comparison aims to assess whether AI-generated educational material is suitable for medical professionals and if it can serve as a reliable supplement to established clinical references.

METHODS

This cross-sectional original research study was conducted at Narendra Modi Medical College Ahmedabad, Gujarat, India and data were collected virtually over a one-week period from April 13 to April 20, 2025. As the study did not involve human participants, patient-identifiable information, biological specimens/clinical interventions, Institutional Ethics Committee approval was not required.

The study evaluated the readability of educational content related to the management of ACS. Six clinically relevant subtopics within ACS management were selected to provide a comprehensive comparison between AI-generated content and a standard evidence-based clinical reference. Educational content for each subtopic was generated using DeepSeek AI (accessed April 13, 2025) with the standardized prompt:

"Write a detailed educational guide for medical professionals on (topic), including definition, clinical features, diagnosis, and treatment options."

For each corresponding topic, educational content was retrieved from UpToDate (accessed April 13, 2025), a widely used evidence-based clinical decision support resource. To ensure uniformity and comparability, only the main educational text was included for analysis, while figures, tables, references, hyperlinks, author information, and supplementary material were excluded.

DeepSeek AI and UpToDate served as the data sources for this study. Each educational text was analyzed independently using the same assessment method. The texts were assessed for readability using the online tool from WebFX.[6] The following parameters were recorded for each response: Word count, sentence count and average words per sentence, FRE, FKGL and SMOG index, difficult word count and difficult word percentage.

The primary outcome of the study was the comparison of readability between DeepSeek AI-generated educational content and UpToDate content using standardized readability metrics. Secondary variables included word count, sentence count, average words per sentence, FRE score, FKGL, SMOG Index, difficult word count, and difficult word percentage.

All data were entered into Microsoft excel and analyzed using R software (version 4.3.2). Continuous variables were summarized using median and interquartile range (IQR). Comparisons between DeepSeek AI and UpToDate were performed using the Mann-Whitney U test. A two-sided $p<0.05$ was considered statistically significant.

RESULTS

DeepSeek AI and UpToDate were used to generate educational content on six subtopics within the management of ACS. The readability of these responses was evaluated using standard metrics: word count, sentence count, average words per sentence, FRE, FKGL, SMOG index, difficult word count, and difficult word percentage.

As shown in Table 1, there was a significant difference in key readability metrics between DeepSeek AI and UpToDate content: word count was significantly lower in DeepSeek AI (Median: 262.5) than UpToDate (Median: 2544.5), p=0.006. Sentence Count followed a similar pattern (DeepSeek AI median: 27.0 vs. UpToDate: 144.0), p=0.016. FRE was significantly lower for DeepSeek AI (Median: 21.4) than UpToDate (Median: 28.2), p=0.025, indicating more difficult text. Difficult word count and difficult word percentage were both significantly higher in DeepSeek AI content, with difficult word percentage reaching 31.6% versus 24.8% in UpToDate (p=0.004).

No significant differences were observed in FKGL and SMOG index (p=1.000 and p=0.297, respectively), though values still indicated content appropriate for a highly educated audience.

Table 1: Characteristics of responses generated by UpToDate and DeepSeek AI.

Variables	Median (IQR)		U statistic	P value
	UpToDate	DeepSeek AI		
Word count	2544.5 (1430.5-3756.0)	262.5 (187.2-333.8)	1	0.006*
Sentence count	144.0 (86.5-248.5)	27.0 (11.5-39.5)	3	0.016*
Word/sentence count	16.5 (14.9-18.0)	9.9 (8.9-15.3)	6	0.053
FRE	28.2 (23.1-41.4)	21.4 (13.9-23.2)	4	0.025*
FKGL	13.5 (11.4-14.3)	12.8 (12.4-14.7)	18	1.000
Smog index	11.8 (10.6-12.6)	10.4 (9.8-12.6)	11.5	0.297
Difficult word count	667.0 (400.5-894.2)	85.5 (59.2-111.8)	3	0.016*
Difficult word percentage	24.8 (22.0-28.2)	31.6 (30.5-37.0)	0	0.004*

*Mann-Whitney U Test. P<0.05 are considered statistically significant.

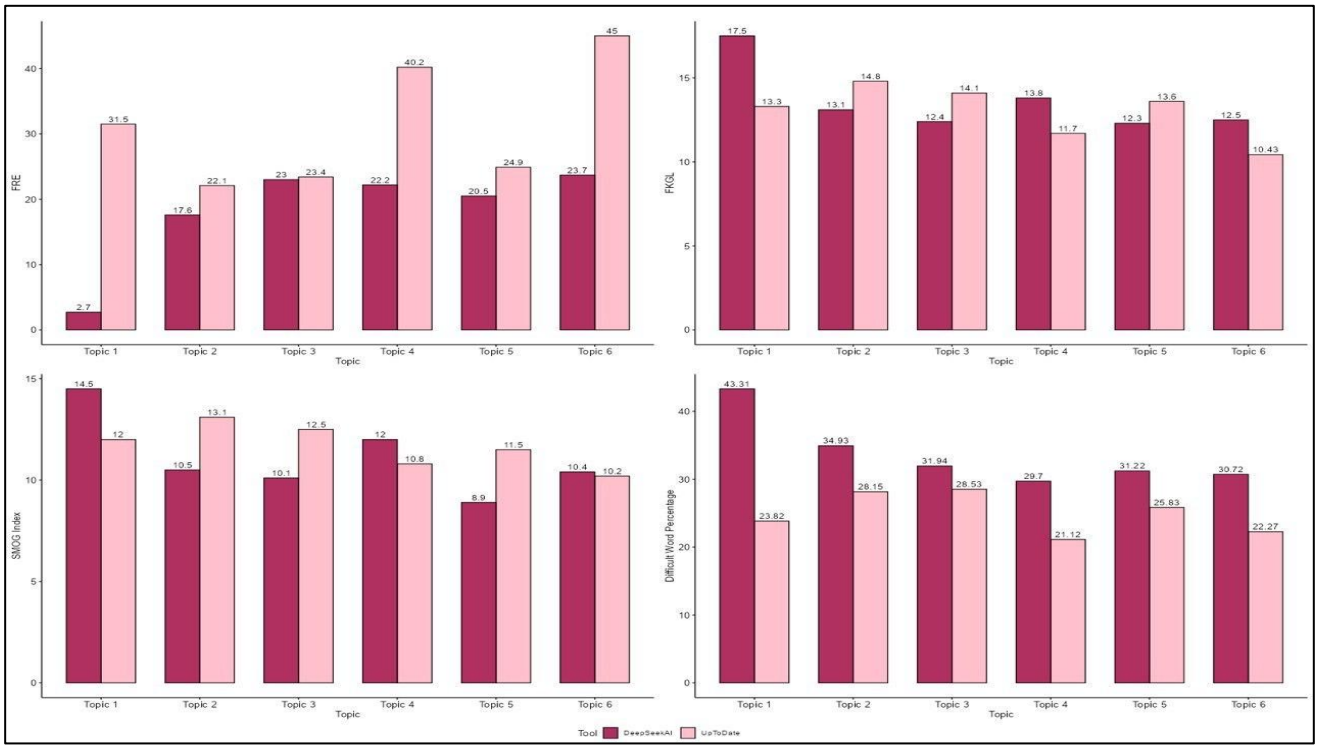

Figure 1: Comparison between FRE, FKGL, SMOG index and difficult word percentage for the patient education guide generated by UpToDate and DeepSeek AI.

Figure 1 illustrates the graphical comparison across the six ACS subtopics. FRE scores were consistently lower for DeepSeek AI, suggesting poorer readability. Topic 1 showed the largest gap (FRE: DeepSeek AI 2.7 vs. UpToDate 40.2). While FKGL and SMOG indices showed variation across topics, DeepSeek AI maintained consistently higher grade levels. Difficult word percentages were markedly elevated in DeepSeek AI responses across all subtopics.

DISCUSSION

This study compared educational content generated by DeepSeek AI with that from UpToDate on ACS, focusing on how readable and useful the material was for healthcare professionals. As shown in Table 1 and Figure 1, there were clear differences in both readability scores and how the content was presented. While both sources were factually accurate and grounded in evidence, DeepSeek AI produced shorter and denser text that was noticeably harder to read. This reflects a broader trend we're starting to see: AI tools may deliver information quickly, but they sometimes do so in less user-friendly ways, especially for clinicians working under pressure.

AI platforms like DeepSeek are becoming more common in medical education. Their speed and scale could be a real asset, particularly in settings where clinicians are stretched thin and don't have time to dig through traditional resources. However, our findings suggest that UpToDate still has the edge when it comes to usability. Its content is laid out with clinical context in mind, offering clear formatting and well-structured guidance. In contrast, DeepSeek AI, while informative, can be hard to navigate and absorb-especially in fast-paced environments like

emergency rooms or hospital wards. This concern isn't new: other studies have pointed out similar challenges when using AI tools for frontline medical decision-making.

When we looked more closely at the readability, the numbers told the story. DeepSeek's content scored lower on the FRE scale and higher on the SMOG index, meaning it used more complex language and sentence structures. For busy clinicians, this could pose a problem. Quick understanding is often crucial, and if the language slows that down, it can get in the way of good decision-making. Fijačko et al found similar results when examining AI-generated material for sepsis management-it was accurate but more difficult to read.[7] Kung et al and Ayers et al also flagged that AI models like ChatGPT can struggle with presenting information clearly and concisely when clinicians need it most.[8-10]

Other studies echo this concern. Lee et al found that GPT-4, even when used to generate patient education materials, often produced content at reading levels above the recommended level.[13] Hancı et al saw a similar issue in the context of palliative care: chatbot-generated text needed editing before it could meet basic literacy standards.[12] This suggests that while AI is good at producing content fast, it still needs help when it comes to tailoring information to the people who will use it.

That said, there's another side to this conversation. Research by Kung et al showed that AI models can perform as well as medical students on licensing exams, which hints at their potential in training environments.[8] Singhal et al similarly demonstrated that LLMs can encode extensive clinical knowledge and generate medically accurate information, though their outputs-like those of DeepSeekAI in our study-often require refinement to achieve optimal clarity and usability for healthcare professionals.[11] Rao et al emphasized how much the quality of AI output depends on how the prompt is written-something Patil et al explored further by showing that prompt design can make a big difference in how understandable and useful the results are.[11] So while there are limitations, there's also a clear path forward if these tools are used thoughtfully.[14,15]

Limitations

This study isn't without its limitations. We looked at only one AI model-DeepSeek-and focused on just one clinical topic: acute coronary syndrome. We didn't test how accurate the clinical recommendations were, or how they might influence decisions in actual practice. And we didn't look at how users, such as doctors or students, experience the material or whether they find it helpful in real-world situations. These are all important areas for future research. Still, our findings highlight a real gap between the potential of AI in medicine and its readiness for everyday use-at least for now.

CONCLUSION

This study compared educational and medical data generated from DeepSeek AI and UpToDate on acute coronary syndrome for medical professionals. Significant differences were found in readability metrics, especially word count and difficulty, with DeepSeek AI producing more concise but denser and complex content. While both follow standard medical knowledge and protocols, DeepSeekAI data needs further refinement to enhance readability and clinical usability compared to UpToDate. It holds promise as a supplement for healthcare professionals but requires improvements in ease of understanding, and additional studies are needed to achieve better accuracy, clarity, and adaptability across various medical disciplines.

Declaration

The authors declare the use of DeepSeek and Chatsonic as research tools in the methodology of this study. The questions were entered into these models to generate data for comparative analysis. The human authors were responsible for the study design, data collection, analysis, and validation of all results. The authors did not use any generative AI or AI-assisted technology for the writing or editing of this manuscript.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: The study was approved by the Institutional Ethics Committee

REFERENCES

1. Nori H, King N, McKinney SM, Carignan D, Horvitz E. Capabilities of GPT-4 on medical challenge problems. arXiv. 2023;arXiv:2303.13375.
2. Rao A, Pang M, Kim J, Schnickel R, Merrill JA, Ensley CL, et al. Assessing the utility of ChatGPT throughout the entire clinical workflow: development and usability study. JMIR Med Inform. 2023;11:e48659.
3. UpToDate. About Us. Available at: https://www.uptodate.com/contents/search. Accessed on 13 April 2026.
4. DeepSeek AI. DeepSeek homepage. Available at: https://chat.deepseek.com/. Accessed on 13 April 2026.
5. DeepSeek AI. DeepSeek Medical GitHub repository. Available at: https://github.com/deepseek-ai. Accessed on April 13, 2025.
6. Salvagno M, Taccone FS, Gerli AG. Can artificial intelligence help for scientific writing? Crit Care. 2023;27(1):75.
7. WebFX. Readable: Free readability test tool. Available at: https://www.webfx.com/tools/read-able/. Accessed on 13 April 2026.
8. Fijačko N, Gosak L, Štiglic G, Picard CT, Douma MJ. Can ChatGPT pass the life support exams without

entering the American Heart Association course? Resuscitation. 2023;185:109732.
9. Kung TH, Cheatham M, Medenilla A, Chavez C, Hicks L, Foster N, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health. 2023;2(2):e0000198.
10. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med. 2023;183(6):589-96.
11. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. Nature. 2023;620(7972):172-80.
12. Hancı V, Ergün B, Gül Ş, Uzun Ö, Erdemir İ, Hancı F. Assessment of readability, reliability, and quality of ChatGPT®, BARD®, Gemini®, Copilot®, and Perplexity® responses on palliative care. Medicine (Baltimore). 2024;103(33):e39305.
13. Lee M, Cadiente A, Chen J, Zhou Y, Greenwald B. An evaluation of the reliability and readability of large language models in the dissemination of traumatic brain injury information. Digit Health. 2025;11:20552076251350760.
14. Patil R, Heston TF, Bhuse V. Prompt engineering in healthcare. Electronics. 2024;13(15):2961.
15. MacDonald S, Steven K, Trzaskowski M. Interpretable AI in healthcare: enhancing fairness, safety, and trust. In: Raz M, Nguyen TC, Loh E, editors. Artificial Intelligence in Medicine. Singapore: Springer. 2022;85-110.

Cite this article as: Sabuwala NM, Mufti TA, Iqbal HZ, Rathod RJ, Pathuri S. A cross-sectional study to assess and compare artificial intelligence-generated content for medical professionals with UpToDate on management of acute coronary syndrome. Int J Res Med Sci 2026;14:xxx-xx.