International Journal of Research in Medical Sciences
Harini SS et al. Int J Res Med Sci. 2026 Nov;14(11):5143-5147
www.msjonline.org
pISSN 2320-6071 | eISSN 2320-6012
DOI: https://dx.doi.org/10.18203/2320-6012.ijrms20264170

Original Research Article

AI in patient education: a cross-sectional comparative analysis of patient information guides produced by ChatGPT versus Google Gemini for common haematology conditions

Sri Harini S.[1], Himabindu Dande[2]*, Sujitha Nagabhyru[3], Ayesha Shafiya[4], Anjana Kurian[5], Eda Sai Venkata Teja[6], Nithin Reji[7]

[1]Department of Internal Medicine, Government Thiruvannamalai College, Thiruvannamalai, Tamil Nadu, India
[2]Department of General Medicine, Dinesh Medical Centre (Dr. E. C. Gangi Reddy Hospital), Islampuram, Pulivendula, Andhra Pradesh, India
[3]Department of Internal Medicine, Mother Theresa Multi-Speciality Hospital, Narasaraopet, Andhra Pradesh, India
[4]Department of Internal Medicine, St. Martha's Hospital, Bengaluru, Karnataka, India
[5]Department of Plastic Surgery, Aster MIMS Kannur, Chala, Keralam, Kerala, India
[6]Acute Medical Care Unit, RK Hospital, Visakhapatnam, Andhra Pradesh, India
[7]Department of Cardio Thoracic and Vascular Surgery, Malankara Orthodox Syrian Church Medical College, Kochi, Keralam, Kerala, India

Received: 28 July 2026
Accepted: 13 October 2026

***Correspondence:**
Dr. Himabindu Dande,
E-mail: bindu.hima6934@gmail.com

Copyright: © the author(s), publisher and licensee Medip Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

Background: Haematological disorders encompass a diverse range of conditions affecting blood and blood-forming organs. Educating Patients is crucial in managing these disorders, enabling individuals to understand their conditions, associated complications, and preventive measures. As artificial intelligence (AI) continues to progress, chatbots such as ChatGPT (CG) and Google Gemini (GG) offer readily available patient information.
Methods: This cross-sectional comparative study was performed from January 15, 2025, to January 22, 2025. Six haematological conditions were selected. Patient education materials were created using CG (version 4.0) and GG, following standardized prompts. The generated texts were analysed for readability, similarity, and reliability with the use of Flesch-Kincaid Calculator (FKC), Quill Bot Plagiarism Tool (QPT) and Modified DISCERN score, respectively.
Results: The study demonstrated no significant difference between the CG and GG in word count (WC), sentence count (SC), readability score, or reliability score in the statistical evaluation. The CG demonstrated a slightly higher ease-of-understanding score, but this difference was not statistically significant (p=0.175).
Conclusions: AI tools including CG and GG have an important role in educating patients by delivering quick and structured information on haematological conditions. Future studies should explore the role of AI in other medical domains and investigate ways to enhance AI-generated patient education materials.

Keywords: Artificial intelligence, ChatGPT, Google Gemini, Haematology, Iron deficiency anaemia, Patient education, Vitamin B12 deficiency

INTRODUCTION

Haematological disorders are diverse range of disorders involving blood and blood forming organs, which can range from deficiency causing anaemia and excess causing malignancies like leukaemia.[1] Iron deficiency anaemia refers to a deficiency in haemoglobin secondary to iron scarcity in the body, resulting in hypochromic microcytic

blood picture, presenting with nonspecific complaints like fatigue, dyspnoea on exertion.[2] Vitamin B12 is a water-soluble vitamin deficiency which can lead to various haematological and neurological disorders.[3] Sickle cell anaemia is a type of autosomal recessive disorder caused by the abnormal haemoglobin synthesis causing haemolysis and chronic organ damage.[4] Thalassemia is a heterogenous group of disorders resulting from decreased alpha or beta globin synthesis.[5] Haemophilia is X linked recessive disorder resulting from deficiency of clotting factors 8 and 9, characterised by prolonged and excessive bleeding.[6] Leukaemia is a haematological malignancy due to abnormal leukocyte production involving hematopoietic stem cells.[7]

In this study, the authors intend to understand how artificial intelligence (AI) is utilized in public education, as it serves as a valuable tool which enables patients to understand about their disease, the complications associated with it and assist them in following the precautions and maintain good health.

The AI tools that we have extensively studied are ChatGPT (CG) and google Gemini (GG). CG is an AI-driven chatbot that employs natural language processing to generate conversations that resemble human interaction.[8] GG is a family of AI models, similar to Open AI's GPT. These models are all multi-modal, meaning they possess the capability to comprehend and produce text like a typical large language model (LLM). Additionally, they have the inherent ability to understand, manipulate, and integrate various types of information, such as images, audio, videos, and code [9]

In the era of technology and availability of resources and access many people are able to utilise these AI tools. They provide rapid access to the basic and reliable information in simple term which can be understood by the general public. The aim of this study was to compare the effectiveness of these AI chatbots in producing patient information guides for prevalent haematological conditions (iron deficiency anaemia, thalassemia and sickle cell anaemia), based on metrics like readability and ease of understanding of the content produced by both AI systems.

METHODS

An original research study with a cross-sectional design was carried out over a one-week period from January 15, 2025 to January 22, 2025. As this research did not involve any human subjects, approval from the Ethics Committee was not needed.[10,11]

Six prevalent diseases in haematology were selected such as iron deficiency anaemia, sickle cell disease, thalassemia, haemophilia, leukaemia, vitamin B12 deficiency. Two Al tools, namely on January 16, ChatGPT (CG) version 4.0 and Google Gemini (GG) were selected to develop patient education guides.[12,13] Prompts were given to both the Al tools for all six diseases separately [example: create a patient information guide for iron deficiency anaemia]. The responses generated by these tools were compiled into a Microsoft Word Document. The Flesch-Kincaid Calculator (FKC) was utilized to evaluate responses based on word count (WC), sentence count (SC), ease of understanding and readability of the generated content.[14] The Quill bot Plagiarism Tool (QPT) was utilized to conduct a similarity check.[15] The reliability of the scientific text was evaluated using the Modified DISCERN score. The Modified DISCERN score consists of a 5-question survey designed to evaluate the reliability of scientific and health-related information. Each item receives a score of either 0 or 1. A total score of 5 indicates high reliability and a score of 0 reflects poor reliability.[16]

Finally, the information was transferred to a Microsoft Excel spreadsheet for statistical evaluation. The statistical analysis was conducted using R version 4.3.2, as developed by the R core team in 2023. This software, titled "R: A Language and Environment for Statistical Computing," is provided by the R Foundation for Statistical Computing, located in Vienna, Austria.

The Unpaired t-test was applied to analyse the outcomes generated by CG and GG. A P value below 0.05 was deemed significant. Pearson's Correlation Coefficient was used to analyse the relationship between the ease score and the reliability score.

RESULTS

ChatGPT (CG) and Google Gemini (GG) were utilized to create patient education guides for various common haematological conditions, including iron deficiency anaemia, vitamin B12 deficiency anaemia, sickle cell anaemia, thalassemia, haemophilia, and leukaemia.

The mean values, standard deviation (SD)and p values of words, sentences, Average Words Per Sentence (AWPS), Average Syllables Per Word (ASPW), Grade Level (GL), Ease Score (ES), similarity percent and reliability score (RS) which are the characteristics of responses generated by CG and GG, are shown in Table 1.

An independent sample t-test was performed to assess the mean differences between CG and GG. The Shapiro-Wilk Test was employed to evaluate the normality of the variables, and equality of variances was tested using Levene’s test. Based on the p values obtained in Table 1, no significant difference found between the responses generated by CG and GG statistically.

Nevertheless, the reliability score was relatively better for responses produced by CG compared to GG (p=0.175). The graphical comparison of GL, ES, similarity% and RS for the patient education guide produced by CG and GG was represented in Figure 1.

The correlation between CG and GG was calculated with an ease score and reliability score using Pearson's coefficient of correlation which represents that CG has better correlation coefficient (r=0.31) than GG (r=0) (Table 2).

Figure 2 illustrates that there is a slight positive correlation between the ES and the RS for CG. In contrast, for GG, there is no correlation between the ES and the RS.

Table 1: Comparison of responses produced by CG and GG.

Variable	ChatGPT		Google Gemini		P value*
	Mean	SD	Mean	SD	
Words	468.50	149.11	503.83	326.53	0.814
Sentences	56.83	16.48	62.50	33.32	0.717
AWPS	8.45	2.52	7.92	0.98	0.639
ASPW	1.88	0.08	1.95	0.12	0.282
GL	9.92	1.62	10.50	1.34	0.511
ES	38.93	7.96	33.82	9.98	0.349
Similarity %	52.98	19.18	43.73	8.62	0.307
RS	2.67	0.52	3.00	0.00	0.175

*t-test, p values <0.05 are deemed statistically significant

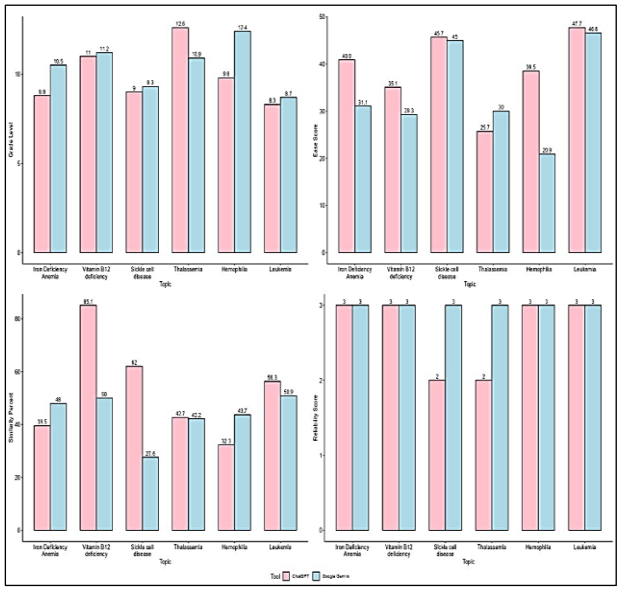

Figure 1: Comparison of GL, ES, similarity % and RS for the patient education guide produced by CG and GG.

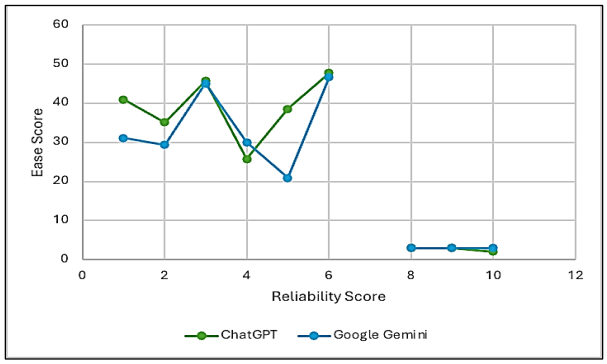

Figure 2: Correlation for CG and GG was compared with ES and RS.

Table 2: Relationship between ChatGPT and Google Gemini.

Chat bots	r value
ChatGPT	0.31
Google Gemini	0

DISCUSSION

A study with cross sectional design was carried out to evaluate the outputs generated by two different AI tools, ChatGPT (CG) and Google Gemini (GG), for patient education brochures on various haematological conditions, including iron deficiency anaemia, sickle cell anaemia, vitamin B12 deficiency anaemia, haemophilia, thalassemia and leukaemia indicated no significant difference in word count(WC), sentence count (SC) ease score (ES), reliability score (RS) or grade level (GL) between the responses generated by CG and GG in statistical evaluation.

Natural Language Processing (NLP) is a crucial component of AI that allows machines to comprehend, process, and develop human like language. Large Language Models (LLMs), such as GPT-4, are built using deep learning techniques and vast neural networks, allowing them to carry out various activities, such as responding to inquiries, creating text, and translating languages, analysing sentiment, and classifying information. These models are trained on extensive datasets, making them highly effective in applications like chatbots, virtual assistants, and automated content creation.[17]

The rise of freely accessible AI dialogue platforms provides a unique opportunity to adapt existing patient education materials to align with recommended readability levels.[18] AI chatbots can help patient's process medical information by delivering real-time, accurate answers to

their questions, serving as an easily accessible, on-demand resource for healthcare guidance.[17]

Readability plays an important role in understanding patient education materials. Health materials should match the reader's education level, ensuring clarity and accessibility. To improve readability, it is suggested to use concise sentences, ideally between 8 to 10 words, making health information easier to comprehend.[19] In this study, average words per sentence remains at the desirable level for both the AI tools studied.

ES with this respective study calculated with Flesch Kincaid calculator (FKC) demonstrates the mean ES being at the college level for both the AI tools. Even though CG is slightly better than GG, it remains statistically insignificant.

In a study assessing the precision and readability of AI chatbot answers to questions about Vasectomy indicated that, on average, all five AI chatbots had Flesch Reading Ease (FRE) scores below 50, corresponding to a reading level above the 10th grade.[20]

In a study conducted in 2024 study, Ghanem, et al investigated the quality of AI tool's answers to questions about medical information related to appendicitis. Upon evaluating the readability of each platform through FRES and FKGL, it was discovered that responses generated by all AI platforms were challenging to read, with Google Bard being the easiest to understand among them.[21]

CG enhanced the FRE score of the 10 most challenging online articles, raising it from a university level to a 10th grade level (19.7±4.4 vs. 56.9±5.9, $p<0.001$). The article was easier to read than 48.5% of articles (38.9 vs. 39.4±12.4) and surpassed 94% in quality (51.0 vs. 37.6±6.1).[22]

The issue of plagiarism in AI chatbots is becoming increasingly significant, particularly when they generate responses that closely resemble copyrighted or proprietary material without appropriate attribution. Plagiarism in medical science is particularly harmful because it undermines the integrity, credibility, and progress of healthcare research and practice. Innovation in medicine comes from novel ideas and rigorous research.

The DISCERN tool, which is frequently cited in academic literature, is a well-established method for evaluating the quality of generated texts. It has been employed in numerous studies to assess healthcare information intended for patients.[21] In this study, the DISCERN score of CG and GG are 2.67 and 3 respectively but there's no statistical significance.

A study by Golan et al evaluated the quality of information on shock wave therapy available online for erectile dysfunction and found that CG's ability to evaluate text content quality was insufficient and did meet the standards established by human reviewers and reliable tools.[23]

A different study, which involved inquiries about cancers in kidney cancer, bladder cancer, prostate cancer, urinary stones and benign prostatic hypertrophy through CG, found that all treatment responses received a medium quality score according to DISCERN. The researchers emphasized the need for caution when interpreting information from CG, as the chatbot does not reveal its information sources and may exhibit bias, even with straightforward questions about fundamental urological diseases.[24]

Numerous studies have found that while chatbot responses were generally accurate, they were often composed at a reading level that was challenging unless specific prompts were used to improve readability.[25]

In Ghanem et al, study, evaluators used the Likert scale to measure the accuracy of chatbot responses in relation to the standard of care. In terms of accuracy, all platforms showed similar performance, with a minimum mean score of 4, indicating 'Somewhat Accurate'. Despite CG-3.5 achieving the top accuracy score, statistical analysis indicated that there were no significant differences among the five AI chatbot engines. This suggests that while CG led in accuracy, other platforms remained close in performance. Notably, CG's higher accuracy appeared to come at the expense of readability, requiring a higher level of comprehension compared to other platforms.[21]

This study has few limitations. This original research analyses between only two of the recent popular chatbots, where analysis of other AI chatbots is to be done by many more original researchers. This study conducted on a small scale, aimed to compare two AI platforms which resulted in no significant differences but upcoming studies need to explore similar study with a larger dataset so that any statistical significance may be obtained. Additional limitations are that the DISCERN score that is used for reliability testing is a subjective measure. Though AI tools can integrate data from various sources, their accuracy depends on the reliability and timeliness of sources they access. Thus, AI tools may fail to adapt to rapidly evolving medical knowledge without human oversight, requiring continuous validation and expert supervision for patient education and clinical decision-making support.

CONCLUSION

Future studies should evaluate various AI tools across multiple medical conditions, focusing on their adherence to current medical guidelines as well as their accuracy, reliability, and readability. Incorporating patient feedback on usefulness may be essential in enhancing the quality and applicability of patient education guides. While AI improves the readability of patient education materials, it remains a challenge to maintain accuracy. Therefore, future studies should assess content that includes the latest

guidelines and meets readability standards to ensure effective patient education.

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

REFERENCES

1. Van Stavern GP, Smith LD, James HR. Hematologic disorders. Albert Jakobiec's Princi Pract Ophthalmol. 2021:1-40.
2. 2. Warner MJ, Kamran MT. Iron deficiency anaemia. In: StatPearls. Treasure Island (FL): StatPearls Publishing; 2025.
3. Ankar A, Kumar A. Vitamin B12 deficiency. In: StatPearls. Treasure Island (FL): StatPearls Publishing; 2025.
4. Mangla A, Ehsan M, Agarwal N. Sickle cell anaemia. In: StatPearls. Treasure Island (FL): StatPearls Publishing; 2025.
5. Bajwa H, Basit H. Thalassemia. In: StatPearls. Treasure Island (FL): StatPearls Publishing; 2023.
6. Mehta P, Reddivari AKR. Hemophilia. In: StatPearls. Treasure Island (FL): StatPearls Publishing; 2023.
7. Chennamadhavuni A, Iyengar V, Mukkamalla SKR, Shimanovsky A. Leukemia. In: StatPearls. Treasure Island (FL): StatPearls Publishing; 2023.
8. OpenAI. ChatGPT: Optimizing language models for dialogue, 2023. Available at: https://openai.com/chatgpt. Accessed 02 April 2026.
9. Guinness H. What is Google Gemini? What you need to know, 2024. Available at: https://zapier.com/blog/google-gemini/. Accessed 02 April 2026.
10. Morrissey M, McLaughlin J, O'Connor Y. Reporting of ethical considerations in qualitative research using social media data. J Med Internet Res. 2024;26:e51496.
11. Deoghare S. Virtual research designs and IRB requirements: clarifying what truly needs ethics approval. Minerva Med. 2026;117(2):96-7.
12. OpenAI. ChatGPT, 2025. Available at: https://chatgpt.com/. Accessed 02 April 2026.
13. Google. Gemini-Chat to supercharge your ideas, 2024 Available at: https://gemini.google.com/. Accessed 02 April 2026.
14. Good Calculators. Flesch Kincaid Calculator, 2019 Available at: https://goodcalculators.com/flesch-kincaid-calculator/. Accessed 02 April 2026.
15. Quill Bot, Plagiarism checker, 2026. Available at: https://quillbot.com/plagiarism-checker. Accessed 02 April 2026.
16. Li HOY, Bailey A, Huynh D, Chan J. YouTube as a source of information on COVID-19: a pandemic of misinformation? BMJ Glob Health. 2020;5(5):e002604.
17. Didier AJ, Moss G, Sutton JM. Applications of artificial intelligence for cancer patient education. J Cancer Educ. 2025;40(3):313.
18. Kirchner GJ, Kim RY, Weddle JB, Bible JE. Can artificial intelligence improve the readability of patient education materials? Clin Orthop Relat Res. 2023;481(11):2260-7.
19. Gül Ş, Erdemir İ, Hancı V, Aydoğmuş E, Erkoç YS. How artificial intelligence can provide information about subdural hematoma: assessment of readability, reliability, and quality of chatbot responses. Medicine (Baltimore). 2024;103(18):e38009.
20. Carlson JA, Cheng RZ, Lange A, Nagalakshmi N, Rabets J, Shah T, et al. Accuracy and readability of artificial intelligence chatbot responses to vasectomy-related questions: public beware. Cureus. 2024;16(8):e67996.
21. Ghanem YK, Rouhi AD, Al-Houssan A, Saleh Z, Moccia MC, Joshi H, et al. Dr Google to Dr ChatGPT: assessing the content and quality of artificial intelligence-generated medical information on appendicitis. Surg Endosc. 2024;38(5):2887-93.
22. Warn M, Meller LL, Chan D, Torabi SJ, Bitner BF, Tajudeen BA, et al. Assessing the readability, reliability, and quality of AI-modified patient education materials for endoscopic skull base surgery. Am J Rhinol Allergy. 2024;38(6):396-402.
23. Golan R, Ripps SJ, Reddy R, Loloi J, Bernstein AP, Connelly ZM, et al. ChatGPT's ability to assess quality and readability of online medical information: a cross-sectional study. Cureus. 2023;15(7):e42214.
24. Szczesniewski JJ, Tellez Fouz C, Ramos Alba A, Diaz Goizueta FJ, García Tello A, Llanes González L. ChatGPT and most frequent urological diseases: analysing the quality of information and potential risks for patients. World J Urol. 2023;41(11):3149-53.
25. Paluszek O, Loeb S. Artificial intelligence and patient education. Curr Opin Urol. 2025;35(2):126-31.

Cite this article as: Harini SS, Dande H, Nagabhyru S, Shafiya A, Kurian A, Teja ESV, et al. AI in patient education: a cross-sectional comparative analysis of patient information guides produced by ChatGPT versus Google Gemini for common haematology conditions. Int J Res Med Sci 2026;14:5143-7.