Disseminating Steatotic Liver Disease Patient Education: The Emerging Role of Large Language Models

Yee Hui Yeo, MD¹, Ju Dong Yang, MD¹,²,³

¹Karsh Division of Gastroenterology and Hepatology, Cedars-Sinai Medical Center, Los Angeles, California, USA
²Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA
³Comprehensive Transplant Center, Cedars-Sinai Medical Center, Los Angeles, California, USA

Corresponding author
Ju Dong Yang, MD
Medical Director, Liver Cancer Program.
Karsh Division of Gastroenterology and Hepatology,
Samuel Oschin Comprehensive Cancer Institute,
Comprehensive Transplant Center
Cedars-Sinai Medical Center,
Los Angeles, California, USA
JuDong.Yang@cshs.org
We thank the LLM-Liver Investigators for commenting on our recent publication [1]. Their commentary presents a comprehensive assessment of the capability of various Large Language Models (LLMs) in providing patient education on liver diseases. The authors' work to assess various LLMs' performance in providing patient education on liver diseases is a timely study that contributes to understanding the role of multiple recently-introduced LLMs in healthcare dissemination. In the study, LLM's high performance in delivering appropriate responses underscores the potential for AI to support healthcare providers in disseminating accurate medical information. By setting a benchmark for LLM performance, the study not only contributes to the academic field but also lays the groundwork for the development of AI-driven patient education tools. Their findings could play a crucial role in bridging the gap in health literacy and ensuring equitable access to medical information across diverse patient populations[2].

There are several aspects that need clarification for proper interpretation of the results. First, the term "steatotic liver disease" is a recently developed nomenclature [3]. The reliability of LLMs to provide up-to-date information depends on their training with current datasets. As the authors used “fatty liver disease” instead of steatotic liver disease, the study would benefit from disclosure of the end dates of the datasets used to train each LLM to ensure that the information provided meets contemporary standards. Furthermore, using the same terminology as that used in recent guidelines would enhance the study's applicability and clarity. Second, the methodology behind the question selection process remains unclear in the authors' study. An explanation of how the 30 questions for each LLM were chosen, potentially from clinical guidelines and the authors' experience, would solidify the study's robustness. It would pre-emptively address concerns regarding the selection of questions that might disproportionately favor the capabilities of LLMs.

Additionally, the "washout" period used in the study to minimize recall bias may raise concerns. Recalling responses from previous rounds may influence subsequent
evaluations. Finally, it is unsure if there is any statistically significant difference in the performance among the LLMs as there was no analysis performed.

In conclusion, the authors' study represents an important contribution to the field of AI in patient education. By addressing the areas outlined above, the study can achieve greater validity and provide a more reliable framework for assessing the capability of LLMs in patient education. It is with great anticipation that the medical community looks forward to additional research that builds on this work. We again congratulate the authors for performing the study to enhance our ability to understand and harness AI's potential in enhancing patient outcomes and health literacy.

References


Acknowledgments

None.

Conflict of interests

All other authors report no potential conflicts.

Funding
None.