

Comparative accuracy of chatbots for queries related to methods of tobacco cessation: A systematic review: Reg No: 242

Sonal Bhatia, Harsh Priya, Nitesh Tiwari, Bharathi M. Purohit, Partha Haldar

All India Institute of Medical Sciences

CORRESPONDING AUTHOR

Dr Sonal Bhatia, All India Institute of Medical Sciences

Email: jamsonalbhatia@gmail.com

CITATION

Bhatia Sonal, Priya Harsh, TiwariNitesh, Purohit Bharathi M., Haldar Partha. Comparative accuracy of chatbots for queries related to methods of tobacco cessation: A systematic review: Reg No: 242. Journal of the Epidemiology Foundation of India. 2025;3(2Suppl):68.

DOI: <https://doi.org/10.56450/JEFI.2025.v3i2Suppl.068>

ARTICLE CYCLE

Received: 10/11/2025; Accepted: 20/12/2026; Published: 31/12/2025

This work is licensed under a Creative Commons Attribution 4.0 International License.

©The Author(s). 2026 Open Access

ABSTRACT

Background: Artificial intelligence (AI)-based conversational agents, including chatbots and machine learning (ML) platforms, are increasingly being used to disseminate health information and assist in tobacco cessation counselling. However, their comparative accuracy in providing evidence-based tobacco cessation advice remains unclear. **Objective:** This systematic review aimed to evaluate and compare the accuracy of responses provided by chatbots and ML platforms to queries related to methods of tobacco cessation. **Methods:** Following PRISMA guidelines, a comprehensive electronic search was performed in October 2025 across PubMed, EMBASE, Web of Science, and Scopus. The search strategy was structured using the PICO framework. Grey literature was identified through Google Scholar and OpenGrey, and reference lists of included studies were hand-searched. Eligible studies included descriptive, observational, cohort, cross-sectional, and case-control designs that compared the accuracy of chatbots or ML-based systems for tobacco cessation queries. Data were extracted and synthesized narratively, with quantitative measures of accuracy expressed as frequencies, proportions, or mean differences. **Results:** A total of 2,516 records were identified (PubMed=1,587; EMBASE=371; Web of Science=558). After removing duplicates and screening, preliminary inclusion suggests that most studies evaluated chatbots such as ChatGPT, Gemini, and Perplexity for smoking cessation counselling and pharmacotherapy recommendations. Early evidence indicates substantial variability in accuracy across platforms, with large language models (LLMs) giving similar accuracies compared to earlier or rule-based systems. Factors such as prompt phrasing, question domain, and language significantly influenced accuracy levels. **Discussion:** The findings highlight an evolving but uneven performance landscape among AI conversational agents for tobacco cessation. While newer generative AI models show improved reliability and contextual understanding, inconsistencies persist, particularly in region-specific or non-English responses. Further standardization of benchmarking methods and continuous model training on verified cessation guidelines are recommended to enhance clinical applicability and public health integration.

KEYWORDS