



Evaluating ChatGPT's efficacy in assessing the safety of non-prescription medications and supplements in patients with kidney disease

DIGITAL HEALTH
Volume 10: 1–9
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076241248082
journals.sagepub.com/home/dhj



Mohammad S. Sheikh¹, Erin F. Barreto², Jing Miao¹, Charat Thongprayoon¹ ,
James R Gregoire¹, Benjamin Dreesman², Stephen B. Erickson¹,
Iasmina M. Craici¹ and Wisit Cheungpasitporn¹ 

Abstract

Background: This study investigated the efficacy of ChatGPT-3.5 and ChatGPT-4 in assessing drug safety for patients with kidney diseases, comparing their performance to Micromedex, a well-established drug information source. Despite the perception of non-prescription medications and supplements as safe, risks exist, especially for those with kidney issues. The study's goal was to evaluate ChatGPT's versions for their potential in clinical decision-making regarding kidney disease patients.

Method: The research involved analyzing 124 common non-prescription medications and supplements using ChatGPT-3.5 and ChatGPT-4 with queries about their safety for people with kidney disease. The AI responses were categorized as “generally safe,” “potentially harmful,” or “unknown toxicity.” Simultaneously, these medications and supplements were assessed in Micromedex using similar categories, allowing for a comparison of the concordance between the two resources.

Results: Micromedex identified 85 (68.5%) medications as generally safe, 35 (28.2%) as potentially harmful, and 4 (3.2%) of unknown toxicity. ChatGPT-3.5 identified 89 (71.8%) as generally safe, 11 (8.9%) as potentially harmful, and 24 (19.3%) of unknown toxicity. GPT-4 identified 82 (66.1%) as generally safe, 29 (23.4%) as potentially harmful, and 13 (10.5%) of unknown toxicity. The overall agreement between Micromedex and ChatGPT-3.5 was 64.5% and ChatGPT-4 demonstrated a higher agreement at 81.4%. Notably, ChatGPT-3.5's suboptimal performance was primarily influenced by a lower concordance rate among supplements, standing at 60.3%. This discrepancy could be attributed to the limited data on supplements within ChatGPT-3.5, with supplements constituting 80% of medications identified as unknown.

Conclusion: ChatGPT's capabilities in evaluating the safety of non-prescription drugs and supplements for kidney disease patients are modest compared to established drug information resources. Neither ChatGPT-3.5 nor ChatGPT-4 can be currently recommended as reliable drug information sources for this demographic. The results highlight the need for further improvements in the model's accuracy and reliability in the medical domain.

Keywords

Drug safety, non-prescription medications, supplements, kidney diseases, ChatGPT, artificial intelligence

Submission date: 24 December 2023; Acceptance date: 28 March 2024

Background

In contemporary times, the utilization of search engines has brought about a revolutionary change in the way people acquire information, particularly concerning health-related matters.¹ The convenience and ease of searching for drug

¹Department of Nephrology, Mayo Clinic Minnesota, Rochester, MN, USA

²Department of Pharmacy, Mayo Clinic Minnesota, Rochester, MN, USA

Corresponding author:

Wisit Cheungpasitporn, Division of Nephrology and Hypertension,
Department of Medicine, Mayo Clinic, Rochester, MN, 55905, USA.
Email: wcheungpasitporn@gmail.com



safety information online have rendered search engines highly popular among individuals seeking knowledge about medication safety. Over-the-counter medications and supplements, often regarded as innocuous and readily accessible, are extensively used by the general population.² Nevertheless, over-the-counter medications and supplements are associated with serious adverse effects in individuals with underlying kidney ailments.^{3,4} Kidney diseases, including chronic kidney disease, are prevalent conditions that affect a substantial proportion of the global population.⁵ In order to ensure optimal patient care and avert potential harm, it is crucial to accurately interpret the safety of medications within the context of kidney disease.

While healthcare practitioners have traditionally turned to tertiary drug information resources such as Micromedex to determine the safety of non-prescription medications and supplements, the advent of advanced artificial intelligence large language models, including ChatGPT-3.5 and ChatGPT-4, has generated considerable interest in their potential to enhance clinical decision-making processes.^{6–10} These models employ state-of-the-art natural language processing techniques, with ChatGPT-4 demonstrating enhanced proficiency compared to its predecessor, to generate responses that closely resemble human-generated ones.¹¹ ChatGPT-3.5 and ChatGPT-4 have garnered attention due to its ability to comprehend and respond to intricate queries, making it a prospective resource for individuals seeking drug-related information.^{12,13} However, it is uncertain the extent to which these large language models can be used for guiding non-prescription medication and supplement safety considerations in patients with kidney disease.¹⁴ We therefore sought to evaluate ChatGPT-3.5 and ChatGPT-4's performance in assessing the safety of over-the-counter medications and supplements in patients with kidney disease, while simultaneously comparing its performance to that of Micromedex.

Method

Study design

This study aimed to evaluate the accuracy of ChatGPT in assessing the safety of commonly used non-prescription medications and supplements in individuals with kidney disease. The study was designed as a comparative analysis, aiming to evaluate the efficacy of artificial intelligence language models, specifically ChatGPT-3.5 and ChatGPT-4, in this regard. The research was conducted over a period of 3 months, from March 2023 to May 2023. This timeframe was selected to ensure ample opportunity for data collection and analysis, allowing for a comprehensive evaluation of the AI models' performance over a sustained period. The study was carried out utilizing virtual environments for accessing the OpenAI API and Micromedex

database, thereby not being limited to a physical location. However, the coordination and oversight of the study activities, including the formulation of queries to the AI models and the categorization of responses, were managed from our research institution located in Mayo Clinic, Minnesota. A systematic approach was employed, comparing the outputs from ChatGPT with the categorization of safety provided by the Micromedex database. Additionally, the findings were confirmed by experienced pharmacists. This study does not involve human or animal subjects, nor does it include patient information or identifiable personal data; thus, ethical approval and informed consent were not needed.

Inclusion criteria encompassed medications and supplements commonly used by the general population and those of particular interest or concern for individuals with kidney disease. This included, but was not limited to, over-the-counter pain medications, common cold preparations, gastrointestinal upset remedies, topical treatments for skin conditions, and a broad range of dietary supplements known for their potential renal implications. Exclusion criteria were designed to narrow the focus of our study to non-prescription items, thereby excluding any prescription medications, medications primarily used for the treatment of conditions other than those for which over-the-counter versions are intended, and supplements not widely available or recognized by authoritative bodies such as the FDA or similar regulatory agencies. Additionally, products with scarce or ambiguous information regarding their safety in kidney disease were also excluded to maintain the integrity and reliability of our findings.

Accessing ChatGPT

Access to ChatGPT was facilitated through the OpenAI API, offering programmatic access to both the ChatGPT-3.5 and ChatGPT-4 language models. The specific version of ChatGPT-3.5 utilized in this study was the OpenAI version known for its proficiency in natural language processing tasks. Similarly, the GPT-4 version, which represents a more advanced iteration, was employed to ensure a comprehensive evaluation. The series of queries was conducted on in from March 2023 to May 2023 utilizing OpenAI versions of both ChatGPT-3.5 and ChatGPT-4, providing a robust comparison of their performance in assessing the safety of non-prescription medications and supplements in patients with kidney disease.

Selection of non-prescription medications and supplements

A comprehensive evaluation encompassing a cohort of 124 widely utilized non-prescription medications and supplements

was selected for evaluation. The selection of these entities was grounded in their prevalent utilization within clinical practice, coupled with a discerning consideration for their potential relevance to individuals affected by kidney disease.

Query and categorization process

The inquiry presented to both ChatGPT-3.5 and ChatGPT-4 for each non-prescription medication and supplement was structured as follows: “Is [Name of the Medication/Supplement] safe in people with kidney disease?” Subsequently, the generated output from both ChatGPT-3.5 and ChatGPT-4 underwent thorough examination, and the outcomes were meticulously categorized into one of three classifications: “generally safe,” “potentially harmful,” or “unknown toxicity.” This method ensured a comprehensive evaluation of the safety assessments provided by both language models for non-prescription medications and supplements in the context of kidney disease.

Comparison with micromedex

To establish a foundation for comparison, the safety evaluation of the same non-prescription medications and supplements was conducted using the Micromedex database. Micromedex, a widely recognized and authoritative reference database, offers comprehensive information on drug interactions, adverse effects, and safety precautions. The safety categorization in Micromedex mirrored the process outlined earlier for ChatGPT-3.5 and ChatGPT-4.

The results derived from both ChatGPT-3.5 and ChatGPT-4, as well as Micromedex, underwent additional scrutiny by experienced pharmacists. These professionals meticulously reviewed the assessments provided by both language models and cross-referenced them with the information available in Micromedex. Their expertise and knowledge in pharmaceuticals served as an additional layer of validation, ensuring the robustness and reliability of the study’s findings.

Concordance assessment

The concordance between the outputs from both ChatGPT-3.5 and ChatGPT-4 and Micromedex was determined. The level of agreement was then summarized to comprehensively assess the accuracy of ChatGPT, considering both iterations, in evaluating the safety of non-prescription medications and supplements for individuals with kidney disease. This approach allowed for a nuanced understanding of the performance of both language models in comparison to the widely used Micromedex database.

To calculate the Overall Concordance Percentage for both ChatGPT-3.5 and ChatGPT-4:

1. Determine the “Number of times Micromedex, ChatGPT-3.5, and ChatGPT-4 agree for all medications.”
2. Divide that number by the “Total number of medications assessed.”
3. Multiple the result by 100 to convert it into a percentage.

This method ensures a comprehensive assessment of the agreement between Micromedex, ChatGPT-3.5, and ChatGPT-4 across all evaluated non-prescription medications and supplements, providing a robust measure of overall concordance.

In addition, to enhance the accuracy of our agreement assessment between ChatGPT versions 3.5 and 4.0 and Micromedex, we have adopted the Cohen’s Kappa test (κ) as the statistical method of choice. This approach calculates the agreement between the two sources by taking into account the observed agreement (P_o) and the expected agreement by chance (P_e), according to the formula: $\kappa = (P_o - P_e) / (1 - P_e)$. The results of this test provide a more nuanced understanding of the agreement, categorized as follows: 0.10–0.20 indicating slight agreement, 0.21–0.40 indicating fair agreement, 0.41–0.60 indicating moderate agreement, 0.61–0.80 indicating substantial agreement, 0.81–0.99 indicating near-perfect agreement, and 1.00 indicating perfect agreement.

Results

In the comprehensive analysis of 124 commonly used non-prescription medications and supplements, a thorough examination was conducted across distinct subclasses, including pain medications (7), common cold medications (15), gastrointestinal (GI) upset medications (26), topical skin condition medications (8), and supplements (68) (Supplemental material1). Micromedex identified 85 (68.5%) of the assessed medications as “generally safe,” while designating 35 (28.2%) as “potentially harmful,” and 4 as “unknown.” Conversely, the language models, ChatGPT-3.5 and ChatGPT-4, presented nuanced perspectives, with ChatGPT-3.5 identifying 89 (71.7%) as “generally safe,” 11 (8.8%) as “potentially harmful,” and 24 (19.3%) as “of unknown toxicity.” The more advanced ChatGPT-4 identified 82 (66.1%) as “generally safe,” 29 (23.4%) as “potentially harmful,” and 13 (10.5%) as “of unknown toxicity” (Figure 1). Examining the overarching agreement, ChatGPT-3.5 demonstrated a concordance of 64.5% with Micromedex, while ChatGPT-4 exhibited a higher concordance of 81.4%. Delving into the subclass concordances for ChatGPT-3.5, notable figures emerged: 57.1% for pain medications, 66.6% for common cold medications, 65.4% for gastrointestinal upset medications, 87.5% for topical skin condition medications, and 60.3% for supplements. The more advanced ChatGPT-4 showcased heightened

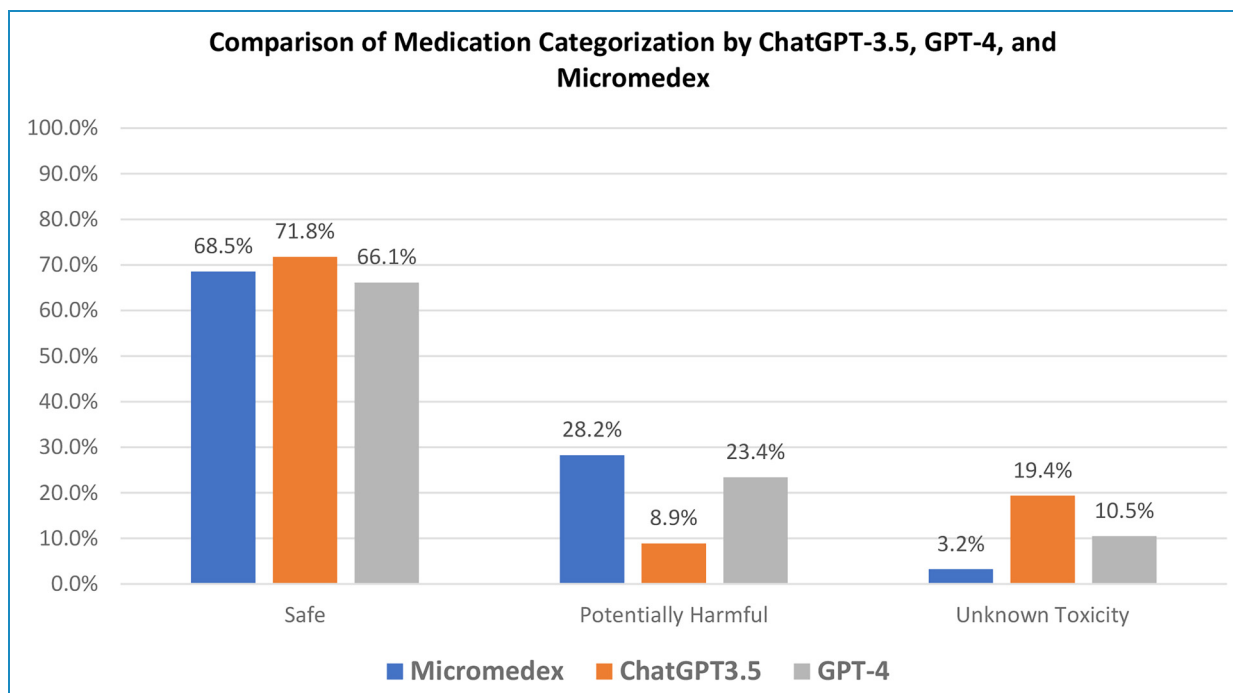


Figure 1. Comparison of micromedex and ChatGPT assessments for 124 commonly used medications and supplements.

subclass concordances: 85.7% for pain medications, 66.6% for common cold medications, 80.8% for gastrointestinal upset medications, 100% for topical skin condition medications, and 82.3% for supplements (Figure 2).

A granular analysis of 85 medications classified as “generally safe” by Micromedex revealed insightful findings. ChatGPT-3.5 identified 68 (80.0%) as “generally safe,” flagged 3 (3.5%) as “potentially harmful,” and classified 14 (16.5%) as “unknown toxicity” (Figure 3). In comparison, ChatGPT-4 marked a more substantial proportion as “generally safe,” identifying 74 (87.1%), and designating 4 (4.7%) as “potentially harmful,” with only 7 (8.2%) deemed “of unknown toxicity” (Figure 4). Shifting focus to the 35 medications classified as “potentially harmful” by Micromedex, ChatGPT-3.5 identified 8 (22.8%) as such, labeled 21 (60%) as “generally safe,” and assigned 6 (17.1%) to the category of “unknown toxicity” (Figure 5). Intriguingly, ChatGPT-4 demonstrated a nuanced perspective, designating 8 (22.9%) as “generally safe,” marking 26 (74.3%) as “potentially harmful,” and classifying only 1 (2.9%) as “of unknown toxicity” (Figure 6).

Upon applying the Cohen’s Kappa test to assess the level of agreement between ChatGPT versions and Micromedex regarding the safety categorizations of the evaluated non-prescription medications and supplements, we observed the following:

- The agreement between ChatGPT-3.5 and Micromedex yielded a Cohen’s Kappa value of $\kappa = 0.62$, which falls

into the range of substantial agreement. This suggests that, despite the observed limitations, ChatGPT-3.5 significantly aligns with the established drug safety assessments provided by Micromedex, with room for improvement.

- For ChatGPT-4, the Cohen’s Kappa test indicated a higher level of agreement with Micromedex, achieving a value of $\kappa = 0.83$. This result falls within the range of near-perfect agreement, underscoring the marked improvement in ChatGPT-4’s ability to assess the safety of non-prescription medications and supplements for individuals with kidney disease in alignment with Micromedex’s recommendations.

Discussion

The principal aim of this study was to evaluate the efficacy of ChatGPT, an advanced artificial intelligence language model, in gauging the safety of non-prescription medications and supplements for individuals with kidney disease.¹⁵ The investigation revealed that ChatGPT’s proficiency in discerning the safety of these medications, when juxtaposed with Micromedex, a widely utilized tertiary drug information reference, was relatively modest. This assessment was extended to include the latest version, ChatGPT-4, to gauge any improvement in performance.

Micromedex, recognized for its credibility and expansive database, has long been a cornerstone for healthcare professionals seeking comprehensive and up-to-date drug-related information.¹⁶ Its robustness positioned it as

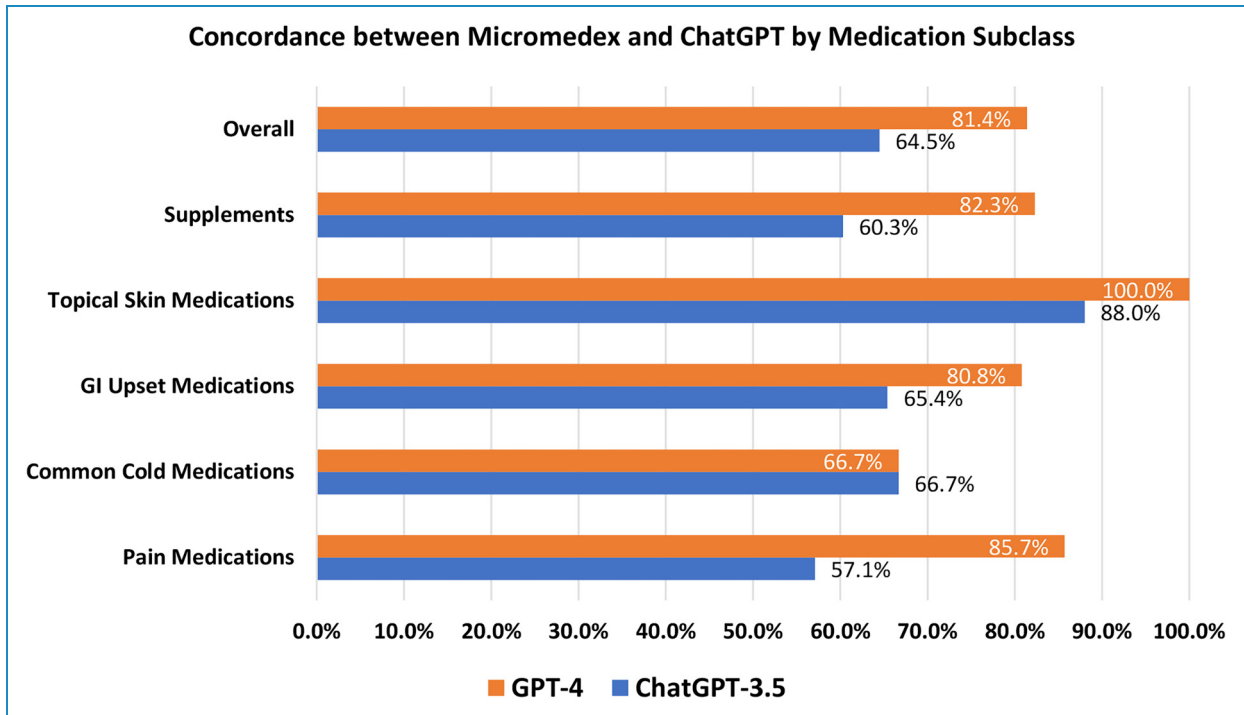


Figure 2. Concordance between micromedex and ChatGPT assessments by medication subclasses.

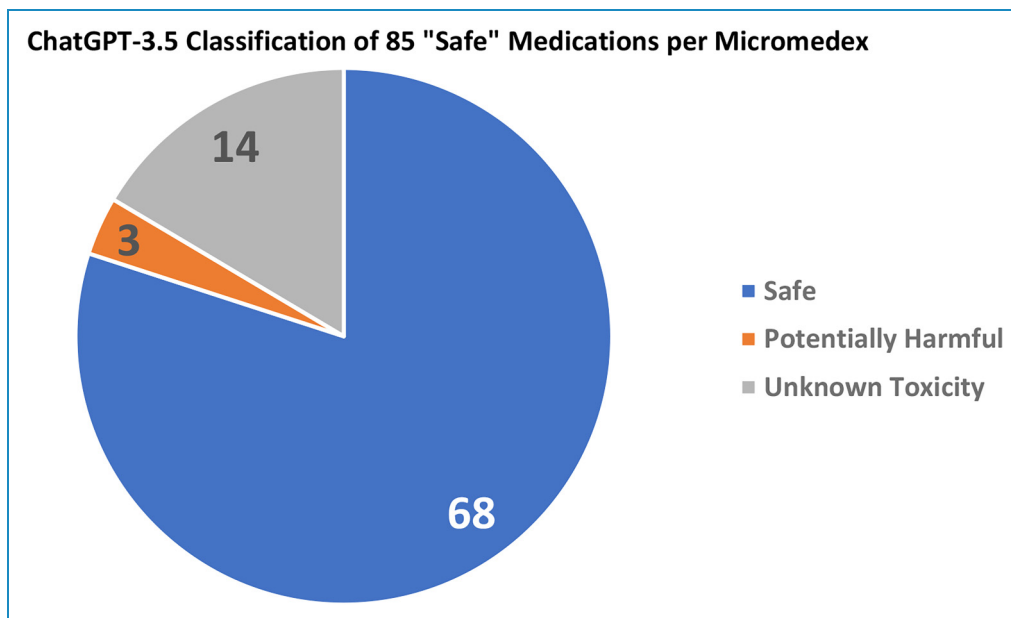


Figure 3. ChatGPT assessment of medications classified as "Safe" by micromedex.

an apt benchmark for assessing the performance of ChatGPT. By comparing the outputs generated by ChatGPT-3.5 and ChatGPT-4 with the assessments provided by Micromedex, valuable insight was gained regarding ChatGPT's ability to discern medication safety in individuals with kidney disease.

In a comparative analysis between Micromedex and both ChatGPT-3.5 and its updated version, ChatGPT-4, it was observed that Micromedex exhibited superior performance in the identification of medications with potential harm, particularly concerning patients with kidney disease. While both platforms predominantly categorized medications as

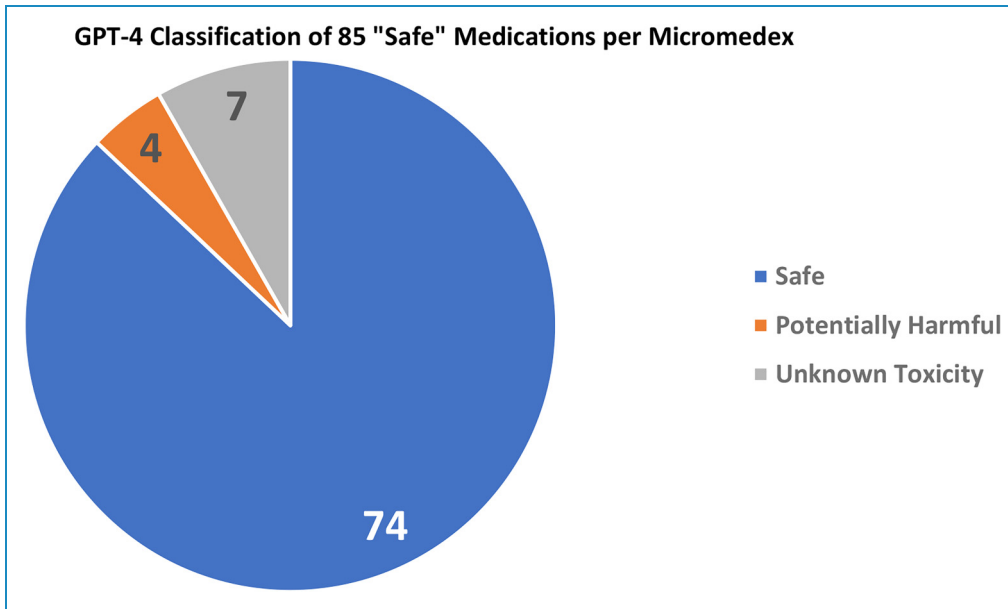


Figure 4. ChatGPT assessment of medications classified as "Potentially Harmful" by micromedex.

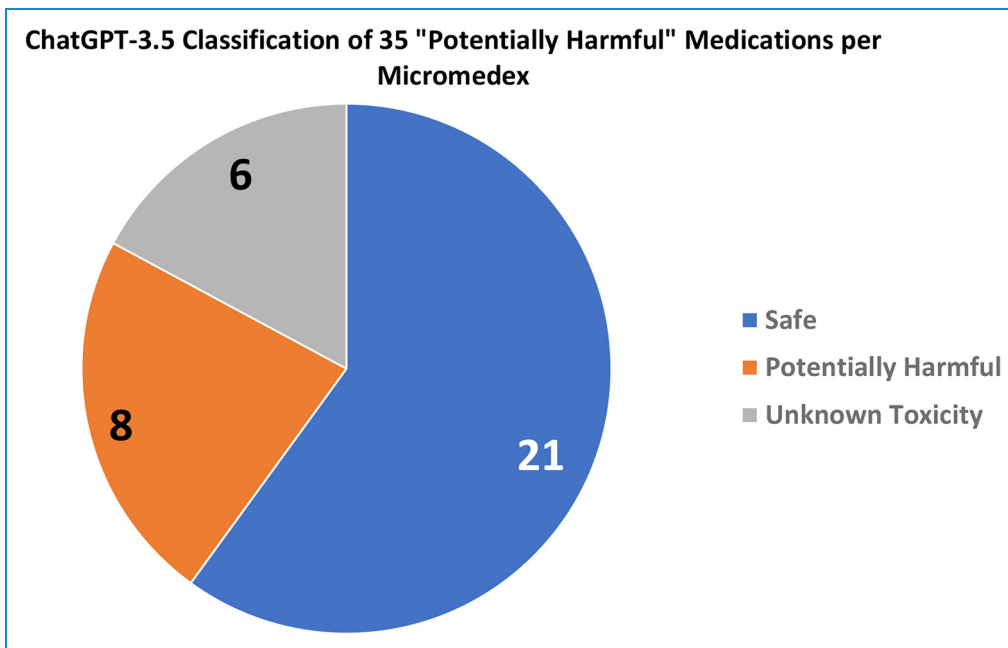


Figure 5. Classification of medications deemed "Potentially Harmful" by micromedex as evaluated by ChatGPT-3.5.

generally safe, Micromedex maintained a more stringent classification, identifying a greater proportion of medications as potentially harmful. In contrast, both ChatGPT-3.5 and ChatGPT-4 displayed reduced sensitivity in flagging medications that could be detrimental to this patient population. The concordance between Micromedex and ChatGPT-3.5 was quantified at 65%, indicating a suboptimal alignment of ChatGPT-3.5's

evaluations with those of Micromedex. A salient factor contributing to this divergence was the limited dataset pertaining to dietary supplements. These supplements were predominantly categorized by ChatGPT-3.5 as having indeterminate toxicity, and they accounted for a substantial fraction of the medications that exhibited low agreement between the two platforms. Therefore, the paucity of data on supplements significantly influenced the observed

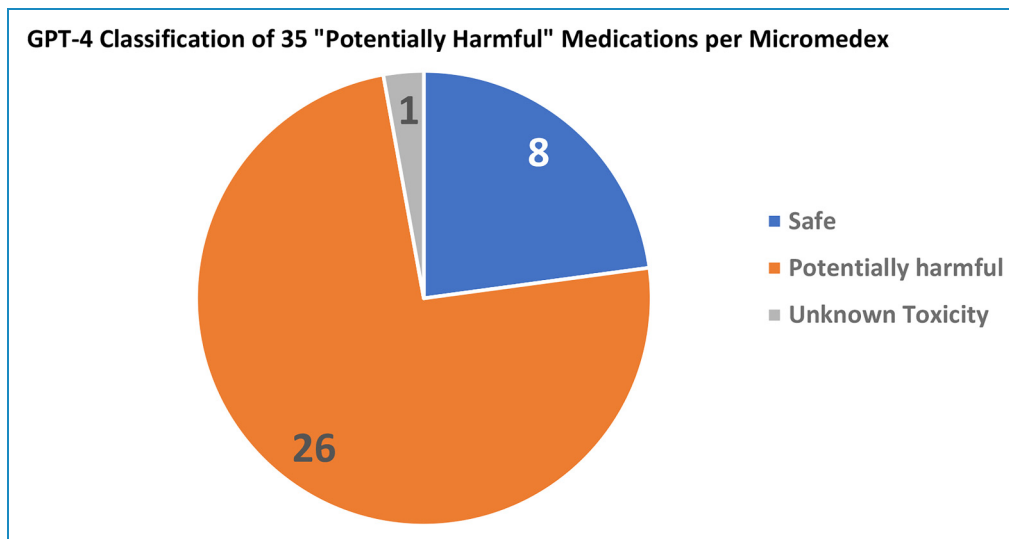


Figure 6. ChatGPT-4's reassessment of medications classified as "Potentially Harmful" by micromedex.

discrepancy in performance metrics. While ChatGPT-3.5 showed potential in improving clinical decision-making, ChatGPT-4 presented a notable improvement in performance with a concordance of 81%. In addition, findings derived from the Cohen's Kappa statistical analysis further enrich our study by providing a quantitatively robust measure of agreement between the AI models and the Micromedex database. They corroborate our initial observation of improved performance in ChatGPT version 4.0 over version 3.5 and highlight the potential of AI in augmenting healthcare decision-making processes, with an emphasis on the continuous enhancement of AI technologies for better accuracy and reliability in clinical settings. However, caution is warranted, and further refinement and enhancement of these models are still necessary to ensure their accuracy and reliability in providing drug information specific to patients with kidney disease.^{17,18} Incorporating comprehensive and updated data on medications, including supplements, is crucial for enhancing the performance of AI models like ChatGPT.¹⁹

Future studies should focus on addressing the limitations identified in this study to enhance the performance of AI language models like ChatGPT in evaluating the safety of non-prescription medications and supplements for patients with kidney disease.^{20,21} Expanding the dataset available to the model, particularly with regard to supplements, is crucial.²² Additionally, investigating the reasons behind ChatGPT's lower sensitivity in identifying potentially harmful medications compared to Micromedex and identifying strategies to improve this sensitivity are necessary. This may involve refining the model's algorithms and training it on a larger and more diverse dataset of drug safety information specifically tailored to patients with kidney disease. Furthermore, it would be valuable to examine the

performance of ChatGPT in other disease contexts beyond kidney disease, as different diseases may have unique considerations and interactions with medications. Such investigations would provide a more comprehensive understanding of ChatGPT's capabilities and limitations, enabling researchers to identify areas where the model excels and areas that require further improvement.

Several limitations should be considered in interpreting the findings of this study. Firstly, the focus on non-prescription medications and supplements excludes the evaluation of prescription medications commonly used by patients with kidney disease. Therefore, the generalizability of the findings to the broader context of medication safety in kidney disease may be limited. Secondly, the study relied on a single drug information resource, Micromedex, as the comparator for ChatGPT. While Micromedex is widely used and trusted, utilizing multiple drug information references or expert consensus could offer a more robust evaluation and mitigate potential biases or limitations associated with a single reference. Furthermore, the study evaluated ChatGPT's performance based on a specific query format, which might have influenced the model's performance. Exploring alternative query formats or refining the query structure could enhance the model's understanding and interpretation of drug safety information for individuals with kidney disease. In our study, we selected Micromedex as a reference for assessing the safety of non-prescription medications and supplements, given its broad acceptance and utilization as a drug information resource. However, it is crucial to clarify that while Micromedex is respected for its comprehensive database and commitment to high evidence standards, our comparison did not extend to validating Micromedex's precision against an independent comparative database or reference. Therefore, the mention of

Micromedex's 'superiority' in our discussion aimed not to assert its infallibility but to contextualize the performance of ChatGPT versions against a recognized standard in drug information. Additionally, the study solely focused on ChatGPT's ability to determine medication safety, neglecting other important aspects such as dosing recommendations, potential drug interactions, and contraindications. Expanding the scope to encompass a broader range of drug information aspects relevant to patients with kidney disease would provide a more comprehensive evaluation. Moreover, the study did not involve direct clinical validation or assess the impact of ChatGPT's recommendations on patient outcomes. Evaluating the real-world clinical impact and effectiveness of AI language models like ChatGPT would provide valuable insights into their practical implications and utility in healthcare decision-making.^{14,23} Furthermore, future studies should focus on optimizing input prompts to guarantee that they are carefully crafted to extract the most pertinent and precise data from large language models.^{24,25} This may entail creating standardized prompt structures specifically tailored for medical queries, thus improving the accuracy of the generated responses. Finally, the study did not address the challenges and barriers to the implementation of AI language models in clinical practice, including user acceptance, integration into existing healthcare systems, and considerations related to data privacy and security.²⁶ Future research should investigate these implementation-related factors to ensure successful integration and adoption of AI language models in the healthcare setting. Despite these limitations, this study offers valuable insights into ChatGPT's performance in assessing the safety of non-prescription medications and supplements for individuals with kidney disease. Acknowledging these limitations lays the foundation for future research to build upon these findings, address the identified gaps, and enhance the accuracy, reliability, and applicability of AI language models in the field of drug safety and clinical decision-making.

In conclusion, this study, encompassing the analysis of both ChatGPT-3.5 and the updated ChatGPT-4, underscores that ChatGPT's performance in assessing the safety of non-prescription medications and supplements in patients with kidney disease is modest compared to a well-established drug information resource such as Micromedex. While AI language models, including ChatGPT-4, hold promise in various healthcare applications, including drug safety assessment, the findings emphasize that further advancements are necessary to enhance their accuracy and reliability in providing drug information tailored to the specific needs of patients with kidney disease. Continued research and development in this area, incorporating improvements based on ChatGPT-3.5 and ChatGPT-4 insights, will be essential to leverage the full potential of AI in improving clinical decision-making and enhancing patient care within the realm of medication safety.

Contributorship: Conceptualization: MSS, EFB, CT, JM, WC, JG, BD, SBE, and IMC. Data curation: MSS, CT, and WC. Formal analysis: MSS and WC. Funding acquisition: EFB and WC. Investigation: MSS, EFB, CT, WC, SBE, and IMC. Methodology: MSS, EFB, JM, WC, and BD. Project administration: CT, JM, and JG. Resources: WC and BD. Software: MSS and WC. Supervision: EFB, JM, WC, and IMC. Validation: MSS, EFB, WC, and JG. Visualization: MSS, EFB, and WC. Writing – original draft: MSS. Writing – review & editing: MSS, EFB, CT, JM, WC, JG, BD, and SBE. All authors have read and agreed to the published version of the manuscript.


Data availability statement: The data underlying this article will be shared on reasonable request to the corresponding author.

Declaration of conflicting interests: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethics approval: This study does not require Ethics Committee or Institutional Review Board approval because it does not involve human or animal subjects, nor does it include patient information or identifiable personal data. Consequently, participant consent was waived for the same reasons.

Funding: The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health, (grant number K23AI143882).

Guarantor: W.C.'

ORCID iDs: Charat Thongprayoon  <https://orcid.org/0000-0002-8313-3604>

Wisit Cheungpasitporn  <https://orcid.org/0000-0001-9954-9711>

Supplemental material: Supplemental material for this article is available online.

References

1. Simmering JE, Polgreen LA and Polgreen PM. Web search query volume as a measure of pharmaceutical utilization and changes in prescribing patterns. *Res Social Adm Pharm* 2014; 10: 896–903.
2. Barrenberg E and Garbe E. Use of over-the-counter (OTC) drugs and perceptions of OTC drug safety among German adults. *Eur J Clin Pharmacol* 2015; 71: 1389–1396.
3. Perazella MA. Pharmacology behind common drug nephrotoxicities. *Clin J Am Soc Nephrol* 2018; 13: 1897–1908.
4. Leonberg-Yoo AK, et al. Use of dietary supplements in living kidney donors: a critical review. *Am J Kidney Dis* 2020; 76: 851–860.

5. Abramov D, et al. Cardiovascular mortality trends in patients with chronic kidney disease compared to the general population. *J Nephrol* 2023; 36: 1489–1491.
 6. Subramanian T, Pociavsek L and Alverdy JC. To operate or not? Balancing advanced imaging, machine learning, and the doctor-patient relationship in complex clinical decision making. *Acad Med* 2023; 98: S34–S36.
 7. Chung EM, et al. Feasibility and acceptability of ChatGPT generated radiology report summaries for cancer patients. *Digit Health* 2023; 9: 20552076231221620.
 8. Kim SH, et al. Changes in patient perceptions regarding ChatGPT-written explanations on lifestyle modifications for preventing urolithiasis recurrence. *Digit Health* 2023; 9: 20552076231203940.
 9. Zhu Z, et al. ChatGPT's potential role in non-English-speaking outpatient clinic settings. *Digit Health* 2023; 9: 20552076231184091.
 10. Sheikh MS, et al. Personalized medicine transformed: ChatGPT's contribution to continuous renal replacement therapy alarm management in intensive care units. *J Pers Med* 2024; 14: 233.
 11. Miao J, et al. Innovating personalized nephrology care: exploring the potential utilization of ChatGPT. *J Pers Med* 2023; 13: 1681.
 12. Abu Hammour K, et al. ChatGPT in pharmacy practice: a cross-sectional exploration of Jordanian pharmacists' perception, practice, and concerns. *J Pharm Policy Pract* 2023; 16: 115.
 13. Miao J, et al. Performance of ChatGPT on nephrology test questions. *Clin J Am Soc Nephrol* 2023; 19: 35–43.
 14. Singh S, et al. Artificial intelligence and machine learning in pharmacological research: bridging the gap between data and drug discovery. *Cureus* 2023; 15: e44359.
 15. Gao Z, et al. Examining the potential of ChatGPT on biomedical information retrieval: fact-checking drug-disease associations. *Ann Biomed Eng* 2023 Oct 19. Epub ahead of print 20.
 16. Abbas A, et al. Determination of potential drug-drug interactions in prescription orders dispensed in a community pharmacy setting using micromedex((R)) and lexicomp((R)): a retrospective observational study. *Int J Clin Pharm* 2022; 44: 348–356.
 17. Haverkamp W, Tennenbaum J and Strodthoff N. ChatGPT fails the test of evidence-based medicine. *Eur Heart J Digit Health* 2023; 4: 366–367.
 18. Fraser H, et al. Comparison of diagnostic and triage accuracy of Ada health and WebMD symptom checkers, ChatGPT, and physicians for patients in an emergency department: clinical data analysis study. *JMIR Mhealth Uhealth* 2023; 11: e49995.
 19. Zhao D, et al. Current progress in artificial intelligence-assisted medical image analysis for chronic kidney disease: a literature review. *Comput Struct Biotechnol J* 2023; 21: 3315–3326.
 20. Daugirdas JT. OpenAI's ChatGPT and its potential impact on narrative and scientific writing in nephrology. *Am J Kidney Dis* 2023; 82: A13–A14.
 21. Miao J, Thongprayoon C and Cheungpasitporn W. Assessing the accuracy of ChatGPT on core questions in glomerular disease. *Kidney Int Rep* 2023; 8: 1657–1659.
 22. Claire-Del Granado R and Espinosa-Cuevas M. Herbal nephropathy. *Contrib Nephrol* 2021; 199: 143–154.
 23. Krisanapan P, et al. Revolutionizing chronic kidney disease management with machine learning and artificial intelligence. *J Clin Med* 2023; 12: 3018.
 24. Miao J, et al. Chain of thought utilization in large language models and application in nephrology. *Medicina (Kaunas)* 2024; 60: 148.
 25. Miao J, et al. Integrating retrieval-augmented generation with large language models in nephrology: advancing practical applications. *Medicina (B Aires)* 2024; 60: 445.
 26. Layton AT. "Hi, how can i help you?": embracing artificial intelligence in kidney research. *Am J Physiol Renal Physiol* 2023; 325: F395–F406.
-