



Can ChatGPT be Used as a Drug Interaction Checker in Psychiatric Inpatients?

ChatGPT Psikiyatri Servisinde İlaç Etkileşimi Denetleyicisi Olarak Kullanılabilir mi?

Esra Büyük Gezer¹, Aytaç Gül², Ertan Yılmaz³

¹Hatay Mustafa Kemal University Tayfur Ata Sökmen Faculty of Medicine, Department of Medical Pharmacology, Hatay, Türkiye

²Hatay Mustafa Kemal University Tayfur Ata Sökmen Faculty of Medicine, Department of Medical Biology, Hatay, Türkiye

³Hatay Mustafa Kemal University Tayfur Ata Sökmen Faculty of Medicine, Department of Psychiatry, Hatay, Türkiye

ABSTRACT

Objective: ChatGPT is a chatbot used in many fields. Recently, it has also been used in health science. The present study investigated the clinical usefulness of ChatGPT as a drug interaction checker in a psychiatric inpatient clinic.

Methods: This retrospective cross-sectional study was conducted at a psychiatric inpatient clinic in Hatay, Türkiye. Drug-drug interactions (DDIs) were analyzed using UpToDate and ChatGPT version 4.0 based on 126 psychiatric inpatient prescriptions collected between July and October 2024. The results were compared quantitatively, and Pearson correlation analysis was performed. Interaction mechanisms were evaluated using an interrater agreement test to assess accuracy and consistency.

Results: This study evaluates DDIs in a psychiatric inpatient clinic using ChatGPT version 4.0. ChatGPT identified 93% of DDIs, in 93% of the 126 prescriptions analyzed while UpToDate identified DDIs in 92%. UpToDate identified 1127 DDIs, categorized as follows: 57 (5.1%) B, 943 (83.6%) C, 120 (10.6%) D, and 7 (0.6%) X. ChatGPT detected 1694 DDIs, and 0 (0.0%) B, 1102 (65.0%) C, 584 (34.5%) D, and 8 (0.5%) X. ChatGPT demonstrated a weak correlation with UpToDate, and the mechanisms of interaction identified by the two tools were inconsistent.

Conclusion: Although ChatGPT demonstrates strong search capabilities and facilitates the comparison of multiple drug interactions, it still requires further improvement to be considered a reliable tool for drug interaction checking.

Keywords: ChatGPT, drug interactions, drug prescription, UpToDate

ÖZ

Amaç: ChatGPT, birçok alanda kullanılan bir sohbet robotudur. Son zamanlarda sağlık bilimlerinde de kullanılmaya başlanmıştır. Bu çalışma, psikiyatri servisinde ilaç etkileşimi denetleyicisi olarak ChatGPT'nin klinik yararını araştırmıştır.

Yöntemler: Bu retrospektif kesitsel çalışma, Türkiye'nin Hatay ilindeki bir psikiyatri yatan hasta kliniğinde gerçekleştirilmiştir. Temmuz-Ekim 2024 tarihleri arasında elde edilen 126 psikiyatri yatan hasta reçetesinde UpToDate ve ChatGPT 4.0 sürümü kullanılarak ilaç-ilâç etkileşimleri (İİE) analiz edilmiştir. Sonuçlar nicel olarak karşılaştırılmış ve Pearson korelasyon analizi yapılmıştır. Etkileşim mekanizmaları, doğruluk ve tutarlılığı değerlendirmek için bir değerleyici güvenirliği testi kullanılmıştır.

Bulgular: Bu çalışma, ChatGPT 4.0 sürümü kullanılarak bir psikiyatri yatan hasta kliniğinde İİE'yi değerlendirmektedir. ChatGPT analiz edilen 126 reçetenin %93'ünde İİE tespit ederken, UpToDate %92'sinde İİE tespit etmiştir. UpToDate aşağıdaki şekilde kategorize edilen 1127 İİE tespit etmiştir: 57 (%5,1) B, 943 (%83,6) C, 120 (%10,6) D ve 7 (%0,6) X. ChatGPT 1694 İİE tespit etmiş ve 0 (%0,0) B, 1102 (%65,0) C, 584 (%34,5) D ve 8 (%0,5) X grubu olarak sınıflandırılmıştır. ChatGPT, UpToDate ile zayıf bir korelasyon göstermiş ve ikisi tarafından tespit edilen etkileşim mekanizmaları tutarsız bulunmuştur.

Sonuç: ChatGPT güçlü arama yetenekleri sergilemekle birlikte çoklu ilaç etkileşimlerinin karşılaştırılmasını kolaylaştırmaktadır. Ancak, güvenilir bir ilaç etkileşimi denetleyicisi olarak kabul edilebilmesi için daha fazla iyileştirme gerekmektedir.

Anahtar Kelimeler: ChatGPT, ilaç etkileşimleri, ilaç reçetelendirme, UpToDate

Corresponding Author/Sorumlu Yazar: Asst. Prof., Esra Büyük Gezer, Hatay Mustafa Kemal University Tayfur Ata Sökmen Faculty of Medicine, Department of Medical Pharmacology, Hatay, Türkiye

E-mail: esabuyuk@gmail.com

ORCID ID: orcid.org/0000-0001-7881-7414

Received/Geliş Tarihi: 22.09.2025 Accepted/Kabul Tarihi: 26.12.2025 Publication Date/Yayınlanma Tarihi: 31.12.2025

Cite this article as/Atıf: Büyük Gezer E, Gül A, Yılmaz E. Can ChatGPT be used as a drug interaction checker in psychiatric inpatients? J Health Inst Turk. 2025;8(3):55-61

INTRODUCTION

ChatGPT is one of the large language models (LLMs) that enable searching, writing, and analyzing. It has gained popularity due to its wide range of applications. Students commonly use it to access basic information, engage in casual conversations, and receive academic assistance and tutoring in their daily lives (1). Moreover, it is being explored in the medical field, such as academic writing, student and patient learning (2), making a diagnosis (3), safe prescribing (4), drug discovery (5), and therapy management (6).

ChatGPT can generate patient handouts, assess their readability, and potentially supplement traditional research methods. Researchers have presented clinical scenarios to various LLMs, including ChatGPT, Gemini, Claude, and Llama, to evaluate their performance. They tested the models for dose checks, recommendations based on given pharmacogenetic information, drug-drug interactions (DDIs), and drug monitoring. LLMs showed limited performance in identifying dosing regimens and therapeutic drug monitoring. However, they evaluated potential drug interactions well and provided pharmacogenomic-based recommendations (6). In another study, ChatGPT demonstrated consistency in reporting adverse drug reactions and generating patient handouts but showed limitations in interpreting data for safe prescribing (4). It achieved a 79% success rate in responding to 264 questions posed by clinical pharmacists to assess its clinical usefulness (7). Additionally, ChatGPT 4.0 was tested on 39 patient management scenarios of varying complexity levels. Two clinical pharmacists evaluated the responses based on the criteria of drug interaction, contraindication, and alternative drug recommendation. The accuracy of ChatGPT was defined as over 70%, and in some cases, it discovered drug interactions that pharmacists did not mention. However, it consistently avoided recommending specific drug doses (8). In a separate study on geriatric patient management, ChatGPT was queried about polypharmacy which identified seven inappropriate drugs for geriatrics and suggested deprescribing measures.

ChatGPT correctly detected 5 out of 6 DDIs and 3 out of 8 drug-disease interactions. However, it was unable to recognize an ineffective medication and fabricated two irrelevant drug-disease interactions (9). In another study, ChatGPT 4.0 was evaluated for its ability to analyze DDIs across 15 treatment regimens, successfully identifying 93% of all interactions. ChatGPT and the conventional method identified clinically significant DDIs as 86% and 53% of cases, respectively (10). Among the 40 drug interaction lists compiled from the literature, ChatGPT analyzed all and initially scored 39 out

of 40. However, the final score was 20 out of 40. When the reason for each interaction was assessed, 17 were classified as conclusively true, 22 as inconclusive, and 10 as true (11). In a retrospective study, 120 patient prescriptions were randomly selected from a total of 3,360, and a pharmacist analyzed DDIs with Stockley's interaction checker. At the same time, a second, blinded researcher performed the same analysis using ChatGPT version 3.5. ChatGPT achieved only 24% of the detection rate compared to the pharmacist's results. The researchers suggested that using improved artificial intelligence (AI) programs, e.g., Bing, Bard, MedPalm, or ChatGPT 4.0, would be beneficial (12).

Recent studies have focused on real clinical samples and the latest versions of LLMs. In one such study, ChatGPT version 4.0 was used to analyze 301 discharge prescriptions, and its performance was compared with that of Micromedex. ChatGPT demonstrated high accuracy, achieving a 100% detection rate for DDIs. However, it demonstrated limited accuracy in describing the severity of DDIs (37.3%) and moderate accuracy in identifying their onset (65.2%) (13). With the introduction of ChatGPT version 4.0, several studies have compared its performance to version 3.5. One study evaluated diagnostic accuracy and reported an accuracy score of 0.86 for version 4.0, compared to 0.63 for version 3.5 (14). A survey assessing different chatbots for detecting DDIs involved 255 drug interaction scenarios analyzed by ChatGPT-3.5, ChatGPT-4, Microsoft Bing AI, and Google Bard and compared their sensitivity, specificity, and accuracy. They used Drugs.com and Micromedex as conventional drug interaction checkers. Microsoft Bing AI was the most sensitive, specific, and accurate among chatbots. ChatGPT-4 outperformed better than ChatGPT-3.5; different specificity and accuracy values were observed for pharmacologic groups of drugs. However, the study methodology did not cover rank or interaction mechanism-based analysis (15).

The aforementioned databases, such as Stockley's, Micromedex, Drugs.com, UpToDate and Medscape, are considered the standard for their drug interaction checkers. Among them, UpToDate is an evidence-based clinical database that provides current information supported by under Wolters Kluwer publication. UpToDate has achieved the highest scope score, reflecting its strong sensitivity in identifying and distinguishing drug interactions (16). In this study, we used the UpToDate drug interaction checker as the reference standard. This retrospective cross-sectional study aimed to evaluate the performance of ChatGPT in identifying DDIs and to compare its results with a validated clinical tool. A psychiatric inpatient clinic was selected as the study setting due to the high likelihood of polypharmacy and associated DDIs (17). The prescriptions from the clinic were analyzed

for DDIs using both the UpToDate drug interaction checker and ChatGPT version 4.0. The results from ChatGPT were compared with those from UpToDate in terms of accuracy and consistency to assess ChatGPT's potential as a drug interaction checker.

MATERIALS AND METHODS

The retrospective cross-sectional study was performed in the psychiatric inpatient clinic of Hatay Mustafa Kemal University Tayfur Ata Sokmen Faculty of Medicine from July to October 2024, with the approval of the research Ethics Committees of Hatay Mustafa Kemal University Tayfur Ata Sokmen Faculty of Medicine (approval no: 26, date: 30.10.2024). As the study was retrospective in nature, patient consent was not required.

A clinical pharmacologist analyzed all prescriptions without exclusions for potential DDIs using the UpToDate drug interaction checker. Independently, a second researcher, blinded to the first analysis, evaluated the same prescriptions using ChatGPT (version 4.0). The results were compared quantitatively, correlation analysis was conducted, and interaction mechanisms were assessed using an interrater agreement test to increase precision. Data were stratified and analyzed according to patient age, sex, clinical indication, severity rankings of DDIs provided by UpToDate, and identified interaction mechanisms.

Prompt Adaptation

The analysis focused on determining ChatGPT's accuracy, consistency, and alignment with the risk categorizations provided by the UpToDate framework. The evaluation targeted specific interaction mechanisms such as central nervous system depression, QT prolongation, serotonin syndrome, and metabolic interference, ensuring a comprehensive assessment of clinical outcomes.

ChatGPT was first introduced to the UpToDate Risk Rating system to establish a consistent understanding of interaction categories. The system was explained using a series of prompts that defined the categories. The following prompt was used to ensure ChatGPT understood these categories:

- "These are the UpToDate risk rating categories: A means no known interaction; B means no action is needed; C means monitor therapy; D means consider therapy modification; and X means avoid combination. Do you understand?"

The study analyzed drug interactions after confirming that ChatGPT had accurately assimilated these definitions. The analysis was conducted in two phases. In the first phase, ChatGPT was prompted to analyze drug combinations

using standardized queries, such as:

- "Analyze the interactions between (drug A), (drug B), and (drug C). Provide detailed descriptions, classify their severity using the predefined UpToDate categories, and justify your classification."

In addition to this primary prompt, supplementary prompts were used to enhance the depth of the analysis:

- "Describe the mechanisms of interaction between (drug A) and (drug B), and explain their clinical consequences."
- "Why would the interaction between (drug A) and (drug B) necessitate therapy modification or monitoring?"
- "Classify the interaction between (drug A), (drug B), and (drug C) using UpToDate risk rating categories, and explain the reasoning behind your classification."

These prompts ensured that ChatGPT provided structured outputs, including the interaction descriptions, risk classifications, and justifications for each classification. Responses were collected and organized into structured tables with columns for drug combinations, interaction descriptions, risk ratings, and justifications. The results from ChatGPT were compared directly with those from UpToDate to evaluate agreement, discrepancies, and potential gaps in ChatGPT's analysis. The findings were analyzed descriptively to assess the consistency and accuracy of ChatGPT's classifications compared to UpToDate. This comparison aimed to determine the extent to which ChatGPT could serve as a supplementary tool for identifying and classifying DDIs in clinical practice.

Statistical Analysis

Baseline characteristics and numerical comparisons were given as a mean and standard deviation or percentage (%). The Pearson correlation coefficient was used to test the accuracy of ChatGPT's ranking system (rank C and D). Cohen's Kappa analysis was used to evaluate the consistency of the drug interaction mechanisms. Microsoft Excel (2021) and GraphPad Prism (version 10, USA) were used for all calculations and analyses. A p-value below 0.05 is considered significant.

RESULTS

The study included 126 patient prescriptions, with 74 patients (59%) male. The mean age was 38.6 ± 16.2 years, as presented in Table 1. The most common clinical indication was depression, observed in 52 cases (41%). The three most interacted drugs were olanzapine, quetiapine, and risperidone, as given in Table 2. A total of 552 medications were evaluated for potential DDIs. The analysis of DDIs identified by UpToDate and ChatGPT is presented in Table 3, including group-level results

and average DDIs per patient. UpToDate identified a total of 1.127 DDIs, with the following severity rank distribution: 57 (5.1%) classified as B, 943 (83.6%) as C, 120 (10.6%) as D, and 7 (0.6%) as X. This corresponds to an average of 8.9 DDIs per patient. In contrast, ChatGPT detected 1.694 DDIs, with the following distribution: 0 (0.0%) classified as B, 1,102 (65.0%) as C, 584 (34.5%) as D, and 8 (0.5%) as X. The average number of DDIs per patient was 13.4. ChatGPT identified a higher number of interactions than UpToDate, likely due to its ability to evaluate multiple drugs simultaneously and compare beyond two-drug combinations, unlike UpToDate. ChatGPT's internal ranking distribution for DDIs (separate from UpToDate's scale) was: 420 (98%) classified as X, 3 (0.7%) as D, and 6 (1.3%) as C. These values were not directly comparable with UpToDate's scoring system. To assess accuracy, the DDIs classified as C and D by ChatGPT were compared to UpToDate's corresponding ranks using Pearson correlation analysis. For rank C, a moderate and statistically significant correlation was found ($r=0.69$, $p<0.001$), as shown in Figure 1. For rank D, the correlation was weak and not statistically significant ($r=0.05$, $p=0.33$), as shown in Figure 2. The consistency of ChatGPT's

identification of interaction mechanisms was also evaluated. As presented in Table 4, Cohen's Kappa coefficient was calculated as -0.475, indicating poor agreement and a lack of statistical significance.

DISCUSSION

The present study explores ChatGPT version 4.0's performance in detecting DDIs in the psychiatric inpatient clinic. In a similar study involving 511 patients, Lexicomp identified an average of 8.5 ± 5.1 DDIs per patient (18). UpToDate found 8.9 DDIs per patient, while ChatGPT found 13.4. This finding aligns with the report by Roosan et al. (8) who observed that ChatGPT tends to detect more DDIs than conventional tools.

Several factors may explain this discrepancy. First, ChatGPT often counts overlapping mechanisms such as sedation and respiratory depression as separate interactions, whereas UpToDate typically merges them into a single entry. Second, side effects like weight gain, commonly associated with certain antidepressants, are listed as distinct interactions by ChatGPT, while UpToDate may either group them under a general advisory or omit them entirely. Lastly, ChatGPT occasionally assigns multiple interaction counts to a single mechanism. Similarly, Al-Ashwal et al. (15) reported that ChatGPT versions showed the highest rate of false-positive DDIs and the lowest accuracy and specificity among the LLMs evaluated. They explained this difference by noting that ChatGPT processes a vast amount of general information compared to the structured and curated content used in clinical databases such as Micromedex and Drugs.com (15).

UpToDate and ChatGPT revealed a significant correlation in identifying rank C DDIs (Figure 1) and failed to show a significant correlation for rank D interactions (Figure 2). When we examined the numbers in Table 3, ranks C and X appeared relatively compatible between the two tools. However, ChatGPT tends to identify DDIs as rank D, compared to UpToDate. ChatGPT identified more weight gain-related interactions and classified them as rank D. Secondly, ChatGPT tended to count more DDIs than were described in the interaction mechanisms. Additionally, as presented in Table 4, there was no agreement between the two platforms regarding

Table 1. Baseline characteristics

Characteristics	Frequency n (%)
Gender	
Male	74 (59%)
Female	52 (41%)
Age (mean \pm SD)	38.6 \pm 16.2
Indications	Depression 52 (41%) Bipolar disorder 32 (25%) Psychosis 30 (24%)

SD: Standard deviation

Table 2. The three most interacted drugs in UpToDate and ChatGPT

Number	UpToDate (total DDIs 1127)	ChatGPT (total DDIs 1694)
1	Olanzapine (332, 29%)	Olanzapine (231, 14%)
2	Quetiapine (217, 19%)	Quetiapine (188, 11%)
3	Risperidone (205, 18%)	Risperidone (185, 11%)

DDIs: Drug-drug interactions

Table 3. Comparison of DDIs of UpToDate and ChatGPT

	Database	Total	B	C	D	X
	Rank					
UpToDate	(n, %)	1127 (100%)	57 (5.1%)	943 (83.6%)	120 (10.6%)	7 (0.6%)
	DDIs per patient (n)	8.9	0.4	7.5	0.9	0.05
ChatGPT	(n, %)	1694 (100%)	0 (0.0%)	1102 (65.0%)	584 (34.5%)	8 (0.5%)
	DDIs per patient (n)	13.4	0	8.7	4.6	0.06

DDIs: Drug-drug interactions

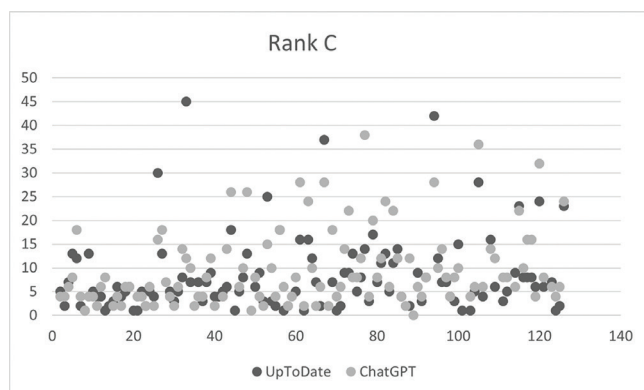


Figure 1. Pearson-correlation analysis of rank C DDIs UpToDate and ChatGPT [coefficient (r): 0.69, p=0.00]

DDIs: Drug-drug interactions

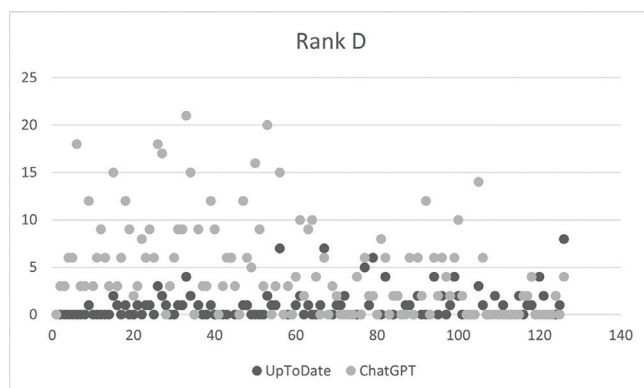


Figure 2. Pearson-correlation analysis of rank D DDIs UpToDate and ChatGPT [coefficient (r): 0.05, p=0.33]

DDIs: Drug-drug interactions

Table 4. Consistency of UpToDate and ChatGPT drug interaction mechanisms

		UpToDate	
		Agree	Disagree
ChatGPT	Agree	0.34684	0.294874
	Disagree	0.19365	0.164636
Weighted Cohen's Kappa analysis. Kappa <0: No agreement, Kappa =-0.475 [confidence interval (95%) = (-0.510 to -0.441)]			

the underlying mechanisms of the DDIs. The most interacted drugs, which correlated with prescription numbers, are given in Table 2. When compared to interaction numbers UpToDate versus (vs.) ChatGPT: Olanzapine (332 vs. 231), Quetiapine (217 vs. 188), and Risperidone (205 vs. 185). ChatGPT missed some DDIs, and this result also addresses the inconsistency between UpToDate and ChatGPT.

Previous research has reported that version 3.5 has a low intra-rater agreement with pharmacists (12). In our study, version 4.0 also demonstrated inconsistency in this regard. Unfortunately, the study did not include a detailed analysis

of the causes behind these discrepancies, which represents a limitation. One notable issue was that ChatGPT did not recognize the seizure-threshold-lowering effects of the drugs, a difference from UpToDate. This discrepancy may stem from UpToDate's access to a comprehensive range of proprietary scientific literature, whereas ChatGPT primarily relies on open-access sources.

Additionally, Medscape and Epocrates databases identified less interactions with biperiden (18). UpToDate reported a limited number of DDIs with biperiden, while ChatGPT reported a higher number and frequently classified them as rank D. Juhi et al. (11) also reported that although ChatGPT provided 22 accurate responses, these were ultimately considered inconclusive in their study. ChatGPT 4.0 demonstrated a sensitivity of 0.747, a specificity of 0.523, and an overall accuracy of 0.592 when compared to conventional drug interaction checkers (15).

A similar study reported that Micromedex identified 60.13% of DDIs from 301 discharge prescriptions, ChatGPT's accuracy achieved 100%, and guessed the onset (rapid, delayed, or not specified) of the interactions of 65.2%. However, it showed a weak performance in determining the severity of DDIs (37.3%) and could not document the relationship of DDIs (20.6%) (13). Another study investigated the pharmacology of drugs by comparing outputs from ChatGPT versions 3.5 and 4.0. The DrugBank database was used as a reference, with version 3.5 predicted 64.64% and version 4 predicted 64.33% of DDIs of some selected drugs (19). As a chatbot, ChatGPT lacks analytical depth and consistency. Several studies have indicated that its performance varies depending on the drug group. ChatGPT and the other LLMs showed different sensitivity, specificity, and accuracy scores according to drug type (15). For example, one study reported that ChatGPT failed to predict the properties of dequalinium, a large molecule compound (19). ChatGPT showed high accuracy (100%) and a weak sensitivity in determining the severity of DDIs (37.3%), which comprises respiratory system drugs 26.05%, and followed by several other pharmacological groups (13). Additionally, ChatGPT could analyze prescribed drugs such as haloperidol, chlorpromazine, and olanzapine and rank them X for QT prolongation. UpToDate identified interactions only in pairwise combinations, typically classifying them as rank C or D. This capability of ChatGPT to assess multi-drug interactions can provide clinicians with more comprehensive guidance and potentially save time in clinical decision-making. It is also important to predict the cumulative effect of concurrently administered drugs, particularly when they act as CYP3A4 substrates, inhibitors, or inducers, as these can significantly influence the pharmacokinetics and overall therapeutic outcome. While most conventional drug interaction checkers

assess interactions in pairwise combinations, Drugs.com and ChatGPT can evaluate multiple drug interactions simultaneously. After prompting ChatGPT with complex drug regimens, it provided detailed and ranked interaction data, offering valuable insights into potential risks associated with polypharmacy.

Study Limitations

This study has several limitations that warrant careful consideration. First, the retrospective cross-sectional design restricts the assessment of temporal consistency in ChatGPT's performance and limits conclusions about causality or clinical impact. Second, although we used standard comparison metrics such as Pearson correlation and Cohen's Kappa to evaluate the agreement between ChatGPT outputs and established references, the observed low concordance indicated fundamental discrepancies. Given this, we did not proceed with more advanced or outcome-focused statistical tests, as these would likely not yield meaningful additional insights at this stage of evaluation. The aim was to provide an initial benchmark of agreement between LLM outputs and UpToDate databases. Besides, medazepam was excluded from both lists due to its exclusion in the UpToDate drug interaction checker. Due to UpToDate's limitation to dual comparisons, ChatGPT was used to evaluate multiple drug interactions separately. Rank A, B, and X DDIs do not have sufficient data for correlation analysis.

CONCLUSION

ChatGPT demonstrates strong search capabilities, the ability to perform multiple drug interaction comparisons, and offers informative guidance, which may be beneficial in clinical settings and contribute to time efficiency. However, it still requires substantial improvement before it can be reliably used as a standalone drug interaction checker. This study focused on psychiatric medications; therefore, the findings may vary depending on the drug class involved.

Ethics

Ethics Committee Approval: The retrospective cross-sectional study was performed in the psychiatric inpatient clinic of Hatay Mustafa Kemal University Tayfur Ata Sokmen Faculty of Medicine from July to October 2024, with the approval of the Research Ethics Committees of Hatay Mustafa Kemal University Tayfur Ata Sokmen Faculty of Medicine (approval no: 26, date: 30.10.2024).

Informed Consent: As the study was retrospective in nature, patient consent was not required.

Footnotes

Authorship Contributions

Concept: E.B.G., Design: E.B.G., A.G., Data Collection or Processing: E.Y., Analysis or Interpretation: E.B.G., A.G., Literature Search: E.B.G., Writing: E.B.G., A.G., E.Y.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

REFERENCES

1. Yu SC, Huang YM, Wu TT. Tool, threat, tutor, talk, and trend: college students' attitudes toward ChatGPT. *Behav Sci (Basel)*. 2024;14(9):755.
2. Sridharan K, Sivaramakrishnan G. Investigating the capabilities of advanced large language models in generating patient instructions and patient educational material. *Eur J Hosp Pharm*. 2025;32(6):501-7.
3. Younis HA, Eisa TAE, Nasser M, Sahib TM, Noor AA, Alyasiri OM, et al. A systematic review and meta-analysis of artificial intelligence tools in medicine and healthcare: applications, considerations, limitations, motivation and challenges. *Diagnostics (Basel)*. 2024;14(1):109.
4. Bull D, Okaygoun D. Evaluating the performance of ChatGPT in the prescribing safety assessment: implications for artificial intelligence-assisted prescribing. *Cureus*. 2024;16(11):e73003.
5. Wang R, Feng H, Wei GW. ChatGPT in drug discovery: a case study on anticocaine addiction drug development with chatbots. *J Chem Inf Model*. 2023;63:7189-209.
6. Sridharan K, Sivaramakrishnan G. Unlocking the potential of advanced large language models in medication review and reconciliation: a proof-of-concept investigation. *Explor Res Clin Soc Pharm*. 2024;15:100492.
7. van Nuland M, Erdogan A, Açar C, Contrucci R, Hilbrants S, Maanach L, et al. Performance of ChatGPT on factual knowledge questions regarding clinical pharmacy. *J Clin Pharmacol*. 2024;64(9):1095-100.
8. Roosan D, Padua P, Khan R, Khan H, Verzosa C, Wu Y. Effectiveness of ChatGPT in clinical pharmacy and the role of artificial intelligence in medication therapy management. *J Am Pharm Assoc (2003)*. 2024;64(2):422-8.e8.
9. Cheng HY. ChatGPT's attitude, knowledge, and clinical application in geriatrics practice and education: exploratory observational study. *JMIR Form Res*. 2025;9:e63494.
10. Most A, Chase A, Sikora A. Assessing the potential of ChatGPT-4 to accurately identify drug-drug interactions and provide clinical pharmacotherapy recommendations [Internet]. 2024 [cited 2024 Oct 21]. Available from: <http://medrxiv.org/lookup/doi/10.1101/2024.06.29.24309701>
11. Juhi A, Pipil N, Santra S, Mondal S, Behera JK, Mondal H. The capability of ChatGPT in predicting and explaining common drug-drug interactions. *Cureus*. 2023;15(3):e36272.
12. Radha Krishnan RP, Hung EH, Ashford M, Edillo CE, Gardner C, Hatrick HB, et al. Evaluating the capability of ChatGPT in predicting drug-drug interactions: real-world evidence using hospitalized patient data. *Br J Clin Pharmacol*. 2024;90(12):3361-6.
13. Thapa RB, Karki S, Shrestha S. Exploring potential drug-drug interactions in discharge prescriptions: ChatGPT's effectiveness in assessing those interactions. *Explor Res Clin Soc Pharm*. 2025;17:100564.
14. Kaboudi N, Firouzbakht S, Shahir Eftekhari M, Fayazbakhsh F, Joharivarnoosfaderani N, Ghaderi S, et al. Diagnostic accuracy of ChatGPT for patients' triage: a systematic review and meta-analysis. *Arch Acad Emerg Med*. 2024;12:e60.

15. Al-Ashwal FY, Zawiah M, Gharaibeh L, Abu-Farha R, Bitar AN. Evaluating the sensitivity, specificity, and accuracy of ChatGPT-3.5, ChatGPT-4, Bing AI, and bard against conventional drug-drug interactions clinical tools. *Drug Healthc Patient Saf.* 2023;15:137-47.
16. Shariff A, Belagodu Sridhar S, Abdullah Basha NF, Bin Taleth Alshemeil SSH, Ahmed Aljallaf Alzaabi NA 4th. Assessing consistency of drug-drug interaction-related information across various drug information resources. *Cureus.* 2021;13(3):e13766.
17. Sinclair LI, Davies SJ, Parton G, Potokar JP. Drug-drug interactions in general hospital and psychiatric hospital in-patients prescribed psychotropic medications. *Int J Psychiatry Clin Pract.* 2010;14(3):212-9.
18. Ranković A, Milentijevic I, Jankovic S. Factors associated with potential drug-drug interactions in psychiatric inpatients. *Eur J Hosp Pharm.* 2024;31(2):127-34.
19. Zhang Y, Ren S, Wang J, Lu J, Wu C, He M, et al. Aligning large language models with humans: a comprehensive survey of ChatGPT's aptitude in pharmacology. *Drugs.* 2025;85(2):231-54.