LLM-generated PA letters improve efficiency but are not yet reliable without clinician review.

Assessing Quality of ChatGPT-5 in Writing a Prior Authorization Letter in Nephrology

Noppawit Aiumtrakul[1], Charat Thongprayoon[1], Chutawat Kookanok[2], Methavee Poochanasri[3], Wisit Cheungpasitporn[1]

[1]Division of Nephrology and Hypertension, Department of Medicine, Mayo Clinic, Rochester, MN, USA,

[2]One Brooklyn Health, Interfaith Medical Center, Brooklyn, NY, USA, [3]Bhumibol Adulyadej Hospital, Thailand

BACKGROUND

- Writing prior authorization (PA) letters is a time-consuming task. Surveys show 94% of physician report PA-related delays, sometimes leading to serious adverse events, and 95% report burnout from the process.
- Large language models (LLMs), such as ChatGPT, can facilitate this work. However, the quality of artificial intelligence (AI)-generated letters must be carefully assessed.
- We aim to evaluate the quality of PA letters drafted by ChatGPT-5 for commonly used medications requiring PA in Nephrology. Quality was evaluated based on correctness and strength of clinical reasoning.

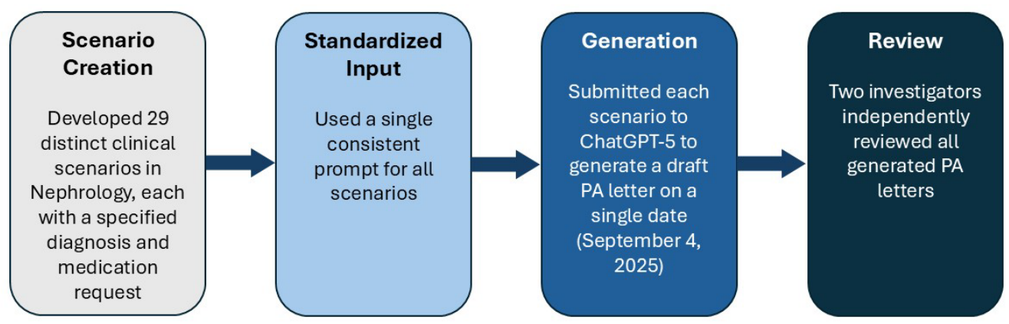
METHODS

- We created a single standardized prompt and used in 29 different Nephrology scenarios to generate PA letters. Each PA letter was reviewed against four criteria: absence of false statements or hallucinations, correctness of ICD-10 coding, presence and validity of citations, and clinical reasoning.
- Clinical reasoning was rated on a 4-point Likert scale (illogical, weak, adequate and strong). FDA drug labels, KDIGO guidelines and related randomized controlled trials were used as reference standards.

RESULTS

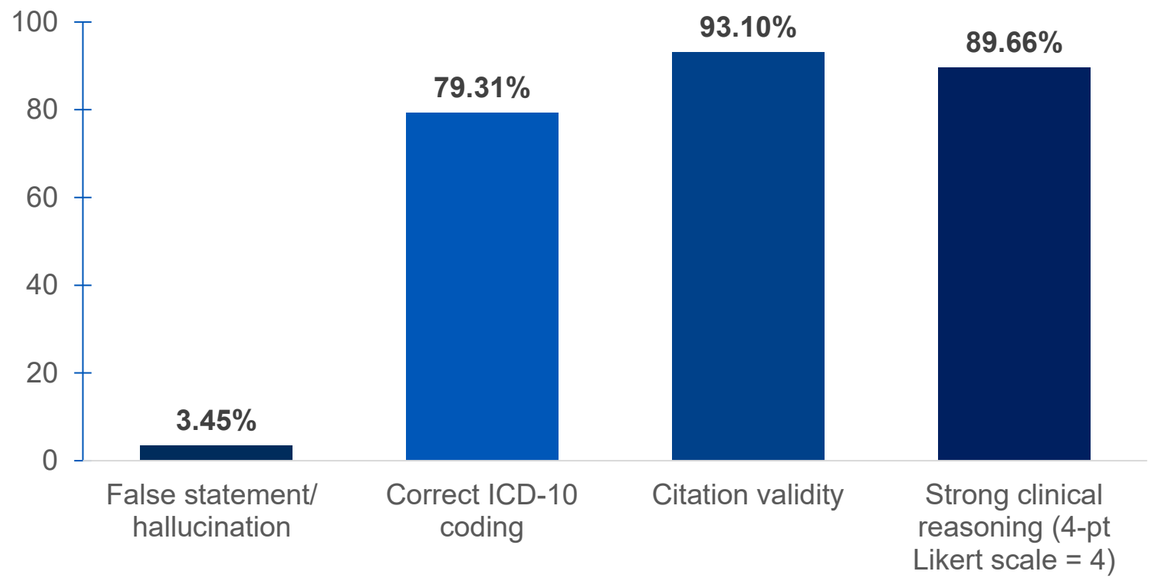
- Across 29 letters, one letter (3.45%) contained false statements mentioning an irrelevant clinical trial.
- The 10th revision of the International Classification of Diseases (ICD-10) coding was correct in 23 letters (79.31%), most errors were related to CKD staging (stage 3a vs 3b) or to inconsistent definitions (one code with renal involvement and another code without).
- 27 letters (93.10%) cited valid references, with one letter citing an incorrect trial and another one citing a correct KDIGO guideline with inaccessible link.
- Twenty-six of the twenty-nine letters (89.66%) demonstrated strong clinical reasoning, supported by guideline-oriented or FDA label–aligned justification. The remaining 3 letters still rated as adequate reasoning.
- The main areas for improvement involved citing relevant references and emphasizing special considerations, for example REMS compliance for eculizumab.

Figure 1: Workflow for the evaluation of ChatGPT-5–produced PA letters



We used a structured prompt as follows:
“You are a board-certified Nephrologist writing a prior authorization (PA) letter to health plan medical reviewers in a professional tone. Task: Draft a ≤350-word PA letter for the scenario below. Requirements: 1) Clear statement of the indication and requested regimen/dose. 2) Diagnosis with ICD-10 code(s): choose the most specific and appropriate code(s). 3) Clinical reasoning: why this medicine is medically necessary for this patient. 4) References section with at least one clinical guideline or high-quality source. Provide hyperlinks as full URLs.”.

Figure 2: Performance of ChatGPT-5-Generated PA letters



Bars represent the proportion of PA letters meeting predefined quality criteria across four domains. Among 29 evaluated letters, only one contained a false statement (3.5%), while most demonstrated correct ICD-10 coding (79.3%), valid citations (93.1%), and strong clinical reasoning (89.7%).

CONCLUSION

- ChatGPT-5 generated overall accurate PA letters drafts with acceptable clinical reasoning.
- The errors mainly involved coding reference selection, and incomplete safety discussion.
- Our findings highlight the need for careful review before use. With appropriate oversight, LLM generated drafts may help reduce administrative burden, but they are not yet reliable enough to be used without clinician verification.