

Artificial Intelligence in Chronic Myeloid Leukemia: Assessing the Accuracy and Clinical Utility of Large Language Models

Berkha Rani • Rachita Kour • Aakanksha Pitiliya • Rajesh Thirumaran • Manzoor Rather

Mercy Catholic Medical Center, Darby, PA

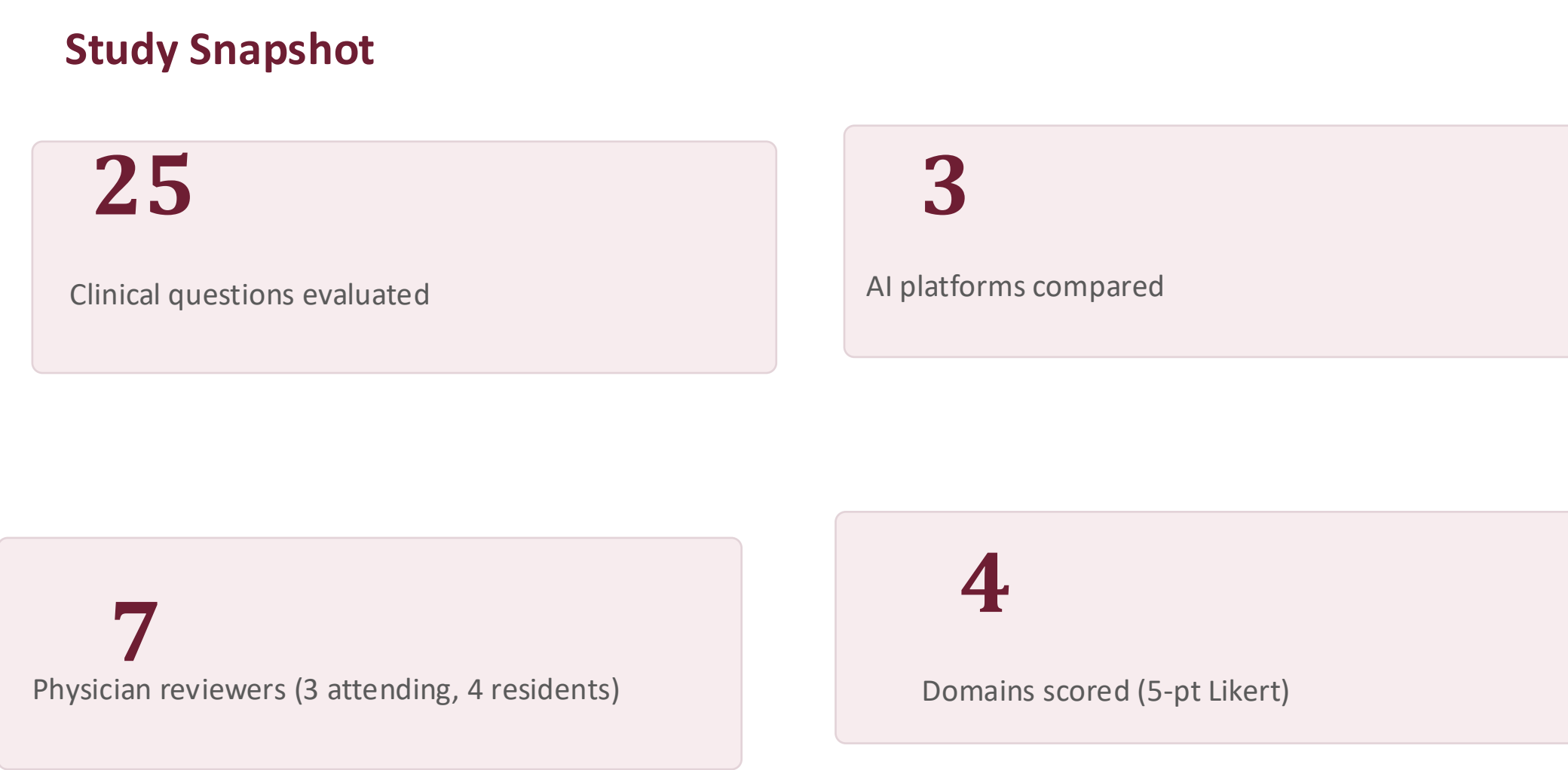


Introduction

- AI-based large language models (LLMs) are increasingly explored as tools to support clinical decision-making and medical education.
- In chronic myeloid leukemia (CML), treatment selection, molecular monitoring, and long-term disease management require guideline-based precision, making accurate interpretation of evidence-based recommendations essential.
- Despite growing use of AI platforms, their comparative performance on clinically relevant CML questions has not been well established.
- This study compared three leading AI language models — ChatGPT (OpenAI), Copilot (Microsoft), and Gemini (Google) — on accuracy, relevance, response quality, and timeliness for CML-related questions.

Methods

- A set of 25 clinically relevant CML questions was developed, spanning common diagnostic, therapeutic, and monitoring scenarios.
- Each question was submitted verbatim to ChatGPT, Copilot, and Gemini; responses were anonymized and evaluated single-blind by physician reviewers blinded to chatbot source.
- Reviewers scored each response across four domains — accuracy, clarity, relevance, and clinical usefulness — using a 5-point Likert scale.
- Mean domain scores were compared using ANOVA and pairwise t-tests, with $p < 0.05$ considered significant.



Results

- Responses from all three platforms were independently evaluated by 7 physician reviewers (3 attending, 4 residents) across accuracy, clarity, relevance, and clinical utility.
- ChatGPT showed the strongest performance in relevance and overall clinical utility; Copilot received the highest clarity ratings; Gemini performed comparably on accuracy.
- Overall differences among the three platforms were not statistically significant (ANOVA $p > 0.05$; pairwise comparisons $p > 0.05$).
- Reviewer scoring was consistent between residents and attending physicians, reflecting strong inter-rater agreement.

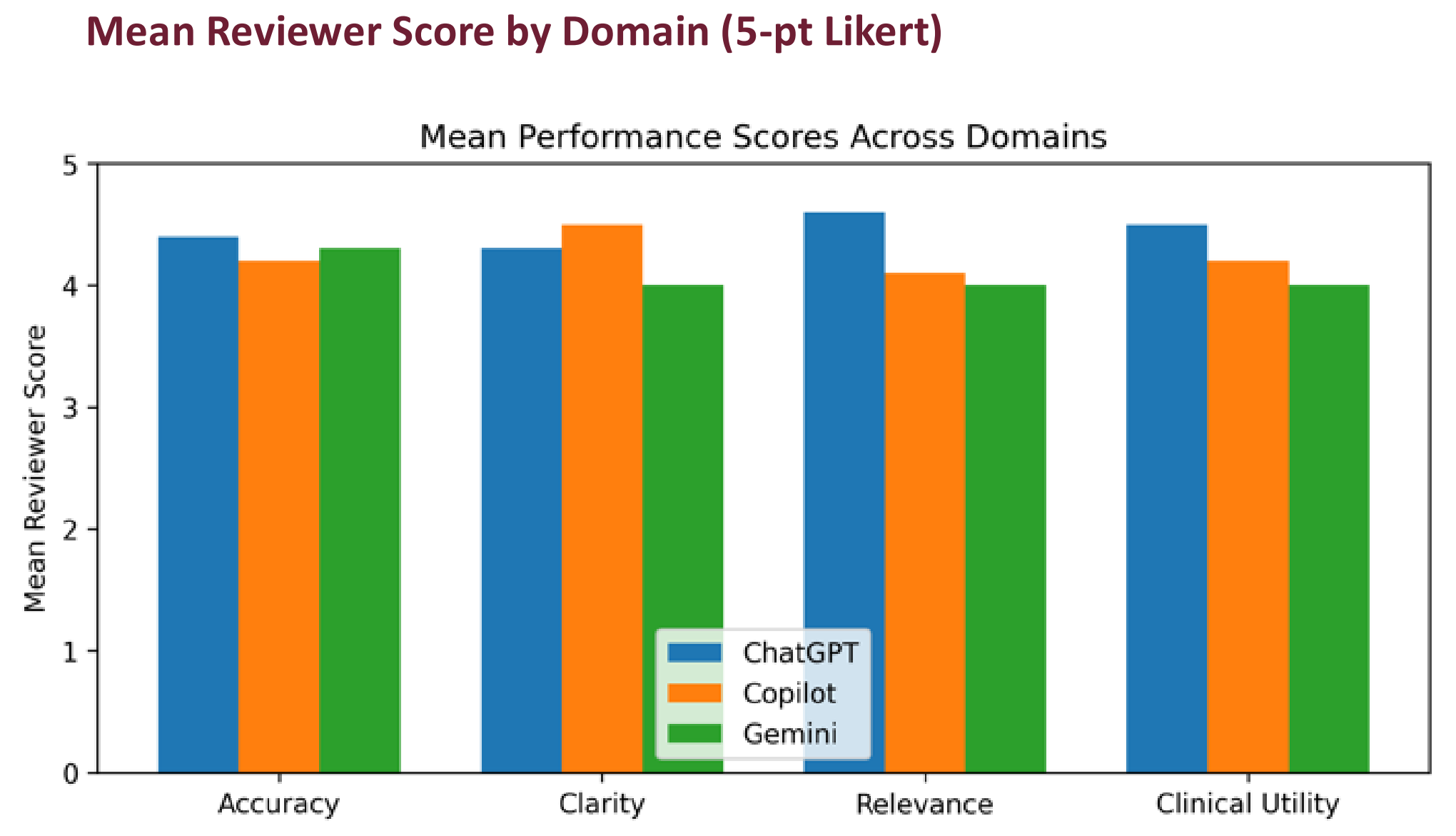
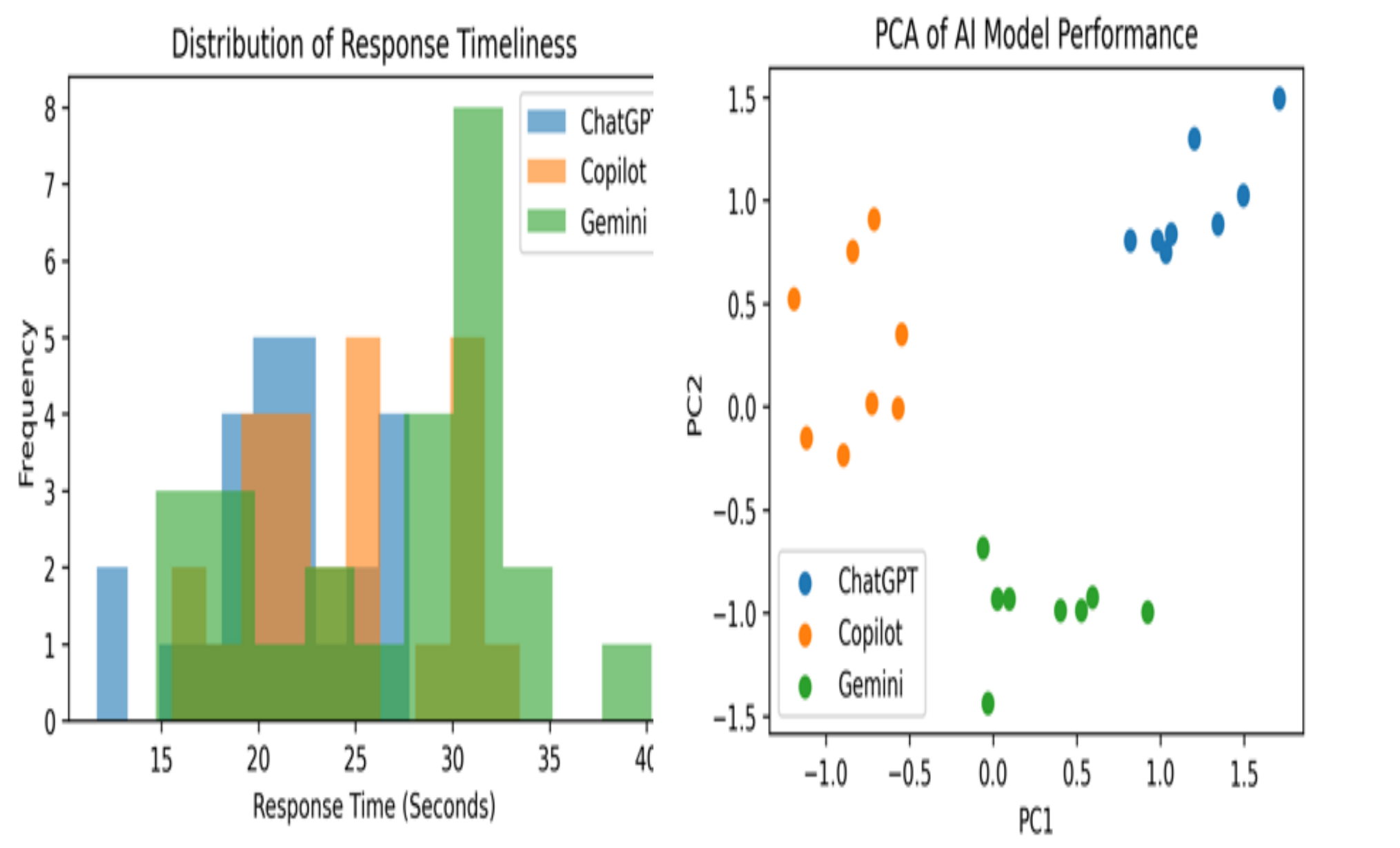


Table 1. Mean Reviewer Scores

Domain	ChatGPT	Copilot	Gemini	ANOVA p
Accuracy	4.4	4.2	4.3	> 0.05
Clarity	4.3	4.5	4.0	> 0.05
Relevance	4.6	4.1	4.0	> 0.05
Clinical Utility	4.5	4.2	4.0	> 0.05

Gold shading marks the highest-scoring platform per domain. None of the differences reached statistical significance (ANOVA and pairwise t-tests, $p > 0.05$).



Response time (seconds) and a PCA projection of response characteristics per platform, from the same reviewer dataset used for domain scoring.

Discussion & Conclusion

- Contemporary AI chatbots can generate clinically meaningful, patient-oriented responses to common CML questions.
- ChatGPT and Copilot showed comparatively stronger performance, particularly in clarity, relevance, and clinical utility.
- Although differences among platforms were modest, variability in response quality highlights the need for continued physician oversight in hematologic oncology practice.
- These findings suggest AI language models may serve as valuable adjuncts in CML education and clinical decision support.

Appendix. The 25 Clinical Questions Evaluated

- | | | | | |
|--|---|--|--|--|
| 1. What is Chronic Myeloid Leukemia (CML)? | 6. How is CML diagnosed? | 11. How common is CML? | 16. Which TKIs are commonly used (imatinib, dasatinib, nilotinib, etc.)? | 21. What is the prognosis and survival rate for CML? |
| 2. What causes CML and what is the Philadelphia chromosome? | 7. What blood tests and procedures are used to confirm CML? | 12. What is the difference between CML and other types of leukemia? | 17. When is a stem cell/bone marrow transplant recommended for CML? | 22. How often is monitoring (PCR tests) needed during treatment? |
| 3. What are the symptoms of CML? | 8. What is the BCR-ABL gene and why is it important? | 13. How is CML staged or risk-stratified (Sokal, EUTOS scores)? | 18. What is treatment-free remission (TFR) in CML? | 23. Can CML be cured? |
| 4. Can CML occur without any symptoms? | 9. Who is at risk for developing CML? | 14. What are the main treatment options for CML? | 19. What are the common side effects of TKI therapy? | 24. What happens if CML becomes resistant to treatment? |
| 5. What are the different phases of CML (chronic, accelerated, blast)? | 10. Is CML hereditary or passed down in families? | 15. What are tyrosine kinase inhibitors (TKIs) and how do they work? | 20. How does treatment differ between chronic phase and advanced phases? | 25. Where can I find support and resources for CML patients? |

Acknowledgments: Editorial support and guidance were provided by Dr. Aakanksha Pitiliya, resident physician at Mercy Fitzgerald Hospital, Dr. Thirumaran, Chair of the Oncology Department, and Dr. Rather, Director of Medicine at Mercy Fitzgerald Hospital. No compensation was given beyond their regular salaries.

Conflict of interest: None to disclose. **Keywords:** AI, Artificial intelligence, CML, Hematological malignancies.