# Comparative Analysis of History of Present Illness Generation: AI versus Physicians

Paige Long, **Jenna Howard**, Jagoda Pasternacka, Aisha Nour, Shebna Unes Kunju, MD
Med Center Health Bowling Green, KY
University of Kentucky College of Medicine

# Faculty Disclosure

- I, Jenna Howard have nothing to disclose

# Introduction

- The history of present illness (HPI) is a vital part of clinical documentation traditionally written by clinicians to capture the narrative of patients' health concerns.
- Increasing patient access to electronic medical records makes note readability more important.
- Emerging artificial intelligence (AI) tools for clinical documentation present potential benefits including improved efficiency and quality of medical records, as well as decreased physician burnout (1-3).
- Prior studies have shown that AI-generated clinical text, such as electronic messages and medical summaries, performed better in terms of completeness and was preferred by participating patients (4).

# Study Objective

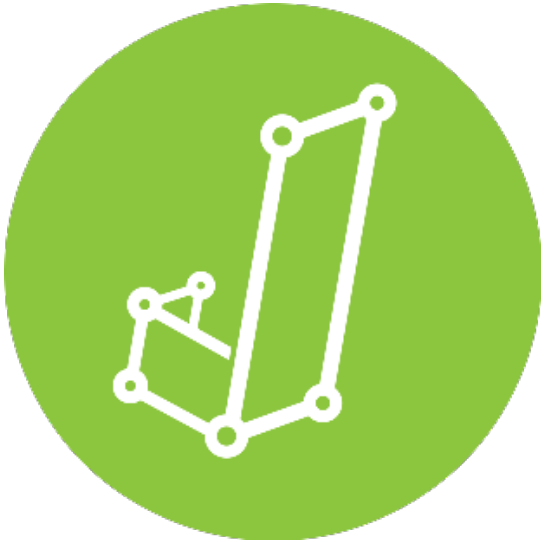
- Limited literature exists evaluating the accuracy, errors, and patient understandability of AI-generated HPIs as compared to physician-written HPIs.
- Study objective: To compare AI application-generated HPIs to provider-generated HPIs in terms of accuracy, errors, understandability, and overall preference.

# Methods– Data Collection

- IRB-approved by WKHL Research Foundation
- Outpatient gynecology setting
- 50 English-speaking patients age 18-80
- Two HPIs were produced:
  - Form A: Provider
  - Form B: Abridge AI
- Single blinded: participants were unaware of form identity
- Post-visit questionnaire was given to each patient to assess:
  - Perceived accuracy (1-5 scale)
  - Errors
  - Readability
  - Overall preference

# Methods– Statistical Analysis

- Statistical analysis was completed using JASP version 0.95.1 software
- Confidence interval 95%
- p-value <0.05 significant
- Analysis of accuracy rankings
    - Paired t-test with Wilcoxon signed-rank
- A Classical binomial test analyzing frequencies was used for categorical variables (errors, clarity, and preference)



# Results– Chi Squared Goodness of Fit

- In comparison to physician-generated, AI-generated HPIs:
  - Had fewer errors ($p = 0.002$)
  - Were easier to understand ($p < 0.001$)
  - Were overall preferred ($p < 0.001$)

| Outcome (A, B, or No Difference) | $\chi^2$ | df | P |
|---|---|---|---|
| Errors | 12.04 | 2 | 0.002 |
| Understanding | 32.92 | 2 | <.001 |
| Preference | 47.32 | 2 | <.001 |

**Table 1:** Chi-square analysis for each of the outcomes analyzed, each comparing the results of Form A vs Form B

# Results: Wilcoxon Signed-Rank

- Accuracy was rated on a 1 (most inaccurate) to 5 (most accurate) scale
- AI-generated HPIs were deemed by patients as more accurate (p = 0.005) with a mean accuracy score of **4.36**
- Provider-generated HPIs received a mean score of **3.82**

| | N | Mean | SD | SE |
|---|---|---|---|---|
| Form A Accuracy | 50 | 3.82 | 1.466 | 0.207 |
| Form B Accuracy | 50 | 4.36 | 1.396 | 0.197 |

**Table 2:** Descriptive statistics for wilcoxon signed-rank accuracy analysis

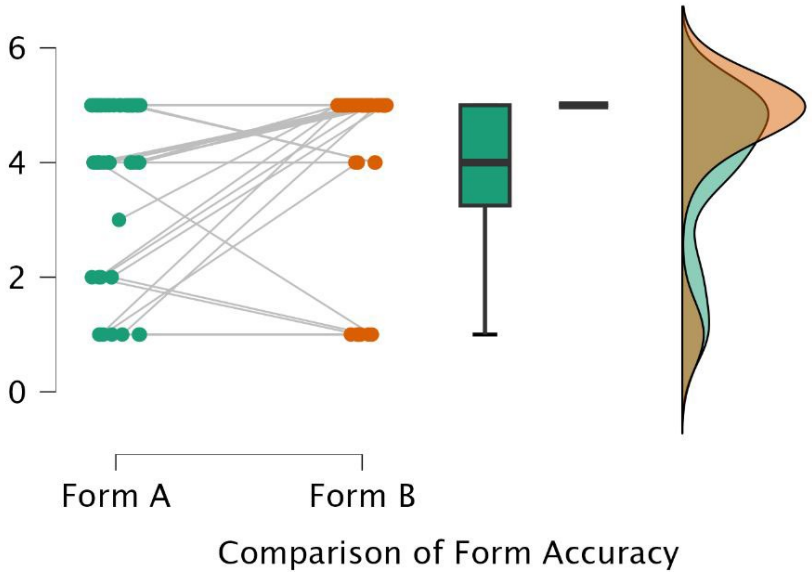



**Figure 1:** A raincloud illustration demonstrating the distribution of accuracy ratings for Forms A and B

# Discussion

- AI-generated HPIs were more accurate, easier to understand, less error-prone, and overall preferred by patients, as compared to provider-generated HPIs.
- Existing literature revealed similar findings, suggesting that AI-generated notes performed better in terms of overall impression, completeness, usefulness, and organization (5).
- These outcomes support the exciting potential of using AI tools for streamlining history taking and generating more comprehensible, patient-friendly notes.

# Limitations

- Small sample size
- Single institution
- Variation in physician documentation practices

# Conclusion

- Cross-sectional study comparing AI-generated vs. physician-generated HPIs, as assessed by patients.
- AI-generated notes were found to be more accurate, less error-prone, easier to understand, and overall preferred.
- Future studies with larger sample sizes and clinicians as note evaluators are warranted to evaluate generalizability and support clinical application.

# References

1. Lin SY, Shanafelt TD, Asch SM. Reimagining clinical documentation with artificial intelligence. Mayo Clin Proc. 2018;93(5):563-565. doi:10.1016/j.mayocp.2018.02.016

2. Lee C, Vogt KA, Kumar S. Prospects for AI clinical summarization to reduce the burden of patient chart review. Front Digit Health. 2024;6:1475092. doi:10.3389/fdgth.2024.1475092

3. Albrecht M, Shanks D, Shah T, et al. Enhancing clinical documentation with ambient artificial intelligence: a quality improvement survey assessing clinician perspectives on work burden, burnout, and job satisfaction. JAMIA Open. 2025;8(1):ooaf013. doi:10.1093/jamiaopen/ooaf013

4. Robinson EJ, Qiu C, Sands S, Khan M, Vora S, Oshima K, Nguyen K, DiFronzo LA, Rhew D, Feng MI. Physician vs. AI-generated messages in urology: evaluation of accuracy, completeness, and preference by patients and physicians. World J Urol. 2024 Dec 27;43(1):48. doi: 10.1007/s00345-024-05399-y. PMID: 39729119; PMCID: PMC11680670.

5. Almario CV, Chey W, Kaung A, Whitman C, Fuller G, Reid M, et al. Computer-generated vs physician-documented history of present illness (HPI): results of a blinded comparison. Am J Gastroenterol. 2015;110(1):170-179. doi:10.1038/ajg.2014.356

# Questions?