Automating Anesthesiology Antenatal Consultation Notes for Postpartum Hemorrhage Risk: A Feasibility Study Using Large Language Models

Domenic J. Pedulla, MD; Vesela Kovacheva, MD, PhD

Department of Anesthesiology, Perioperative and Pain Medicine, Brigham and Women's Hospital, Boston, MA

Antenatal Anesthesiology Consultations for Postpartum Hemorrhage (PPH):

- Can improve care coordination, patient experience, and outcomes for patients with high risk
- PPH remains a leading cause of morbidity and mortality

Large Language Models (LLMs):

- Useful in automation of clinical workflows
- Recent studies suggest LLMs can produce high-quality handoff notes

Generated Using GPT

Question:

- Can LLMs be used to generate consistent, high-quality antenatal anesthesiology consultation notes for patients with elevated risk of PPH?

Hypothesis:

- LLMs will successfully generate accurate, standardized antenatal anesthesiology consultation notes

Methods

Mock Patient Scenario
- Based on known risk factors for PPH
- Designed by OB anesthesiologist or fellow

Mock Consultation Conducted
- Simulated patient/physician conversation with each role performed by anesthesiologist
- Recorded and Transcribed using Whisper large-v3 1550M

Mock OB Note
- Written by OB anesthesiologist or fellow
- Based on real OB prenatal visit notes

Sample Antenatal Anesthesiology Consult Note
- Written by OB anesthesiologist or fellow
- Framework for all other notes

LLM- Generated Note
- GPT4 (Open AI) utilized
- Prompted with chain-of-thought technique
- Evaluated using qualitative framework (Likert scale 1-5)

Criteria	Description
Readability	Well-written with use of professional medical language
Completeness	Overall key information from transcription and OB note are included in the consult note
Curation	Exclusion of irrelevant information
Correctness: Hallucination	Invention of sentences with no context in the source documents
Correctness: Knowledge Gap	Generation of sentences that are inconsistent with knowledge from source documents
Correctness: Faulty Logic	Inferring logically incorrect sentences based on information from source documents
Correctness: Bias	Demonstrates biases toward or against the patient
Correctness: Overall	Aggregate of individual correctness scores
Patient Safety Risk	Physician rating of overall safety risk based on errors or incomplete information in the consult note
Usefulness	Degree to which consult note require minimal corrections or adjustments by a physician before being entered into the medical record

Adapted from Hartman et al., 2024

Results

Interim Analysis:

- Mean Quality Score:
  5 (4-5 IQR) out of 5

- 60% Sufficient quality and accuracy to be directly included in medical record

- 1 note with inadequate quality (inaccurate OB history, incomplete plan)

Patient Scenario	Correctness (Overall)	Patient Safety Risk	Usefulness
1	5	5	5
2	5	5	5
3	5	5	5
4	5	5	4
5	5	5	4
6	4	5	4
7	4	4	3
8	5	5	5
9	5	5	5
10	5	5	5

Conclusions

LLM-Generated Anesthesiology Consultation Notes:

- High quality and accuracy scores
- May improve standardization and note quality
- Combine with checklist for common HROAC indications

Next Steps:

- Apply same protocol to 10 additional mock patient scenarios
- Develop objective, automated evaluation framework
- Develop concurrent PPH risk checklist
- Conduct follow up study utilizing LLM for real patient scenarios

Generated Using GPT

References

1- Uwubamwen, Verma. *Anaest Intens Care M*. 2019.

2- Corbetta-Rastelli et al. *Obstet Gynecol*. 2023.

3- Williams et al. *MedRXIV*. 2024.

4- Hartman et al. *JAMA Netw Open*. 2024.