American Journal of Psychiatric Rehabilitation

Expert Opinion Article

SMART NOTES IN THE OPERATION THEATRE: A CROSS-SECTIONAL COMPARISON OF CHATGPT VERSUS SURGEON GENERATED OPERATIVE DOCUMENTATION IN LAPAROSCOPIC CHOLECYSTECTOMY

Akshai Kumar[1*], Naveed Ali Khan[2], Raazia Ramzan[3], Abdul Khalique[4], Javeria Khalid[5], Mujeeb Rehman Malik[6]

[1*]Affiliation: Dow University of Health Sciences Email: dr_akshai@ymail.com
[2]Affiliation: Dow University of Health Sciences Email: dr.naveedduhs@gmail.com
[3]Affiliation: Dow University of Health Sciences Email: natharaazia@gmail.com
[4]Affiliation: Dow University of Health Sciences Email: ak.mahar@duhs.edu.pk
[5]Affiliation: Dow University of Health Sciences Email: javeria.khalid529@gmail.com
[6]Affiliation: Dow University of Health Sciences Email: mujeebur.rehman@duhs.edu.pk

***CORRESPONDING AUTHOR:** Akshai Kumar

ABSTRACT

Objective: To compare the completeness, guideline adherence, time efficiency and user satisfaction of operative notes generated by ChatGPT versus those written by general surgery residents.

Methods: This study included 118 patients undergoing Laparoscopic Cholecystectomy in the Department of General Surgery, Dow University Hospital over a period of three months from August to October 2025. For each case, operative notes were generated using ChatGPT (open AI platform) with a standardized prompt and compared with conventional surgeon written notes. AI generated data was reviewed for factual inconsistencies. Using Royal College of Surgeons guidelines, completeness, accuracy and guideline adherence was assessed. Time for each note completion was recorded. User satisfaction was also assessed. Using SPSS version 27, data was assessed.

Results: A total of 118 patients were enrolled in this study with the mean age of 43.75 ± 14.21 years. There were 54.2% females and 45.8% males. Symptomatic Cholelithiasis was the most common diagnosis. ChatGPT generated operative notes took considerably less amount of time than human written notes (6.6 seconds vs 5.43 minutes; p value of < 0.001). A 100% guideline adherence was observed in AI generated notes to 59.3% in human written notes. Factual inaccuracies were noted in 32 (27.11%) of the cases but those became less evident with each new entry. Documentation of anticipated blood loss and DVT prophylaxis was more significant in AI generated notes. For AI-generated notes, 93.3% of the surgeons reported being very satisfied, and for conventional notes, and 80% reported their satisfaction. A majority of respondents (96.7%) stated that they would recommend AI generated operative notes for clinical documentation.

Conclusion: Large Language Models such as ChatGPT may facilitate structured, comprehensive and time efficient operative note writing, reducing burden of documentation in busy surgical settings. However, the accuracy and reliability of AI generated notes depend on precise surgeon input, and human oversight remains essential to ensure clinical accountability.

Keywords: Large Language Models (LLM), ChatGPT, AI-generated operative note, Conventional operative note, Laparoscopic Cholecystectomy

INTRODUCTION

Operation notes play a pivotal role in capturing a comprehensive account of a surgical procedure. It is the responsibility of every surgeon to maintain a detailed operative record, delineating individual patient details, providing a technical description of the surgical team's activity, stating indications for the procedure, documenting operative findings, and outlining postoperative instructions [1]. The minutiae of accurate record-keeping are essential for ensuring quality healthcare delivery [1,2].

Studies show that up to 45% of operative notes are incomplete, and nearly 30% are illegible, posing significant medico-legal risks and affecting continuity of care [3]. The quality of operative notes thus serves as a crucial conduit of communication among various healthcare providers, both from therapeutic and legal standpoints [4]. However, due to issues like illegibility and incompleteness, surgeons often face difficulties in documentation during high-pressure clinical situations [5,6].

To address these concerns, the American College of Surgeons (ACS) and Royal College of Surgeons (RCS) have published comprehensive guidelines outlining essential components of operative notes [7]. Despite the availability of such guidelines, studies report

persistent gaps in implementation, with frequent delays or errors in documentation [7].

In recent years, artificial intelligence (AI), particularly Large Language Models (LLMs), has shown promise in improving the accuracy, fluency, and efficiency of clinical documentation. LLMs like ChatGPT (Chat Generative Pretrained Transformer) are trained on large datasets including textbooks, scientific literature, and clinical narratives, enabling them to generate human-like responses to diverse prompts [1,3,4].

ChatGPT, launched in November 2022, has undergone continuous refinement through supervised fine-tuning and reinforcement learning [5,6]. Its integration into medical writing, academic research, and clinical documentation has rapidly increased, with tools like Microsoft Copilot, Google Bard, and Claude being used alongside it [7,8,9]. Traditional search engines offer generalized information, but LLMs have garnered attention due to their ability to generate tailored, context-aware responses, making them increasingly relevant for daily medical practice [8,9,10].

With this context, the present comparative cross-sectional study aimed to assess the efficiency, completeness, and accuracy of operative notes generated by ChatGPT versus those written by surgical professionals.

METHODOLOGY

We conducted a comparative cross-sectional study in the department of General Surgery at Dow University Hospital on 118 patients undergoing Laparoscopic Cholecystectomy for a period of three months from August 2025 to October 2025. This study was conducted to compare the completeness, guideline adherence, time efficiency and user satisfaction of operative notes generated by LLM versus those written by general surgery residents.

Ethical approval for the study was obtained from the Institutional Review Board of Dow University of Health Sciences (Ref: IRB-4131/DUHS/Approval/2025/299). Informed consent was obtained from all the patients priors to inclusion in the study.

Patients undergoing laparoscopic cholecystectomy were consecutively enrolled in the study. Only those with incomplete intra-operative data or procedures complicated by deviations from standard laparoscopic cholecystectomy were excluded. Operative notes were generated using ChatGPT and conventional hand-written method for each participant. For AI-generated group, free account on the OpenAI platform (https://chat.openai.com) was used and a standardized written prompt was entered into the text field. The prompt was "Create a Royal College-compliant operative note for a laparoscopic cholecystectomy for [indication], performed by [surgeon's name], with [intraoperative findings], and [complications or issues if any]. Include post-op plan and specimen handling." The AI generated group was reviewed for factual inconsistencies but was not substantively altered. For the conventional group, hand written notes were completed by the general surgery residents without any assistance from the LLM, using the department's routinely used operative note format.

The primary outcome was completeness and accuracy of operative notes, assessed using Royal College of Surgeons 18-point operative note criteria. Each component was scored dichotomously, with one point given for adequate documentation and zero for insufficient one. Notes acquiring a score of 16 or more out of 18 were classified adherent to guidelines, while those scoring less than this threshold were considered non-adherent. Time taken to complete each operative note was recorded through a stopwatch and perceived usefulness was evaluated through a feedback form using a Likert scale ranging from 'very satisfied' to 'very dissatisfied'.

Data analysis was performed using IBM SPSS base Licensed version 27. Categorical variables were presented as frequency and percentages while continuous variables were reported as mean ± standard deviation. A p-value of less than 0.05 was considered statistically significant.

RESULTS

A total of 118 operative notes for Laparoscopic Cholecystectomy were analyzed. The mean age of the patients included in this study was 43.75 ± 14.21 years. There were 54 (45.8%) males and 64 (54.2%) females. The most common indication for surgery was Symptomatic Cholelithiasis (35.6%), followed by Acute Cholecystitis (27.1%), Chronic Cholecystitis (18.6%), Empyema gallbladder (13.6%) and Mucocele (5.1%).

The mean time required to complete operative notes via AI-generated system was significantly lower as compared to conventional surgeon written notes with AI notes being completed in 6.6 ± 0.14 seconds and conventional notes completed in 5.43 ± 1.33 minutes (P value of < 0.001). Among AI generated notes, 82 (69.5%) required only one attempt to complete an accurately written note whereas 32 (27.1%) required two edits and 4 (3.4%) required more than two edits before final approval.

Parameter		Statistical metrics
Age		43.75 ± 14.21 years
Gender	Male	54 (45.8%)
	Female	64 (54.2%)
Diagnosis	Symptomatic Cholelithiasis	42 (35.6%)
	Acute Cholecystitis	32 (27.1%)
	Chronic Cholecystitis	22 (18.6%)
	Mucocele	6 (5.1%)
	Empyema gallbladder	16 (13.6%)
Time taken for operative note completion	AI generated note	6.6 ± 0.14 seconds
	Conventional note	5.43 ± 1.33 minutes
No. of edits for AI generated operative note	One	82 (69.5%)
	Two	32 (27.1%)
	More	4 (3.4%)

Table 1 shows demographic parameters of the cases included in the study

AI generated notes demonstrated higher adherence level to the Royal College of Surgeon's guidelines. Documentation of anticipated blood loss (35.6% for conventional and 100% for AI notes) and DVT prophylaxis (0% for conventional and 100% for AI notes) was more significant in AI generated notes when compared to conventional notes. The RCS guideline components and adherence to each component in both AI generated and conventional notes is given in table 2.

Operative notes' components	Frequency (%)	
	Conventional	AI generated
Date	118 (100%)	118 (100%)
Time	104 (88.1%)	118 (100%)
Elective/emergency procedure	92 (78%)	118 (100%)
Name of theatre anesthetist	118 (100%)	118 (100%)
Anesthesia	118 (100%)	118 (100%)
Names of operating surgeon and assistant	118 (100%)	118 (100%)
Operative procedure carried out	118 (100%)	118 (100%)
Incision	118 (100%)	118 (100%)
Operative diagnosis	118 (100%)	118 (100%)
Operative findings	118 (100%)	118 (100%)
Problems/complications	118 (100%)	118 (100%)
Extra procedure performed and why it was performed	118 (100%)	118 (100%)
Details of tissue removed, added or altered	110 (93.2%)	118 (100%)
Anticipated blood loss	42 (35.6%)	118 (100%)
Antibiotic prophylaxis	118 (100%)	118 (100%)
Details of closure technique	100 (84.7%)	118 (100%)
DVT prophylaxis	0	118 (100%)
Detailed post-operative instructions	118 (100%)	118 (100%)
Signature	112 (94.9%)	118 (100%)
Mean compliance score (out of 18)	15.69 ± 0.87	18 ± 0
RCS adherence (16 or more)	70 (59.3%)	118 (100%)

Table 2 showing the RCS guideline components and adherence to each component in both conventional and AI generated notes

Based on the dichotomous scoring of each component of the RCS guideline, all operative notes were scored out of 18, and the notes achieving a score of 16 or more was classified as guideline adherent. The mean compliance score for AI generated operative note was significantly higher than conventional notes with mean score of 18 in AI written notes and 15.69 ± 0.87 in conventional notes with a p value of <0.001. A total of 118 (100%) AI generated notes were found to be guideline adherent and 70 (59.3%) surgeon written notes fell in this category with a p value of <0.001. Factual inaccuracies were noted in 32 (27.11%) of the cases but those became less evident with each new entry.

A total of 30 members of the surgical team were also asked for their feedback on the use of LLM in surgical documentation. Of these 8 (26.7%) were consultants and 22 (73.3%) were surgical trainees. All participants reported prior use of AI for clinical documentation.

All participants rated their satisfaction of AI-written and conventional operative notes on a five point Likeart scale (very dissatisfied to very satisfied). For AI-generated notes, 28 (93.3%) of the participants reported being very satisfied, and for conventional notes, 24 (80%) reported their satisfaction. A statistically significant difference was observed between ratings of AI generated notes and conventional notes (p value 0.006). Table 3 demonstrates satisfaction of surgeons on clarity and quality of AI generated and conventional written operative notes.

Operative note format	Satisfaction level	Frequency (%)
AI generated operative note	Very satisfied	21 (70%)
	Satisfied	7 (23.3%)
	Neutral	2 (6.7%)
	Dissatisfied	0
	Very dissatisfied	0
Conventional operative note	Very satisfied	11 (36.7%)
	Satisfied	13 (43.3%)
	Neutral	6 (20%)
	Dissatisfied	0
	Very dissatisfied	0

Table 3 showing satisfaction level of surgical team with AI generated versus conventional operative notes

The majority of participants i.e., 23 (76.7%) reported no clinically inaccurate statements or assumptions in the AI generated operative notes whereas 7 (23.3%) identified inaccuracies. Of them, 2 (6.7%) surgeons preferred conventional notes, 1 (3.3%) preferred AI notes and 27 (90%) preferred both as means for clinical documentation.

A majority of respondents 29 (96.7%) stated that they would recommend AI generated operative notes for clinical documentation. AI generated notes were perceived as useful by 29 (96.7%) participants. Similarly, all participants considered AI generated notes to be a helpful tool for teaching correct operative note structure and content.

DISCUSSION

Operative notes are the sole account detailing all events occurring during a surgical procedure however, complex or delicate the procedure might be. (1) Well documented operative notes play a vital role in patient care, as they have a direct influence on their post-operative course and follow up care. Their importance in medico-legal litigation is also well known but despite of it, only 55% operative notes are found to be of use in the court of law. (11) While the style and language of an operative note is often a personal preference of a surgeon, some recommended information must be included in the note. The Royal College of Surgeons of England has therefore introduced comprehensive guidelines to write complete and concise operative notes for postoperative management of the patients as well as early recognition of complications. (12)

Despite the availability of more advanced alternatives, many resource limited settings have resorted to relying on hand-written operative notes for documenting surgical procedures. Electronic record systems with pre-formed templates have also surfaced but due to lack of intelligent features such as pattern recognition and development, these systems fail to incorporate the complexities encountered during surgery.

LLMs provide a globally accessible and easy-to-use option that allows AI to assist with clinical documentation. In this context, our experimentation with ChatGPT for operative note documentation was promising with a 100% guideline adherence to RCS recommendations when compared to human written note with guideline adherence of 59.3%. This result is also in accordance to other studies showing similar outcomes. (10, 13) Factual inaccuracies were noted in 27.11% of the cases in our study, which is quite significant and confers with the findings of another study on inaccuracies and fabrication found in medical related documentation obtained from ChatGPT. (14) Despite the need for editing of the operative notes produced by ChatGPT, it remained a faster method than traditional method for producing operative notes. It is noteworthy that commands

given for editing were incorporated and remembered when subsequent notes were produced, however this update was only limited to the same interaction, as the forum lacked the capability to reproduce the same update in a different interaction.

Despite its extensive training data, ChatGPT required iterative refinement on multiple occasions. For example, across multiple operative procedure description, Veress needle insufflation was included as a technique for establishing of pneumoperitoneum, despite the Hasson technique being more commonly used. In another instance, an operative note for Acute Cholecystitis mentioned routine drain placement, although this is not standard practice in all cases. Once this preference was clarified, subsequent notes appropriately omitted mention of drain placement unless specified. This highlights how prompt engineering with fine tuning of surgery-specific datasets could significantly enhance performance when used appropriately. Overall, these observations demonstrate that ChatGPT is most effective when used as an 'intelligent' assistant rather than an autonomous body.

Other instances where fine tuning of LLMs has been used include Chat Doctor LLM, which scrapes data in real time to present reliable answers to patients' medical related questions. (15) High accuracy and time efficiency have also been demonstrated in radiology reporting. (16)

If similar training of LLM for operative documentation is undertaken, with more accurately crafted prompts that include essential keywords, more precise surgical notes tailored to individual procedures and surgeons can be created.

It is imperative to note that this modality was accepted by the vast majority of the surgical team, as reflected in their feedback in which majority of the surgeons expressed their satisfaction with AI generated notes. This is contrary to findings reported by Abdelhady et al in 2023, where older patients and surgeons alike were reluctant toward this technology and failed to show optimism in its ability to produce optimal notes. (10)

The speed with which operative notes are generated might be variable depending on internet connectivity and the version of ChatGPT employed for the note creation however, the difference is barely of seconds which is still far less when compared to human written notes. Another limitation of using ChatGPT without full access to patient details is that it may lack sufficient context to generate a more precise document.

While our study highlights insights into the administrative burden that could be alleviated if LLMs are adopted by healthcare departments, it still lacks certain qualities to be widely accepted. Its adoption might not be applicable to all types of surgeries especially more complex specialized ones. Other than that, legal and ethical concerns surrounding use of AI such as patient privacy and protection of patient information is still a grey area. Our use of ChatGPT did not conflict with current General Data Protection Regulation and data governance rules because no specific patient information was added to the platform, but a cautious approach is required when giving access of patient data to these AI software causing a confidentiality risk and impact patient's trust in care providers negatively. (17) Furthermore, logistical challenges remain that may delay the integration of large language models into healthcare IT infrastructure.

CONCLUSION

LLM such as ChatGPT have the potential to serve as a useful adjunct for producing structured and comprehensive operative notes in a time efficient manner reducing the cognitive burden and errors associated with the fast paced surgical settings. Nevertheless, the quality, accuracy and reliability of AI generated surgical documentation are entirely contingent upon the detailed prompt and context specific input from the operating surgeons. It is still important to acknowledge the inability of AI systems to independently verify and adapt to unexpected encounters during surgery and clinical decision making. Surgeon oversight, confirmation and final authorship remains essential to ensure procedural accuracy, relevance and integrity of clinical documentation.

FUNDING

None.

CONFLICT OF INTEREST

There is no conflict of interest.

REFERENCES

1. Ghosh A. An audit of orthopedic operation notes: what are we missing? *Clin Audit*. 2010;2:37-40. doi:10.2147/CA.S9665
2. Mathew J, Baylis C, Saklani A, Al-Dabbagh A. Quality of operative notes in a district general hospital: a time for change? *Internet J Surg*. 2002;5:10.
3. Desiree RM, Kabir R. Wrist fracture management and the role of surgical care practitioner through the patient's journey. *J Perioper Pract*. 2022;32:115–22.
4. Mathioudakis A, Rousalova I, Gagnat AA, et al. How to keep good clinical records. *Breathe (Sheff)*. 2016;12:369–73.

5. Ma GW, Pooni A, Forbes SS, et al. Quality of inguinal hernia operative reports: room for improvement. *Can J Surg*. 2013;56:393.
6. Singh R, Chauhan R, Anwar S. Improving the quality of general surgical operation notes in accordance with the Royal College of Surgeons guidelines: a prospective completed audit loop. *J Eval Clin Pract*. 2012;18(3):578–80. doi:10.1111/j.1365-2753.2010.01626.
7. De Angelis L, Colaprico A, Carcagnì A, et al. ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health. *Front Public Health*. 2023;11:1180842.
8. Kasneci E, Seßler K, Küchemann S, Bannert M, Dementieva D, Fischer F, et al. ChatGPT for good? On opportunities and challenges of large language models for education. *Learn Individ Differ*. 2023;103:102274.
9. Webb M. A generative AI primer [Internet]. Bristol (UK): Jisc; 2023 [cited 2025 May 24]. Available from: https://nationalcentreforai.jiscinvolve.org/wp/2023/05/11/generative-ai-primer/
10. Abdelhady AM, Davis CR. Plastic surgery and artificial intelligence: how ChatGPT improved operation note accuracy, time, and education. *Mayo Clin Proc Digit Health*. 2023;1(3):299–308.
11. Lefter LP, Walker SR, Dewhurst F, Turner RWL. An audit of operative notes: Facts and ways to improve. ANZ J Surg. 2008; 78(9):800-2.
12. Cutting J, Hossain T, Maude K. Quality of operation note documentation in general surgical patients: Re-audit results. Int J Surg 2014; 12:S50.
13. Robinson A, Aggarwal S, Aggarwal Jr S. When precision meets penmanship: ChatGPT and surgery documentation. Cureus. 2023 Jun 17;15(6).
14. Bhattacharyya M, Miller VM, Bhattacharyya D, Miller LE, Miller V. High rates of fabricated and inaccurate references in ChatGPT-generated medical content. Cureus. 2023 May 19;15(5).
15. Li Y, Li Z, Zhang K, Dan R, Zhang Y. ChatDoctor: a medical chat model fine-tuned on LLaMA model using medical domain knowledge. Preprint. Published online March. 2023;24.
16. Ma C, Wu Z, Wang J, et al. ImpressionGPT: An Iterative Optimizing Framework for Radiology Report Summarization with ChatGPT. Preprint. Published online April 17, 2023.
17. Powles J, Hodson H. Google DeepMind and healthcare in an age of algorithms. Health Technol (Berl). 2017;7(4):351-367.