



Contents lists available at ScienceDirect

The Journal of Arthroplasty

journal homepage: www.arthroplastyjournal.org

Evaluating the Impact of Artificial Intelligence Scribes on Clinical Workflow and Documentation Quality: A Randomized Controlled Trial

Abdu Etagiuri, MD ^a, Daniel Chaudhry ^b, Jano van der Merwe ^b,
Michaela Nickol, MSc ^c, Johannes M. van der Merwe, MBChB, FRCSC ^{c,*}

^a College of Medicine, University of Saskatchewan, Saskatoon, SK, Canada

^b Division of Orthopaedics, Department of Surgery, University of Saskatchewan, Saskatoon, SK, Canada

^c Department of Orthopaedic Surgery, College of Medicine, University of Saskatchewan, Saskatoon, SK, Canada

ARTICLE INFO

Article history:

Received 14 November 2025

Received in revised form

23 March 2026

Accepted 25 March 2026

Available online xxx

Keywords:

artificial intelligence

scribe

arthroplasty

workflow

office efficiency

ABSTRACT

Background: The adoption of artificial intelligence (AI) scribes has grown rapidly in recent years, aiming to improve clinical workflow, increase efficiency, and reduce administrative burden by transcribing doctor–patient interactions. These systems capture in-office conversations and generate summaries using either standard or surgeon-specific templates. However, to date, no randomized controlled trials have rigorously assessed the proposed benefits of AI scribes. This randomized controlled trial focused on assessing AI scribe efficiency, error rate, accuracy, and patients' perceptions in comparison to traditional clinical dictation-transcription methods.

Methods: A randomized controlled trial was conducted involving 304 patients in a single-surgeon adult reconstruction practice. Patients were allocated into the following two groups: a conventional documentation group (using dictation transcribed by the office administrator or preconstructed templates: 140 patients) and an AI scribe group (139 patients). Timing data were recorded from when the surgeon entered a consultation room to when they began with the next patient. Additional data included administrative processing times, such as transcription and faxing of medical letters. All documentation was reviewed for medical errors by the surgeon in both groups. Grammar, spelling, and wrong wording were assessed by unblinded researchers. Letter quality was independently assessed using ChatGPT, focusing on grammar, clarity, and completeness.

Results: Use of AI scribes resulted in a statistically significant reduction in both physician and administrative processing times ($P < 0.0001$). There was no significant difference in spelling or grammatical errors between groups. However, letters produced using AI scribes were rated as significantly more complete, professional, and clear. There were 19 medical errors in the AI scribe group compared to only nine errors in the conventional group. Patients were not negatively affected by using an AI scribe, and the vast majority found it helpful.

Conclusions: The use of AI scribes can effectively enhance clinical efficiency and reduce administrative workload without compromising documentation quality or patient safety. Their implementation offers meaningful benefits to both physicians and office staff, allowing more focus on patient care and other critical tasks.

© 2026 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to <https://doi.org/10.1016/j.arth.2026.03.069>.

* Address correspondence to: Johannes M. van der Merwe, MBChB, FRCSC, Division of Orthopaedics, Department of Surgery, University of Saskatchewan, 514 Queen Street, Saskatoon, Saskatchewan S7K 3J6, Canada.

<https://doi.org/10.1016/j.arth.2026.03.069>

0883-5403/© 2026 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Physicians devote a large portion of their day dictating patient interactions and focusing on other administrative tasks, for example, dictating referral letters and generating tasks for their office administrators. Despite this, many physicians drastically underestimate the actual time they spend on such activities. A comparison of electronic health record (EHR) log data with self-reported surveys revealed that clinicians spent substantially

more time on documentation, averaging 13.6 minutes more per patient and 5.63 hours of after-hours work, than what the physicians reported [1].

The increase in paperwork and transcription of doctor–patient interactions is a factor contributing to decreased efficiency, clinical workflow, and an increase in physician burnout. It is reported that 74.5% of physicians who reported symptoms of burnout identified the usage of EHRs as the main contributor [1]. Furthermore, among 1,792 physicians studied, 70% of physicians who use EHRs reported stress related to administrative work. Alongside this, those who reported spending a moderately high or excessive amount of time using electronic records at home were 1.9 times more likely to burn out than those who did not use them [2].

Burnout has also been linked to an increase in administrative tasks and has clear and measurable consequences for patient care quality. High levels of burnout have been associated with increased medical errors and poorer clinical outcomes [3]. Notably, malpractice suits were strongly associated with physicians who have reported burnout, depression, or suicidal ideations [4]. In another study, it was shown that burned-out physicians included a differential diagnosis in 31% of their patient encounters, compared to 73% among non–burned-out peers, showcasing a clear impact on diagnostic thoroughness [5]. Quality of care is inversely related to burnout, and administrative stress remains a critical component to this relationship. Though clinical documentation is an essential component for medical practice, the current administrative burden on physicians represents a clear systematic inefficiency.

To date, physicians have addressed administrative workload issues by introducing dictating software, hiring transcription services, and creating specific templates that meet their patients' needs. Although such methods may enhance efficiency, they remain limited by the need for administrative review, manual editing, and interruptions between patient encounters. They are also limited by the fact that to make notes detailed, there naturally must be more time spent on each patient's note. Physicians who provided higher details in their notes, whether through typing, copying, and pasting, or using a template, all noted higher burden rates and lower efficiency overall [6].

In recent years, there has been a growing interest in the use of artificial intelligence (AI) scribes. Through capturing real-time conversations with patients, these AI systems can automatically generate relevant clinical notes in seconds [7]. This technology aims to allow physicians to see more patients, concentrate more on clinical care, reduce administrative workload, and combat burnout [8]. Among 968 physicians who have incorporated AI scribes in their clinical practice, many have noted reduced work hours and more meaningful interactions with their patients. This is also advantageous for patients, as they described improved interactions with their physicians through the usage of AI [7]. In terms of the quality of AI-generated notes, Kernberg et al. [9] concluded that further revisions of medical notes need to be taken by physicians and/or the office administrators before completely relying on the AI scribe. The majority of errors were omissions, accounting for 71% of all identified errors [10]. Even when minor revisions are required, the initial first draft created by the AI scribe can still greatly reduce documentation time, allowing physicians to identify any errors [11].

While the benefits of AI are well known, there have yet to be any randomized controlled trials that rigorously assessed AI scribes' impact on clinical accuracy and efficiency. Introducing this technology will have drastic impacts on health care by reducing administrative tasks and burnout, and increasing clinic efficiency. We have therefore conducted this research to compare the proposed benefits of AI scribe with evidence collected directly from a clinical setting. This randomized controlled trial focused on

assessing AI scribe efficiency, error rate, accuracy, and patients' perceptions in comparison to traditional clinical dictation–transcription methods. The null hypothesis stated that using the AI scribe will not reduce the time for note completion compared to the conventional manner. Secondary objectives included assessment of quality metrics of the medical documentation and patient satisfaction with using an AI scribe.

Methods

Study Design

After obtaining ethics (BIO 5718 NER) and registering the trial on [ClinicalTrials.gov](https://clinicaltrials.gov) (NCT07489469), we conducted a randomized controlled trial from May 2025 to July 2025. There were 304 eligible patients during the study period. There were 13 patients who declined to participate in the study. Of the remaining 291 enrolled patients, 12 were excluded due to missing data, leaving 279 randomized equally to AI scribe ($n = 139$) and conventional groups ($n = 140$) (Figure 1). Baseline characteristics, including age, body mass index (BMI), sex, Charlson Comorbidity Index, diagnosis (e.g., total knee arthroplasty, total hip arthroplasty, meniscal injury, osteoarthritis of the hip or knee, fracture follow-up), and visit complexity (e.g., follow-up visit, new consult, complex consult [e.g., total hip arthroplasty or total knee arthroplasty revision referrals, multiple complaints]), were comparable between groups ($P > 0.05$) (Table 1).

The remaining 279 patients were randomized equally into a conventional documentation group (140 patients) or an AI scribe group (139 patients) (Heidi AI Medical Scribe, Heidi Health, New York [12]). Informed consent was obtained by a researcher prior to randomization. Patients were informed that an AI scribe might be used in the consultation room. Patients were blinded, seeing that a computer or cellphone was used for the AI scribe and the preconstructed templates, respectively. Patients were also informed that their blinded consultation letters were going to be uploaded to ChatGPT for assessment. All the data were collected in an office setting. It was a single surgeon, adult reconstruction practice, without any trainees, nurse practitioners, or physician assistants during the study period. Randomization of patients was performed using a randomization app (Randomizer). Before the surgeon entered the consultation room for each patient, randomization was performed by the research team, and patients were allocated to their respective group. The AI scribe platform requires a yearly subscription costing \$696 USD more than conventional methods. The research team was not blinded to group assignments for the primary objective (processing times) due to the nature of the study design, but was blinded for the secondary objectives (quality metrics).

Eligibility Criteria

Eligible participants included all new and follow-up patients attending the clinic during the length of the study period. All

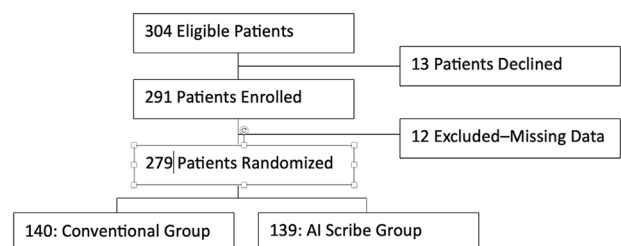


Figure 1. Flow diagram of enrollment. AI, artificial intelligence.

Table 1
Multivariable Regression Analysis Between AI Scribe and Conventional Group.

Variable	AI Scribe Group	Conventional Group	P-Value	Odds Ratio	95% CI
Age (mean; SD) in years	63.59 (14.3)	62.55 (13.69)	0.30	2.69	0.97–1.00
Sex (women)	98	90	0.37	2.22	0.48–1.31
BMI (mean, SD)	31.36 (6.80)	31.82 (7.44)	0.44	2.75	0.97–1.04
CCI (median, IQR)	0 (1.0)	0 (1.0)	0.49	2.87	0.90–1.23
Diagnosis	n/a	n/a	0.35	2.87	0.94–1.19
Consult complexity (follow-up visit, new consult, complex consult, multiple concerns)	62;81,0,9	56,79,1,17	0.05	4.04	0.99–1.96
Laterality (right)	55	53	>0.05	n/a	n/a

None of the Predictors Reached Conventional Significance ($P > 0.05$).

AI, artificial intelligence; BMI, body mass index; CCI, Charlson Comorbidity Index; IQR, interquartile range; n/a, not applicable.

patients were approached for the inclusion of the study regardless of diagnosis. The majority of cases were either related to primary total hip or knee arthroplasties. Other diagnoses such as meniscal injuries, fractures, complex arthroplasty, revision procedures, and various combinations were also included in this study. Exclusion criteria included phone appointments or patients declining to be a part of the study due to being uncomfortable with the usage of AI and being recorded in the office.

Clinical Documentation Methods

Patients who were assigned to the conventional documentation group are part of the senior author's usual practice. Preconstructed templates were filled in by the surgeon during the patient encounter in the consultation room, with a summary and management plan dictated thereafter, which was subsequently transcribed by the office administrator. The finalized transcribed letter was then emailed/faxed to the referring physician. The conventional documentation is part of an EHR (Accuro EMR; AccuroEMR, QHR Technologies Inc, Canada), with which the senior surgeon has more than 10 years of experience.

It was an easy implementation of an AI scribe into an existing practice. The preconstructed templates that were used in the EHRs were transferred to the AI-scribe documentation. The consultation (history and physical examination) was summarized and filled into the correct spots in the templates by the AI scribe. The surgeon talked through the physical examination so that it could be recorded. In addition, imaging and medical history were verbally reviewed before entering the consultation room. Following the patient encounter, the clinical note was automatically created. The generated letter was reviewed by the surgeon to verify data correctness, and any mistakes were immediately edited. The office administrator then reviewed the letter to identify grammar or spelling errors prior to emailing or faxing the finalized consultation letter. Because the letter was immediately available, it allowed the surgeon's office to supply a copy of the letter to the patient after the consultation, prior to the patient leaving the office.

Outcome Measures and Data Collection

The primary outcome for this study was the time (recorded in minutes) spent by the surgeon on each consultation letter (preconstructed templates or AI scribe) independent from location (surgeon's office or consultation room) until the letter is sent to the office administrator to be transcribed. The secondary outcomes included the time it took for the office administrator to finalize the consultation letter and to fax or email it to the referring physician and/or patient. Other outcomes assessed included any technical failures of the software, note completeness, accuracy, and identification of spelling, grammar, or wording errors. In the AI scribe group, the patient's comfort, perceived usefulness,

understanding of information, and satisfaction were also assessed through a study-specific questionnaire that occurred immediately after the consultation.

A sign was posted in front of the clinic that highlighted the use of AI in the clinic with patient encounters. The office administrator also communicated to patients that an AI scribe was going to be used, the reason for usage, and the ability to opt out if the patient chooses to. A senior author was present with the surgeon at all times, and timing data were recorded from when the surgeon started reviewing the radiographs and medical history until the end of the consultation. In the conventional group, the timing was stopped once the dictation was finished or after completion of the preconstructed template, and the letter was sent to the office administrator for transcribing. For the AI scribe group, timing was stopped immediately after the surgeon finished the patient encounter and reviewed the letter for correctness. Immediately after the office visit in the AI Scribe group alone, another senior author would enter the room and verbally survey the patient about their overall satisfaction and their comfort with being recorded during the patient encounter. Patients were asked to rate comfort with recording, perceived usefulness, understanding, and clarity, and overall satisfaction using a 5-point Likert scale ranging from very favorable (5 points) to very unfavorable (1 point).

Documentation quality was independently evaluated by three senior authors, identifying any spelling (e.g., coxa vara spelled as coxa vara), grammar (commas used incorrectly or wrong capitalization), or wording errors (e.g., not using Mrs. or Mr. in front of the patient's last name). Group allocation to the senior authors was revealed only after evaluations of the letters were documented in the dataset. We utilized ChatGPT to grade the consultation letters out of 100 (0 being the worst) for clarity, clinical detail, completeness, postoperative planning, and professional tone. ChatGPT was also asked to give an overall rating out of 100 for each letter. ChatGPT was used to rule out any bias that can occur from the research team evaluating the notes. Even with blinding the researchers, it is fairly easy to distinguish the letters between the conventional and treatment group. The surgeon then reviewed every letter in both the conventional and AI scribe groups, highlighting if any medical concerns (e.g., omissions of relevant patient data, incorrectness of patient encounter summaries, or treatment plans) were present in the medical notes. BMI, Charlson Comorbidity Index, and the complexity of the visit were all documented for adjustment in subsequent analyses.

Data Analyses

We described continuous measures using means $\pm$ SDs when they appeared symmetric and medians with interquartile ranges when skewed, and we reported categorical data as simple counts and percentages. To compare groups, we chose nonparametric tests—Mann–Whitney *U*-tests for continuous outcomes and

Table 2

With Effect Sizes of Approximately 0.58 and 0.66 Respectively, the Study Has Essentially 100% Power to Detect These Differences at the 0.05 Significance Level.

Metric	Conventional Group	AI Scribe Group	Mean	Mean	Cohen's d	Power
Surgeon-to-note completion time	151	153	13.9	11.76	0.577	0.999
Office staff processing time	143	150	4.71	3.46	0.662	1

Fisher exact tests for binary comparisons—because many of our variables were skewed or had small cell counts. Also, we fit a multivariable binomial logit model by maximum-likelihood estimation to estimate adjusted odds ratios for assignment to the AI scribe versus template group across patient age, BMI, visit type, diagnosis, sex, comorbidity index, and visit complexity. Power analysis was performed prior to initiating the study and after analysis and is summarized in Table 2. Preoperatively, we determined to have at least 100 patients per group to reach a statistical power of 0.80 with alpha at 0.05.

Results

We did find a statistically significant difference in favor of the AI group for the surgeon's note completion time (11.76 minutes $\pm$ 3.55) compared to the conventional group (13.90 minutes $\pm$ 3.86) ($P < 0.001$). This was a similar finding comparing the office staff's processing time (AI group 3.46 $\pm$ 1.37; conventional group 4.70 minutes $\pm$ 2.30) ($P < 0.001$) (Table 3 and Figure 2). Comparing the quality metrics (number of spelling or grammatical errors or wrong words used) between the two groups, we did not find a statistical significance (AI scribe group 122 errors; conventional group: 101 errors) (Table 4). When we assessed the quality of the notes between the two groups, we did find a statistical significance in favor of the AI scribe group (AI scribe group weighted mean 91.8%; conventional group weighted mean 89.15%) ($P < 0.001$) (Table 5). The senior author's assessment of the clinical notes for information accuracy revealed 19 errors in the AI scribe group compared to nine errors in the conventional group. This was not statistically significant ($P = 0.072$). We did not experience any technical failures with the AI scribe software during the study period (e.g., need to reset, no recording, recording lost, and so forth). The mean value for patient satisfaction, knowing that their office visit was recorded, was 3.6 of 5 (SD 1.95). Their overall satisfaction with the office visit using the AI scribe was 4.8 of 5 (SD 0.45).

Discussion

To the authors' knowledge, this is the largest randomized controlled trial to date comparing AI scribes to conventional methods. There are a wide variety of studies investigating the role of AI scribes in physician well-being, efficiency, and accuracy of documentation. Most of these studies are nonrandomized trials and suffer from small sample sizes [13–20]. The current study is thus very relevant and delivers important information that can

help clinicians safely adopt new technology and understand the benefits and shortcomings of new technology when integrating it into current practice.

The main findings of our study demonstrated a time savings effect using an AI scribe compared to conventional methods without compromising documentation quality. We thus successfully reject our null hypothesis. This time savings did not lead to more grammatical or spelling errors, and the AI-generated consultation letters were considered to be of better overall professional quality. We did not find a statistical significance in the number of medical errors between the AI scribe and the conventional group.

By using an AI scribe, the total processing time (surgeon and office administrator) from when the surgeon starts assessing the patient until the letter is emailed or ready to be mailed to the patient is on average three minutes and 39 seconds less than using the conventional methods (preconstructed templates or dictation and transcribing). The time saved using an AI scribe can be allocated to increase the number of consultations daily, subsequently reducing patient wait times to be seen, spending more time with patients having complex concerns, or simply having more free time. In addition, the administrative bottleneck can be avoided, freeing up the surgeon's administration so that they can focus on other important tasks, for example, returning phone calls, scheduling patients, and managing clinics. In addition, the instant production of consultation letters by AI scribe allows quicker communication with the referring physicians, further increasing physician collaboration. Consistent with these findings, prior studies have demonstrated major enhancements in documentation efficiency among physicians and administrative personnel, with time reductions ranging from 0.57 to 2.5 minutes per patient encounter [13,14,16,18–20].

Surgeon burnout is a work-related syndrome and is characterized by emotional exhaustion, depersonalization, and a reduced sense of personal accomplishment. Of the many causes of surgeon burnout, one is working long hours. Burnout can have negative consequences for both the surgeon and the associated health care system, leading to higher rates of medical errors, physician attrition, and a decrease in quality of patient care [21–24]. While time savings alone may not be sufficient to prevent burnout, improving workflow efficiency and reducing administrative burdens represent meaningful steps toward mitigating its impact and promoting physician well-being. This has been seen in multiple studies assessing physicians' well-being. Several studies have reported that the implementation of AI scribes leads to improved clinician satisfaction and a measurable reduction in physician burnout

Table 3

Processing Time for the Surgeon and Office Staff.

Group	Surgeons Note Completion Time	95% CI Lower	95% CI Upper	Office Staff Note Processing Time	95% CI Lower	95% CI Upper	P-Value
AI scribe	11.76 $\pm$ 3.55	11.2	12.33	3.46 $\pm$ 1.37	3.24	3.68	<0.001
Conventional group	13.90 $\pm$ 3.87	13.28	14.53	4.71 $\pm$ 2.3	4.33	5.09	<0.001

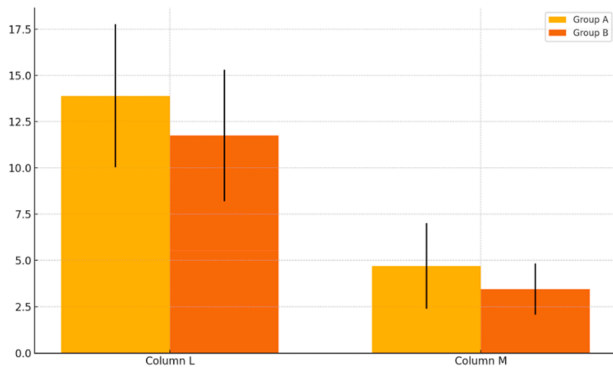


Figure 2. Box and Whisker plot. Group A (CG), Group B (AI). Column L = Surgeon's time to complete office visit; Column M = Office staff's time to complete the letter until it is emailed to the patient. Column L and M were both statistically significant ($P < 0.001$). AI, artificial intelligence; CG, convention group.

[15,18,20,25]. These findings suggest that automation of clinical documentation may alleviate some of the cognitive and administrative burdens contributing to professional exhaustion. In addition, other studies have demonstrated reductions in after-hours work, translating into enhanced work–life balance and overall physician well-being [14–16].

We have found a statistically significant improvement in documentation quality (Table 4). The documentation provided by the AI scribe was considered to be more complete, clear, and professional compared to conventional methods. Interestingly, the results vary in the literature, with certain studies noting concerns with note length and editing of documents [15]. Some AI scribes offer the ability to change the amount of detail in the consultation letters. While it lessens the burden of editing a longer letter, valuable information could be lost. From a medicolegal perspective, thorough documentation is critical in defending against malpractice claims. Courts often rely on the medical record as the most reliable account of events. Well-documented, timely, and accurate notes can demonstrate that the standard of care was met [26–28].

In our study, we did see spelling and/or grammatical errors (Table 3) in both the conventional and AI groups, which were not statistically significant. A cross-sectional observational study found a 7.4% error rate in their documentation that improved substantially with physician and transcriptionist review [17]. Even though these errors might not be clinically meaningful, it is important to have professional and high-quality documentation for patients in your practice.

More importantly is the accuracy of the medical information. The senior author found 19 errors in the AI scribe group compared to only nine errors in the conventional group. Even though it was found to be not statistically significant, it is clinically relevant. Wrong interpretations by the AI scribe from the patient encounter can alter diagnosis and treatment plans

in the patient's official documentation, which could lead to serious consequences for the patient and surgeon alike. It is thus imperative for the physician to review the AI documentation prior to distributing it to the referring physician and/or patient. Certain colleges or boards mandate physicians to review their documentation, and every physician planning to use an AI scribe in their practice should be aware of the recommendations by their respective college or board prior to instituting. There are two systematic reviews that noted concerns of the accuracy of information in the official physician documentation [18,19]. Even though these reviews consisted of nonrandomized observational studies, which were heterogeneous in design and quality, they represent the best available evidence up to date. The present randomized controlled trial therefore makes an important contribution to the existing body of literature by providing higher-level evidence on the impact of AI-assisted scribing in a real-world clinical setting. At present, the current evidence does not allow surgeons to trust AI scribes completely without reviewing the documentation before distributing.

While the vast majority of patients (96%) were satisfied with their overall office visit, the lower satisfaction rate (72%) regarding the recording of their visit likely reflects concerns about privacy, data security, or discomfort being recorded. Improving patients' comfort with office recordings can be achieved by providing clear, informed consent; digital resources; and opt-out options, which in turn can increase trust and autonomy. Educating patients about the technology's purpose, data security, and intended use increases comfort and willingness to consent [29–31]. The Health Insurance Portability and Accountability Act requires that any recordings of protected health information of patients fulfill certain criteria: patients should give their informed consent (written or verbal) before a recording is made, and this should be documented; physicians should consider individual state laws (e.g., some states require two-party or all-party consent); if recordings will be used for education or research, specific consent is required, and data should be deidentified; consent may not be needed if surveillance is necessary for patient security [32].

Potential Limitations

This study has limited external validity, as it was conducted with a single adult reconstruction surgeon in a single clinical setting. While this design restricts the generalizability of the findings, it also reduces interclinician variability, thereby minimizing differences in dictation style, phrasing, and template use that could otherwise confound the results. Future research should include multiple surgeons across diverse geographic locations and practice environments to enhance the applicability of the findings. Additionally, the study may have been influenced by the Hawthorne effect, wherein patients alter their behavior simply because

Table 4
Quality Measures Between the Two Groups Using the Mann-Whitney *U* Tests.

Quality Metrics	AI Scribe (Actual Number)	Conventional Group (Actual Number)	AI Scribe Group (Mean, SD)	Convention Group (Mean, SD)	<i>P</i> -Value	Mann- Whitney <i>U</i> Tests
Number of spelling errors	1	4	0.006 ± 0.081	0.027 ± 0.165	0.161	10,953.5
Number of grammatical errors	117	91	0.780 ± 1.104	0.636 ± 1.091	0.199	9,921.5
Wrong words/wrong content used	4	6	0.026 ± 0.198	0.041 ± 0.201	0.284	10,957.5

Table 5
Quality Measures Between the Two Groups Using the Mann-Whitney *U* Tests.

Quality Metrics	AI Scribe	Conventional Group	<i>P</i> -Value	Mann-Whitney <i>U</i> Tests
Overall rating	91.86 ± 3.58	89.21 ± 4.42	<0.001	6,745.0
Clarity	92.24 ± 3.68	88.47 ± 5.27	<0.001	5,915.5
Clinical detail	92.17 ± 4.55	89.89 ± 4.92	<0.001	7,411.5
Completeness	91.31 ± 4.30	88.80 ± 5.18	<0.001	7,196.5
Postoperative planning	90.43 ± 3.79	87.46 ± 4.54	<0.001	6,578.0
Professional tone	93.12 ± 3.59	91.08 ± 4.45	<0.001	7,827.5

Bold values indicate statistically significant.

they are aware of being recorded, rather than behaving naturally. This effect is expected to diminish over time as AI scribe technology becomes more commonplace and patients grow increasingly comfortable with its use. Clear communication regarding the purpose of recording and data handling can further mitigate this concern. Also, the unblinded nature of the study introduces a potential source of bias, as surgeons may have modified their behavior—such as conducting consultations more rapidly—when using the AI scribe system. The evaluation of medical notes may have been subject to observer bias, as the reviewing surgeon was aware of which notes were generated by the AI scribe and which were conventional. This lack of blinding could have influenced the surgeon's assessment of medical concerns, potentially affecting the objectivity of the comparison between groups.

Conclusions

AI scribes can effectively enhance clinical efficiency and reduce administrative workload without compromising documentation quality or patient safety. Their implementation offers meaningful benefits to both physicians and office staff, allowing more focus on patient care and other critical tasks. The authors believe that an AI scribe can safely be implemented in an office.

CRedit authorship contribution statement

Abdu Etagiuri: Writing – original draft, Project administration, Investigation, Data curation. **Daniel Chaudhry:** Writing – original draft, Investigation, Data curation. **Jano van der Merwe:** Writing – original draft, Investigation, Data curation. **Michaela Nickol:** Visualization, Supervision, Funding acquisition. **Johannes M. van der Merwe:** Writing – original draft, Visualization, Validation, Supervision, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

References

- [1] Tajirian T, Stergiopoulos V, Strudwick G, Sequeira L, Sanches M, Kemp J, et al. The influence of electronic health record use on Physician Burnout: Cross-Sectional Survey. *J Med Internet Res* 2020;22:e19274. <https://doi.org/10.2196/19274>.
- [2] Gardner RL, Cooper E, Haskell J, Harris DA, Poplau S, Kroth PJ, et al. Physician stress and burnout: the impact of health information technology. *J Am Med Inform Assoc* 2019;26:106–14. <https://doi.org/10.1093/jamia/ocy145>.
- [3] Williams ES, Manwell LB, Konrad TR, Linzer M. The relationship of organizational culture, stress, satisfaction, and burnout with physician-reported error and suboptimal patient care: results from the MEMO study. *Health Care Manage Rev* 2007;32:203–12. <https://doi.org/10.1097/01.HMR.0000281626.28363.59>.
- [4] Balch CM, Oreskovich MR, Dyrbye LN, Colaiano JM, Satele DV, Sloan JA, et al. Personal consequences of malpractice lawsuits on American surgeons. *J Am Coll Surg* 2011;213:657–67. <https://doi.org/10.1016/j.jamcollsurg.2011.08.005>.
- [5] Sullivan EE, Khazen M, Arabadjis SD, Mirica M, Ramos JM, Olson APJ, et al. Exploring relationships between physician stress, burnout, and diagnostic elements in clinician notes. *Diagnosis (Berl)* 2023;10:309–12. <https://doi.org/10.1515/dx-2022-0118>.
- [6] Apathy NC, Rotenstein L, Bates DW, Holmgren AJ. Documentation dynamics: note composition, burden, and physician efficiency. *Health Serv Res* 2023;58:674–85. <https://doi.org/10.1111/1475-6773.14097>.
- [7] Tierney AA, Gayre G, Hoberman B, Mattern B, Balesca M, Kipnis P, et al. [Internet]. *NEJM Catalyst*. Available at: <https://catalyst.nejm.org/doi/full/10.1056/CAT.23.0404>. [Accessed 1 October 2025].
- [8] Shah SJ, Devon-Sand A, Ma SP, Jeong Y, Crowell T, Smith M, et al. Ambient artificial intelligence scribes: physician burnout and perspectives on usability and documentation burden. *J Am Med Inform Assoc* 2025;32:375–80. <https://doi.org/10.1093/jamia/ocae295>.
- [9] Kernberg A, Gold JA, Mohan V. Using ChatGPT-4 to create structured medical notes from audio recordings of physician-patient encounters: comparative Study. *J Med Internet Res* 2024;26:e54419. <https://doi.org/10.2196/54419>.
- [10] Arko IV L, Hudelson C, Kumar J, Badlani S, Stoffel M, Markowitz R, et al. Documenting care with AI: a comparative analysis of commercial scribe tools. *Stud Health Technol Inform* 2025;329:337–41. <https://doi.org/10.3233/SHTI250857>.
- [11] van Buchem MM, Kant IMJ, King L, Kazmaier J, Steyerberg EW, Bauer MP. Impact of a digital scribe System on clinical documentation time and quality: usability Study. *JMIR AI* 2024;3:e60020. <https://doi.org/10.2196/60020>.
- [12] Heidi AI Scribe website. https://www.google.com/url?sa=i&source=web&rct=j&url=https://www.heidihealth.com/&ved=2ahUKEwjat8HQhrSRAxUvEDQIHQOzOtyQy_kOegQIBhAB&opi=89978449&cd&psig=AOvVaw1uRmGETN5YhpiBh0aBk2nx&ust=1765491536805000. [Accessed 28 April 2025].
- [13] Pearlman K, Wan W, Shah S, Laiteerapong N. Use of an AI scribe and electronic health record efficiency. *JAMA Netw Open* 2025;8:e2537000. <https://doi.org/10.1001/jamanetworkopen.2025.37000>.
- [14] Ma SP, Liang AS, Shah SJ, Smith M, Jeong Y, Devon-Sand A, et al. Ambient artificial intelligence scribes: utilization and impact on documentation time. *J Am Med Inform Assoc* 2025;32:381–5. <https://doi.org/10.1093/jamia/ocae304>.
- [15] Shah SJ, Crowell T, Jeong Y, Devon-Sand A, Smith M, Yang B, et al. Physician perspectives on ambient AI scribes. *JAMA Netw Open* 2025;8:e251904. <https://doi.org/10.1001/jamanetworkopen.2025.1904>.
- [16] Duggan MJ, Gervase J, Schoenbaum A, Hanson W, Howell 3rd JT, Sheinberg M, et al. Clinician experiences with ambient scribe technology to assist with documentation burden and efficiency. *JAMA Netw Open* 2025;8:e2460637. <https://doi.org/10.1001/jamanetworkopen.2024.60637>.
- [17] Zhou L, Blackley SV, Kowalski L, Doan R, Acker WW, Landman AB, et al. Analysis of errors in dictated clinical documents assisted by speech Recognition software and professional transcriptionists. *JAMA Netw Open* 2018;1:e180530. <https://doi.org/10.1001/jamanetworkopen.2018.0530>.
- [18] Hassan H, Zipursky AR, Rabbani N, You JG, Tse G, Orenstein E, et al. Clinical implementation of artificial intelligence scribes in health care: a systematic review. *Appl Clin Inform* 2025;16:1121–35. <https://doi.org/10.1055/a-2597-2017>.
- [19] Sasseville M, Yousefi F, Ouellet S, Naye F, Stefan T, Carnovale V, et al. The impact of AI scribes on streamlining clinical documentation: a systematic review. *Healthcare (Basel)* 2025;13:1447. <https://doi.org/10.3390/healthcare13121447>.
- [20] Sarraf B, Ghasempour A. Impact of artificial intelligence on electronic health record-related burnouts among healthcare professionals: systematic review. *Front Public Health* 2025;13:1628831. <https://doi.org/10.3389/fpubh.2025.1628831>.
- [21] Galaiya R, Kinross J, Arulampalam T. Factors associated with burnout syndrome in surgeons: a systematic review. *Ann R Coll Surg Engl* 2020;102:401–7. <https://doi.org/10.1308/rcsann.2020.0040>.
- [22] Williford ML, Scarlet S, Meyers MO, Luckett DJ, Fine JP, Goettler CE, et al. Multiple-Institution comparison of resident and faculty perceptions of burnout and depression during surgical training. *JAMA Surg* 2018;153:705–11. <https://doi.org/10.1001/jamasurg.2018.0974>.
- [23] Shanafelt TD, Balch CM, Bechamps GJ, Russell T, Dyrbye L, Satele D, et al. Burnout and career satisfaction among American surgeons. *Ann Surg* 2009;250:463–71. <https://doi.org/10.1097/SLA.0b013e3181ac4dfd>.
- [24] Hu YY, Ellis RJ, Hewitt DB, Yang AD, Cheung EO, Moskowitz JT, et al. Discrimination, abuse, harassment, and burnout in surgical residency training. *N Engl J Med* 2019;381:1741–52. <https://doi.org/10.1056/NEJMs1903759>.

- [25] Shin HS, Williams H, Braykov N, Jahan A, Meller J, Orenstein EW. The influence of artificial intelligence scribes on clinician experience and efficiency among pediatric subspecialists: a rapid, randomized quality improvement trial. *Appl Clin Inform* 2025;16:1041–52. <https://doi.org/10.1055/a-2657-8087>.
- [26] Lyons 3rd JM, Martinez JA, O'Leary JP. Medical malpractice matters: medical record M & Ms. *J Surg Educ* 2009;66:113–7. <https://doi.org/10.1016/j.jsurg.2008.12.002>.
- [27] Albano GD, Bertozzi G, Maglietta F, Montana A, Di Mizio G, Esposito M, et al. Medical records quality as prevention tool for healthcare-associated infections (HAIs) related litigation: a case series. *Curr Pharm Biotechnol* 2019;20:653–7. <https://doi.org/10.2174/1389201020666190408102221>.
- [28] Weintraub MI. Documentation and informed consent. *Neurol Clin* 1999;17: 371–81. [https://doi.org/10.1016/s0733-8619\(05\)70140-7](https://doi.org/10.1016/s0733-8619(05)70140-7).
- [29] Lawrence K, Kuram VS, Levine DL, Sharif S, Polet C, Malhotra K, et al. Informed consent for ambient documentation using generative AI in ambulatory care. *JAMA Netw Open* 2025;8:e2522400. <https://doi.org/10.1001/jamanetworkopen.2025.22400>.
- [30] Li W, Liu X. Anxiety about artificial intelligence from patient and doctor-physician. *Patient Educ Couns* 2025;133:108619. <https://doi.org/10.1016/j.pec.2024.108619>.
- [31] Ogundare O, Owadokun T, Ogundare T, Ekpo P, Nguyen HL, Bello S. Integrated artificial intelligence in healthcare and the patient's experience of care. *Sci Rep* 2025;15:21879. <https://doi.org/10.1038/s41598-025-07581-7>.
- [32] HIPAA - Patient Consent Requirements. <https://www.cloudtalk.io/blog/hipaa-call-recording-requirements/#:~:text=Under%20HIPAA%2C%20any%20voice%20recordings,What%20To%20Consider>. [Accessed 28 April 2025].