

COMMENTARY

Ambient Artificial Intelligence Scribes to Alleviate the Burden of Clinical Documentation

Aaron A. Tierney, PhD, Gregg Gayre, MD, Brian Hoberman, MD, MBA, Britt Mattern, MBA, Manuel Ballesca, MD, Patricia Kipnis, PhD, Vincent Liu, MD, MS, Kristine Lee, MD

Vol. 5 No. 3 | March 2024

DOI: 10.1056/CAT.23.0404

Clinical documentation in the electronic health record (EHR) has become increasingly burdensome for physicians and is a major driver of clinician burnout and dissatisfaction. Time dedicated to clerical activities and data entry during patient encounters also negatively affects the patient-physician relationship by hampering effective and empathetic communication and care. Ambient artificial intelligence (AI) scribes, which use machine learning applied to conversations to facilitate scribe-like capabilities in real time, has great potential to reduce documentation burden, enhance physician-patient encounters, and augment clinicians' capabilities. The technology leverages a smartphone microphone to transcribe encounters as they occur but does not retain audio recordings. To address the urgent and growing burden of data entry, in October 2023, The Permanente Medical Group (TPMG) enabled ambient AI technology for 10,000 physicians and staff to augment their clinical capabilities across diverse settings and specialties. The implementation process leveraged TPMG's extensive experience in large-scale technology instantiation and integration incorporating multiple training formats, *at-the-elbow* peer support, patient-facing materials, rapid-cycle upgrades with the technology vendor, and ongoing monitoring. In 10 weeks since implementation, the ambient AI tool has been used by 3,442 TPMG physicians to assist in as many as 303,266 patient encounters across a wide array of medical specialties and locations. In total, 968 physicians have enabled ambient AI scribes in ≥ 100 patient encounters, with one physician having enabled it to assist in 1,210 encounters. The response from physicians who have used the ambient AI scribe service has been favorable; they cite the

technology's capability to facilitate more personal, meaningful, and effective patient interactions and to reduce the burden of after-hours clerical work. In addition, early assessments of patient feedback have been positive, with some describing improved interaction with their physicians. Early evaluation metrics, based on an existing tool that evaluates the quality of human-generated scribe notes, find that ambient AI use produces high-quality clinical documentation for physicians' editing. Further statistical analyses after AI scribe implementation also find that usage is linked with reduced time spent in documentation and in the EHR. Ongoing enhancements of the technology are needed and are focused on direct EHR integration, improved capabilities for incorporating medical interpretation, and enhanced workflow personalization options for individual users. Despite this technology's early promise, careful and ongoing attention must be paid to ensure that the technology supports clinicians while also optimizing ambient AI scribe output for accuracy, relevance, and alignment in the physician-patient relationship.

KEY TAKEAWAYS

- » Ambient artificial intelligence (AI) scribes show early promise in reducing clinicians' burden, with a regional pilot noting a reduction in the amount of time spent constructing notes among users.
- » Ambient AI scribes were found to be acceptable among clinicians and patients, largely improving the experience of both parties, with some physicians noting the transformational nature of the technology on their care.
- » Although a review of 35 AI-generated transcripts resulted in an average score of 48 of 50 in 10 key domains, AI scribes are not a replacement for clinicians. They can produce inconsistencies that require physicians' review and editing to ensure that they remain aligned with the physician-patient relationship.
- » Given the incredible pace of change, building a dynamic evaluation framework is essential to assess the performance of AI scribes across domains including engagement, effectiveness, quality, and safety.

The Challenge

The burden of clinical documentation has become an enormous challenge for health care systems and produces clinician burnout, job dissatisfaction, and impaired physician-patient interactions, which are linked to potentially worsened quality of care.¹⁻⁵ The increasing

complexity of documentation in the electronic health record (EHR) and the competing demands on clinicians have also resulted in more “pajama time” — i.e., time in clerical activities outside working hours — for physicians.^{6,7} The threat of the documentation burden has grown so grave that some believe it even threatens the stability of the physician workforce.⁸

The Goal

To address this urgent challenge, The Permanente Medical Group (TPMG), a multidisciplinary physician group including >9,000 physicians working in the Kaiser Permanente Northern California integrated health care delivery system, executed a rapid regional pilot assessment of recently enabled ambient artificial intelligence (AI) scribe technology. The AI scribe was trained by, and acquired from, an external vendor and then implemented within TPMG. The ambient AI technology uses machine learning to produce real-time transcripts of clinician–patient encounters to rapidly convert speech collected from microphones on clinicians’ TPMG-provided secure smartphones into text and apply natural language processing techniques to summarize key clinical content to reduce the burden of clerical documentation and initial editing required by clinicians. Deploying the technology using smartphones is meant to enhance ease of use for clinicians and reflects current trends in AI scribe products. The AI scribe technology was not meant to replace human scribes, which are rarely used in TPMG, but meant to reduce clerical documentation burden.

The goal of this pilot was to identify optimal approaches that would facilitate the effective and safe use of ambient AI scribes to support clinicians’ workflows with the main objectives to:

- *facilitate engagement* by demonstrating growing and sustained adoption of ambient AI by number of clinicians and percentage of patient encounters across diverse specialties and settings;
- *aim for effectiveness* by reducing the burden of documentation within and outside of direct patient encounters;
- *enhance the physician–patient relationship* by increasing the amount of time physicians spend interacting with patients by improving engagement and reducing time spent interacting with a computer⁴; and
- *maintain documentation quality* by developing approaches to assess and safely use ambient AI technology capabilities in transcription and summarization.

The Execution

Setting and Licensing

TPMG had previously deployed software-based medical dictation technology in clinical settings to improve clinical documentation workflows and, by 2022, nearly 6,000 physicians had adopted

standard dictation using software assistants for their own clinical documentation. However, the emergence of ambient AI technology that could unobtrusively transcribe natural language from physicians and patients in real time during an encounter and summarize key clinical aspects heralded a transformative approach to jointly reducing the clerical burden of documentation and enhancing the physician-patient relationship through improved communication. To capitalize on this opportunity, on August 14, 2023, the organization initiated a limited, 2-week pilot with 47 physician test users; at the end of this trial period, physicians reported favorable impressions of the AI scribe's capabilities. Based on the limited pilot, TPMG secured and enabled product access licenses for 10,000 physicians and staff across diverse clinical specialties and settings to assess the value of the AI scribe technology in a larger regional pilot implementation. We describe TPMG's early experience in the regional pilot including data between October 16, 2023, and December 24, 2023.

“*Our teams trained clinicians to use The Permanente Medical Group-issued secure smartphones to record the encounter and generate summarized documentation that was automatically uploaded to and stored on a secure cloud.*”

Clinician Training and Support

Building on existing technology training infrastructure that includes regional and local physician and staff trainers, TPMG developed a live and recorded 1-hour virtual interactive webinar with questions and answers to familiarize all TPMG physicians with the technology; we deployed multiple training sessions with this webinar regionally and locally across our 21 medical centers in northern California. The training focused on the use of the ambient AI scribe including key practices for making the technology effective and safe. Because of the pace of implementation, we were not able to achieve direct integration within our EHR. Our teams trained clinicians to use their TPMG-issued secure smartphones to record the encounter and generate summarized documentation that was automatically uploaded to and stored on a secure cloud, accessible on their desktop computer through secure Web interfaces for incorporation and text editing into the EHR by the clinician.

During the training process, we also educated physicians on gaining permission from patients to use the AI medical scribe for in-person or virtual encounters and implemented a standardized EHR template to document patients' verbal consent for the use of the ambient AI scribe. To facilitate adoption, we leveraged internal support sites to address help tickets and provided at-the-elbow physician technology leads who could assist users facing challenges in installing or using the AI scribe. Finally, our training material reinforced the ongoing responsibility of physicians to review and revise all clinical documentation, whether or not it had been assisted by the AI scribe.

Patient-Facing Materials

To familiarize our patients with the use of the ambient AI scribe, we developed educational handouts for patients and made them available at each of the clinics and settings in which the technology was used. The handouts contained information that was verbally summarized by the physician to patients during the permission process, which took, in total, approximately 30 seconds. Additional cues, such as laminated posters, were placed in visible clinic locations to inform patients that physicians could use AI scribe technology in their visits with their permission.

Evaluation

Although the AI scribe showed early promise in the limited 2-week pilot in August 2023, we recognized that the underlying technology and concomitant workflow would continue to evolve rapidly, necessitating an ongoing adaptive assessment strategy to inform future Plan-Do-Study-Act cycles.⁹ For the *study* portion of the cycle, we considered a variety of adoption, effectiveness, and quality metrics to assess the value of the AI scribe technology both across the organization and within individual physician-patient interactions. For our larger, regional pilot evaluation, we focused on data from the first 10 weeks after implementation of the AI scribe between October 16 and December 24, 2023. In alignment with our overall objectives, we focused primarily on the tool's ability to assist in improving the quality of care and interaction, while also reducing clinician workloads arising from documentation. This includes patient surveys adapted from Mishra et al.⁴ analyzing how AI scribes affect encounters and patient comfort with AI tools being used to support their health care delivery. Early insights from these assessments were then intended to be used to inform future enhancements and investments in ambient AI scribe tools and workflows.

Quality and Safety Considerations

Given the novel technology used in ambient AI scribes, an objective of the implementation was to maintain the quality/accuracy of clinical documentation while also identifying, minimizing, or mitigating potential safety risks introduced by the use of the AI scribe technology.^{4,10} Prior studies of human medical scribes provided a framework for assessing the documentation quality, and we assessed samples of transcripts and clinical summaries across an array of clinical specialties using a modified version (Table 1) of the Physician Documentation Quality Instrument (PDQI-9).¹¹

To adapt the PDQI-9 for AI scribes, we removed attribute 1, the *up-to-date* domain (used to score whether the note contains the most recent test results and recommendations) and instead added two attributes to assess freedom from *hallucinations* (false information provided without sound basis) and *bias* (biased results based on use of discriminatory data, algorithms, or faulty heuristics), because these have been noted to potentially occur in outputs from large language models.^{12,13} We retained the 5-point Likert scale scoring from the original PDQI-9 instrument (with 1 representing the worst score of being *not at all* true or present and the best score, 5, being *extremely* true or present), but with our 10 attributes we rated notes on a scale with a maximum value of 50.

Table 1. Description of the Modified PDQI-9 Scribe Quality Assessment Tool (10 Domains)

Attribute	Description of Ideal Note
Accurate	The note is true. It is free of incorrect information.
Thorough	The note is complete and free from omission and documents all of the issues of importance to the patient.
Useful	The note is extremely relevant, providing valuable information and/or analysis.
Organized	The note is well-formed and structured in a way that helps the reader understand the patient's clinical course.
Comprehensible	The note is clear, without ambiguity or sections that are difficult to understand.
Succinct	The note is brief, to the point, and without redundancy.
Synthesized	The note reflects the AI scribe's understanding of the patient's status and ability to develop a plan of care.
Internally Consistent	No part of the note ignores or contradicts any other part.
Free from Hallucination	The note is free of hallucination and only contains information verifiable by the transcript.
Free from Bias	The note is free of bias and contains only information verifiable by the transcript and not derived from characteristics of the patient or visit.

Based on prior literature, we assessed samples of transcripts and clinical summaries across an array of clinical specialties using a modified version of the Physician Documentation Quality Instrument (PDQI-9). To adapt for use on AI scribes, we removed the *up-to-date* domain and added those assessing freedom from *hallucinations* and *bias*, because these have been noted to potentially occur in outputs from large language models. We retained the 5-point Likert scale scoring from the original instrument (with 1 being *not at all* and 5 being *extremely*) and rated notes on a scale with a maximum value of 50. Source: The authors

The Team

The team responsible for overseeing the implementation of the ambient AI scribe included seven TPMG regional technology leaders and staff who oversaw the limited, 2-week pilot in August 2023, supported the contracting process involving the vendor of the technology product, and developed the training and support infrastructure for the ambient AI scribe. Three lead physician champions helped to support the regional pilot deployment in October 2023 with collaboration from regional teams for innovation and evaluation, ultimately adding 21 physician champions who then recruited 200 champions located within individual medical centers and specialty groups.

“*Our training material reinforced the ongoing responsibility of physicians to review and revise all clinical documentation, whether or not it had been assisted by the AI scribe.*”

Metrics

Engagement

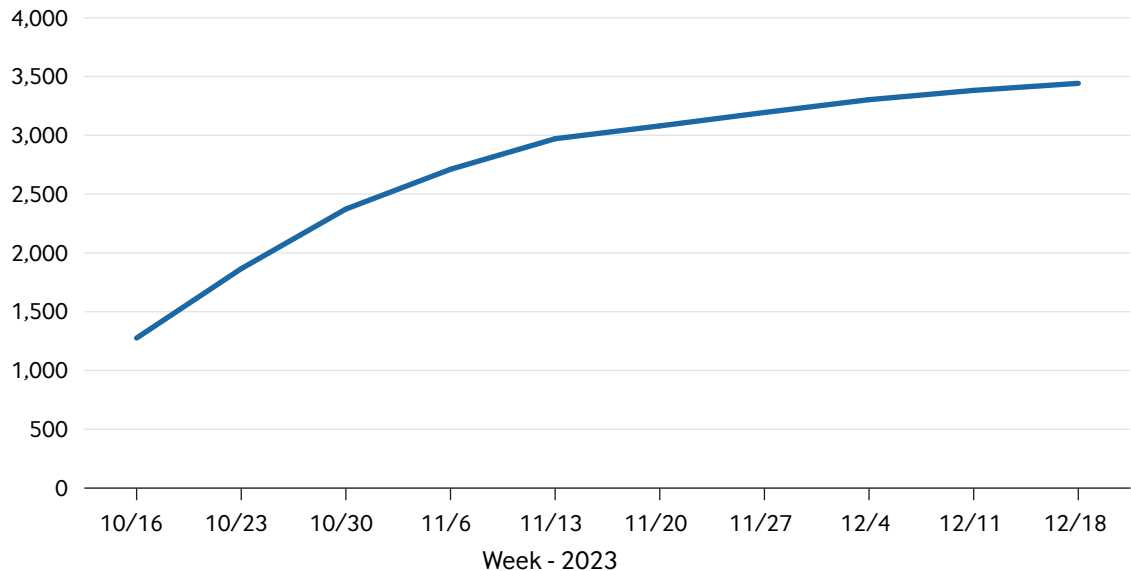
In the 10 weeks since implementation of the regional pilot (October 16–December 24, 2023), of the 10,000 physicians and staff provided access to accounts, 3,442 TPMG physicians enabled the tool in a total of 303,266 encounters across a wide array of medical specialties and locations (Figure 1).

FIGURE 1

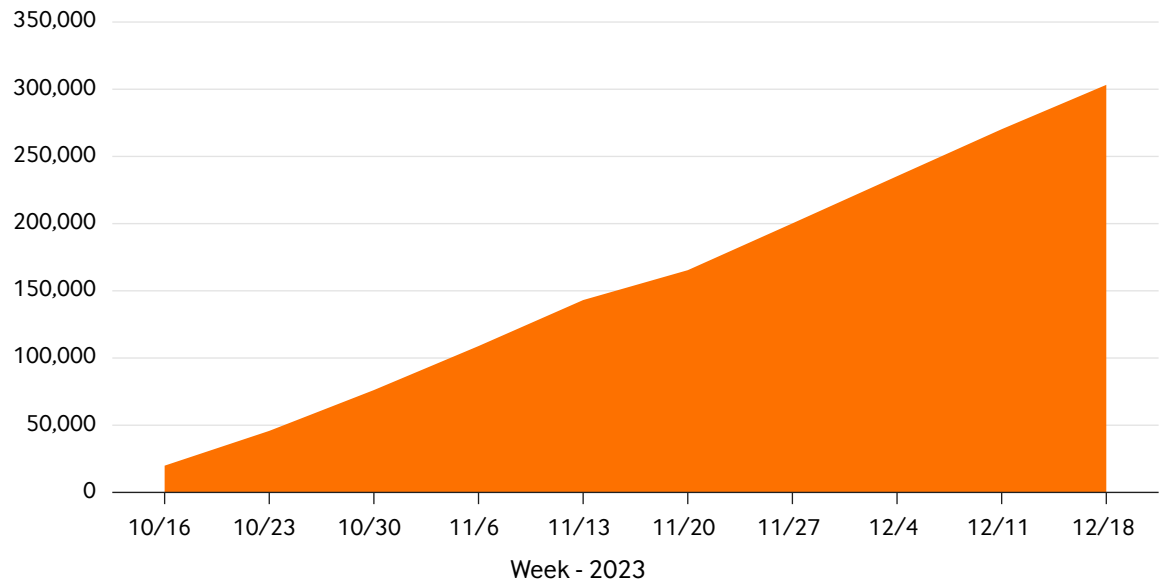
Cumulative Use of the Ambient Artificial Intelligence (AI) Scribe Tool, October 16–December 24, 2023

Between go-live on October 16, 2023, and December 24, 2023, there were 3,442 unique physician and staff users (Panel A) and a total of 303,266 patient–physician encounters in which the AI scribe was enabled and in which the encounter lasted at least 2 minutes (Panel B).

Panel A. Unique Physicians Ever Using AI Scribe



Panel B. Cumulative AI Scribe Visits



AI = artificial intelligence.

Source: The authors

NEJM Catalyst (catalyst.nejm.org) © Massachusetts Medical Society

Through December 24, 968 physicians had enabled ambient AI in ≥ 100 patient encounters, with one physician having enabled it to assist in 1,210 encounters. The number of weekly visits in which the AI scribe was enabled was 19,911 at the outset and exceeded 30,000 in 7 of the 10 weeks.

Effectiveness

To assess the effectiveness of the technology, we identified several aggregate EHR workload metrics (e.g., “pajama time,” time outside working hours, time spent in notes per chart) to assess changes before and after AI use among AI scribe users, compared with nonusers, and adjusted for clinician, specialty, and appointment volume (i.e., difference-in-differences analysis). Focusing on primary care physicians, the group that would be most likely to benefit from the technology, we found statistically significant ($P < 0.05$) associations between the use of the ambient AI scribe with larger decreases in the time spent in the EHR outside 7 a.m. and 7 p.m. and in the time spent in notes during appointments among AI scribe users (Figure 2). We focused on relative changes because the heterogeneity in the EHR workload metrics made interpretation of the absolute values challenging. As an example, in unadjusted analyses of time in notes we found decreases, before and after implementation, in both AI scribe users (mean, 5.3 to 4.8 minutes) and nonusers (5.0 to 4.7 minutes).

Our analysis also showed that there were no differences in other EHR workload metrics, such as time spent in clinical review or in-basket, which we would not expect to change as a result of AI scribe use. Intriguingly, when we stratified primary care providers by their usage rates (i.e., high, medium, and low tertiles), the data suggested a dose-response relationship whereby physicians with higher use also appeared to exhibit the most time reductions.

Clinicians’ and Patients’ Experience

The anecdotal response from physicians who have engaged regularly with the ambient AI scribe has been favorable, with comments such as “It makes the visit so much more enjoyable because now you can talk more with the patient and concentrate on their concerns.” Anecdotally, we also heard from clinicians that the AI scribe was particularly useful for very long appointments (>60 minutes) including one in which a 50-page transcript could easily be condensed and edited by the clinician to document their interaction. A new physician user noted, “I use it for every visit I can and it is making my notes more concise and my visits better. I know I’m gushing, but this has been the biggest game changer for me.” Those who used the AI scribe also commented on its capability to facilitate more personal, meaningful, and effective patient interactions. The same physician said, “I even had a patient praise the fact that I could listen instead of type during the visit.”

Early assessments of patient feedback have also been positive, with a small proportion of patients opting out of ambient AI use and many reporting improved interaction with their physicians. Based on preliminary data in a sample of 21 patient surveys from a single clinic site, 71% reported they spent more time speaking with their physician, while one said they spent

less time. Overall, 81% of patients reported that their physician spent less time looking at the computer screen than in their previous visits. All patients stated that the AI scribe either had no effect or enhanced their visit. All patients reported feeling *neutral to very comfortable* about an AI tool being used in their visit. Evaluations of patients through surveys and focus groups are ongoing to facilitate future improvements in the use and integration of the technology.

FIGURE 2

Ambient Artificial Intelligence (AI) Scribe Intervention: Before-and-After Metrics

(Panel A) We reviewed aggregate EHR metrics from before and after the use of the ambient AI scribe, comparing users versus nonusers with adjusted difference-in-difference (DiD) methods. The four activities were for time spent during “pajama time,” on unscheduled days, at times outside scheduled hours, and at times outside of 7 a.m. to 7 p.m. DiD point estimates are shown along with 95% confidence intervals, where values above the dotted line indicate larger relative decreases in time among AI scribe users than among nonusers (there was a universal decrease during this time period for all clinicians). For example, there is a significantly larger decrease in time spent in the EHR outside 7 a.m. to 7 p.m. among users starting to use the ambient AI scribe than among those who did not use the tool. (Panel B) We reviewed four visit-level EHR metrics per appointment, including time in clinical review, time in in-basket, time in notes, and time in orders. There was a statistically significant reduction in time in notes per appointment, comparing users (before and after) versus nonusers (before and after) with DiD. Unsurprisingly, there were no differences noted in other metrics unrelated to visit documentation.

Panel A. Primary Care Physician Time Spent in the EHR-Related Activities

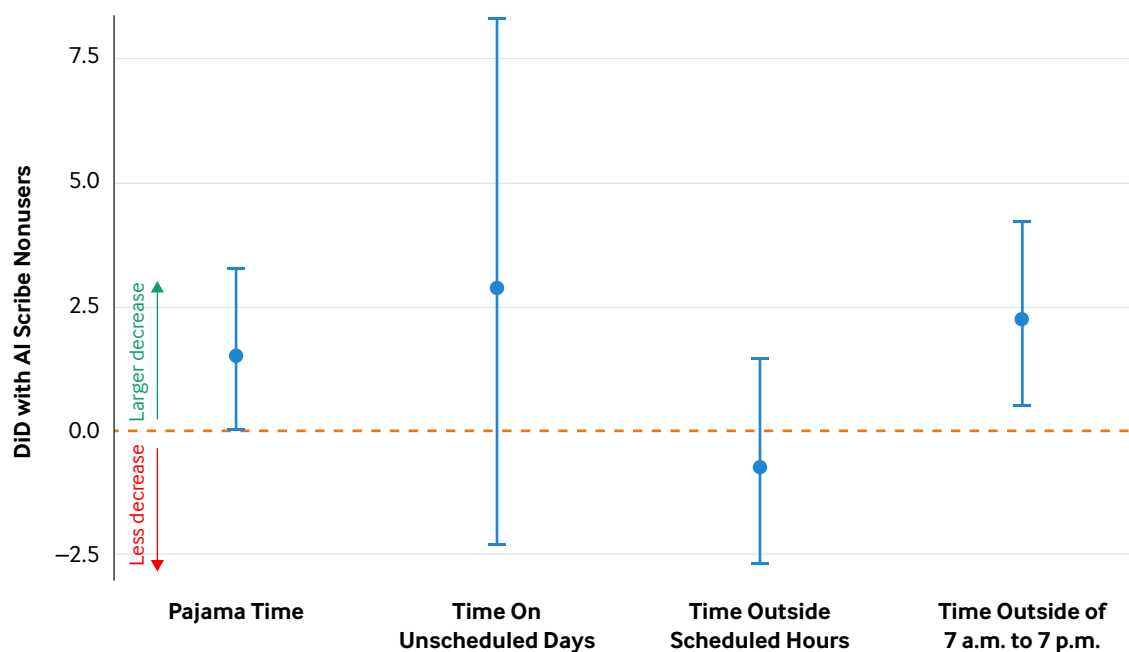
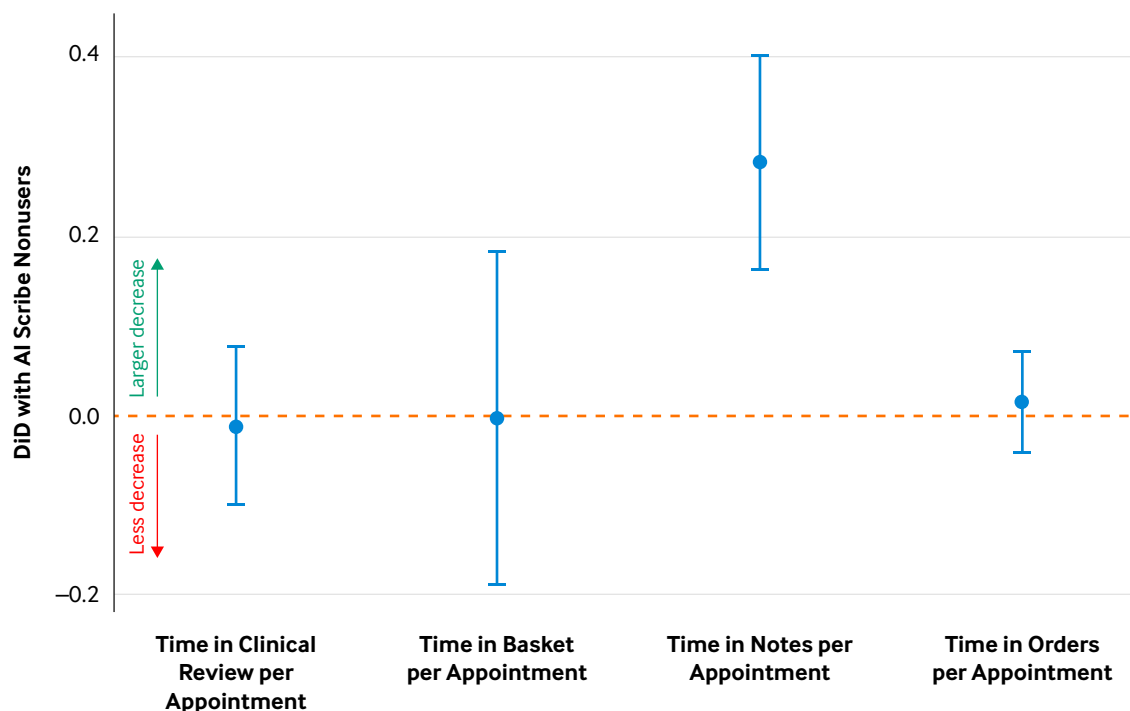


FIGURE 2

Ambient Artificial Intelligence (AI) Scribe Intervention: Before-and-After Metrics: Continued

Panel B. Primary Care Physician Time Spent in Appointment-Related Activities



Note: In each panel, a coefficient greater than 0 means that users of the ambient AI scribe were more likely (aligned with the coefficient, whether 0.1 or 2.5, etc.) to see a larger decrease in time spent than those who never used the AI scribe. A coefficient less than 0 means that users of the AI scribe were more likely to see a smaller decrease in time spent than those who never used the AI scribe. The zero line represents where there was no difference in the shift of time spent between AI scribe users and nonusers.

AI = artificial intelligence, EHR = electronic health record.

Source: The authors

NEJM Catalyst (catalyst.nejm.org) © Massachusetts Medical Society

“

Although the AI scribe showed early promise in the limited 2-week pilot in August 2023, we recognized that the underlying technology and concomitant workflow would continue to evolve rapidly.”

Quality

To assess the quality of the AI scribe, we examined transcripts and encounter summaries generated by the technology, including those from primary care, pediatric, hospitalist, mental health, surgical, and ED clinicians across three different medical centers using the modified

PDQI-9 instrument. Each summary was rated based on the transcript of the encounter. In total, we assessed 35 examples selected to be representative of each specialty for common themes until saturation was achieved. Overall, the transcripts averaged a score of 48 out of a possible 50 points with the highest average ratings (average value, >4.95) achieved in the *free from bias*, *synthesized*, *internally consistent*, and *succinct* domains (Figure 3). Modestly lower, but still considered *excellent*, average ratings (average, 4.6 to 4.7) were noted in the *thorough*, *organized*, and *accurate* domains.

There were a few instances of hallucination noted. In one example, the physician mentioned scheduling a prostate examination for the patient and the AI scribe summarized that a prostate examination had been performed. In another, the physician mentioned issues with the patient's hands, feet, and mouth and the AI summary recalled the patient being diagnosed with hand, foot, and mouth disease. There were also a few instances where the summary was missing some details, such as missing chest pain and anxiety assessments. In some cases where performance was lower, the summarized clinical content did not appear to fit cleanly within preexisting note templates, which resulted in inconsistencies in summarization.

Hurdles

We have encountered several hurdles during implementation that still need to be addressed. First, the current AI scribe technology works only for encounters spoken entirely in English. Although the software has the potential for support of different languages, regulations about certified medical interpretation services limit the use of the ambient AI scribe in these settings, as of early January 2024. Other factors may also limit patients' comfort with their physicians using the tool and will need to be better understood and addressed.

“*A major goal is to directly integrate artificial intelligence scribe tools into the electronic health record so that barriers to ease of access and use are eliminated and the documentation can be more seamlessly blended into existing workflows.*”

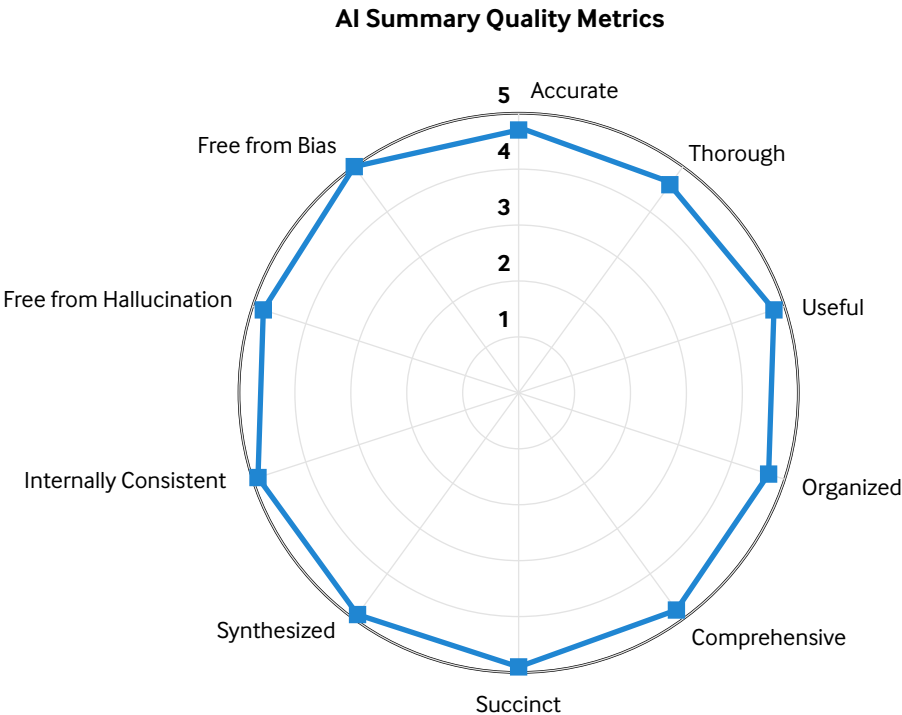
Second, in an early, focused survey of 45 physicians at a single site, half of whom were not using the AI scribe at the time despite having access, the most common reasons for not using the tool were typical of those seen with the diffusion of new technology¹⁴ (i.e., too many steps to activate the tool, a lack of familiarity with the tool, initial technical difficulties with accessing the tool, other workflow solutions not integrated with the tool). A major goal is to directly integrate AI scribe tools into the EHR so that barriers to ease of access and use are eliminated and the documentation can be more seamlessly blended into existing workflows.

Due to this lack of direct integration with the EHR, our evaluation approach continues to be refined to facilitate direct linkages between AI scribe transcripts and summaries with EHR data from individual encounters. We were unable to verify directly what edits and what proportion of

FIGURE 3

Radar Plot of Modified PDQI-9 (10 Domains) Based on Physicians’ Review of Ambient Artificial Intelligence (AI) Scribe Transcripts and Notes

Our analysis of the quality of the ambient AI-generated summaries of the visit transcripts showed high levels of performance across all 10 of our metrics on a subsample of 35 randomly assessed notes across multiple clinical specialties. Our regional pilot phase suggests that using AI scribes resulted in high-quality — but not perfect — notes generated by the technology.



AI = artificial intelligence.
Source: The authors
NEJM Catalyst (catalyst.nejm.org) © Massachusetts Medical Society

note revisions to the AI-generated summary occurred. Although we identified some significant findings in EHR workload metrics among AI scribe users, these findings should be interpreted as exploratory given uncertainty and heterogeneity in the metrics, which are collected passively during EHR use. Nonetheless, because our goal was to reduce clinical documentation burden, not to reduce the duration of physician–patient visits, our findings related to the quality of documentation and of the clinical interaction remain key, which will be enhanced by further integration. Future enhancements could also allow clinicians to customize their AI scribe for their documentation style, which is not currently available.

Third, the approaches to robustly evaluate the quality and safety of AI technologies, including tools such as large language models, remain incompletely defined. The underlying algorithms

and relevant regulations are also continuing to evolve rapidly, which will necessitate ongoing benchmarking, evaluation, and monitoring as the technology improves and vendors bring new software to market. Adoption rates and usage patterns are also expected to change as new user groups and application domains are identified and tested.

Where to Start

For health care organizations looking to implement AI-based medical scribe solutions, we found that it was possible to do a rapid, large-scale regional pilot implementation that aims to reduce the clerical burden of documentation, maintain the quality of notes, and improve the physician-patient experience. It is essential to have clinical champions to overcome barriers to adoption and to foster an organizational culture that supports innovation, while also iterating carefully with a limited pilot followed by a regional or larger-scale pilot and opportunities for clinician and patient feedback that result in ongoing improvement that is tangible to stakeholders. Finally, as is the case with all emerging technologies, it is important to develop benchmarking and monitoring processes that offer proactive assessment of the tools and their impact on meaningful outcomes.

Aaron A. Tierney, PhD

Clinical Informatics Delivery Science Postdoctoral Research Fellow, The Permanente Medical Group, Kaiser Permanente Northern California Division of Research, Oakland, California, USA

Gregg Gayre, MD

Chief Technology Officer, The Permanente Medical Group, Oakland, California, USA

Brian Hoberman, MD, MBA

Executive Vice President, Information Technology and Chief Information Officer, The Permanente Federation, Oakland, California, USA

Chief Information Officer, The Permanente Medical Group, Oakland, California, USA

Britt Mattern, MBA

Managing Director, Kaiser Permanente HealthConnect, The Permanente Medical Group, Oakland, California, USA

Manuel Balleca, MD

Internal Medicine Physician, The Permanente Medical Group, Oakland, California, USA

Patricia Kipnis, PhD

Principal Statistician, The Permanente Medical Group, Kaiser Permanente Northern California Division of Research, Oakland, California, USA

Vincent Liu, MD, MS

Regional Director of Hospital Advanced Analytics and Research Scientist and Intensivist, The Permanente Medical Group, Kaiser Permanente Northern California Division of Research, Oakland, California, USA

Kristine Lee, MD

Associate Executive Director of Virtual Medicine, The Permanente Medical Group, Oakland, California, USA

Internal Medicine Physician, San Francisco Medical Center, San Francisco, California, USA

Disclosures: Aaron A. Tierney, Gregg Gayre, Brian Hoberman, Britt Mattern, Manuel Balleca, Patricia Kipnis, Vincent Liu, and Kristine Lee have nothing to disclose.

References

1. Centers for Medicare & Medicaid Services. CMS Unveils Major Organizational Change to Reduce Provider and Clinician Burden and Improve Patient Outcomes. Centers for Medicare & Medicaid Services. June 23, 2020. Accessed January 4, 2024. <https://www.cms.gov/newsroom/press-releases/cms-unveils-major-organizational-change-reduce-provider-and-clinician-burden-and-improve-patient>.
2. Cheung K, Rogoza C, Chung AD, Kwan BYM. Analyzing the administrative burden of competency based medical education. *Can Assoc Radiol J* 2022;73:299-304 <https://journals.sagepub.com/doi/10.1177/08465371211038963> <https://doi.org/10.1177/08465371211038963>.
3. “Reducing Clinician Burden” Project. Health Level Seven (HL7) Electronic Health Record Work Group (EHR WG). January 16, 2023. Accessed January 4, 2024. <https://confluence.hl7.org/pages/viewpage.action?pageId=204280676&preview=%2F204280676%2F204280705%2FReducing+Clinician+Burden-Overview-20230116.pdf>.
4. Mishra P, Kiang JC, Grant RW. Association of medical scribes in primary care with physician workflow and patient experience. *JAMA Intern Med* 2018;178:1467-72 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6248201/> <https://doi.org/10.1001/jamainternmed.2018.3956>.
5. Tajirian T, Stergiopoulos V, Strudwick G, et al. The influence of electronic health record use on physician burnout: cross-sectional survey. *J Med Internet Res* 2020;22:e19274 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7392132/> <https://doi.org/10.2196/19274>.
6. Kane L. ‘Death by 1000 Cuts’: Medscape National Physician Burnout & Suicide Report 2021. Medscape. January 22, 2021. Accessed January 5, 2024. <https://www.staging.medscape.com/slideshow/2021-lifestyle-burnout-6013456>.
7. Saag HS, Shah K, Jones SA, Testa PA, Horwitz LI. Pajama time: working after work in the electronic health record. *J Gen Intern Med* 2019;34:1695-6 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6712097/> <https://doi.org/10.1007/s11606-019-05055-x>.
8. Manzione L. Clinician Turnover and the EHR Experience. Klas Research. April 8, 2022. Accessed December 8, 2023. <https://klasresearch.com/archcollaborative/report/clinician-turnover-and-the-ehr-experience/418>.

9. Nicolay CR, Purkayastha S, Greenhalgh A, et al. Systematic review of the application of quality improvement methodologies from the manufacturing industry to surgical healthcare. *Br J Surg* 2012; 99:324-35 <https://academic.oup.com/bjs/article/99/3/324/6138712> <https://doi.org/10.1002/bjs.7803>.
10. Medtech Pulse. The Promise and Potential Pitfalls of AI Medical Scribes. Medtech Pulse. April 18, 2023. Accessed December 11, 2023. <https://www.medtechpulse.com/article/insight/the-promise-and-potential-pitfalls-of-ai-medical-scribes>.
11. Stetson PD, Bakken S, Wrenn JO, Siegler EL. Assessing electronic note quality using the Physician Documentation Quality Instrument (PDQI-9). *Appl Clin Inform* 2012;3:164-74 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3347480/> <https://doi.org/10.4338/ACI-2011-11-RA-0070>.
12. Drake K. We Fact-Checked ChatGPT's Medical Advice. Healthnews. December 9, 2023. Accessed December 11, 2023. <https://healthnews.com/news/we-fact-checked-chatgpt-medical-advice/>.
13. Kotek H, Dockum R, Sun D. Gender bias and stereotypes in large language models. Proceedings of The ACM Collective Intelligence Conference. November 5, 2023. Accessed January 4, 2024. <https://dl.acm.org/doi/10.1145/3582269.3615599> <https://doi.org/10.1145/3582269.3615599>.
14. Rogers EM. Diffusion of innovations. 4th ed. New York: Free Press, 2010.