OXFORD

Research and Applications

A patient-centered digital scribe for automatic medical documentation

Jesse Wang (iD)[1]**, Marc Lavender**[2]**, Ehsan Hoque**[3]**, Patrick Brophy**[4] **and Henry Kautz**[3]

[1]Department of Translational Biomedical Science, University of Rochester Medical Center, Rochester, New York, USA, [2]Department of Medicine and Pediatrics, University of Rochester Medical Center, Rochester, New York, USA, [3]Department of Computer Science, University of Rochester, Rochester, New York, USA and [4]Department of Pediatrics, University of Rochester Medical Center, Rochester, New York, USA

Corresponding Author: Jesse Wang, Department of Translational Biomedical Science, University of Rochester Medical Center, 601 Elmwood Ave, Rochester, NY 14622, USA (jesse_wang@urmc.rochester.edu)

ABSTRACT

Objective: We developed a digital scribe for automatic medical documentation by utilizing elements of patient-centered communication. Excessive time spent on medical documentation may contribute to physician burnout. Patient-centered communication may improve patient satisfaction, reduce malpractice rates, and decrease diagnostic testing expenses. We demonstrate that patient-centered communication may allow providers to simultaneously talk to patients and efficiently document relevant information.

Materials and Methods: We utilized two elements of patient-centered communication to document patient history. One element was summarizing, which involved providers recapping information to confirm an accurate understanding of the patient. Another element was signposting, which involved providers using transition questions and statements to guide the conversation. We also utilized text classification to allow providers to simultaneously perform and document the physical exam. We conducted a proof-of-concept study by simulating patient encounters with two medical students.

Results: For history sections, the digital scribe was about 2.7 times faster than both typing and dictation. For physical exam sections, the digital scribe was about 2.17 times faster than typing and about 3.12 times faster than dictation. Results also suggested that providers required minimal training to use the digital scribe, and that they improved at using the system to document history sections.

Conclusion: Compared to typing and dictation, a patient-centered digital scribe may facilitate effective patient communication. It may also be more reliable compared to previous approaches that solely use machine learning. We conclude that a patient-centered digital scribe may be an effective tool for automatic medical documentation.

Key words: digital scribe, physician burnout, medical documentation, patient-centered communication

Lay Summary

Burnout is affecting physicians at epidemic levels, largely due to excessive time spent documenting patient visits. It has consequences for both physicians and patients. including increased risk of medical errors, malpractice lawsuits, and physician suicide. To help reduce physician burnout, we developed a digital scribe capable of automatically generating medical documentation by listening to physician–patient conversations. The system requires physicians use patient-centered communication, a style of interaction that minimizes miscommunication and improves patient satisfaction. We tested the system with two medical students, who took turns acting as the physician and the patient. Our study showed that documenting visits with the digital scribe was at least two times faster than typing and dictation. Both participants required minimal training to use the digital scribe and improved at using the system over the course of the study. By relying on patient-centered communication, our system allowed providers to simultaneously talk to patients and document information. Patient-centered communication also enabled our system to overcome the challenges of relying solely on machine learning. We conclude that the patient-centered digital scribe may be an effective tool for automatic medical documentation.

INTRODUCTION

Patient-centered communication

Patient-centered medical interviewing begins with the provider introducing everyone present and, in nonurgent situations, making small talk to build rapport.[1] The provider then elicits the patient's agenda with open-ended questions. After listing and prioritizing all the patient's concerns, the provider discusses each item in more detail. These discussions also begin with open-ended questions but may later be clarified with closed-ended questions. After discussing a topic, the provider summarizes the relevant information, then transitions to another topic by signposting. The provider repeats the process of summarizing and signposting to confirm and obtain important information.

King and Hoppe[2] reviewed multiple studies that support the use of patient-centered communication. One group of studies showed a positive association between patient-centered communication and patient satisfaction.[3–13] Another group of studies showed a positive association between patient-centered communication and patient recall, patient understanding, and patients' adherence to therapy.[7,8,14–22] Other studies showed that physicians with higher malpractice rates had twice as many patient complaints about communication, and physicians with poor communication scores on the Canadian medical licensing exam had higher malpractice claims; however, physicians with fewer malpractice claims encouraged patients to talk, checked patients' understanding, and solicited patients' opinions.[23–29] Epstein et al[30] demonstrated that patient-centered communication may be associated with lower diagnostic testing expenses but also increased visit times. Incorporating patient-centered communication into a digital scribe may mitigate longer visit times by reducing documentation times while facilitating effective provider–patient communication.

System overview

To illustrate how patient-centered communication can facilitate automatic medical documentation, imagine an adult male patient who is presenting to his primary care provider with complaints of chest pain. As the provider talks to the patient, she periodically recaps information to confirm her understanding and to allow the patient to correct any mistakes. For example, she says, "To recap, you've been having chest pain for about a month. It feels worse when you walk and climb the stairs. Is that right?" The digital scribe parses the summary, expands contractions, and inflects the second person to the

third person so that the note-ready summary reads, "He has been having chest pain for about a month. It feels worse when he walks and climbs the stairs." The digital scribe adds information to the history of present illness section by default. To document other history sections, the provider signposts with a section name. For example, to document family history, she asks, "Could you tell me about your family history?" Then, after collecting relevant information, she summarizes, and the digital scribe adds the note-ready summary to the family history section. Verbal cues for summaries and section names can be customized by the provider. The digital scribe only considers the provider's speech when generating notes. By summarizing and signposting, the provider can specify which information should be included in a note, where the information should be added, and how the information should be written. The digital scribe assumes that providers clearly structure their conversations to minimize the risk of miscommunication.

To document the physical exam, the provider verbalizes her findings while examining the patient. For example, during the cardiovascular portion, she says, "I'm going to listen to your heart. Breathe normally please. Normal S1 and S2. No murmurs, rubs, or gallops. Can you hold your breath? No carotid bruits. I'm glad you had a good vacation." The digital scribe parses the relevant exam findings by removing imperatives and questions. It also removes small talk using a classifier trained on medical notes and film scripts. This pipeline produces the following text for the physical exam: "Normal S1 and S2. No murmurs, rubs, or gallops. No carotid bruits." The provider specifies sections of the exam by using the "section" cue. For example, she says, "section pulmonary," so that the digital scribe adds subsequent findings to the pulmonary section. She begins and stops documenting the exam by saying, "begin exam" and "stop exam," respectively. This method for documenting the physical exam requires providers be considerate of patients' feelings when saying aloud potentially concerning findings. Verbalizing findings is a routine communication pattern between providers and human scribes.[31]

Physician burnout

Burnout is a syndrome characterized by emotional exhaustion, depersonalization, and a reduced sense of personal accomplishment.[32,33] The condition has increased to epidemic proportions among physicians, with a prevalence of about 50% in national studies.[34–38] It negatively impacts physicians' health, increasing the risk

of alcohol dependence and suicidal ideation.[39–41] It also presents potential consequences for patient care, including increased risk of medical errors, malpractice lawsuits, and post-discharge recovery times.[42–44] A significant contributor to physician burnout is excessive time spent documenting encounters, which decreases time spent interacting with patients.[45–48] In addition to being a health concern, physician burnout also has financial consequences. Han et al[49] estimated that burnout costs the healthcare industry $4.6 billion each year due to physician turnover and reduced clinical hours. The objective of a digital scribe is to decrease time spent on documentation, increase time spent interacting with patients, and decrease the prevalence of physician burnout.

Human scribes, individuals who chart medical encounters in real time, have reportedly enabled physicians to spend less time on documentation.[50–52] However, human scribes present several challenges: costly salaries, substantial up-front investment for training, and a fast turnover rate due to many scribes pursuing full-time medical studies.[53–55] Other considerations include interpersonal compatibility with physicians and patient comfort when human scribes are present during sensitive discussions.[56,57] In addition to being present in-person during medical encounters, human scribes may be present virtually from an offsite location.[55]

Our patient-centered digital scribe aims to alleviate the documentation burden without the drawbacks of an in-person or virtual human scribe. Because providers can verbalize summaries and physical exam findings in their own styles of writing, our digital scribe does not need to be retrained for different providers. It also avoids problems with turnover and interpersonal compatibility by functioning without a human intermediary. In addition, patients may feel more comfortable without the presence of an additional person.

Related works

Related works focus on training machine learning models to structure, parse, and convert provider-patient conversations. Park et al[58] trained models on 279 encounters to classify talk-turn segments, achieving 67.42–78.37% F-score. Example topics included patient history, physical exam, and small talk. Rajkomar et al[59] trained a model on 2547 encounters for extracting symptoms, achieving 73.6% F-score. Du et al[60] trained models on 2950 encounters for extracting symptoms, achieving 50–80% F-score depending on the model and medical condition. However, they also reported low inter-labeler agreement between scribes due to ambiguous and informal discussions of symptoms. In a separate paper, Du et al trained models on the same dataset for extracting relations with symptoms and medications. They achieved 34–57% F-score depending on the entity and relation.[61] Selvaraj and Konam[62] trained models on 6692 encounters for extracting medication dosages and frequencies, achieving 89.57% and 45.94% F-scores, respectively. Shafran et al[63] trained models on about 6000 encounters for extracting medications at 90% F-score, symptoms at 72% F-score, and conditions at 57% F-score. However, inconsistent and incorrect annotations complicated the calculation and interpretation of these scores. Enarvi et al[64] trained transformer models on approximately 800 000 orthopedic conversations and notes. Depending on the note section, they achieved 19.2–65.4% improvement in ROUGE-L score compared to baseline; however, they did not report the performance of their baseline model. Note repetitions also decreased the quality of their generated notes. Despite significant progress in automatically interpreting provider–patient conversations, to the best of our knowledge, there have been no published evaluations of an end-to-end system in either real or simulated patient encounters.

Several challenges may affect the performance of machine learning for a digital scribe.[65] One challenge is the highly nuanced and potentially ambiguous nature of provider–patient conversations.[66] This characteristic complicates the annotation process for a large corpus of medical encounters. It also complicates the review process for notes generated with sequence-to-sequence learning. Another challenge is the limited availability of medical transcripts and notes due to strict privacy protections.[60,67] Manually deidentifying a large corpus of protected health information is time-consuming and expensive.[68] A third challenge is the absence of a reliable metric for note quality. Different providers, specialties, and practices may have different expectations for how notes should be written. Our digital scribe side-steps these issues by using patient-centered communication to write the history sections of a note and verbalized findings to write the physical exam sections.

The main advantage of our digital scribe is that it does not require annotated transcripts or notes for training. It uses patient-centered communication, specifically summarizing and signposting, to allow providers to simultaneously talk to patients and document history. Moreover, summarizing information is already a routine part of healthcare workflows. For example, medical students in the United States are required to orally summarize information during their medical licensing exam, and providers orally summarize information when presenting patient cases.[69,70] Our digital scribe also uses publicly available, unannotated data sets to distinguish medically relevant sentences from small talk. This model allows providers to simultaneously perform and document the physical exam by verbalizing observations, a routine aspect of human scribe workflows.[31] In addition, unlike regular dictation, our system allows providers to focus on talking to patients. By using existing communication patterns, we demonstrate an approach for a digital scribe that avoids the challenges of relying solely on machine learning.

MATERIALS AND METHODS

System architecture

Audio recordings of simulated patient encounters were captured with an earset microphone. They were transcribed, diarized, and punctuated using an automatic speech-to-text service. Transcript preprocessing involved identifying providers by counting verbal cues, converting numbers to written form, and removing certain disfluencies (Supplementary Tables A1 and A2). Certain medical homophones were also replaced. History summaries and physical exam findings were parsed from the providers' preprocessed transcripts and synthesized into notes. To help providers efficiently edit

Table 1: Completion speed for each note section and writing method

Section	Method	WPM	
		$\bar{x}$	s
History	Scribe	207.37	39.01
	Typing	75.88	11.14
	Dictation	76.41	10.95
Physical Exam	Scribe	110.16	24.77
	Typing	50.75	8.54
	Dictation	35.22	6.47

notes, an interface was developed to automatically format edited punctuations.

Audio capture

Audio was captured using the Countryman H6 earset microphone and transmitted in the 584–608 MHz frequency range using a Shure BLX1 Wireless Bodypack Transmitter paired to a Shure BLX4 Single Channel Receiver. The receiver was plugged into an Aspire E 15 laptop running Windows 10 with a Tisino XLR-to-3.5 mm adapter cable and a headset splitter cable. Audio was recorded at a sample rate of 48 kHz in the Waveform Audio File Format (WAV) using the Recorder.js plugin. Transmitter batteries were checked for sufficient charge before each recording using a D-FantiX Battery Tester. Both the laptop and transmitter were adjusted for maximum gain. The RealTek HD Audio Manager was used to boost gain by an additional 20 dB. The microphone was fitted to sit at the corner of the provider's mouth, where it was able to capture both the provider's and the patient's speech.

Automatic speech recognition

The Google Speech-to-Text video model was used to transcribe provider–patient conversations, diarize speakers, and automatically add punctuations. Selection of this service was motivated by a study that compared the Google Assistant to Siri and Alexa, which showed that the Google Assistant was more accurate for both branded and generic medication names. It also showed that the Google Assistant was equally accurate for multiple accents.[71] These findings suggested that Google Speech-to-Text would be suitable for transcribing provider–patient conversations.

History documentation

Providers summarized information to document the history sections of a note. The verbal cues "recap" and "summarize" indicated the start of a summary. The verbal cues "is that right" and "is that correct" indicated the end of a summary. Text between the start and stop cues were processed by replacing contractions and inflecting second person to third person (Supplementary Tables A3–A5). If a stop cue was not mentioned, then the entire text between the start cue and the end of the provider's talk-turn segment was processed.

String matching without context was used to identify section names. For example, when a provider signposted with the question, "Any family history of heart disease," the phrase "family history" indicated that subsequent summaries should be added to the family history section. Similarly, when a provider signposted with the statement, "Now I'd like to ask about your social history," the phrase "social history" indicated that subsequent summaries should be added to the social history section. The following section names were recognized: "medical history," "surgical history," "family history," and "social history." Verbal cues for signposting and summarizing were customizable. For example, a provider could have used the phrase "personal habits" as a substitute for "social history." Summaries were added to the history of present illness section by default. To avoid manually editing misplaced information, providers were asked to only mention section names when signposting.

Physical exam documentation

Providers simultaneously performed and documented the physical exam by verbalizing findings. They used the verbal cues "begin exam" and "stop exam" to indicate when the exam was being performed. To isolate relevant information, a rule-based approach was used to remove questions and imperatives. A text classifier was used to remove small talk. To specify which part of the physical exam to document, providers used the "section" cue. For example, saying "section musculoskeletal" indicated that subsequent findings should be added to the musculoskeletal section. A list was used to identify multi-word section names immediately after the "section" cue. If a multi-word section name was not found, then the first word after the "section" cue was used as the section name.

Questions and imperatives are typically used to provide patient instructions. For example, the sentences "Could you take a deep breath," "Kick your leg out," and "Please turn your head" instruct the patient but do not convey relevant information for a note. Questions were identified by the presence of question marks automatically added by Google Speech-to-Text. To complement Google's automatic punctuation insertion, which may have performed poorly on physical exam findings due to uncommon vocabulary and syntax, periods were inserted for silent pauses of duration greater than 2 s. Providers were asked to pause at sentence boundaries when verbalizing findings. Imperatives were identified by the presence of an infinitive verb at the start of a sentence or by the presence of "please" at the start or end of a sentence.

In addition to removing questions and imperatives, small talk was removed using a text classifier trained on a corpus of unannotated medical notes and film scripts. A total of 2 083 180 medical notes were downloaded from the MIMIC-III project and 142 255 film scripts were web scraped from the *Springfield* website.[72,73] Sentence tokenization produced about 41 million sentences from medical notes and about 99 million sentences from film scripts. To reduce class imbalance, each medical sentence was duplicated at least once. About 22% of medical sentences were duplicated twice. With oversampling, a total of about 92 million medical sentences were collected. The data was split so that 80% of sentences were used for training, 10% for testing, and 10% for validation. Sentences were pre-processed to replace slashes, hyphens, and other nonalphanumeric characters with spaces; replace symbols and abbreviations for "positive," "negative," and "normal" with their full spellings; and insert spaces between digits in multidigit numbers.

Several classification models were trained using the Scikit-learn library, including naive Bayes, logistic regression, and support vector machine (SVM).[74] Logistic regression achieved 98.9% F-score for medical sentences, outperforming both naive Bayes at 97.13% F-score and SVM at 98.85% F-score. Because all medical sentences identified by the classifier were included in physical exam documentations, providers were asked to avoid discussing unrelated medical information, such as explaining exam findings, collecting additional history, and discussing treatment options. If providers needed to have these discussions, they could pause the exam with the "stop exam" cue and later resume with the "begin exam" cue. The use of a text classifier for identifying medical information avoided the need to use start and stop cues for each finding.

Note editor

To help providers efficiently edit notes, an interface was developed to automatically format edited periods and commas. For example, if a user inserted a period at the end of a word, then the first character of the next word was capitalized. If a user inserted a period at the beginning of a word, then the first character of the edited word was capitalized, and the period was moved to the end of the preceding word. Comma insertions were treated similarly except the first character of the next word was lowercased. The first character of the

next word was also lowercased when a user deleted a period. Removing unnecessary keystrokes for editing periods and commas aimed to reduce editing time if punctuations were inaccurately inserted by Google Speech-to-Text. The Google Chrome Web Speech API was used to allow providers to dictate notes. The editor automatically monitored the amount of time providers spent editing history sections and physical exam sections.

Study design

To evaluate the digital scribe, simulated encounters were conducted with two medical students at the University of Rochester Medical Center (URMC). Each participant took turns acting as a provider and as a standardized patient. They were trained to use the digital scribe with a 10-min presentation and a practice encounter. After training, each provider completed 32 encounters based on cases from *First Aid for the USMLE Step 2 CS*, a review book for the clinical skills portion of the United States Medical Licensing Exam.[69] Psychiatric and pediatric cases were excluded. Case information was available as a paper chart. Student providers obtained information for the history sections supported by the digital scribe, including history of present illness, past medical history, past surgical history, family history, and social history. They also performed and documented a focused physical exam. They were given a list of suggested history and physical exam elements to investigate based on the patient's chief complaint. After each encounter, student providers edited a digital scribe note by typing, including any information missed by the digital scribe. They then typed and dictated copies of the edited digital scribe note. They were not observed when interacting with patients or when editing notes. The times required to finish notes were captured automatically in the note editor. Participants were native English speakers between the ages of 25 and 34. They had experience typing notes for their clinical clerkships, but minimal experience dictating notes. They were already trained in patient-centered communication as part of their medical curriculum. The study was approved by the Institutional Review Board at URMC.

RESULTS

A total of 64 simulated patient encounters were conducted with two medical students. Note completion speed was measured in words per minute (WPM). An alpha level of 0.05 was used for statistical analyses. The digital scribe successfully generated notes using speaker diarization for all encounters, except one where the standardized patient spoke very softly to portray a patient with chest pain. The digital scribe note for this encounter was generated without speaker diarization and was included in study results.

One-way within subjects ANOVA indicated at least one significant difference between note completion speeds grouped by writing method and note section ($P < .0001$). Post hoc comparisons were performed using two-tailed pairwise t-tests for paired groups with Bonferroni correction (Table 1, Figure 1). For history sections, editing notes generated by the digital scribe ($\bar{x} = 207.37$, $s = 39.01$) was about 131.49 WPM faster than typing ($\bar{x} = 75.88$, $s = 11.14$, $P < .0001$) and about 130.96 WPM faster than dictation ($\bar{x} = 76.41$, $s = 10.95$, $P < .0001$). Typing was not significantly different than dictation ($P = 1.00$). For physical exam sections, the digital scribe ($\bar{x} = 110.16$, $s = 24.77$) was about 59.41 WPM faster than typing ($\bar{x} = 50.75$, $s = 8.54$, $P < .0001$) and about 74.94 WPM faster than dictation ($\bar{x} = 35.22$, $s = 6.47$, $P < .0001$). Typing was about 15.53 WPM faster than dictation ($P < .0001$). Mild outliers were observed for the digital scribe and dictation. Each writing method was faster for history sections compared to physical exam sections: the digital scribe was about 97.21 WPM faster ($P < .0001$), typing was about 25.13 WPM faster ($P < .0001$), and dictation was about 41.19 WPM faster ($P < .0001$).

Pearson correlation coefficients were calculated to assess the relationship between the number of completed encounters and note completion speed (Figure 2). For history sections, the digital scribe
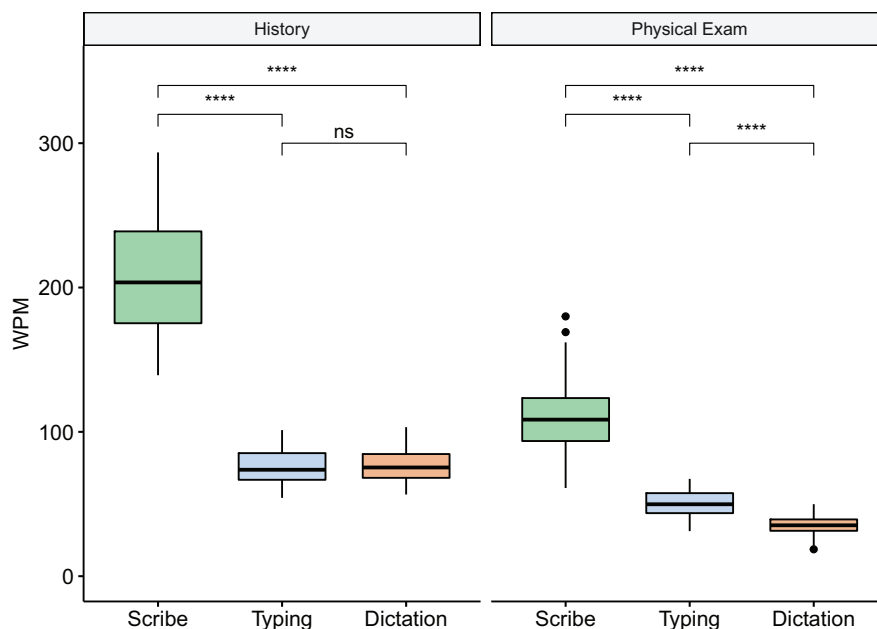

Figure 1: Comparison of completion speed for each note section and writing method.
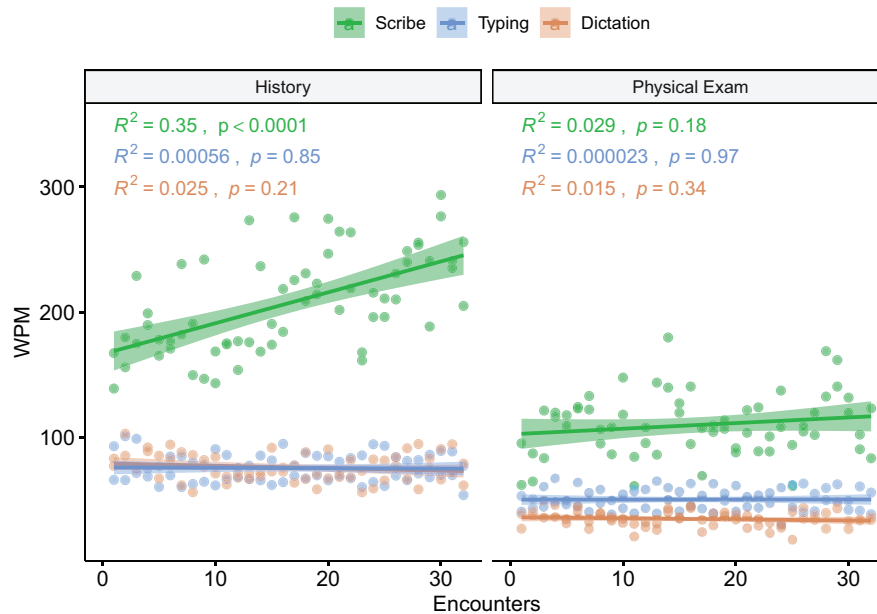
Figure 2: Completion speed for each note section and writing method as providers completed more encounters.

showed a positive correlation between the number of encounters and completion speed ($R^2 = 0.35$, $P < .0001$), but neither typing ($R^2 = 0.00056$, $P = .85$) nor dictation ($R^2 = 0.025$, $P = .21$) showed a correlation. For physical exam sections, no correlation was observed for the digital scribe ($R^2 = 0.029$, $P = .18$), typing ($R^2 = 0.000023$, $P = .97$), and dictation ($R^2 = 0.015$, $P = .34$).

DISCUSSION

Our study suggests that patient-centered communication techniques, such as summarizing and signposting, can be used to document patient history. For history sections, the digital scribe was about 2.7 times faster than both typing and dictation. Our study also suggests that providers can simultaneously perform and document the physical exam by verbalizing findings. For physical exam sections, the digital scribe was about 2.17 times faster than typing and about 3.12 times faster than dictation. The performance of the digital scribe may be underestimated in our study when compared to typing and dictation. Because typing and dictation speeds were measured by writing copies of edited digital scribe notes, they did not reflect the cognitive challenges of organizing and editing information. The digital scribe may be even more efficient than typing and dictation in busy clinical settings, where providers have limited time to write notes between seeing patients. By using the digital scribe to document information during patient encounters, providers may more easily recall encounter events when editing multiple notes. The performance of the digital scribe may be more accurately assessed in a real clinical setting.

In addition, our study suggests that providers may require minimal training to use the digital scribe, and that they may improve at using the system to document patient history. Study participants were trained with a short presentation and a practice session. After 32 patient encounters, study participants became significantly more efficient at documenting history sections. Given that the average family physician in the United States sees about 19 patients per day, this finding suggests that providers may improve at documenting

history in as few as 2 days.[75] Moreover, because the digital scribe does not rely solely on machine learning, providers can identify the reason for an unexpected result. For example, they can identify that a summary is missing because they did not say "recap" or "summarize." By using patient-centered communication, the digital scribe may facilitate both high-quality patient interaction and efficient medical documentation.

While the digital scribe was faster than typing and dictation in general, it was slower for physical exam sections compared to history sections. The difference may be due to increased typing errors for physical exam sections, as suggested by decreased typing speed. It may also be due to increased speech recognition errors, as suggested by decreased dictation speed. In addition, the difference may be due to increased errors with automatic punctuation insertion since physical exam findings tend to be described in phrases rather than complete sentences. Speech recognition and punctuation insertion models specifically trained for physical exam findings may improve the performance of the digital scribe for physical exam sections.

Note completion speed served as an overall metric for digital scribe performance. It reflected the accuracy of Google Speech-to-Text and text processing pipelines since participants edited missing or incorrect information. It also reflected participants' abilities to speak clearly, summarize information, and minimize disfluencies. In general, these elements performed sufficiently for the digital scribe to be faster than typing and dictation. A noticeable area for improvement is speaker diarization, which was disabled for one encounter where the standardized patient spoke very softly to simulate chest pain. This error suggests that providers and patients may be required to speak above a certain volume for adequate speaker diarization. Despite this error, assessing note completion speed was prioritized over the performance of individual elements. The digital scribe may be a useful application as long as it helps providers complete notes faster than typing and dictation.

Our study has several limitations. Simulated encounters with a small number of medical students may not capture the diversity of provider–patient interactions, such as psychiatric and pediatric encounters, or the types of information that providers must docu-

ment, such as medications, allergies, and the assessment and plan. Psychiatric and pediatric cases were excluded because they may be difficult for medical students to portray as standardized patients. Moreover, simulated patient interactions may not accurately portray real-life conversations. Medications and allergies were excluded because this information may be documented as a structured table rather than narrative text. The assessment and plan section was excluded because medical students may not have enough clinical experience to summarize this section while interacting with patients. In addition, because our study participants typed at an average speed of 75.88 WPM for history sections, much faster than the average typing speed of 34.11 WPM for physicians, our results may not generalize to providers with average or below-average typing speeds.[76] Future studies involving practicing physicians in real clinical settings may provide additional insights into our system.

CONCLUSION

Patient-centered communication helps provide high-quality patient care. By simulating patient encounters with a small group of medical students, we demonstrate that providers may use specific elements of patient-centered communication, such as summarizing and signposting, to simultaneously interact with patients and document the history sections of a note. We also demonstrate that providers may verbalize findings to simultaneously perform and document the physical exam. In addition, we demonstrate that providers may quickly improve with the digital scribe after minimal training. Because editing notes generated by the patient-centered digital scribe was significantly faster than typing and dictation, we conclude that the system may be an effective tool for automatic medical documentation.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

CONTRIBUTORS

J.W. conceived the idea for the patient-centered digital scribe, developed the source code, and analyzed the data. All authors were involved in designing the study and writing the paper.

FUNDING

DATA AVAILABILITY STATEMENT

Conversation recordings and transcripts cannot be shared publicly for the privacy and confidentiality of study participants.

REFERENCES

1. Hashim MJ. Patient-centered communication: basic skills. *Am Fam Physician* 2017; 95 (1): 29–34.
2. King A, Hoppe RB. Best practice for patient-centered communication: a narrative review. *J Grad Med Educ* 2013; 5 (3): 385–93.
3. Korsch BM, Gozzi EK, Francis V. Gaps in doctor-patient communication. 1. Doctor-patient interaction and patient satisfaction. *Pediatrics* 1968; 42 (5): 855–71.
4. Bertakis KD. The communication of information from physician to patient: a method for increasing patient retention and satisfaction. *J Fam Pract* 1977; 5 (2): 217–22.
5. Stiles WB, Putnam SM, Wolf MH, *et al.* Interaction exchange structure and patient satisfaction with medical interviews. *Med Care* 1979; 17 (6): 667–81.
6. Buller MK, Buller DB. Physicians' communication style and patient satisfaction. *J Health Soc Behav* 1987; 28 (4): 375–88.
7. Roter DL, Hall JA, Katz NR. Relations between physicians' behaviors and analogue patients' satisfaction, recall, and impressions. *Med Care* 1987; 25 (5): 437–51.
8. Stewart ME, Roter DE, eds. Communicating With Medical Patients. Newbury Park, CA: Sage Publications; 1989.
9. Rowland-Morin PA, Carroll JG. Verbal communication skills and patient satisfaction. A study of doctor-patient interviews. *Eval Health Prof* 1990; 13 (2): 168–85.
10. Wanzer MB, Booth-Butterfield M, Gruber K. Perceptions of health care providers' communication: relationships between patient-centered communication and satisfaction. *Health Commun* 2004; 16 (3): 363–83.
11. Brédart A, Bouleuc C, Dolbeault S. Doctor-patient communication and satisfaction with care in oncology. *Curr Opin Oncol* 2005; 17 (4): 351–4.
12. Roter DL, Frankel RM, Hall JA, *et al.* The expression of emotion through nonverbal behavior in medical visits. *J Gen Intern Med* 2006; 21 (S1): S28–34.
13. Tallman K, Janisse T, Frankel RM, *et al.* Communication practices of physicians with high patient-satisfaction ratings. *Perm J* 2007; 11: 19–29.
14. Francis V, Korsch BM, Morris MJ. Gaps in doctor-patient communication. Patients' response to medical advice. *N Engl J Med* 1969; 280 (10): 535–40.
15. Ley P, Bradshaw PW, Kincey JA, *et al.* Increasing patients' satisfaction with communications. *Br J Soc Clin Psychol* 1976; 15 (4): 403–13.
16. Garrity TF. Medical compliance and the clinician-patient relationship: a review. *Soc Sci Med E* 1981; 15 (3): 215–22.
17. Bartlett EE, Grayson M, Barker R, *et al.* The effects of physician communications skills on patient satisfaction; recall, and adherence. *J Chronic Dis* 1984; 37 (9-10): 755–64.
18. Tuckett D, Boulton M, Olson C, Williams A. *Meetings Between Experts: An Approach to Sharing Ideas in Medical Consultations*. London, England: Tavistock Publications; 1985.
19. Kjellgren KI, Ahlner J, Säljö R. Taking antihypertensive medication—controlling or co-operating with patients? *Int J Cardiol* 1995; 47 (3): 257–68.
20. Stewart MA. Effective physician-patient communication and health outcomes: a review. *CMAJ* 1995; 152 (9): 1423–33.
21. .Silk KJ, Westerman CK, Strom R, Andrews KR. The role of patient-centeredness in predicting compliance with mammogram recommendations: an analysis of the Health Information National Trends Survey. *Commun Res Rep.* 2008; 25 (2): 131–144
22. Zolnierek KBH, Dimatteo MR. Physician communication and patient adherence to treatment: a meta-analysis. *Med Care* 2009; 47 (8): 826–34.
23. Beckman HB. The doctor-patient relationship and malpractice. Lessons from plaintiff depositions. *Arch Internal Med* 1994; 154 (12): 1365–70.
24. Hickson GB. Obstetricians' prior malpractice experience and patients' satisfaction with care. *JAMA* 1994; 272 (20): 1583.
25. Levinson W, Roter DL, Mullooly JP, *et al.* Physician-patient communication. The relationship with malpractice claims among primary care physicians and surgeons. *JAMA* 1997; 277 (7): 553–9.
26. Ambady N, Laplante D, Nguyen T, *et al.* Surgeons' tone of voice: a clue to malpractice history. *Surgery* 2002; 132 (1): 5–9.
27. Whitlock EP, Orleans CT, Pender N, *et al.* Evaluating primary care behavioral counseling interventions: an evidence-based approach. *Am J Prev Med* 2002; 22 (4): 267–84.

28. Tamblyn R, Abrahamowicz M, Dauphinee D, *et al*. Physician scores on a national clinical skills examination as predictors of complaints to medical regulatory authorities. *JAMA* 2007; 298 (9): 993–1001.

29. Mazor KM, Roblin DW, Greene SM, *et al*. Toward patient-centered cancer care: patient perceptions of problematic events, impact, and response. *J Clin Oncol* 2012; 30 (15): 1784–90.

30. Epstein RM, Franks P, Shields CG, *et al*. Patient-centered communication and diagnostic testing. *Ann Fam Med* 2005; 3 (5): 415–21.

31. Miller N, Howley I, McGuire M. Five lessons for working with a scribe. *Fam Pract Manag* 2016; 23 (4): 23–7.

32. West CP, Dyrbye LN, Erwin PJ, *et al*. Interventions to prevent and reduce physician burnout: a systematic review and meta-analysis. *Lancet* 2016; 388 (10057): 2272–81.

33. Maslach C, Jackson SE, Leiter MP, *et al. Maslach Burnout Inventory*. Palo Alto, CA: Consulting Psychologists Press; 1986.

34. Dyrbye LN, Thomas MR, Massie FS, *et al*. Burnout and suicidal ideation among US medical students. *Ann Intern Med* 2008; 149 (5): 334–41.

35. West CP, Shanafelt TD, Kolars JC. Quality of life, burnout, educational debt, and medical knowledge among internal medicine residents. *JAMA* 2011; 306 (9): 952–60.

36. Shanafelt TD, Balch CM, Bechamps GJ, *et al*. Burnout and career satisfaction among American surgeons. *Ann Surg* 2009; 250 (3): 463–71.

37. Shanafelt TD, Boone S, Tan L, *et al*. Burnout and satisfaction with work-life balance among US physicians relative to the general US population. *Arch Intern Med* 2012; 172 (18): 1377–85.

38. Shanafelt TD, Hasan O, Dyrbye LN, *et al*. Changes in burnout and satisfaction with work-life balance in physicians and the general US working population between 2011 and 2014. *Mayo Clin Proc* 2015; 90 (12): 1600–13.

39. Oreskovich MR, Kaups KL, Balch CM, *et al*. Prevalence of alcohol use disorders among American surgeons. *Arch Surg* 2012; 147 (2): 168–74.

40. Shanafelt TD, Balch CM, Dyrbye L, *et al*. Special report: suicidal ideation among American surgeons. *Arch Surg* 2011; 146 (1): 54–62.

41. van der Heijden F, Dillingh G, Bakker A, *et al*. Suicidal thoughts among medical residents with burnout. *Arch Suicide Res* 2008; 12 (4): 344–6.

42. Shanafelt TD, Balch CM, Bechamps G, *et al*. Burnout and medical errors among American surgeons. *Ann Surg* 2010; 251 (6): 995–1000.

43. Balch CM, Oreskovich MR, Dyrbye LN, *et al*. Personal consequences of malpractice lawsuits on American surgeons. *J Am Coll Surg* 2011; 213 (5): 657–67.

44. Halbesleben JRB, Rathert C. Linking physician burnout and patient outcomes: exploring the dyadic relationship between physicians and patients. *Health Care Manage Rev* 2008; 33 (1): 29–39.

45. Shanafelt TD, Dyrbye LN, Sinsky C, *et al*. Relationship between clerical burden and characteristics of the electronic environment with physician burnout and professional satisfaction. *Mayo Clin Proc* 2016; 91 (7): 836–48.

46. Dyrbye LN, West CP, Burriss TC, *et al*. Providing primary care in the United States: the work no one sees. *Arch Intern Med* 2012; 172 (18): 1420–1.

47. Sinsky C, Colligan L, Li L, *et al*. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. *Ann Intern Med* 2016; 165 (11): 753–60.

48. Tai-Seale M, Olson CW, Li J, *et al*. Electronic health record logs indicate that physicians split time evenly between seeing patients and desktop medicine. *Health Aff* 2017; 36 (4): 655–62.

49. Han S, Shanafelt TD, Sinsky CA, *et al*. Estimating the attributable cost of physician burnout in the United States. *Ann Intern Med* 2019; 170 (11): 784–90.

50. Gellert GA, Ramirez R, Webster SL. The rise of the medical scribe industry: implications for the advancement of electronic health records. *JAMA* 2015; 313 (13): 1315–6.

51. Campbell LL, Case D, Crocker JE, *et al*. Using medical scribes in a physician practice. *J AHIMA* 2012; 83: 64–9.

52. Shultz CG, Holmstrom HL. The use of medical scribes in health care settings: a systematic review and future directions. *J Am Board Fam Med* 2015; 28 (3): 371–81.

53. Walker KJ, Dunlop W, Liew D, *et al*. An economic evaluation of the costs of training a medical scribe to work in Emergency Medicine. *Emerg Med J* 2016; 33 (12): 865–9.

54. Earls ST, Savageau JA, Begley S, *et al*. Can scribes boost FPs' efficiency and job satisfaction? *J Fam Pract* 2017; 66 (4): 206–14.

55. Brady K, Shariff A. Virtual medical scribes: making electronic medical records work for you. *J Med Pract Manage* 2013; 29 (2): 133–6.

56. Yan C, Rose S, Rothberg MB, *et al*. Physician, scribe, and patient perspectives on clinical scribes in primary care. *J Gen Intern Med* 2016; 31 (9): 990–5.

57. Dunlop W, Hegarty L, Staples M, *et al*. Medical scribes have no impact on the patient experience of an emergency department. *Emerg Med Austral* 2018; 30 (1): 61–6.

58. Park J, Kotzias D, Kuo P, *et al*. Detecting conversation topics in primary care office visits from transcripts of patient-provider interactions. *J Am Med Inform Assoc* 2019; 26 (12): 1493–504.

59. Rajkomar A, Kannan A, Chen K, *et al*. Automatically charting symptoms from patient-physician conversations using machine learning. *JAMA Intern Med* 2019; 179 (6): 836.

60. Du N, Chen K, Kannan A, *et al*. Extracting Symptoms and their Status from Clinical Conversations. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics; 2019: 915–25.

61. Du N, Wang M, Tran L, *et al*. Learning to Infer Entities, Properties and their Relations from Clinical Conversations. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China: Association for Computational Linguistics; 2019: 4979–90.

62. Selvaraj SP, Konam S. Medication Regimen Extraction from Medical Conversations. In: Shaban-Nejad A, Michalowski M, Buckeridge DL, eds. *Explainable AI in Healthcare and Medicine: Building a Culture of Transparency and Accountability*. New York, NY: Springer; 2021: 195–209.

63. Shafran I, Du N, Tran L, et al. The Medical Scribe: Corpus Development and Model Performance Analyses. In: *Proceedings of the 12th Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association; 2020: 2036–44.

64. Enarvi S, Amoia M, Teba MD-A, *et al*. Generating Medical Reports from Patient-Doctor Conversations using Sequence-to-Sequence Models. In: *Proceedings of the First Workshop on Natural Language Processing for Medical Conversations*. Online: Association for Computational Linguistics; 2020: 22–30.

65. Quiroz JC, Laranjo L, Kocaballi AB, *et al*. Challenges of developing a digital scribe to reduce clinical documentation burden. *NPJ Digit Med* 2019; 2: 114.

66. Lacson RC, Barzilay R, Long WJ. Automatic analysis of medical dialogue in the home hemodialysis domain: structure induction and summarization. *J Biomed Inform* 2006; 39 (5): 541–55.

67. Liu Z, Lim H, Ain, NF, *et al*. Fast Prototyping a Dialogue Comprehension System for Nurse-Patient Conversations on Symptom Monitoring. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Industry Papers)*. Minneapolis, MN: Association for Computational Linguistics; 2019: 24–31.

68. Neamatullah I, Douglass MM, Lehman L-WH, *et al*. Automated de-identification of free-text medical records. *BMC Med Inform Decis Mak* 2008; 8 (1): 32.

69. Le T, Bhushan V. *First Aid for the USMLE Step 2 CS*, 6th ed. New York, NY: McGraw-Hill Education; 2017.

70. Goldberg C. The oral presentation. UCSD's practical guide to clinical medicine. https://meded.ucsd.edu/clinicalmed/oral.html accessed March 4, 2020.

71. Palanica A, Thommandram A, Lee A, *et al*. Do you understand the words that are comin outta my mouth? Voice assistant comprehension of medication names. *NPJ Digit Med* 2019; 2: 55.

72. Johnson AEW, Pollard TJ, Shen L, *et al*. MIMIC-III, a freely accessible critical care database. *Sci Data* 2016; 3 (1): 160035.

73. Springfield! Springfield! https://www.springfieldspringfield.co.uk/ accessed June 16, 2019.

74. Pedregosa F, Varoquaux G, Gramfort A, *et al.* Scikit-learn: machine learning in python. *J Mach Learn Res* 2011; 12: 2825–30.

75. Altschuler J, Margolius D, Bodenheimer T, *et al.* Estimating a reasonable patient panel size for primary care physicians with team-based task delegation. *Ann Fam Med* 2012; 10 (5): 396–400.

76. Blackley SV, Schubert VD, Goss FR, *et al.* Physician use of speech recognition versus typing in clinical documentation: a controlled observational study. *Int J Med Inform* 2020; 141: 104178.

77. Industrial-Strength Natural Language Processing. spaCy. https://spacy.io/ accessed June 16, 2019.

78. Manning CD, Surdeanu M, Bauer J, *et al.* The Stanford CoreNLP natural language processing toolkit. In: Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations. Baltimore, MD: Association for Computational Linguistics; 2014: 55–60.

79. Bird S, Klein E, Loper E. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. Newton, MA: O'Reilly Media, Inc.; 2009.