Automating Clinical Documentation with Digital Scribes: Understanding the Impact on Physicians

Brenna Li
brli@cs.toronto.edu
University of Toronto
Toronto, Canada

Dr. Noah Crampton
noah.crampton@mail.utoronto.ca
University of Toronto
Toronto, Canada

Dr. Thomas Yeates
tom.yeates@mail.utoronto.ca
University of Toronto
Toronto, Canada

Yu Xia
sophie.xia@mail.utoronto.ca
University of Toronto
Toronto, Canada

Xirong Tian
xirong.tian@mail.utoronto.ca
University of Toronto
Toronto, Canada

Khai N. Truong
khai@cs.toronto.edu
University of Toronto
Toronto, Canada

ABSTRACT

Recently, digital scribe systems have been gaining popularity as a possible work-around solution to the Electronic Medical Record (EMR) documentation burden that affects many physicians. The proposed system would automate the clinical summary physicians take by capturing and extracting the patient-physician conversation during the consultation. While promising in concept, how this system would apply to real-world use and its limitations are still not well understood. To examine these issues, we designed a digital scribe prototype to generate notes of different qualities ranging from the reality of current state-of-the-art technology to the potential of future implementations. We conducted a "Wizard of Oz" study with 24 primary care physicians using our digital scribe prototype in 4 simulated medical encounters followed by a semi-structured interview. This exploratory study provides an understanding of physicians' interaction with digitally scribed notes, their perceptions on note quality, their perceived workflow impact and several directions for improvements.

CCS CONCEPTS

• **Human-centered computing** → **HCI design and evaluation methods**; **Empirical studies in HCI**; • **Social and professional topics** → **Medical records**; • **Computing methodologies** → *Natural language processing*.

KEYWORDS

human computer interaction, electronic medical records, audio-speech documentation, digital scribes

ACM Reference Format:
Brenna Li, Dr. Noah Crampton, Dr. Thomas Yeates, Yu Xia, Xirong Tian, and Khai N. Truong. 2021. Automating Clinical Documentation with Digital Scribes: Understanding the Impact on Physicians. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3411764.3445172

CHI '21, May 8–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8096-6/21/05…$15.00
https://doi.org/10.1145/3411764.3445172

1 INTRODUCTION

The last few decades have experienced a global healthcare movement to transform paper-based documentation to Electronic Medical Records (EMR) [42, 47]. There are numerous benefits to digitizing patient records, yet, for many physicians, the use of EMRs has resulted in increased documentation demand, and burden [11, 31]. Populating the clinical record for a single patient can involve numerous steps and various forms of structured and unstructured data, such as patient lab tests, prescriptions, billing codes, and free-form text-based clinical notes [2, 35]. Studies have shown that documentation tasks in the EMR can take up to 25-50% of clinicians' daily in-office working time; in addition, many physicians spend another one to two hours on EMR clerical task outside of the office [25, 49]. In total, physicians can be seen spending up to four hours each day just on EMR documentation, a burden that is contributing to their burnout, affecting their mental well-being and impacting patient's quality of care [16, 31, 47]. To offload this burden, physicians and medical institutions have been employing documentation assistants, also known as medical scribes, to help, however, this effort is too economically costly to widely implement [51, 53].

Recently, advancements in natural language processing and artificial intelligence have provided the opportunity to explore the use of technology to assist in EMR documentation tasks. This concept, known as the digital scribe, has been proposed to replicate the roles of medical scribes by generating clinical notes directly from physician-patient conversations during the consultation [10]. Most of the recent efforts from industry and academia have been focused on developing the algorithms and pipeline to create notes in the SOAP structure [10, 15, 23, 30], a globally adopted note-taking format that consists of four sections — the "Subjective" information that patients report, the "Objective" information found from observations or physical exams, the physician's "Assessment" of the diagnosis, and the "Plan" for further care [5, 18]. However, besides some speculations, little is known about how this system can affect physicians and what steps need to be considered for clinical use [40].

In this work, we explore the use of digital scribe systems in a simulated medical study with 24 primary care physicians. Our objective is to understand 1) How physicians would interact with the digital scribe notes, in particular, how they would perceive and react to the system's strengths and limitations. 2) How the digital scribe system can impact physicians and their current workflows. We implemented the digital scribe system to have three scribe conditions — machine, hybrid and human to reflect three stages — current, intermediary and futuristic implementations of the technology. We included the intermediary and futuristic scribe conditions to investigate participants' experience with digital scribes that are not hindered by current technical limitations, as there is a general understanding that this field is still in its infancy [40]. This also enabled us to understand factors that could influence the adoption of digital scribe systems for clinical use in a progressive manner.

In this paper, we report a user-centered study on the effects of a digital scribe system on physicians which contributes the following:

(1) Understanding the interactions and reactions that physicians have of the notes generated by current digital scribe technology (Machine), intermediary human-assisted digital scribes (Hybrid), and futuristic, doctor-like medical scribes (Human).
(2) Understanding the factors that influence physicians' perceptions of different note qualities.
(3) Understanding the impact of digital scribes on physicians' workflow.
(4) Directions for improving future digital scribe systems and the notes generated.

2 RELATED WORKS

2.1 Unintended consequences of EMR documentation

Research on EMR system's impact has been growing. As we highlighted earlier, there is a growing body of literature studying the effect of EMR use, specifically for physicians. There have been many positive outcomes since EMRs were introduced, such as documentation standardization, interoperability for information sharing, cognitive load reduction, better preventative health, among many others [9, 22, 46]. However, the unintended consequences of EMRs are also significant and only more recently are we learning more about them. These include demanding documentation expectations, impeding clinical workflow and information reviewing, and distracting notification and warning systems [36]. Some studies attribute these consequences to the lack of user-experience considerations for physicians and other health professionals prior to deployment of such systems [6, 24, 45]. Therefore, to minimize the potential unintended consequences of the digital scribe, researchers and designers need to understand what these limitations are and how to adapt to them before using it in clinics.

2.2 Studies on speech documentation for EMR

Several works have looked at clinical usage of speech documentations. Zheng et al. [57] compared the linguistic differences in voice-dictated versus typed clinical entries and found that while the vocabulary is similar, the length of the record is shorter and the usage of acronyms a lot higher in typed documentations. Although they did not look at the context of the notes, their result potentially speaks to the time constraint clinicians have with typed entry as well as the possibility of missing information. Mamykina et al. [33] conducted a time-and-motion study of clinicians interacting with EMR systems, from which they argued that typing should not be replaced because it helps clinicians to synthesize diagnoses. Instead, they suggested that in order to improve efficiency and workflow, tools should be implemented to reduce work fragmentation and provide clinicians with uninterrupted documentation [33]. While the study was thorough in capturing the documentation process, it did not address the constraints clinicians have with time and the pressure to see more patients; therefore, an uninterrupted documentation session may not be practical. A recent study by Willis and Jarrahi [55] looked at the possibility of having automation in clinical documentation. From their observations and interviews, they suggested that the best form of interaction is when clinicians would make sense of patient context, and the agent would assist with decision support and text entry [55]. However, that idea has not been tested in a simulated or real environment.

2.3 Medical scribes

Medical scribes are trained personnel hired by physicians to help with EMR documentations [48]. While their roles may vary depending on the clinic and setting, generally, their job is to generate real-time, free-text, clinical notes based on conversations in the exam room during patient consultations [7, 51]. These notes would then be reviewed by physicians at a later time before being saved into the patient's chart [48, 56]. The benefit of medical scribes includes, increased efficiency, reduced documentation burden, and improved physician-patient interaction and patient's quality of care [20, 34, 48, 56]. As a result, the demand for medical scribes is increasing. However, there are also growing concerns on the sustainability of this workaround solution [7, 48]. One concern is that medical-scribes can be very costly to employ. Their wage, in addition to other costs, such as recruitment, training, and management, can become cost-prohibitive for smaller clinics and institutions [53]. Another concern is the lack of standards in training and job regulations. Depending on the physician they are assisting, medical scribes can be given additional responsibilities, such as retrieving information from patient records, assisting in clinical procedures, providing patient education and generating order entries on the physician's behalf [17, 48]. These tasks blur the boundary between an assistant and a licensed health care practitioner, which can put patients' privacy and safety at risk.

2.4 Digital scribes

The concept of digital scribes, as described by Coiera et al. [10], involves the use of advances in artificial intelligence, speech recognition and natural language processing to automatically enter parts of the EMR documentation for clinicians during the clinical encounters. Like the variety of medical scribe roles, there is a range in ideas for how digital scribes should assist physicians. These can broadly be categorized into human directed systems, machine directed systems, and a mix of human and machine directed systems [10, 50]. Human directed systems provide physicians with the highest level of control and often involves physicians dictating summary notes or EMR interaction orders through speech recognition services that they would later correct afterwards [10, 26]. This strictly speech-based dictation-like interaction has been widely studied and implemented in the medical field for its improved efficiency [26, 37, 52]. However, studies have found this type of interaction is prone to higher error rates because EMR interfaces were traditionally designed for keyboard and mouse input modality and not suitable for voice-interaction [21]. In contrast, machine directed systems are entirely autonomous and require minimum input from users [50]. These are generally adept at well-structured and defined tasks, such as medication alert systems or structured form-like entries, but will struggle in contexts where there is a high degree of variability [40]. The process of generating clinical notes however, requires the system to be both precise and consistent in its information capturing to meet the medical-legal expectations and to adapt to physicians' style, situation, and variety of medical encounters [28, 29]. This requirement outlines the human-machine directed initiative, where human scribes or physicians can provide some instructions for what is expected in the generated output, but the system is intelligent enough to understand and learn the process independently [10, 29]. Several recent projects have explored the technical aspects of building such systems. For example, Enarvi et al. [13] implemented a sequence-to-sequence model with recurrent neural networks to generate notes from orthopedic surgery transcripts that contains four sections — history of illness, physical examination, assessment and plan, and diagnostic imaging results. Their implementation relieves the burden of annotating transcript data to train the model but is restrictive in its application for highly specialized medical fields that often have more structure to their notes. Another method which was used in *Autoscribe* [23, 27] is to train the machine model on medically annotated transcripts to extract clinical entities such as anatomical location, diagnosis, symptoms, medications, reason for visit, etc., which then gets classified into the SOAP sections of a generated note. This method is more suitable for primary care applications because the system can be pre-trained with a variety of knowledge resources to afford better versatility. Finely et al. [15] described a similar process but included an audio speech recognition service to the medical entity extraction model for a more "end-to-end" implementation of a digital scribe. However, because information on the note generation process is fairly sparse and difficult to validate in the work by Finely et al, we decided to use *Autoscribe's* model and integrated a third party audio to speech recognition service, Otter.ai [1], to create our own end-to-end digital scribe system to explore physician user's interactions and perceptions of the system.

3 METHODS

3.1 Study overview

We designed a "Wizard of Oz" digital scribe prototype that is integrated with an EMR. The prototype generates notes of three conditions, Machine, Hybrid and Human, that would reflect the range of output from current state-of-the-art technology to futuristic implementations of the system. We then evaluated the use of this prototype in a simulated medical environment with 24 primary care physicians. Each physician was placed in either Note-Taking or Non-Note-Taking groups and experienced three to four simulated medical scenarios that were each randomly paired with one of the three note conditions as shown in Table 1. These notes were generated "live" behind the scene of the simulated consultation. We then asked physicians to review and edit the notes before concluding with a semi-structured interview.

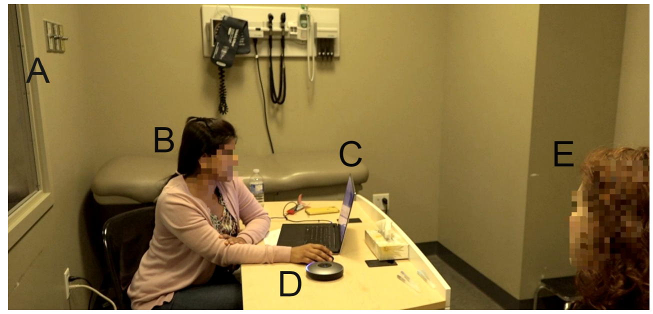


Figure 1: Exam room setup. From the left, A: one-way mirror looking into the room with researchers sitting on the other side, B: physician participant, C: laptop running OscarEMR, D: microphone capturing the consultation, E: standardized patient.

3.2 Study setup

We designed the study to reflect an actual patient encounter to test our digital scribe prototype in an environment that would be as natural as possible. The study was conducted in an exam room at a local teaching hospital to provide participants with a familiar environment. The exam room also had a one-way mirror and audio streaming that allowed researchers, the "wizards", to observe and run the digital scribe system from the outside without interfering with the consultation, as shown in Figure 1. A second-year medical resident on the research team designed the four scenarios to follow the format of the Objective Structured Clinical Examination (OSCE), a qualifying assessment taken by medical students in Canada, to ensure the scenarios contain the depth and complexity of a typical patient visit [19, 43]. We employed standardized patients (SPs), who are professionally trained actors for medical education simulation, and prepared them with these scenarios so that the simulated patient interaction would be realistic and consistent across participants. The digital scribe system was pre-configured to generate clinical notes in the standard Subjective (S), Objective (O), Assessment (A), and Plan (P), the "SOAP" format that most physicians are familiar with [5]. We also pre-loaded background information from these scenarios (such as name, age, gender and primary concerns) into OscarEMR, an open-sourced EMR system, for the participants to preview and use during the consultation [8].

3.3 Study conditions

3.3.1 Note taking variable. We assigned participants to one of two conditions: 1) Note-Taking, where they were asked to take notes in OscarEMR as they normally would during or after each consultation. 2) Non-Note-Taking, where they were asked to not take notes, instead, to only converse with the SPs as they would with their own patients. These conditions allowed us to compare the impact Non-Note-Taking may have on physicians because most of our participants did not have prior experience working with scribes or dictation services.

Participant ID	Note Taking	Order	Scenarios	Digital Scribe Condition
1, 7, 13, 19	Yes	First	Anne Brown	Hybrid
		Second	Jane Robertson	Machine
		Third	Marsha Morris	Human
		Fourth	Jenny Osborne	Human
2, 8, 14, 20	No	First	Jane Robertson	Hybrid
		Second	Marsha Morris	Machine
		Third	Jenny Osborne	Machine
		Fourth	Anne Brown	Human
3, 9, 15, 21	Yes	First	Marsha Morris	Hybrid
		Second	Jenny Osborne	Hybrid
		Third	Anne Brown	Machine
		Fourth	Jane Robertson	Human
4, 10, 16, 22	No	First	Anne Brown	Hybrid
		Second	Jane Robertson	Machine
		Third	Marsha Morris	Human
		Fourth	Jenny Osborne	Human
5, 11, 17, 23	Yes	First	Jane Robertson	Hybrid
		Second	Marsha Morris	Machine
		Third	Jenny Osborne	Machine
		Fourth	Anne Brown	Human
6, 12, 18, 24	No	First	Marsha Morris	Hybrid
		Second	Jenny Osborne	Hybrid
		Third	Anne Brown	Machine
		Fourth	Jane Robertson	Human

Table 1: Participant condition assignment

3.3.2 Clinical scenario variable. The scenarios were written by a second-year family medicine resident, supervised by a senior family medicine physician on our team for realism and validness. Each scenario contains the complexities of a typical patient, such as their family history, health history, detailed symptom development and emotional concerns. Of the four scenarios, three were designed to be straight-forward singular symptom cases, such as coughing, abdominal pain and chest pain. The fourth scenario was more complicated and had two symptoms, insomnia and amenorrhea, which could be related but lacked a clear diagnosis. A full description of the scenarios can be found in the Supplementary Files. We named each scenario after the character the SP portrays, Ms. Brown (coughing), Ms. Morris (abdominal pain), Ms. Osborne (chest pain) and Ms. Robertson (insomnia and amenorrhea). The order of the scenarios for each participant was counterbalanced across the study population and most participants received all four scenarios, unless prevented by time.

3.3.3 Digital scribe variable. Our "Wizard of Oz" digital scribe prototype takes as input the conversation between physician and SP during consultation and outputs a text-based clinical note summary in the SOAP format. All conditions involved some form of human 'wizardry', as researchers were involved in "activating" and "deactivating" the digital scribe recording, as well as copying the generated notes back into the OscarEMR for a seamless, integrated experience. However, the components and steps involved can vary depending on whether it is Machine, Hybrid, or Human, refer to Figure 2.

In the Machine note condition, notes are purely machine-generated. The system followed the design proposed by Quiroz et al.[40]. It used a conference microphone that was placed in the middle of the exam room to stream audio to Otter.ai, an speech to text transcription service equipped with an interface for manual edits [1]. Researchers corrected errors in the transcript before running it through *Autoscribe*, an extraction and summarization process to generate medical notes [23, 27]. All Machine scribe conditions followed this process consistently, and a detailed summary of the *Autoscribe* pipeline, which takes as input a transcript of the consultation and outputs a SOAP note can be found in the fourth figure of the work by Jeblee et al. [23].

In the Hybrid note condition, a second-year medical resident on our team modified the machine-generated note. They were specifically instructed to reorder text to relevant sections of SOAP, remove any unrelated content, correct for inaccurate descriptions, and add details such as date, duration and severity to content that had already been captured by the machine notes. As an example, if the

Machine note captured "patient has stomach pain", the Hybrid note would extend it to "patient has stomach pain for past 3 days", if "3 days" was mentioned in the consultation but was missed in the machine notes.

In the Human note condition, the notes were entirely generated by the same medical resident. The resident was instructed to write the notes based on what they hear from the conversation in the consultation room. We also asked that they try to consistently document at the standard of a medically trained professional for all participants.

For an example of the different scribe conditions, please see Supplementary Files for a sample of the Machine, Hybrid, and Human scribe notes generated from the study. For every participant, the order was Hybrid, Machine and Human. This was due to limitations involving the lengthy amount of time required by the text extraction and summarization algorithms used in our digital scribe prototype. The Hybrid condition was placed first because it required a pass from the Machine and then the resident, which took more time. The Human note condition was placed last because these were generated live by the resident and took less time.

3.4 Participants

In total, we recruited 24 participants from the Greater Toronto Area using a mixture of convenience sampling through word-of-mouth and random sampling from a social media advertisement. The 24 participants include 20 practicing primary care physicians and 4 family medicine residents. In a survey prior to the study, all participants had described themselves as daily and proficient EMR users with the average experience being 6.2 +/- 2.7 years. We received written consent from all participants prior to the study and provided compensation of $300 CAD per session for practicing physicians and $150 CAD for medical residents; an amount that was equal to their typical hourly wage as recommended by the medical professionals on our team.

3.5 Study procedure

In the 2-hour simulation-based user study, participants were presented with four medical scenarios, one at a time. They were asked to treat this session as a walk-in clinic and interact with SPs as they normally would with real patients. They were also told that they

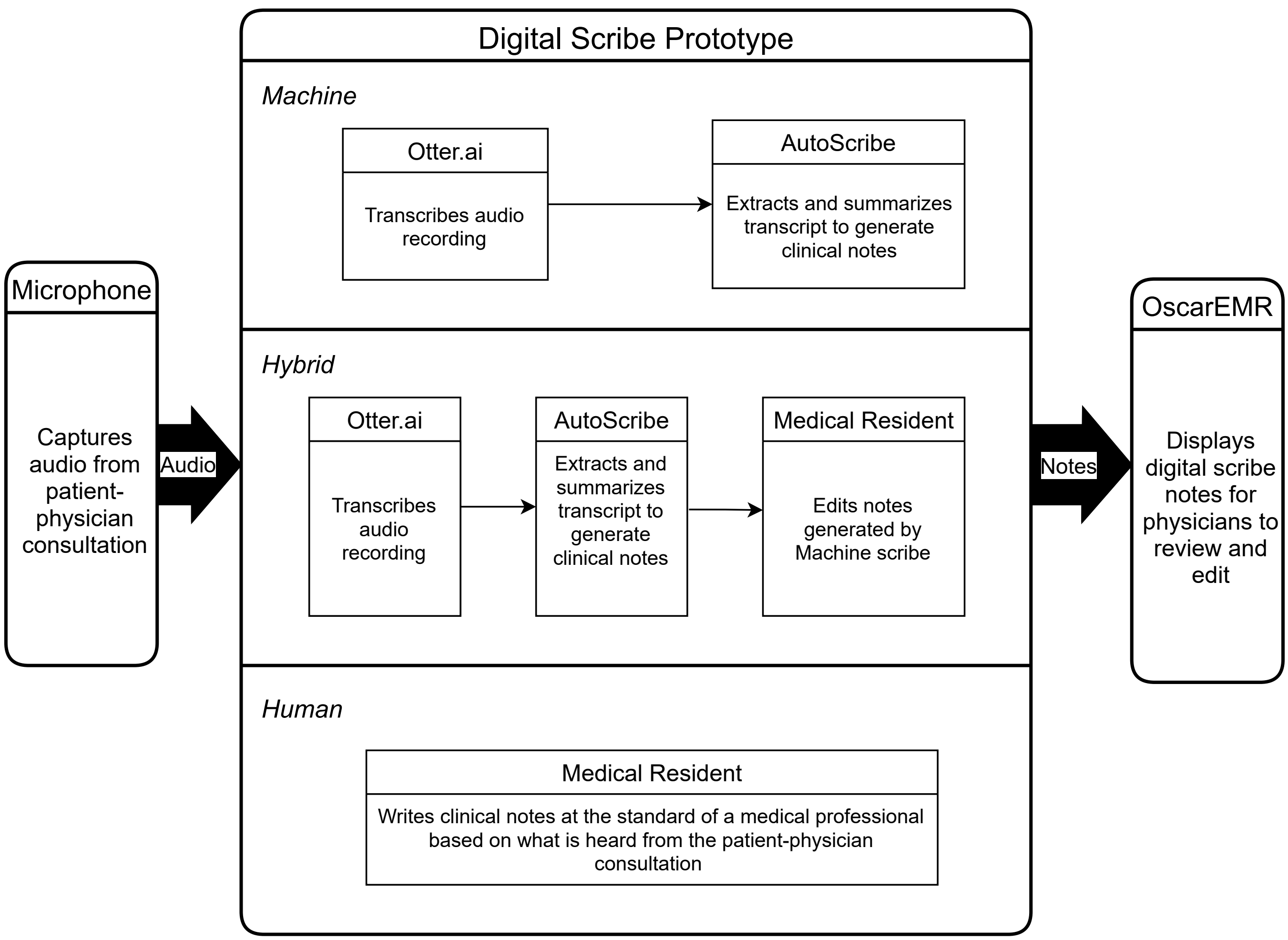


Figure 2: Digital scribe prototype design and setup

had approximately 15-25 minutes with each scenario before they would hear a knock signalling for them to wrap up. Participants were given a dummy interface to "activate" and "deactivate" the digital scribe system but were not told it was managed by researchers, the "wizards", behind the one-way mirror. After completing all patient scenarios, participants were asked to review the digital scribe notes. Participants were not told of the different scribe conditions, only that they were reviewing notes generated by a system from their SP consultations earlier. The generated notes were presented to participants in OscarEMR and we asked each participant to directly edit the generated notes to the standards of their own while they followed a think-aloud procedure to voice their thoughts and reasoning for the changes they made. For each note they reviewed, we also asked them to provide a score on a scale of 1 to 10, with 1 being completely unacceptable, and 10 being on-par with their own notes. After the participant has reviewed all generated notes, we conducted a semi-structured interview asking them to reflect on their experience with the digital scribe.

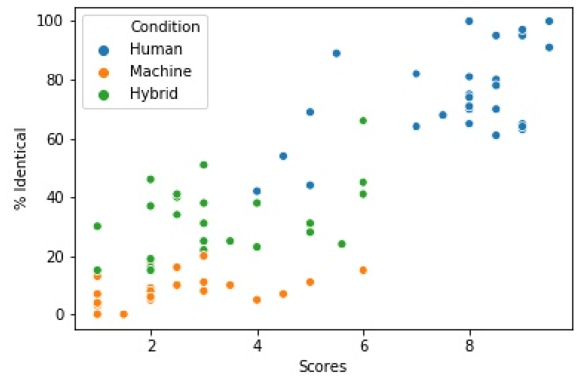


Figure 3: Correlation between the scores and percent identical of the three note conditions

Digital Scribe Condition	Section	Change	Delete	Insert	Total
Machine	S	1	3	12	16
	O	0	4	3	7
	A	0	5	3	8
	P	0	8	4	12
Hybrid	S	1	1	11	13
	O	0	1	3.5	4.5
	A	0	1	2	3
	P	0	1	3	4
Human	S	0	0	2	2
	O	0	0.5	3.5	4
	A	0	0	1	1
	P	0	0	1	1

Table 2: Median number of changes per scribe conditions and note section across all notes

4 RESULTS

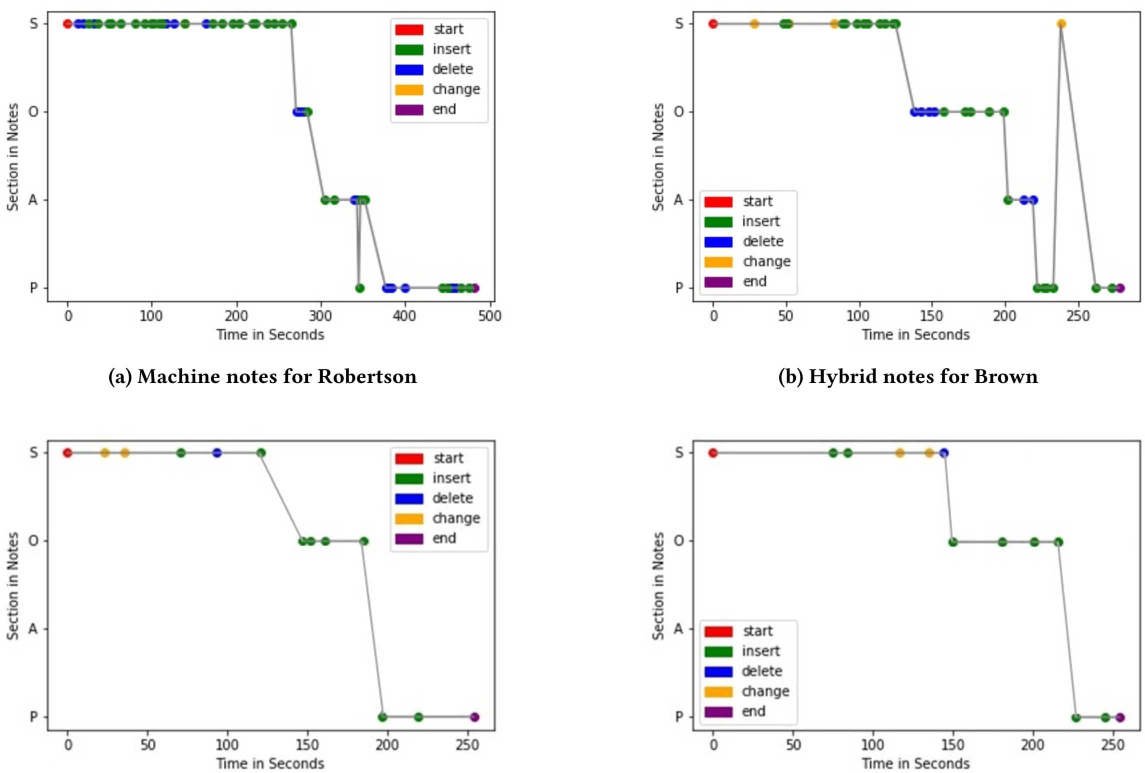
4.1 Interaction with digital scribe notes

The 24 participants in the Note-Taking and Non-Note-Taking groups each reviewed and scored three to four digital scribe notes that were generated from their simulated consultations, for a total of 91 notes reviewed and scored across the Machine, Hybrid and Human digital scribe note conditions. When we calculated the average score that participants gave for the different note conditions, we found that Machine notes ranked the lowest (median of 2/10) followed by Hybrid (median of 3/10) and Human (median of 8/10). We employed a linear mixed model to account for the repeated sampling of an unbalanced design of four scenarios (which our medical professionals recommended for greater variety) with three plus one repeated scribe conditions. Results show that there is significant difference across scribe conditions with p-value « 0.01, and no interaction effect from note-taking and scenario factors. Refer to the Supplementary Files for more details.

4.1.1 How physicians interacted with the notes. To understand what influenced these scores, we looked at how physicians interacted with these notes and the factors that affected their interaction. With a document diff tracker [39], we analyzed the changes participants made to each of the notes they reviewed. The percentage of identical lines is the lowest in Machine notes (median of 8%), followed by Hybrid (median of 31 %) and Human (median of 74%). Refer to Supplementary Files for a table breakdown. As shown in Figure 3, we found a strong correlation between percent identical lines and the scores participants gave to the notes across all scribe conditions. A Pearson correlation score of 0.870 also confirms this.

We further found that the digital scribe note conditions can be characterized by the types of changes that were made. When we analyzed the number of inserted lines, deleted lines, and changed lines (which would include within-line replacements), we found that Machine notes on average had the most deletions (median of 30), Hybrid notes had the most insertions (median of 12.5) and Human notes with the least in both insertion (median of 5) and deletions (median of 1). But the overall number of changes is still the lowest in Human notes (median 6.0), followed by Hybrid notes (median 17.0), and Machine notes (median 48). The difference in the number of insertions, deletions and changes are also significant among the three note conditions (p-value « 0.01 for all). Refer to Supplementary Files for more details.

We then analyzed the screen-recordings of the note review to determine the parts of the notes (i.e., Subjective, Objective, Assessment, and Plan) where different types of changes (i.e., insert, delete, and change) were made. Table 2 provides an outline of the median number of modifications for every note that was edited, broken down into the respective S, O, A, P sections. The Subjective section is where most of the "Total" changes occurred for both the Hybrid and Machine scribe conditions, as both required a significant number of insertions. Both insertions and deletions were the highest for Machine in the Assessment and Plan sections compared to Hybrid and Human. The number of insertions in the Objective sections appears consistent across all scribe conditions, with more deletions in the Machine notes than the other two. In Figure 4, we show a time-series example of the Machine, Hybrid and Human note review process through participant 10. We found that participants interacted with the notes mostly in a linear way, always starting from the Subjective and working down towards the Plan. We also



(a) Machine notes for Robertson

(b) Hybrid notes for Brown

(c) Human notes for Osborne

(d) Human notes for Morris

Figure 4: Examples of a note review process from P10 for all three note conditions and four scenarios

found that participants sometimes jumped back and forth between different sections, particularly in the Machine and Hybrid notes, indicating that they were trying to align the changes they made to the earlier sections with the ones they wrote in the later ones.

4.1.2 Factors that influenced physicians' interaction with the notes. We applied open coding on the think-aloud recordings from participants' note review sessions with no prior themes. Three researchers were involved and coded the first 5 participants together to develop the codes, and the remaining participants were coded by two researchers each. We then organized the codes into four categories: accuracy, completeness, relevance and comprehensibility. Accuracy can be described as how correct the content was perceived to be. Usually, this refers to the correctness in documenting pertinent positives and negatives from the encounter. Completeness describes how well the note's content captures the expected information. This includes having all SOAP sections filled, and details related to the symptom's "o, p, q, r, s, t" (onset, provocation, quality, radiation, severity, and time) documented. Relevance describes how applicable the information is for the encounter. The content should reflect what was discussed in the consultation and organized into the appropriate SOAP sections. Comprehensibility describes how easily the information can be interpreted and understood. Related information is expected to be grouped together and written concisely with the proper medical terminologies and abbreviations. Refer to Table 3 for examples of quotes and the factors for each note type.

When participants reviewed Machine notes, they found accuracy, completeness, relevance, and comprehensibility to all be lacking. They reported the system erred in places where information was clearly and specifically articulated during the encounter. These may include inaccurate labelling of pertinent information, incomplete description of overtly discussed symptoms, repetitive, non-relevant or incoherent writing, and disorganized content that made it difficult to follow and understand what happened during the encounter. Participants stated that they would have no tolerance for and would not trust a system if important findings were incorrect or incomplete because that could mislead them in their diagnosis and treatment plans, putting patients at a greater health risk. To a lesser degree, they also emphasized the importance of comprehensibility, mentioning they prefer notes to be 'point form' because they do not want to use their limited time 'reading long sentences' or 'deciphering' the meaning behind the notes. At the minimum,

> *The system should at least function as a form of a reminder, to help me complete my notes more quickly.* - P2

Digital Scribe Conditions	General Sentiment	Strengths & Flaws		Examples quotes
Machine	*"This is just not helpful, there's a lot I would add. I feel like it's not going to capture the correct information. I would prefer to write my own notes, it would have been faster than editing"*	**Flaws**	Accuracy	*"Some of the negatives were put as pertinent positives and positives as negatives. It's unsafe. If I miss something and then it says something I didn't intend, that could change my management of the patient".*
			Completeness	*"The subjective section actually missed a lot of detail, like "it said pain, but didn't say where the pain was, I would say 'epigastric pain'"*
			Comprehensibility	*"I had to delete more things, took more time, they're all in different lines, made reading through it more difficult as well"*
			Relevence	*" 'patient has fever' this doesn't belong in the Plan"*
				"'patient has burn', 'patient has burn', 'patient burning' that doesn't make sense in "
Hybrid	*"For the subjective section, it wasn't bad. but missing most of the OAP section of a SOAP note, and that's a huge component of the overall picture"*	**Flaws**	Accuracy	*"Might be my fault, I said bloodwork, but I wanted to do TSH"*
				"Not just 'Tum's but 'PPI', and not just 'heart attack' but like the formal diagnosis names"
			Completeness	*"It's still missing some of the 'o p q r s t' descriptions of the symptom"*
				"But it's still missing most of the OAP section of a SOAP note, and that's a huge component of the overall picture"
		Strengths	Accuracy & Completeness	*"it had most of the positives and negatives, to help me push towards or away from a diagnosis"*
			Comprehensibility	*"this is a lot easier to read, and more succinct"*
Human	*"That's pretty good. That covers a lot a lot of what I was looking for and looks like an actual note I'd take. I would use this if the system consistently produced this quality"*	**Flaws**	Completeness	*"it mainly missed the physical exam section"*
				"There are things I make observations and would write in notes but not say out-loud. Such as, when the SP was walking into the room very slowly because her character, Marsha Morris, was supposed to be in a lot of abdominal pain"
			Comprehensilibty	*"But it was just the organization piece that was missing. I would have made a break here and divided it by the number of issues."*
				"I tend to use more short forms. Like, for example, I might use BP for blood pressure? Or I might use like RTC as return to clinic instead. I use short forms so the notes are not so long."
		Strengths	Accuracy & Completeness	*"I think the fact that it was able to capture a lot of the questions I asked, that was really helpful and saves me time"*
			Comprehensibility	*"this is clear, to the point and looks like an actual note I'd take"*
			Relevance	*"the content seems to be in the write sections, despite me going back and forth non-linearly with the questions"*

Table 3: Example quotes from note review session

When participants reviewed Hybrid notes, all participants found the note to be easier to read and consistent with what they remember discussing during the encounter compared to Machine notes. Although the Subjective section was still missing some details, the content was sufficient to remind them of what is missing and needed to be added to the note. However, they indicated that the Objective, Assessment and Plan parts required further improvements. Participants found these sections to lack the professionalism they were expecting because the writing was too colloquial. In particular, for the Assessment and Plan sections, they found the content to be missing the medical synthesis behind what they told the SPs.

> *These did pick up the keywords, but a lot of what I put into my notes is the second order thinking because what I am communicating to the patients is different, and that is missing here* - P9

> *The Plan part that's hard, a lot of it happens in your head and you tend to fill the rest out after the patient has left* - P11

When participants reviewed Human notes, they found the note to meet most of their expectations in terms of accuracy, completeness, relevance and comprehensibility. They found the Subjective section to be mostly complete, and the Assessment and Plan sections to have more of the medical synthesis that was missing from the Hybrid notes. The Objective section was the only section that was still lacking, but participants indicated that it was likely because they did not verbalize the exam findings to the patients and were willing to add that information in themselves.

> *I can just fill the physical exam information in myself. I do it silently now anyways. It would not be a significant change to how I do things now.* - P25

Participants also thought the Human note was more structured, despite their non-linear and fragmented conversation with the SPs during the consultation. However, we found 7 of the participants became more critical of the writing style and pointed out improvements such as the use of capital letters, line breaks, or specific short form notations for high-frequency words that they are more familiar with and comfortable with reading. However, this stylistic preference for better comprehensibility is less important than the other qualities previously mentioned.

> *I know this is being neurotic. I like capital letters when I'm going to start a sentence* - P22

4.2 Impact of the digital scribe

4.2.1 Perceived improved patient engagement and documentation efficiency. In the interview, we found that 21 of the participants were optimistic about the system's potential to improve their patient engagement and documentation efficiency from using our digital scribe prototype. When asked about their day-to-day interaction with patients, participants mentioned that they are aware of the distraction that comes from EMR documentation which is affecting their interactions with patients. As P15 puts it:

> *It's unfortunate the way encounters are now where you have to type while you talk to the patient, or else you'd be spending hours doing your notes. Patients don't like it. They would much rather you sit and just have a conversation with them. I don't want to either, but it's to make sure my documentation is complete.*

If the digital scribe works, with the acceptable note quality, most participants believe that this system could allow them to redirect their focus back on the patients. When asked about their experience as the patient in the room, the SPs agreed that Non-Note-Taking participants were more engaged than the Note-Taking ones. They describe the Non-Note-Taking participants to be more "present" and "listening" than the Note-Taking participants, especially when they were 'typing away in silence'. Along this line, if the digital scribe can produce reliable notes, many participants believe that using this system can greatly reduce the time they spend documenting which could allow them to see more patients. When editing the digital scribe's Human note conditions, P5 said:

> *I don't know what my [reviewing] time was, but it felt faster. I just had to type in the physical findings and add in a point to the plan and it was done. That is just 3 lines as opposed to the entire page.*
> *In an actual visit, you would need to get the patient on the exam table but even if you can get the consultation done in 10 minutes instead of 15, that is significant. You can see more patients.*

4.2.2 Impact on the cognitive load. When observing the behavior of Note-Taking and Non-Note-Taking participants, we found that Note-Taking participants seemed more confident during the encounter while Non-Note-Taking participants seemed more uneasy. They also expressed greater frequencies of pauses and repeated questions during the consultation. Although this could be due to the unfamiliar conditions Non-Note-Taking participants were asked to practice, participants mentioned in the interview that they had a more difficult time because of the added mental load. As P16 describes:

> *Sometimes I use the time for typing for actually thinking what I want to say or do next. When I didn't have that couple of seconds, I find I tend to say "um" a lot more than I probably usually would...It [requires] re-training of how I currently do my [consultations]*

In our prototype, participants were not given the digital scribe notes until the note review stage which is analogous to how physicians interact with medical scribe notes. However, some participants mentioned that the Non-Note-Taking process was more difficult because they also had a difficult time remembering what was discussed and some feedback from the system would have been helpful, if it was not distracting. As P24 and P2 mentions:

> *I do find it helpful to have a visual cue, it's so easy to forget things, especially in family practice because you're seeing such a wide variety of problems that it's just impossible to remember. But I am not sure how that fits in.*

> *If things populate the screen as words are coming in, I might feel tempted to read what it's trying to write to make sure it caught what was discussed. If [the feedback is] like that, it will be distracting and I rather not have it.*

4.2.3 Impact on the clinical workflow. We found that participants also struggled with interacting with the digital scribe prototype from analyzing the screen-capture video of the consultation process. Both Note-Taking and Non-Note-Taking participants were asked to interact with the prototype's interface by clicking "start" when the patient comes in and "stop" when the patient exits to "activate" and "deactivate" the recording. Among all encounters, only 69.14% were "started" and "stopped" *On time*, the rest were either *Late*, started several minutes after the SP has entered or left, or completely forgotten and *Missed*. While familiarity with the digital scribe system may have had an effect, as P19 says:

> *I had to remind myself to start and stop. If I am more used to it, it might be easier.*

This interaction can also be an added burden to doctor's current workflow. As P20 describes:

> *There is a lot going on in the clinic already. It shouldn't be something that I have to personally remember to do. Because if I thought I did and I forgot, then that's the whole encounter gone, which I'll have to rewrite at the end of the day.*

However, all participants agree that the system should not be "on" all the time and that there should be a mechanism for physicians to turn the system "off" easily. This is particularly important, as P24 explains:

> *When patients talk about their mental health, substance abuse, domestic violence or marital affairs, these things can be sticky and patients would wonder where this information gets shared.*

5 DISCUSSION

In the results section we analyzed the interactions participants had with the generated scribe notes. We found that Machine notes had the lowest percentage of identical lines, suggesting that participants felt a greater need to modify these notes, which can be seen in the large number of deletions and insertions that were found throughout the SOAP sections. In particular, what distinguished Machine notes were a large number of deletions that seemed to correspond with the increased instances of "irrelevant" or "inaccurate" content that participants had also reported, suggesting that the two factors are possible explanations for the low score it received. In Hybrid notes, we found the percentage of identical lines to be higher than Machine notes because fewer lines were deleted, yet, there were still a large number of insertions. This suggests the hybrid note was still considered "incomplete" by participants and needed them to add the missing content which can explain its mediocre score. In the Human note we observed the highest percentage of identical lines, suggesting that participants were most satisfied with what was already present in the note. The changes we found were largely insertions in the "Objective" section, which participants rationalized as themselves not vocalizing physical exam findings. The remaining insertions in the note were usually stylistic preferences that participants had to increase "comprehensibility", but all mentioned this is the lowest of their priorities. These factors, with the possibility of fatigue from the order that Machine, Hybrid and Human scribes were presented to participants, could explain why Human scribes received the highest scores.

5.1 Improvements for generated notes

5.1.1 Accuracy and relevance. As we have learned, digital systems built from state-of-the-art components are still far from being ready for use in practice. From the number of changes physicians had to make, the effort needed to modify these notes to the standard physicians find acceptable is substantial. The reason was that these notes lacked in accuracy, completeness, relevance and comprehensibility. Specifically, when compared to Hybrid notes, Machine notes struggled more with accuracy and relevance, which caused participants to make more deletions and changes. This is echoed by the low scores participants gave to Machine notes, with some questioning the purpose of this system if the content is less dependable and requires as much time as their current note-taking process. This suggests that at the minimum, digital scribe systems should focus on improving the accuracy in speech capturing and context relevance for physicians to trust the system. In this regard, this study supports previous works emphasizing the need to greatly improve current audio speech recognition, content extraction, summarization and text generators in the medical domain in order for digital scribe systems to work as intended [10, 28, 40].

5.1.2 Completeness of the "Subjective" section. After accuracy and relevance, completeness was the most common problem participants had when they reviewed digital scribe notes. We learned that many of the inserts in Hybrid notes were because the note lacked details on the "o, p, q, r, s, t" (onset, provocation, quality, radiation, severity, and time) that would be expected in the Subjective section [32]. The Subjective section is essential because its information, along with the Objective findings, is what physicians use to synthesize and justify their Assessments and Plans [5, 38]. In addition, the clinical note is a legal document and can be called upon as evidence for a court case or be used against physicians in a medical lawsuit [4, 12, 41]. Therefore, the information that goes into the Subjective section, which helps inform the rest of the clinical documentation, is crucial, especially if physicians are busy and need to complete their notes at a different time [2].

5.1.3 Medical synthesis. The medical synthesis in the note is what most participants noticed when they reviewed the digital scribe's Human notes. The lack of medical synthesis also explains many of the changes participants made in the Assessment and Plan sections of Hybrid notes. These changes could be as simple as converting the colloquial words used for patients to more medically appropriate terminology, or as complicated as documenting the physician's thought process that leads to their Assessment and Plan (the latter, which physicians have spent many years in training to perfect). Participants perceived this as a more critical barrier than the non-verbalized physical exam findings, because exam findings are usually quick to manually type or are already automatically documented through the electronic devices that are connected to the EMR [3, 14]. Thus, if the purpose of digital scribes is to generate an entire clinical note, then, a complete Subjective, Objective, Assessment and Plan is necessary. However, if the system's goal is to create notes to help physicians generate SOAP notes, we believe a note that is between the output of the Hybrid and Human notes could be more practical and sufficient.

5.2 Designs for future digital scribe systems

While participants are optimistic about the potential of improved efficiency and patient engagement through digital scribes, future designs should still carefully examine its impact on physician's cognitive load. We designed the digital scribe prototype so that physicians did not get to see the generated notes until after the consultation. However, in the interview, we found that participants preferred to see some visual feedback from the system because the notes are sometimes a reminder of what has been discussed and help prompt further questions when needed. But too much feedback, which we were concerned about when implementing the system, can also be distracting. If physicians are constantly checking on the system to see what has been documented, this can remove the proposed benefit of improving patient engagement. Therefore, some questions that designers of future systems need to consider are what kinds of feedback and how much of that feedback the system should provide so that physicians can be reassured it is documenting as expected.

Another point for designers to consider is the changes to clinical workflow that could be introduced if digital scribes were adopted. In our study, we found that if physicians were required to activate and deactivate the system, over a quarter of the encounters could have missing notes. While this potentially can be improved with familiarity to the system, a single missed note could be too many and pose significant risks to both physicians and patients. The system should also not be "on" all the time for ethical and practical reasons. Therefore, who, what, where, and when to trigger the digital scribe system become questions that still needs to be answered.

Finally, as exciting as the concept of digital scribes has been proposed to work, from the study, we have learned that we are still far from achieving those expectations in reality. As a future direction, we would like to explore the possibility of integrating digital scribe systems with note templates that have recently become popular for commonly seen medical symptoms [44, 54]. These note templates, or "macros" referred to by our participants, are pre-generated note outlines that could be filled in quickly and function as a check-list reminder [44]. If integrated with digital scribe systems, these can help structure digital scribe notes to focus on the pertinent medical information in the conversation while also providing physicians with the visual feedback they are accustomed to having. Although more is needed to understand how this would be implemented and evaluated, we think this can be a potential solution to bringing digital scribes and their benefits sooner to practice.

6 CONCLUSION

In this paper, we described an exploratory study on the use of digital scribe systems in a simulated clinical environment. Broadly, our study contributed a better understanding of 1) how physicians would interact with digital scribe notes and 2) the impact of digital scribe systems on physicians' workflow. We experimented with three levels of note quality, entirely machine generated notes, human edits of machine notes and entirely human scribed notes, and showed the interaction efforts needed for each level. We then provided four factors, accuracy, completeness, relevance and comprehensibility, that can influence physician's interaction and perception of the note. We argued that the limitations of current language processing technology make digital scribe systems impractical for actual clinical use. However, physicians are optimistic about using the digital scribe system in the future to improve patient engagement and documentation efficiency. Finally, we suggested a few intermediary directions for future developers and designers to consider.

ACKNOWLEDGMENTS

This work was supported in part by MITACS (IT15421) and NSERC (RGPIN-2016-06326).

REFERENCES

[1] Otter AI. 2021. Otter Voice Meeting Notes. https://otter.ai/

[2] Brian G. Arndt, John W. Beasley, Michelle D. Watkinson, Jonathan L. Temte, Wen-Jan Tuan, Christine A. Sinsky, and Valerie J. Gilchrist. 2017. Tethered to the EHR: Primary Care Physician Workload Assessment Using EHR Event Log Data and Time-Motion Observations. *The Annals of Family Medicine* 15, 5 (Sept. 2017), 419–426. https://doi.org/10.1370/afm.2121

[3] Mohit Arora, Nasrintaj Falsafi, Mohamed Al-Ibrahim, Robert Sawyer, E Siegel, Ashish Joshi, and Joseph Finkelstein. 2005. Evaluation of CoViSTA–an Automated Vital Sign Documentation System–in an Inpatient Hospital Setting. In *AMIA Annual Symposium Proceedings*, Vol. 2005. American Medical Informatics Association, Bethesda, Maryland, USA, 885.

[4] Canadian Medical Protective Association. 2009. The medical record: A legal document - Can it be corrected? https://www.cmpa-acpm.ca/en/advice-publications/browse-articles/2009/the-medical-record-a-legal-document-can-it-be-corrected

[5] Lynn Bickley and Peter G Szilagyi. 2012. *Bates' guide to physical examination and history-taking*. Lippincott Williams & Wilkins, Philadelphia, USA.

[6] Ashly D Black, Josip Car, Claudia Pagliari, Chantelle Anandan, Kathrin Cresswell, Tomislav Bokun, Brian McKinstry, Rob Procter, Azeem Majeed, and Aziz Sheikh. 2011. The impact of eHealth on the quality and safety of health care: a systematic overview. *PLoS med* 8, 1 (2011), e1000387.

[7] Claus Bossen, Yunan Chen, and Kathleen H Pine. 2019. The emergence of new data work occupations in healthcare: The case of medical scribes. *International Journal of Medical Informatics* 123 (2019), 76–83.

[8] Dr. David Chan and McMaster's Family Medicine Department. 2021. OSCAR EMR Connecting Care, Creating Community. https://oscar-emr.com/

[9] Basit Chaudhry, Jerome Wang, Shinyi Wu, Margaret Maglione, Walter Mojica, Elizabeth Roth, Sally C Morton, and Paul G Shekelle. 2006. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. *Annals of internal medicine* 144, 10 (2006), 742–752.

[10] Enrico Coiera, Baki Kocaballi, John Halamka, and Liliana Laranjo. 2018. The digital scribe. *NPJ digital medicine* 1, 1 (2018), 1–5.

[11] Roger Collier. 2017. Electronic health records contributing to physician burnout.

[12] Bridget Dolan. 2004. Medical records: Disclosing confidential clinical information. *Psychiatric Bulletin* 28, 2 (2004), 53–56.

[13] Seppo Enarvi, Marilisa Amoia, Miguel Del-Agua Teba, Brian Delaney, Frank Diehl, Stefan Hahn, Kristina Harris, Liam McGrath, Yue Pan, Joel Pinto, Luca Rubini, Miguel Ruiz, Gagandeep Singh, Fabian Stemmer, Weiyi Sun, Paul Vozila, Thomas Lin, and Ranjani Ramamurthy. 2020. 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*. Association for Computational Linguistics, Online, 22–30. https://doi.org/10.18653/v1/2020.nlpmc-1.4

[14] Vickie K Fieler, Thomas Jaglowski, and Karen Richards. 2013. Eliminating errors in vital signs documentation. *CIN: Computers, Informatics, Nursing* 31, 9 (2013), 422–427.

[15] Gregory Finley, Erik Edwards, Amanda Robinson, Michael Brenndoerfer, Najmeh Sadoughi, James Fone, Nico Axtmann, Mark Miller, and David Suendermann-Oeft. 2018. An automated medical scribe for documenting clinical encounters. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*. Association for Computational Linguistics, New Orleans, Louisiana, 11–15. http://www.aclweb.org/anthology/N18-5003

[16] Rebekah L Gardner, Emily Cooper, Jacqueline Haskell, Daniel A Harris, Sara Poplau, Philip J Kroth, and Mark Linzer. 2019. Physician stress and burnout: the impact of health information technology. *Journal of the American Medical Informatics Association* 26, 2 (2019), 106–114.

[17] George A Gellert, Ricardo Ramirez, and S Luke Webster. 2015. The rise of the medical scribe industry: implications for the advancement of electronic health records. *Jama* 313, 13 (2015), 1315–1316.

[18] William Gossman, Valerie Lew, and Sassan Ghassemzadeh. 2020. SOAP Notes. In *StatPearls [Internet]*. StatPearls Publishing, Treasure Island, FL, USA.

[19] Ronald M Harden and FA Gleeson. 1979. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Medical education* 13, 1 (1979), 39–54.

[20] Heather A Heaton, Ana Castaneda-Guarderas, Elliott R Trotter, Patricia J Erwin, and M Fernanda Bellolio. 2016. Effect of scribes on patient throughput, revenue, and patient and provider satisfaction: a systematic review and meta-analysis. *The American journal of emergency medicine* 34, 10 (2016), 2018–2028.

[21] Tobias Hodgson and Enrico Coiera. 2015. Risks and benefits of speech recognition for clinical documentation: a systematic review. *Journal of the american medical informatics association* 23, e1 (2015), e169–e179.

[22] Michael Z Huang, Candace J Gibson, and Amanda L Terry. 2018. Measuring electronic health record use in primary care: a scoping review. *Applied clinical*

informatics 9, 1 (2018), 15.
[23] Serena Jeblee, Faiza Khan Khattak, Noah Crampton, Muhammad Mamdani, and Frank Rudzicz. 2019. Extracting relevant information from physician-patient dialogues for automated clinical note taking. In *Proceedings of the Tenth International Workshop on Health Text Mining and Information Analysis (LOUHI 2019)*. ACL, Stroudsburg, Pennsylvania, USA, 65–74.
[24] Spencer S Jones, Robert S Rudin, Tanja Perry, and Paul G Shekelle. 2014. Health information technology: an updated systematic review with a focus on meaningful use. *Annals of internal medicine* 160, 1 (2014), 48–54.
[25] Erik Joukes, Ameen Abu-Hanna, Ronald Cornet, and Nicolette F de Keizer. 2018. Time spent on dedicated patient care and documentation tasks before and after the introduction of a structured and standardized electronic health record. *Applied clinical informatics* 9, 1 (2018), 46.
[26] David R Kaufman, Barbara Sheehan, Peter Stetson, Ashish R Bhatt, Adele I Field, Chirag Patel, and James Mark Maisel. 2016. Natural language processing–enabled and conventional data capture methods for input to electronic health records: a comparative usability study. *JMIR medical informatics* 4, 4 (2016), e35.
[27] Faiza Khan Khattak, Serena Jeblee, Noah H Crampton, Muhammad Mamdani, and Frank Rudzicz. 2019. AutoScribe: Extracting Clinically Pertinent Information from Patient-Clinician Dialogues.. In *MedInfo (Studies in health technology and informatics)*. IOS Press, Amsterdam, Netherlands, 1512–1513.
[28] Ahmet Baki Kocaballi, Enrico Coiera, Huong Ly Tong, Sarah J White, Juan C Quiroz, Fahimeh Rezazadegan, Simon Willcock, and Liliana Laranjo. 2019. A network model of activities in primary care consultations. *Journal of the American Medical Informatics Association* 26 (2019), 1071–1082. Issue 10.
[29] A Baki Kocaballi, Kiran Ijaz, Liliana Laranjo, Juan C Quiroz, Dana Rezazadegan, Huong Ly Tong, Simon Willcock, Shlomo Berkovsky, and Enrico Coiera. 2020. Envisioning an artificial intelligence documentation assistant for future primary care consultations: A co-design study with general practitioners. *Journal of the American Medical Informatics Association* 27, 11 (2020), 1695–1704.
[30] Kundan Krishna, Sopan Khosla, Jeffrey P. Bigham, and Zachary C. Lipton. 2020. Generating SOAP Notes from Doctor-Patient Conversations. arXiv:2005.01795 [cs.CL]
[31] Philip J Kroth, Nancy Morioka-Douglas, Sharry Veres, Stewart Babbott, Sara Poplau, Fares Qeadan, Carolyn Parshall, Kathryne Corrigan, and Mark Linzer. 2019. Association of electronic health record design and use factors with clinician stress and burnout. *JAMA network open* 2, 8 (2019), e199609–e199609.
[32] Miriam Lacasse and Dara Maker. 2008. Fishing and history taking: From the net to the line. *Canadian Family Physician* 54, 6 (2008), 891–892.
[33] Lena Mamykina, David K Vawdrey, Peter D Stetson, Kai Zheng, and George Hripcsak. 2012. Clinical documentation: composition or synthesis? *Journal of the American Medical Informatics Association* 19, 6 (2012), 1025–1031.
[34] Pranita Mishra, Jacqueline C Kiang, and Richard W Grant. 2018. Association of medical scribes in primary care with physician workflow and patient experience. *JAMA internal medicine* 178, 11 (2018), 1467–1472.
[35] J Marc Overhage and David McCallie Jr. 2020. Physician time spent using the electronic health record during outpatient encounters: a descriptive study. *Annals of Internal Medicine* 172, 3 (2020), 169–174.
[36] Zachary R Paterick, Nachiket J Patel, and Timothy Edward Paterick. 2018. Unintended consequences of the electronic medical record on physicians in training and their mentors. *Postgraduate medical journal* 94, 1117 (2018), 659–661.
[37] Thomas H Payne, W David Alonso, J Andrew Markiel, Kevin Lybarger, and Andrew A White. 2018. Using voice to create hospital progress notes: description of a mobile application and supporting system integrated with a commercial electronic health record. *Journal of biomedical informatics* 77 (2018), 91–96.
[38] Vivek Podder, Valerie Lew, and Sassan Ghassemzadeh. 2020. SOAP Notes. https://www-ncbi-nlm-nih-gov/books/NBK482263/
[39] prestosoft. 2021. Compare and merge files and folders with ExamDiff Pro. https://www.prestosoft.com/edp_examdiffpro.asp
[40] Juan C Quiroz, Liliana Laranjo, Ahmet Baki Kocaballi, Shlomo Berkovsky, Dana Rezazadegan, and Enrico Coiera. 2019. Challenges of developing a digital scribe to reduce clinical documentation burden. *npj Digital Medicine* 2, 1 (2019), 1–6.
[41] Bevinahalli N Raveesh, Ragavendra B Nayak, and Shivakumar F Kumbar. 2016. Preventing medico-legal issues in clinical practice. *Annals of Indian Academy of Neurology* 19, Suppl 1 (2016), S15.
[42] Louis Raymond, Guy Paré, Ana Ortiz de Guinea, Placide Poba-Nzaou, Marie-Claude Trudel, Josianne Marsan, and Thomas Micheneau. 2015. Improving performance in medical practices through the extended use of electronic medical record systems: a survey of Canadian family physicians. *BMC medical informatics and decision making* 15, 1 (2015), 27.
[43] Stephanie Roy. 2021. OSCE orientation. https://mcc.ca/examinations/osce-orientation
[44] Adam Rule, Isaac H Goldstein, Michael F Chiang, and Michelle R Hribar. 2020. Clinical Documentation as End-User Programming. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–13.
[45] Gordon D Schiff and Laura Zucker. 2016. Medical scribes: salvation for primary care or workaround for poor EMR usability? *Journal of general internal medicine* 9 (2016), 979–981. https://doi.org/10.1007/s11606-016-3788-x
[46] Aviv Shachak, Michal Hadas-Dayagi, Amitai Ziv, and Shmuel Reis. 2009. Primary care physicians' use of an electronic medical record system: a cognitive task analysis. *Journal of general internal medicine* 24, 3 (2009), 341–348.
[47] Tait D. Shanafelt, Lotte N. Dyrbye, and Colin P. West. 2017. Addressing Physician Burnout: The Way Forward. *JAMA* 317, 9 (March 2017), 901–902. https://doi.org/10.1001/jama.2017.0076
[48] Cameron G Shultz and Heather L Holmstrom. 2015. The use of medical scribes in health care settings: a systematic review and future directions. *The Journal of the American Board of Family Medicine* 28, 3 (2015), 371–381.
[49] Christine Sinsky, Lacey Colligan, Ling Li, Mirela Prgomet, Sam Reynolds, Lindsey Goeders, Johanna Westbrook, Michael Tutty, and George Blike. 2016. Allocation of Physician Time in Ambulatory Practice: A Time and Motion Study in 4 Specialties. *Annals of Internal Medicine* 165, 11 (Dec. 2016), 753–760. https://doi.org/10.7326/M16-0961
[50] Eric J Topol. 2019. High-performance medicine: the convergence of human and artificial intelligence. *Nature medicine* 25, 1 (2019), 44–56.
[51] Brian D Tran, Yunan Chen, Songzi Liu, and Kai Zheng. 2020. How does medical scribes' work inform development of speech-based clinical documentation technologies? A systematic review. *Journal of the American Medical Informatics Association* 27, 5 (2020), 808–817.
[52] Markus Vogel, Wolfgang Kaisers, Ralf Wassmuth, and Ertan Mayatepek. 2015. Analysis of documentation speed using web-based medical speech recognition technology: randomized controlled trial. *Journal of medical Internet research* 17, 11 (2015), e247.
[53] Katherine J Walker, Will Dunlop, Danny Liew, Margaret P Staples, Matt Johnson, Michael Ben-Meir, Hamish Gordon Rodda, Ian Turner, and David Phillips. 2016. An economic evaluation of the costs of training a medical scribe to work in Emergency Medicine. *Emerg Med J* 33, 12 (2016), 865–869.
[54] Justin M Weis and Paul C Levy. 2014. Copy, paste, and cloned notes in electronic health records. *Chest* 145, 3 (2014), 632–638.
[55] Matt Willis and Mohammad Hossein Jarrahi. 2019. Automating documentation: a critical perspective into the role of artificial intelligence in clinical documentation. In *Information in Contemporary Society*, Vol. 11420. iConference, Springer, Cham, Springer, Cham, 200–209. https://doi.org/10.1007/978-3-030-15742-5_19
[56] Chen Yan, Susannah Rose, Michael B Rothberg, Mary Beth Mercer, Kenneth Goodman, and Anita D Misra-Hebert. 2016. Physician, scribe, and patient perspectives on clinical scribes in primary care. *Journal of general internal medicine* 31, 9 (2016), 990–995.
[57] Kai Zheng, Qiaozhu Mei, Lei Yang, Frank J. Manion, Ulysses J. Balis, and David A. Hanauer. 2011. Voice-Dictated versus Typed-in Clinician Notes: Linguistic Properties and the Potential Implications on Natural Language Processing. *AMIA Annual Symposium Proceedings* 2011 (2011), 1630–1638. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3243272/