MEDINFO 2021: One World, One Health – Global Partnership for Digital Innovation P. Otero et al. (Eds.) © 2022 International Medical Informatics Association (IMIA) and IOS Press. This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0). doi:10.3233/SHTI220054

How the Content of Progress Notes Affects Readers' Perceptions of Their Usefulness and the Burden Placed on Writers: Improving Interprofessional Communication

Rina Kagawa MD, PhD^a, Hideo Tsurushima MD, PhD^b

^a Department of Biomedical Informatics and Management, University of Tsukuba, Tsukuba, Ibaraki, Japan, ^b Department of Neurosurgery, University of Tsukuba, Tsukuba, Ibaraki, Japan

Abstract

Electronic health records should efficiently store the information required for clinical decision-making and contain progress notes that reference this information. However, beyond the inclusion of subjective data, objective data, assessment, and plan framework, the content required to make progress notes useful for readers with diverse specialties has not been clarified. Moreover, the documentation burden that including additional content places on medical doctors (MDs) has not been determined. We conducted a questionnaire with 74 MDs, nurses, and other clinical professionals to determine whether they found progress notes with varying specific contents useful. In addition, the degree of the burden of writing progress notes that contain specific content was measured when 25 MDs were instructed to add specific content. Our results reveal that progress notes are more useful for clinical reasoning for readers other than MDs when more specific information is included; this can be achieved without increasing the documentation burden

Keywords:

Electronic Health Records, Clinical Reasoning, Clinical Decision-Making

Introduction

Electronic health records (EHRs) should support clinical decision-making as well as billing and data display. Detailed personal data, such as patients' emotional status, recommended future tests, and treatment plans, should be efficiently stored and referenced in EHRs alongside a range of other necessary medical data for clinical reasoning, which is the basis of clinical decision-making. Since there is no widely accepted terminology for clinical reasoning [1], it is currently difficult to store the information it requires as structured data in EHRs. Therefore, a practical solution is to include the information that is useful for clinical reasoning and decision making in the progress notes. Progress notes are medical documents written by medical doctors (MDs) as part of their daily clinical practice. These are regularly read by clinical professionals and patients from diverse backgrounds who use them in their own clinical reasoning and clinical decisions (e.g., nursing diagnoses). The widespread use of EHRs and the implementation of standards for data exchange have made it easier for multiple clinical professionals to read progress notes both inside and outside of the hospital. We believe that progress notes will play an increasingly important role as foundational documents for communication among clinical professionals and patients.

Documents consist of content and arrangement [2]. The subjective data, objective data, assessment, and plan (SOAP) framework, which is based on the problem-oriented medical record [3], is widely accepted worldwide as the ideal for progress notes [4]. However, some research has pointed out issues with the SOAP framework, or has proposed new systems for determining the ideal content and arrangement of progress notes, although they have not been widely accepted [5–9]. In this paper, we focus on whether or not to describe more specific and detailed content (e.g., the reason for deciding on a treatment policy [10]) beyond the contents of S, O, A, P, which is not defined in this framework and depends on the writers, even when following the SOAP framework. To our knowledge, there is no experimentally determined consensus on what content will be considered useful for clinical reasoning and decision-making by readers with diverse specialties. Mnemonics, such as OPQRST (onset, palliation/provocation, quality, region/radiation, symptom/severity, and timing) [11] and SAMPLE (signs and symptoms, allergies, medications, past medical history, last oral intake, and events leading up to the injury or illness) [12], are used by MDs to gather the information required for clinical reasoning. However, they only represent the information that should be collected, and the progress notes that document such information are not always useful for clinical reasoning for readers with different background knowledge. There is also a detailed analysis of the contents of patients' care context data described in the actual progress notes [13]; however, their usefulness to the reader has not been examined.

In addition to considering the readers of progress notes, we also focus on the writers of progress notes. This is because, in general, it is not easy to write documents that readers with different backgrounds will all find useful. In busy clinical practices, even if deeper, more detailed content that is useful for readers' clinical reasoning and decision-making is revealed, if it differs from conventional practice, the MDs bear the burden of documenting this useful content [14]. Therefore, it is important to develop appropriate support systems that take into account the burden that MDs face when writing progress notes and discuss what kind of support should be developed, such as, providing appropriate instructions for writers or mechanical and automatic supporting systems (for example, an authoring support system based on the sentence generation technique).

Our basic idea is that progress notes that balance the burden on writers and the utility for readers are the ideal documents for improving interprofessional communication in the EHR era. In this research, as a first step in proposing progress note design and description support in the EHR era, we focus on the content of progress notes and conduct experiments to test how the content of progress notes affects both the burden on writers and the utility for readers. The objectives of this study are as follows:

- Clarify what kind of content is useful for various clinical professionals in the progress note.
- Clarify the documentation burden on MDs by requesting certain content—that is useful for various clinical professionals—be added in the progress notes and discuss what kind of support should be developed that will be effective for both writers and readers.

Materials and Methods

Materials

Two datasets were used for this study. First, we used the progress notes, medication data, laboratory test results, and injection records of actual 56 Japanese patients (mean age 65.8 [13.0], female 51.2%). The sample was randomly selected from 53,246 patients who visited University of Tsukuba Hospital between January 1, 2013 and September 30, 2018. All the data were preprocessed to ensure the protection of patient privacy before this study was carried out [15]. Second, we randomly selected 25 out of 9,856 public pseudo progress notes1 [16].

Methods

This study consists of three experiments, and the threshold for significance was p < 0.01 for all three. All the experiments were conducted in Japanese but have been translated into English for this manuscript. Perl 5.30 and Python 3.7.2 were used.

Experiment 1: The contents to be analyzed and the descriptions of these contents in the actual progress notes

The contents examined in this study were determined by referring to 26 relation concepts [1] between clinical terms, which were clarified based on a review of clinical research papers (clinical problem-solving cases published in the *New England Journal of Medicine*). Seven types of content (Table 1) were selected from the 26 relation concepts and were selected based on the following: the types of content that were extremely difficult to record in a structured manner² and the types of content that were described regardless of specific medical situation³. It could be concluded from the results of our study that a template should be created that will enable a statement to be written that is useful to the reader. Given this, we selected less than nine [17] types of content, as that is the number of elements that humans can memorize using short-term memory.

For the seven types of content, two MDs (the authors of this manuscript) annotated 40 actual progress notes from the University of Tsukuba Hospital (20 notes for inpatients and 20 notes for outpatients) and 25 public pseudo progress notes with a phrase meaning each of the seven types of content. The phrases were separated by punctuations or conjunctions, If the annotations did not match, the final annotation was decided through discussion. Examples of descriptions for each type of content are shown in Table 1. For 40 actual progress notes, we used Fisher's exact test to compare the percentage of notes that had phrases of each content between notes of inpatients and notes of outpatients, and used Welch's t-test to compare the

mean number of descriptions per note for each content between notes of inpatients and notes of outpatients.

<i>Table 1 – Examples of each type of content and the corre-</i>
sponding phrase in the progress notes. The phrases annotated
with the content are in italics.

Content	Sentence in progress notes
Absence of findings	Glucose was 75-196-147-, and
	there are no subjective symptoms of
	hypoglycemia.
Alteration by activity	Bradycardia has been improved.
Causes	The results of yesterday's blood test
	showed some improvement, sug-
	gesting that the patient has re-
	sponded to the antibiotics.
Description of	CEA is less sensitive to early-stage
knowledge/experience	colorectal cancer and is not suita-
	ble for screening.
Intention of intervention	Insulin administration was discon-
	tinued because blood glucose levels
	had improved significantly.
Progress and prognosis	The condition would change rapidly
	due to both thrombosis and tumor
	growth.
Recommendation	CT will be scheduled next Monday.

Experiment 2: The effects of the presence or absence of specific content on the usefulness perceived by the readers

We investigated which content increased the usefulness perceived by readers with diverse specialties.

For each of the 25 pseudo progress notes in which all the content types were annotated in Experiment 1, the experimental notes were created in which the phrases annotated with each type of content were deleted from the original pseudo progress notes (hereafter called the "original notes"). If the original note contained phrases for two types of content, two experimental notes were created from the original note (Figure 1). If there were multiple phrases with the same content in one original note, one experimental note was created in which all the phrases of the content were deleted from the original note. For experimental notes, minimal adjustments were made only when the connections between postpositions were unnatural due to the deletion of phrases of each content.

The participants were as follows: 11 MDs (mean age 39.5 [7.9], female 27.2%), 30 nurses (mean age 38.3 [9.3], female 80.0%), and 33 other clinical professionals (mean age 36.5 [9.5], female 36.4%). The specialties of the MDs were, cardiology (4 MDs), ophthalmology, pediatric surgery (2 MDs for each), nephrology, neurosurgery, and orthopedics. Each participant viewed randomly selected an original note and its corresponding experimental note, and answered seven questions for the original note and the corresponding experimental note. Each participant repeated this task 50 times. Seven subjective questions were used to analyze whether the usefulness perceived by the readers differed between the original notes and the experimental notes (Table 2). All questions were answered with a ranking on an

¹ This dataset does not include information about each pseudo patient's sex, age, or the medical department where the pseudo notes were written.

² *Time*, which is one of the 26 relation concepts, was excluded because the structured recording of time is not extremely difficult.

³ Discontinue therapy, which is one of the 26 relation concepts, was excluded from our analysis because discontinue therapy is not relevant when the medical examination results in continuing treatment. In other words, *discontinue therapy* is related to a specific medical situation in which medical examination has determined that the treatment is discontinue.

11-point Likert scale. One question "Do you think that this document is useful in clinical practice?" (10: useful – 0: useless) is for usefulness. There are a wide variety of factors that can contribute to perceptions of usefulness in the readers. Therefore, referring to IEEE Std830-1998 [18], which is the standard that software requirements specifications should meet, and research on accounting documents [19], other six questions were also used. The score for each of the seven questions for each type of note is called "the usefulness score" in this manuscript. NASA-TLX [20], which is a famous index that measures workers' perceptions of usefulness and mental workload do not necessarily match.

For each type of content and question, the null hypothesis that the usefulness scores would be normally distributed was rejected using Shapiro-Wilk test. For each type of content and question, Welch's t-tests were performed on the null hypothesis that the mean value of the usefulness scores for the original note and the corresponding experimental note would be equal. We also performed multiple comparisons (Tukey-Kramer test) between specialties for each type of content. Answers that were missing a usefulness score, either for the original note or the experimental note, were excluded from the analysis. The readability scores of each original and experimental note were also evaluated by calculating the Japanese readability scores [21]⁴, which are divided into 13 grade levels (1: 1st year of elementary school [6 years old] – 13: university [18 years old and above]).



Figure 1 – Example of the original note and experimental notes for two types of content.

Experiment 3: The effects of the instruction to describe each type of content on the writers' documentation burden

We examined whether the burden on the MDs would change with the obligation to cover useful content when compared to cases where the requirements for the notes were open-ended.

The participants were 19 MDs (mean age 37.5 [6.5], female 26.3%). The specialties of participants were, cardiology (6 MDs), general internal medicine, ophthalmology, pediatric surgery, dermatology, urology (2 MDs for each), nephrology, neurosurgery, and orthopedics. The experimental procedure was as follows: 1) the participants described the dummy progress notes for the patient, who were randomly selected from University of Tsukuba Hospital, on a screen prepared for the experiment that simulated actual EHRs in Japan, 2) the participants answered the question "Do you think that this document was easy to write?" according to an 11-point Likert scale (10: easy to write -0: difficult to write; hereafter, the resulting score is called the "writability score"), 3) the participants were instructed to add one type of content that was randomly selected from the seven types of content (e.g., the participant may receive the message "please add information about the absence of findings"), 4) participants followed the instructions and added the information to the dummy progress note they wrote in the first part of the procedure, and 5) the participants answered the same question

shown in the second part of the procedure. Each participant repeated this task eight times. For each type of content, the null hypothesis that the writability scores would be normally distributed was rejected using Shapiro-Wilk test. For each type of content, Welch's t-tests were performed on the null hypothesis, wherein the mean values of the writability scores before and after instruction would be equal. Two participants whose descriptions did not change after all the instructions and answers that were missing a writability score, either for before instruction or after instruction, were excluded from the analysis.

Moreover, other six MDs (mean age 48.7 [9.9], female 0.0%) participated in another experiment. The specialties of each participant were, general internal medicine, general surgery, pediatrics, psychiatry, ophthalmology, and palliative care. The same experimental procedure as the one described above was used, but in the third part of the procedure, the participants were instructed to add an arbitrary number that was randomly selected from the seven content types (e.g., a participant could receive the following instruction: "please add information about the absence of findings, alteration by activity, and causes").

Table 2 – Characteristics of usefulness perceived by the readers and questions presented to the participants. All questions began with "Do you think that this document --?"

Characteristics	Question for the readers
(Q1) usefulness	is useful in clinical practice
(Q2) understandability	is easy to understand
(Q3) correctness	does not contain any descriptions that are not related to medical treatment
(Q4) unambiguity	is not ambiguous
(Q5) completeness	contains everything you want to know
(Q6) consistency	has no contradiction
(Q7) verifiability	can be verified

Results

Experiment 1: The contents to be analyzed and the descriptions of these contents in the actual progress notes

The percentages of the progress notes that had phrases of each content were as follows: absence of findings = 44.6%, alteration by activity = 41.5%, causes = 27.7%, description of knowledge/experience = 6.2%, intention of intervention = 53.8%, progress and prognosis = 35.3%, and recommendation = 32.3%. The null hypothesis that the percentages of notes that had phrases of each content would be equal between notes of inpatients and notes of outpatients was not rejected for all types of content. The mean (standard deviation) numbers of descriptions per note were 0.78 (1.13), 0.83 (1.38), 0.52 (1.01), 0.14 (0.64), 0.84 (1.13), 0.52 (0.82), and 1.00 (1.94), respectively. The null hypothesis that the mean number of descriptions per note would be equal between notes of inpatients and notes of outpatients was not rejected for all types of content. These results can be interpreted to mean that there was not sufficient inclusion of the types of content that were considered to be useful for clinical reasoning in the progress notes, regardless of whether the patient was inpatient or outpatient.

⁴ Satoh S. [Internet]. NagoyaObi 3.0.1. Available from: http://kotoba.nuee.nagoya-u.ac.jp/sc/obi3/.

Experiment 2: The effects of the presence or absence of specific content on the usefulness perceived by the readers

Figure 2 showed the differences between the usefulness scores of the original notes and the corresponding experimental notes for each type of content and question.

When the readers were MDs, the null hypothesis that the mean usefulness score of the original note and that of the experimental note would be equal was not rejected for all types of content types and all questions. While, when the readers were nurses, the null hypothesis was rejected for absence of findings (Q01, Q02, Q05), alteration by activity (Q01, Q02, Q04, Q05), causes (Q01, Q02, Q04, Q05, Q07), intention of intervention (Q01, Q02, Q04, Q05, Q07), and recommendation (Q01, Q02, Q04, Q05, Q07). When the readers were other clinical professionals, the null hypothesis was rejected for absence of findings (Q01, Q02, Q04, Q05), alteration by activity (Q01, Q02, Q04, Q05), causes (Q01, Q02, Q04, Q05, Q07), intention of intervention (Q01, Q02, Q04, Q05, Q07), and recommendation (Q01, Q02, Q04, Q05, Q07). All effect sizes were under 0.5. In these significance tests, Q05, which focused on whether the note contained what the readers wanted to know, had the same results as Q01, which focused on usefulness.



Figure 2 – Differences of the usefulness scores of the original notes compared to those of the corresponding experimental notes. Others means other clinical professionals.

Among the combinations of content types and questions that were found to be significantly better by nurses and other clinical professionals, for Q07, the null hypothesis that the mean usefulness score of the original notes and that of the experimental notes would be equal was rejected for alternation by activity, recommendation, and intention of intervention by the Tukey-Kramer test. This suggested that the degree to which progress notes are rendered useful based on the presence of certain content might differ depending on specialty.

The mean differences of the readability scores for original notes compared to the experimental notes for each type of content were as follows: absence of findings = 0.00 (0.78), alteration by activity = -0.18 (1.38), causes = -0.07 (0.26), description of knowledge/experience = 0.00 (0.00), intention of intervention = 0.16 (0.96), progress and prognosis = 0.00 (0.00), and recommendation = -0.24 (1.04). These results revealed that the readability of the notes did not change by the inclusion of additional content; however, the readers' perceptions of the usefulness of the notes for clinical reasoning improved.

Experiment 3: The effects of the instruction to describe each type of content on the writers' documentation burden

Figure 3 showed the differences between the writability scores before and after the instruction. The null hypothesis that the writability scores before and after instruction would be equal was not rejected for all types of content. The p-values and effect sizes were as follows: absence of findings = 0.85 (0.07) (n =16), alteration by activity = 0.87 (0.07) (n = 21), causes = 0.77(0.12) (n = 14), description of knowledge/experience = 0.55 (0.21) (n = 17), intention of intervention = 0.94 (0.03) (n = 17), progress and prognosis = 0.47 (0.23) (n = 21), and recommendation = 0.45 (0.21) (n = 26). In the experiment in which seven MDs were instructed to add an arbitrary number of content types, the null hypothesis that the writability scores before and after instruction would be equal was not rejected for all number of content types instructed to be described. When the number of types of content instructed to be described is one to seven, the p-values and effect sizes were as follows: one = 1.00(0.00)(n = 4), two = 0.03 (1.90) (n = 5), three = 0.87 (0.09) (n = 8), four = 1.00(0.00)(n = 8), five = 0.68(0.18)(n = 12), six = 0.86(0.09) (n = 7), and seven = 0.64 (0.34) (n = 4).

There were seven cases in total in which the medical department of the patient with the dummy progress notes and the specialty of the MD who wrote the notes matched. There was no difference in the tendency of answers between such cases and other cases. The dummy progress notes were confirmed by one MD (first author), and no descriptions that were clearly contrary to the additional instructions were found. These results showed that asking MDs to add useful content after freely writing the progress notes did not increase documentation burden significantly independent of types of content and the number of contents to be instructed to be added.



Figure 3 – Differences of the writability scores after instruction compared to those before instruction.

Discussion

This study found that the usefulness of progress notes as perceived by nurses and other clinical professionals increased significantly with the addition of absence of findings, alteration by activity, causes, intention of intervention, and recommendation, even if the usefulness perceived by MDs did not increase. Our results also showed that this could be achieved without increasing the documentation burden.

Applicability

The finding in regards to asking MDs to add useful content after freely writing the progress notes might result in progress notes that are more useful for various readers without increasing the documentation burden revealed that even very simple authoring support is effective and can be immediately applied in various clinical situations. In the near future, we will try to continue the experiments by assuming more specific clinical reasoning, clarifying the impact of the how the progress notes are arranged, and discussing the effects of rhetoric.

Generalizability

One of the potential reasons for the difference between nurses' or other clinical professionals' perceptions of the usefulness of particular content additions and that of MDs is that MDs can use their background knowledge to supplement the information in progress notes, while nurses and other clinical professionals may not be able to do so. In experiment 2, the question that had the same results as usefulness (Q01) was completeness (Q05). This result was considered to support the above consideration. This study suggested that the readers' background knowledge has a strong influence on their perception of usefulness. In the future, we would like to propose an automatic description support system for personalized descriptions for writers and readers by examining their background knowledge in detail.

Limitations and future work

The data used in this study were acquired exclusively from Japan, and the EHR narrative notes were written in Japanese, so there may be unforeseen problems caused by the language. Since these experiments can be carried out regardless of the language or health care system, we expect that additional experiments will be conducted in many countries and regions.

Conclusion

This study found that simply requesting that MDs add useful content after freely writing the progress notes might result in progress notes that are more useful to readers other than MDs without increasing the documentation burden. This understanding is crucial for designing a description support system for progress notes that is effective for the better interprofessional communication.

Ethics

This research was approved by the Ethics Committee of the University of Tsukuba Hospital (permission number: H30-145, R2-310). Our research was carried out with the opt-out consent of patients. Opt-in informed consent was obtained from clinical professionals who engaged in the experiments.

Acknowledgements

This study was supported by the Japan Science and Technology Agency (JST)-Mirai, Program Grant Number JPMJMI19G8, the Japan Society for the Promotion of Science (JSPS), Grantsin-Aid for Scientific Research (Nos. JP18H06363, JP19K19347), and the R&D Center for Frontiers of MIRAI in Policy and Technology. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. Makoto Ohara MD, PhD, Yuji Kozasa RN, Masako Sato, and Mika Sumimoto contributed to the proofreading procedures for privacy protection. Emiko Nishida contributed to the calculation of the readability scores.

References

- Cimino JJ et al., An exploration of the terminology of clinical cognition and reasoning. *AMIA Annu Symp Proc 2018* (2018), 321– 329.
- [2] Reiter E et al., Building Natural Language Generation Systems. Cambridge: Cambridge University Press. (2000)
- [3] Weed LL. Medical records, medical education and patient care: the problem-oriented record as a basic tool. Cleveland (OH): Press of Case Western Reserve University; 1970.
- [4] Wright A et al., Bringing science to medicine: an interview with Larry Weed, inventor of the problem-oriented medical record. *JAMIA* **21** (2014), 964–968.
- [5] Lenert LA, Toward medical documentation that enhances situational awareness and learning, *AMIA Annu Symp Proc 2016* (2016), 763–771.
- [6] Lenert LA et al., Rethinking the discharge summary: a focus on handoff communication. Academic Medicine 89 (2014), 393-398.
- [7] Meyers KC et al., Problem focused knowledge navigation: implementing the problem focused medical record and the O-HEAP note, *AMIA Annu Symp Proc 1998* (1998), 325–329.
- [8] Donnelly WJ et al., Why SOAP is bad for the medical record, Archives of Internal Medicine 152 (1992), 481–484.
- [9] Tange H et al., Towards an overarching model for electronic medical-record systems, including problem-oriented, goal-oriented, and other approaches. *European Journal of General Practice* 23 (2017), 257–260.
- [10] Cimino JJ, Putting the "why" in "EHR": capturing and coding clinical cognition. *JAMIA* 26 (2020), 1379–1384.
- [11] Cutrer WB et al., Educational strategies for improving clinical reasoning. *Current problems in pediatric and adolescent health care* **43** (2013), 248–257.
- [12] Pollak AN et al., Stratford Emergency Care and Transportation of the Sick and Injured (9th ed.), Jones and Bartlett, Sudbury, MA (2005), 148–149.
- [13] Colicchio TK et al., Formal representation of patients' care context data: the path to improving the electronic health record. *JAMIA* 27 (2020), 1648–1657.
- [14] Weir CR et al., G. Feeling and thinking: can theories of human motivation explain how EHR design impacts clinician burnout? *JAMIA* 28 (2021), 1042–1046.
- [15] Kagawa R et al., Generating publicly available progress notes of authentic quality: A crowdsourcing-based approach. *JMIR Preprints* (2020), 1-37. Available from: http://doi.org/10.2196/preprints.24790 [cited 2021 May 15]
- [16] Kagawa R et al., Publicly available medical text data with authentic quality, Available from: https://zenodo.org/record/4064153 [cited 2021 May 15]
- [17] Gentner D et al., Structure mapping in analogy and similarity. Am Psychol 52 (1997), 45–56.
- [18] Software Engineering Standards Committee of the IEEE Computer Society. IEEE recommended practice for software requirements specifications. 830 (1998)
- [19] Malcolm S et al., Readability and understandability: Different measures of the textual complexity of accounting narrative. Accounting, Auditing Accountability Journal 5 (1992), 84–98.
- [20] Hart SG, Nasa-Task Load Index (NASA-TLX); 20 Years Later. HFES Annu Meet Proc 50 (2006), 904–908.
- [21] Sato S et al., Automatic assessment of Japanese text readability based on a textbook corpus. CLRE Proc 2008 (2008), 654–660.

Address for correspondence

Rina Kagawa is the corresponding author on this report. The email address is kagawa-r@md.tsukuba.ac.jp.