Data from clinical notes: a perspective on the tension between structure and flexible documentation

S Trent Rosenbloom, Joshua C Denny, Hua Xu, Nancy Lorenzi, William W Stead, Kevin B Johnson

ABSTRACT
Clinical documentation is central to patient care. The success of electronic health record system adoption may depend on how well such systems support clinical documentation. A major goal of integrating clinical documentation into electronic heath record systems is to generate reusable data. As a result, there has been an emphasis on deploying computer-based documentation systems that prioritize direct structured documentation. Research has demonstrated that healthcare providers value different factors when writing clinical notes, such as narrative expressivity, amenability to the existing workflow, and usability. The authors explore the tension between expressivity and structured clinical documentation, review methods for obtaining reusable data from clinical notes, and recommend that healthcare providers be able to choose how to document patient care based on workflow and note content needs. When reusable data are needed from notes, providers can use structured documentation or rely on post-hoc text processing to produce structured data, as appropriate.

INTRODUCTION
The process and products of documenting clinical care occupy a critical intersection among the diverse domains of patient care, clinical informatics, workflow, research, and quality. For the current manuscript, we define clinical documentation as the process of creating a text record that summarizes the interaction between patients and healthcare providers occurring during clinical encounters. Clinical documents produced in this process may include notes from outpatient visits, inpatient admissions and discharges, procedures, protocols, and testing results. Healthcare providers generate clinical documents to achieve numerous goals, including: create narrative reports of their observations, impressions, and actions related to patient care; to communicate with collaborating healthcare providers; to justify the level of service billed to third-party payers; to create a legal record in case of litigation; and to provide data to support clinical research and quality-assessment programs. Increasingly widespread adoption of electronic health record (EHR) systems has facilitated reuse of clinical documents for research, quality initiatives, and automated decision support, among other uses. A major emphasis for integrating documentation systems with EHR systems is to increase the availability of structured clinical data for automated downstream processes. This has led to a profusion of computer-based documentation (CBD) systems that promote real-time structured clinical documentation.

The myriad requirements imposed on clinical documentation compel healthcare providers to create notes that are simultaneously accurate, detailed, reusable, and readable. As a result, integrating clinical documentation into workflows that contain EHR systems has proven a challenge. The complex interplay among the note characteristics healthcare providers value, the structure and standardization that data reuse requires, and the attributes of various documentation methods each affect the documentation method’s adoption into clinical workflows. The flexibility of a CBD method to allow healthcare providers freedom and ensure accuracy can directly conflict with a desire to produce structured data to support reuse of the information in EHR systems. The challenge is compounded when those implementing CBD systems have different priorities from clinician users working in busy settings. In particular, we have previously demonstrated that healthcare providers prefer the ability to balance using a standardized note structure and having the flexibility to use expressive narrative text.

This viewpoint paper explores the tension between requirements that documentation methods support both structure and expressivity, then reviews two general approaches for obtaining structured data through clinical documentation, and provides our perspective of how best to use CBD systems in a clinical environment.

TENSION BETWEEN STRUCTURE AND EXPRESSIVITY
There exists a tension between the needs of busy healthcare providers documenting clinical care and of those reusing data from healthcare information systems. Busy clinicians generally value flexibility and efficiency, while those reusing data often value structure and standardization. In investigations evaluating healthcare providers’ impressions of documentation systems, subjects articulated that the documentation methods they use should promote the quality and expressivity of notes they generate and should integrate efficiently into busy workflows. Expressivity has been defined as how well a note conveys the patient’s and provider’s impressions, reasoning, and thought process; level of concern; and uncertainty to those subsequently reviewing the note. Expressivity refers to the linguistic nuance necessary for describing aspects of the patient encounter using any words or phrases that

Perspective

Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, Tennessee, USA

Correspondence to
Dr S Trent Rosenbloom, Eskind Biomedical Library, Room 440, 2209 Garland Avenue, Nashville, TN 37232-8340, USA; trent.rosenbloom@vanderbilt.edu

Received 9 July 2010
Accepted 10 December 2010
Published Online First 12 January 2011
Computer programs can then deduce that users document patient care episodes using relatively unstructured approaches, such as using dictation with transcription, speech-recognition software, or typing using a loosely templated CBD system. Once the clinical documentation is complete, post-hoc text processing algorithms can be used to produce structured data. We use the term ‘text processing’ to describe any of a number of methods designed to identify specific text, data, and concepts from the natural language stored in unstructured, narrative-text computer documents. Text-processing tools can then deduce concepts contained in the notes’ text in a subsequent step. Investigators have worked for decades to convert natural language into structured representations of those concepts. Text-processing technologies have been developed with differing levels of machine ‘understanding,’ from simple systems that search for key words or specific text strings to those systems that attempt to capture clinical concepts with their context. Text-processing tools can also serve as adjuncts to structured documentation.
systems, as recently described by Johnson discussing ‘structured narrative’.101

A basic approach to text processing involves searching documents in EHR systems for key narrative-text strings or string patterns. The narrative-text string search has been successfully employed in some large-scale clinical research studies.82–85 For example, researchers have used plain text searches of dictated or typed medical records to find rare physical exam findings82 and post-operative infections,85 and identify patients with certain types of drug-induced liver injury.84 Others have employed focused ‘regular expression’ pattern matches to extract text strings that can represent possible blood pressure85 and common section header labels (eg, ‘chief complaint’) in clinical notes,86 and remove patient identifiers such as names, phone numbers, and addresses to deidentify medical records.87 These methods are best adapted for solving focused problems. Since string-matching algorithms can be highly tuned for a given task (eg, by including manually derived synonyms, common abbreviations, and even misspellings), they can perform very well. Simple text searching also has an advantage of faster processing speed and easier implementation than complex natural-language processing (NLP) systems, as many off-the-shelf database systems and text-indexing tools support preindexed text queries. However, simple text searching is limited by a lack of generalizability and requires substantial customization for each new task. For example, finding all patients with liver injury from a single medication (such as phenytoin) is easily accomplished with a simple text search, but finding all medications that may be associated with liver injury requires a more complicated system that matches text with a controlled vocabulary that would include all medication brand and generic names. In addition, problems associated with directly entered clinical notes, such as misspellings102 and ambiguous abbreviations (eg, ‘pt’ can stand for ‘patient’ or ‘physical therapy’),103 also limit the use of the simple string-matching methods.

A more complex approach, called concept identification or concept indexing, attempts to normalize the text phrases to standardized terms representing concepts. Successful concept indexing systems include those by Miller and Cooper,89 MetaMap76, SAPHIRE90, the KnowledgeMap concept identifier,80 91 the Multi-threaded Clinical Vocabulary Server92 93 104 and the recently released clinical Text Analysis and Knowledge Extraction System (cTAKES).105 A goal of concept indexing is to ‘understand’ the information in natural-language documents by mapping text to standardized concepts using terminologies such as those in the Unified Medical Language System (UMLS).106 These systems have proven effective, mapping natural-language texts to concepts with recall and precision often exceeding 80% for general tasks and near perfect for some highly focused tasks.90 95 107 108

Investigators have extended concept identification systems to combine them with additional algorithms that can extract contextual elements (eg, certainty, values, and temporal information) associated with a concept to form more robust NLP systems. Combining concept identification systems with negation detection algorithms (eg, ‘no chest pain’, which indicates the absence of the finding ‘chest pain’), investigators have created systems that automatically generate problem lists,109 discover gene–disease associations,110 hypothesize new drug effects111 and new drug–drug interactions from the biomedical literature,96 and identify important findings from clinical narratives.97 96 Building on the work of the Linguistic String Project,79 Friedman and colleagues have been developing the MedLEE (Medical Language Extraction and Encoding System) NLP system since the 1990s.78 95 112 It can identify UMLS concepts from clinical documents, and can discern their timing and negation status. Researchers from several institutions have used MedLEE to identify pneumonia and other clinical conditions from chest x-rays,99 detect adverse events,112 and automatically calculate the Charlson comorbidity index.113 Several of these systems have been successfully incorporated into production clinical systems, some at multiple institutions.78 95 109 112 114 115

The uses of NLP in the clinical domain are quite diverse, including encoding clinical notes for billing purposes, facilitating clinical research by automating the data-extraction processes,110 117 118 conducting EHR-based surveillance, 97 119 facilitating clinical research by automating the data-extraction processes,110 117 118 conducting EHR-based surveillance, 97 119 and enriching EHR functionalities, such as to support visualization tools120 and clinical decision-support systems121 122 and clinical decision-support systems. One of the earliest examples is the MedLEE system,123 which has been used to process chest radiographs to generate coded data for decision-support systems at New York Presbyterian Hospital since 1995. Recently, Day and colleagues121 reported the use of the MPluS NLP system to classify trauma patients at a Level 1 trauma center on a daily basis. For specific tasks such as those mentioned above, advanced NLP systems such as MedLEE have shown equivalent performance as domain experts. However, the performance of current clinical NLP systems is still not satisfactory for broader uses. In addition, few NLP systems have been implemented in clinical settings and used in routine workflows in hospitals. For example, Kashyap et al, reported a study of using a commercial product to automatically structure admission notes in which the results were not judged to be acceptable for clinical use.124 With further improvements in advanced NLP technologies and new structures for clinical notes (such as the ‘structured narrative’),101 NLP may be able to structure clinical text such as admission notes automatically with a satisfactory performance.

DISCUSSION

The choice of a documentation method can alter the balance between expressivity and structure in the resultant notes, hamper the healthcare provider’s workflow, influence the process and products of recording clinical information, and influence how well the note can be incorporated into an EHR system in such a way that the note’s contents can be automatically reused and analyzed.15 19 26 36 70 101 While structured documentation systems can facilitate data collection and reuse, they can be cumbersome to use during patient encounters and may lack the flexibility and expressivity required for general medical practices. Transcribed notes create documents useful for text processing, but can require a time delay for the transcription process to occur. The attributes associated with each documentation method influence how they are best used and adopted. While structured entry emphasizes data standardization and structure, human adoption of CBD systems requires an emphasis on expressivity, efficiency, flexibility, and being well adapted to a typical workflow.19 22 27

Both structured clinical documentation and text-processing algorithms for flexible documentation continue to evolve. The rates of evolution may not be the same for each, but we are not aware of clear evidence that shows that one approach is improving faster than the other. Furthermore, we do not intend to convey an impression that one is necessarily more advanced or has greater promise than the other. Each approach offers distinct advantages to the user. For example, because structured
documentation systems can do more than just create data, it may be relatively appealing to clinicians in certain settings. A template-based structured documentation system may be useful for notes that have a standard and predictable content and format, such as those recording pediatric health maintenance visits, preoperative evaluations, formalized disability examinations, and reviews of systems. By contrast, in some settings, clinicians may prefer the flexibility of complex and nuanced narrative text to document the history of a present illness or a diagnostic impression.

We collectively have over 10 decades of personal experience in viewing or using a wide variety of computer-based documentation systems around the country in public, private and academic settings. We gained this experience in the settings of actual clinical use, observations during site visits, and direct reviews of vendor products at trade shows and professional conferences. Based on this experience, we have observed that typically, for a given site’s implementation, computer-based documentation systems are usually configured to take clinical input primarily in narrative or structured form, but not both. Most computer-based documentation systems that the authors have seen do not support hybrid documentation in the way that we recommend in the current manuscript. We note that, in the case where structured clinical data are needed, information entered using structured entry would be immediately available, whereas narrative text would require processing via NLP algorithms to become similarly ‘structured.’ Nevertheless, we observe that in our experience, few nonacademic sites have NLP-based text processing systems readily available to harvest EHR system data, with the possible exception of NLP-type systems that help to extract billing codes from clinical records.

Priorities for structured documentation and text processing will likely vary in different settings and with different user groups, and the priorities may be informed by the tasks the documentation supports. Certain tasks may require that the EHR systems understand only structural information from the clinical document (eg, its metadata or header information), while others may require a deeper concept-level understanding. For example, a quality-assessment program evaluating whether inpatient progress notes are being placed in the chart in a timely manner may require only the document title and date. Any document that has the correct title, including scanned and tagged paper-based notes, could support this need. Tasks that depend on a deeper understanding of the document may require healthcare providers to document using structured documentation tools or to apply computer programs to process the natural language in narrative-text documents. For example, to implement an automated colorectal cancer screening advisor, the advisor system will need to know the patient’s family and past medical history, whether a prior screening test had occurred, and the result of any screening tests. Gathering these data requires that healthcare providers document the relevant information using tools that can support data extraction and analysis, including CBD systems, structured entry tools, or narrative-text typing (either directly or via a dictation/transcription model) with a subsequent application of text processing. In addition, different healthcare providers may value structured data in EHR systems differently. Some providers may be willing to sacrifice a degree of CBD usability or efficiency for the sake of having key elements of their notes available immediately as structured data in EHR systems. Others may preferentially value documentation methods that allow them to express their impressions fluidly or create notes more quickly, without regard to how well the methods support other functions of EHR systems.

**Recommendation**

Given the tension between structure and flexible documentation, those implementing EHR systems should assume that multiple CBD products will be needed to meet the needs of clinician users, rather than attempting to find a single best documentation method. Factors to consider in selecting CBD products include the method’s fitness for a given workflow, the content in the resultant notes, time efficiency to create a note, costs, ease of using the method, flexibility for using the method to document unforeseen clinical findings or in unexpected circumstances, and support for narrative expressivity, machine readability, and document structure. Certain documentation methods feature some attributes at the expense of others, such as promoting narrative expressivity at the expense of formal note structure. Rather than waiting until documentation tools can be created to accommodate all needs and workflows, or force complex clinical workflows to change to accommodate the EHR system rollout, we recommend that healthcare providers be able to access a variety of different documentation methods and select the one that best fits their documentation, data, and workflow needs. EHR system developers and users can weigh different documentation methods in terms of how they impact relevant documentation-related outcomes such as usability, efficiency, quality, and readability against their utility for EHR systems. The value of this approach is that it allows organizations considering EHR systems to prioritize development and implementation efforts around clinical documentation.

**Funding**

The project was supported by a grant from the United States National Library of Medicine (Rosenbloom, 1R01LM009591-01A1).

**Competing interests**

None.

**Provenance and peer review**

Not commissioned; externally peer reviewed.

**REFERENCES**
