MEDINFO 2004
M. Fieschi et al. (Eds)
Amsterdam: IOS Press

Generating Complex Clinical Documents Using Structured Entry and Reporting

S. Trent Rosenbloom[a], Wendy Kiepek[a], John Belletti[b], Perry Adams[a], Kiki Shuxteau[a], Kevin B. Johnson[a], Peter L. Elkin[c], Edward K. Shultz[a]

[a] *Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA*
[b] *School of Medicine, Vanderbilt University, Nashville, TN, USA*
[c] *Mayo Foundation for Medical Education and Research, Rochester, MN, USA*

Abstract

Structured entry and reporting in medicine remains an elusive goal. Poor adoption of clinical structured entry for documentation results in part from the inherent complexity of entering patient histories, which are generally unstructured. The authors have developed a structured entry tool that has been adopted by practicing physicians for documentation of clinical encounters. To evaluate the impact of this tool on clinical documentation, the authors have performed two comparative studies investigating note complexity. Authors compared documents generated with a standard dictation/transcription model with documents generated with structured entry. Overall, documents generated with the structured entry and reporting tool contained 64% more concepts (P<0.01) than dictated documents while maintaining the same complexity. Depth and complexity of documentation with the structured entry and reporting tool varied by clinician user and by note sub-section.

Keywords:

Medical Records Systems, Computerized; Efficiency; Structured Entry and Reporting

Introduction

Stakeholders in basic science research, health services research and medical informatics recognize the importance of information captured directly from episodes of clinical care for research, real time decision support, and patient screening for clinical trials or clinical syndromes.[1-4] Structured entry and reporting systems promise to meet this need by enabling health care providers to document patient clinical encounters through selection from predefined categorical concepts.[5-11]

Structured entry and reporting systems are designed to enhance the process of clinical documentation by simultaneously presenting useful categorical concepts in a user interface and capturing input into the interface as machine-readable data.[Kahn, 1998 #76] Major goals of such systems include producing raw data that enhances patient care by facilitating reminders and alerts while providing an infrastructure for clinical research. Structured entry is achieved as a user navigates through lists of relevant concepts and sets status value (e.g. present/absent, etc.). In parallel with the clinician generating a document, the tool can capture data suitable for research and for real-time clinical decision support and can generate reports in various human-readable formats. Additionally, because structured entry and reporting tools can integrate documentation templates to reduce variability in the quality and quantity of concepts recorded in individual notes, it is also possible for investigators seamlessly to integrate research questions into the workflow of general practitioners.

Despite the efforts of pioneers in the field of biomedical informatics, who have created many structured entry and reporting systems over the past 50 years, few systems outside of isolated subspecialty domains or single institutions have experienced ongoing voluntary usage. Examples include, but are by no means limited to, the applications by Stead and Slack, including TMR, that allow patients to enter their own histories,[9, 10, 13-16] by Ledley in the domain of radiology,[17] the COSTAR project by Barnett,[18] Weed's PROMIS system,[19] the ARAMIS system that collects national data about patients with rheumatoid arthritis,[20] the endoscopy documentation tool, CORI,[7, 21] Wirtschafter[22] and Shortliffe's[23] independent works developing chemotherapy documentation and decision support systems, Johnson's Clictate, which assists with guideline-based documentation of pediatric care,[11] and Musen's T-HELPER[24] which advises community physicians if their patients meet criteria for enrollment in HIV-related clinical trials or modification of medical therapies. These milestones notwithstanding, creation and widespread adoption of computer-assisted clinical documentation tools remains both a major challenge and a central target in informatics.[25]

The described benefits of individual structured entry and reporting systems in healthcare notwithstanding, published studies investigating the reasons for their subsequent limited acceptance remain rare.[25] Possible reasons for poor adoption that have been suggested include system inefficiency,[8] complexity and slow pace of navigating through user interfaces to find relevant content,[26] inflexibility for documenting unforeseen findings, the lack of integration of clinical applications, and deficiéncies in both coverage by and goals of the underlying data model,[25] and decreased overall efficiency for generating complex documents. Interface inflexibility and difficulty locating relevant content has a greater impact on the relatively unstructured note components, such as the narrative history of present illness, than on the relatively structured sections like the physical exam and system review documentation.

Taking into account the speculative reasons for limited adoption of structured entry and reporting systems, the authors have developed a novel structured entry and reporting tool called QUILL (QUestions and Information Logically Linked) at Vanderbilt University Medical Center (VUMC). The major goals in developing QUILL were to create a tool that permitted rapid entry of categorical clinical information, to allow input by generalists and sub-specialists, to support multiple methods of data-entry, and to output data in both human- and machine-readable form. QUILL, a Java-based application that is a component of our electronic health record (EHR), was designed for use across multiple clinical domains, so that it can be used to detail the clinical assessment of a healthy two-month-old child and to document a history and physical exam of an elderly patient with chest pain. QUILL supports a collaborative workflow so that a nurse and physician can work together to create a final comprehensive document. Vital sign and laboratory data may be imported or automatically populated from the EHR. Medications and problems are retrieved from, edited and stored in a longitudinal summary list that resides within the EHR. Documentation using QUILL occurs through the use of user-customizable templates, which are collections of terms and concepts from an underlying pilot clinical interface terminology. Templates prompt documentation of particular problems or to follow specific guidelines.

To date, nine cardiologists and two clinical fellows have used QUILL over the past two years in general cardiology and arrhythmia specialties to generate over 2500 medical encounter summaries in their clinic. Additionally, QUILL has recently been implemented by six neurologists and one nurse practitioner in adult neurology clinics, including stroke and epilepsy specialties, where an additional 1300 medical encounter summaries have been generated. All physician users have been able to achieve documentation times that meet or improve upon their time to complete a note prior to using QUILL.

Methods

This study tested whether QUILL users generated clinical documents containing the same depth and representing the same complexity as standard documentation using a dictation/transcription model, especially for note sections that are generally structured. Depth was defined as the rates of concepts documented per note; complexity as the levels of service and decision-making represented in the notes, as indicated by independent Evaluation and Management code assignments. Using a retrospective pre/post design, the authors compared notes generated by three cardiologists using standard dictation/transcription before the introduction of QUILL with notes generated by the same cardiologists using QUILL after its introduction. For this study, authors (WK, JB, STR) identified a random selection of 99 clinical encounter summaries: (50 generated using dictation, 49 using QUILL; 33 from each cardiologist) representing new evaluations of patients with a similar degree of complexity. Complexity designation was based on the visit type coded in the institutional demographic database. All QUILL notes, taken from the institutional electronic health record in their final form, included physician-entered free-text comments typed in place of structured entry. The study protocol and methods were approved by the VUMC Institutional Review Board (IRB protocol #30493).

To test whether the introduction of QUILL altered the volume of clinical content documented from clinical encounters, the authors counted clinical concepts in each note using the definition put forth by the Canon Group; according to their explanation, concepts consist of meaningful pathophysiologic entities or events which can be represented by words or phrases.[27] For example, the injury and death of heart muscle cells resulting from the prolonged occlusion of a coronary artery is a clinical concept which can evoke the terms 'myocardial infarction' and 'heart attack'. Authors used the definition of modifiers presented by the International Standards Organization (ISO), "strings which, when added to the term, changes the meaning of the term in a clinical sense."[28] The study definitions, as modified for the operational task of chart abstraction, are presented in Table 1.

Concepts	manifestation of a physical normality or abnormality as signs, symptoms, and physical examination findings, etc
Modifiers	details that modify or quantify atomic concepts with additional information related to severity, timing, quality, associations, etc.
Historical Symptoms	abnormal change of structure, function, or sensation freely described by the patient without inquiry by the interviewer
Historical Signs	abnormal change of structure, function, or sensation mentioned by the patient as a result of inquiry
Physical Examination Findings	abnormal change of structure, function, or sensation determined by observation, inspection, auscultation, palpation, percussion, or elicitation by a clinician

Table 1: Definitions used for manual concept extraction

Two members of the research team,(JB, STR) with consultation of a third (KBJ) for cases of ambiguity, together identified and quantified the number and category of clinical concepts and modifiers present in each note. Reviewers were unblinded as to the notes' source. The research team expected variation among the three physicians in terms of both baseline documentation styles and the impact of QUILL. To evaluate the effect of introduction of QUILL on the rates of concepts documented per note, the research team used nonparametric statistics with Bonferroni correction for multiple comparisons to compare the rates of concepts per notes and per subsections, and logistic regression to correct for the impact of individual users' styles on the rates of concepts per note.

To test whether the introduction of QUILL altered the complexity of clinical documents, three professional coders scored the same 99 notes using the 1995 documentation guidelines for evaluation and management (E&M) services, published by the Centers for Medicare and Medicaid Services (CMS).[29] The authors had professional coders re-code these notes rather than use the originally coded and billed level of service to reduce any variability that could have been introduced over time to coding practices in the clinic. The 1995 guidelines score medical docu-

ments based on the number of findings representing the various note sections. Coders were unblinded to the notes' source. The overall score represents the 'level of service,' and provides a framework for billing to Medicare and Medicaid. For example, the presence or absence of a finding under chief complaint, the number of findings and their modifiers under the history of present illness, and the number of body systems listed in the physical examination all contribute points to the document's complexity. For the purpose of the analysis, all level of service assignments were converted to a 5 point scale, with 1 being low level of service (i.e. minimal) and 5 being high level of service (i.e. comprehensive). Additionally, the 1995 guidelines require assignment of a level of 'medical decision making' to restrict clinicians from generating inappropriately detailed notes (i.e. a clinician simply treating a simple upper respiratory infection in an otherwise healthy patient cannot generate a complex note simply to increase the billable level of service). Using their standard procedures, the three professional coders each scored 33 documents, with QUILL notes and authors evenly distributed between them. To evaluate the effect of introduction of QUILL on the complexity of documentation, the research team used non-parametric statistics with Bonferroni correction for multiple comparisons and logistic regression.

Results

The first study compared the numbers of clinical concepts and modifiers in documents generated using traditional dictation/ transcription with documents generated using QUILL. All 99 notes were available for the analysis. The baseline rates of concepts and modifiers per note and the changes in that rate with adoption of QUILL varied by subject (Table 2). Correcting for the impact of individual subjects' documentation styles demonstrated that QUILL was associated with an overall 64% increase in mean concepts per note from 65 to 103. The authors performed sub-analyses of the impact of QUILL on individual sections of notes, including the history of present illness (HPI), the aggregate past medical, family, and social histories (FMSH), the physical examination (PE), and the assessment and plan (AP). These analyses demonstrated an increase in the mean rates of concepts per note using QUILL in the FMSH section (19 with dictation, 40 with QUILL; P<0.01), and the PE section (38 with dictation, 64 from QUILL; P<0.01). By contrast, dictated notes contained more concepts in the HPI section (17 with dictation, 14 with QUILL; P<0.05).

Subject	Concepts			Modifiers		
	Dict.	Quill	P	Dict.	Quill	P
A	50	134	<0.01	15	26	NS
B	39	109	<0.01	13	16	NS
C	120	106	NS	27	16	<0.01

Table 2: Mean concepts and modifiers per note, by author. NS - nonsignificant

The second study compared the complexity of documents generated using traditional dictation/transcription with documents generated using QUILL. A total of 97 notes were available for analysis; two were inpatient notes misidentified as outpatient notes. The average level of service and documented medical decision making that could be coded from the tested notes varied by subject, but for the most part did not change with usage of QUILL (Table 3). A separate analysis investigating the impact of the volume of concepts and modifiers on the coded decision making demonstrated that for every additional HPI concept modifier, there was a small but significant increase in the decision making score (OR 1.06, P<0.01). However, the volume of HPI concepts had no observable impact. The research team performed sub-analysis of the impact of QUILL on the scores generated by individual note sections, including the chief complaint (CC), the HPI, the review of systems (ROS), FMSH, and the PE. These analyses demonstrated an increase in subsection coding using QUILL for the ROS section (mean of coded systems 7.5 with dictation, 9.4 with QUILL; P=0.02). and a decrease in the HPI section (3.7 bullets with dictation, 2.9 with QUILL; P<0.01). There were no significant differences in the proportion of notes with a documented CC or the numbers of coded elements in the FMSH or PE sections.

Subject	Level of Service			Decision Making		
	Dict.	Quill	P	Dict.	Quill	P
A	3.9	4.0	NS	2.8	2.8	NS
B	3.6	3.8	NS	2.6	2.6	NS
C	4.6	3.9	<0.01	3.1	2.8	NS*

*Table 3: Mean level of service and documented complexity of medical decision making per note, by author. NS – nonsignificant; * P=0.06*

Discussion

Successful development of a structured entry and reporting tool that is useful across clinical domains may lead to improved data capture from standard clinical documentation processes. Like that generated from natural language processing and concept-based indexing of free-text documents, categorical data captured from structured entry has the potential to drive real-time decision support algorithms, population-based screening for diseases and for clinical trials eligibility, clinical and health services research, and other documentation processes. Improving the interfaces and usability of structured entry and reporting systems is a key component of ongoing work towards capturing categorical data from clinical documents.

The author's initial evaluation suggests that the QUILL structured entry and reporting environment is useful for documenting clinical encounters with as much detail and complexity as with a standard dictation/transcription model. The impact of the tool on documentation styles varied by clinician and by note section. In general, using the tool, clinicians who generated terse notes were able to increase the depth of their notes while maintaining complexity. The structured entry and reporting tool may have been more useful for documenting the relatively structured ROS, the FMSH, and the PE than for unstructured elements like the HPI. The relatively low rate of concepts documented for HPI with QUILL is consistent with informal feedback received from users, who have commented that the narrative reporting typical of the HPI is difficult to accomplish in a structured entry interface.

There was variability in document length and complexity among the three physician subjects, with one generating at baseline more lengthy and complex notes than the other two. QUILL us-

age was associated with an increase in detail for the two subjects who started with relatively sparse notes, but not for the subject who started at a high level of detail. The authors speculate that this variability is consistent with underlying variability in documentation practices present in the institution. The authors did not separately investigate the cause of the variability in baseline documentation patterns or of the impact of QUILL on documentation detail. It is possible that difference in detail resulted from QUILL improving the efficiency over dictation for selecting and assigning status to the relatively structured note sections. It is also possible that the difference resulted from unmeasured factors, such as differences in the patient populations served by the three physician subjects.

The reported results have four limitations that merit discussion. First, this study evaluated a small number of notes generated by only three physicians in a single institution, using a single structured entry and reporting tool, and at two sequential points in time. It is possible that the outcomes resulted could have been biased by using notes written by authors who are 'early adopters' of technology, and therefore are not representative of all cardiologists or clinicians. Although the authors are not aware of any, it is also possible that an unmeasured change in documentation practices which occurred around the time of the introduction of QUILL caused the outcomes reported by this study, rather than QUILL itself. In addition, this investigation included an equivalency study; it is possible that the observed absence of an impact on the E&M codes by QUILL resulted from inadequate power due to small sample size.

Second, a structured entry and reporting tool depends heavily on its associated interface terminology. The terminology supporting QUILL has not been independently studied, and its impact on the results is unknown. It is possible that terminology deficiencies could attenuate the outcomes from the two studies, decreasing the ability of QUILL to model common phrases. Likewise, it is possible that the terminology models more concepts than a clinician typically considers, increasing the rate of documented concepts per note. Because performance of a structured entry and reporting tool relates to its underlying terminology, the results for both studies may not generalize to situations in which a different terminology supported QUILL. It is also possible that the structured text generated by Quill is easier or harder than free text to map to a clinical concept, potentially biasing the concept counting. However, the reviewer that the mapping was not difficult, decreasing this concern.

Third, the observations reported suggest that QUILL can be used to document common findings, and, when used, can generate more complete documentation. These outcomes are not a proxy for usability. Although actual use of a tool may provide a reasonable initial measure of usability,[30] early adopters may not represent all possible users. Other measurable outcomes that could give a more complete picture of usability include the time required to train to use the tool, the time needed to generate a note, the ease of searching for, selecting, and setting status to a needed concept, the degree to which input data is relevant to a given case, and the quality of the output data and prose.[31] Long-term followup of usage, specific analysis of the impact of each design feature, and surveys of provider satisfaction will also provide further information regarding usability.

Fourth, the reviewers were not blinded to the source of the study notes. QUILL output is readily identifiable as a result of it's sparse structure and outline format. Blinding of note source would have required altering the notes, which could potentially have introduced bias.

Conclusion

A usable and flexible structured entry and reporting tool may increase depth and complexity of clinical documentation, especially of more categorical note sections such as the review of systems, the past medical, family, and social histories, and the physical examination.

Acknowledgments

This work was supported in part by Office of the Dean, Vanderbilt University School of Medicine. The authors would also like to thank Carol Hanschke, Patricia Kendall, and Sandra Miller-Davis for their efforts related to E&M coding.

References

[1] Committee on Quality of Health Care in America, *Using Information Technology. Crossing the Quality Chasm: A New Health System for the 21st Century.* Washington, D.C.: IOM; 2001.

[2] Committee on Improving the Patient Record. *The Computer-Based Patient Record: An Essential Technology for Health Care*, 2 ed. Washington, D.C.: IOM; 1991.

[3] Use of a data-based approach by a health maintenance organization to identify and address physician barriers to pediatric vaccination--California, 1995; 1996.

[4] Tsui FC, Espino JU, Wagner MM, Gesteland P, Ivanov O, Olszewski RT, et al. Data, network, and application: technical description of the Utah RODS Winter Olympic Biosurveillance System. *Proc AMIA Symp* 2002:815-9.

[5] Kahn CE, Jr., Huynh PN. Knowledge representation for platform-independent structured reporting. *Proc AMIA Annu Fall Symp* 1996:478-82.

[6] Yoder JW, Schultz DF, Williams BT. The MEDIGATE graphical user interface for entry of physical findings: design principles and implementation. Medical Examination Direct Iconic and Graphic Augmented Text Entry System. *J Med Syst* 1998;22(5):325-37.

[7] CORI: Clinical Outcomes Research Initiative. In. p. http://www.cori.org.

[8] Lum F, Schein O, Schachat AP, Abbott RL, Hoskins HD, Jr., Steinberg EP. Initial two years of experience with the AAO National Eyecare Outcomes Network (NEON) cataract surgery database. *Ophthalmology 2000*;107(4):691-7.

[9] Stead WW, Brame RG, Hammond WE, Jelovsek FR, Estes EH, Parker RT. A computerized obstetric medical record. *Obstet Gynecol* 1977;49(4):502-9.

[10]Slack WV, Hicks GP, Reed CE, Van Cura LJ. A computer-based medical-history system. *N Engl J Med* 1966;274(4):194-8.

[11]Johnson KB, Cowan J. Clictate: a computer-based documentation tool for guideline-based care. *J Med Syst* 2002;26(1):47-60.

[12]Kahn CE, Jr. Self-documenting structured reports using open information standards. *Medinfo 1998*;9(Pt 1):403-7.

[13]Slack WV, Peckham BM, Van Cura LJ, Carr WF. A computer-based physical examination system. *Jama* 1967;200(3):224-8.

[14]Stead WW, Heyman A, Thompson HK, Hammond WE. Computer-assisted interview of patients with functional headache. *Arch Intern Med* 1972;129(6):950-5.

[15]Stead WW, Hammond WE, Estes EH. Evaluation of an audio mode of the automated medical history. *Methods Inf Med* 1977;16(1):20-3.

[16]Hammond WE. How the past teaches the future: ACMI distinguished lecture. *J Am Med Inform Assoc* 2001;8(3):222-34.

[17]Ledley RS. Computer aids to medical diagnosis. *Jama* 1966;196(11):933-43.

[18]Barnett GO. COSTAR, a computer-based medical information system for ambulatory care. *Proc IEEE 1979*;67:1226-1237.

[19]Stratmann WC, Goldberg AS, Haugh LD. The utility for audit of manual and computerized problem-oriented medical record systems. *Health Serv Res* 1982;17(1):5-26.

[20]Bruce B, Fries JF. The Stanford Health Assessment Questionnaire: a review of its history, issues, progress, and documentation. *J Rheumatol* 2003;30(1):167-78.

[21]CORI: Clinical Outcomes Research Initiative Publications List. In. p. http://cori.ohsu.edu.

[22]Wirtschafter DD, Scalise M, Henke C, Gams RA. Do information systems improve the quality of clinical research? Results of a randomized trial in a cooperative multi-institutional cancer group. *Comput Biomed Res* 1981;14(1):78-90.

[23]Shortliffe EH. Update on ONCOCIN: a chemotherapy advisor for clinical oncology. *Med Inform* (Lond) 1986;11(1):19-21.

[24]Musen MA, Carlson RW, Fagan LM, Deresinski SC, Shortliffe EH. T-HELPER: automated support for community-based clinical research. *Proc Annu Symp Comput Appl Med Care* 1992:719-23.

[25]McDonald CJ. The barriers to electronic medical record systems and how to overcome them. *J Am Med Inform Assoc* 1997;4(3):213-21.

[26]Poon AD, Fagan LM, Shortliffe EH. The PEN-Ivory project: exploring user-interface design for the selection of items from large controlled vocabularies of medicine. *J Am Med Inform Assoc* 1996;3(2):168-83.

[27]Evans DA, Cimino JJ, Hersh WR, Huff SM, Bell DS. Toward a medical-concept representation language. The Canon Group. *J Am Med Inform Assoc* 1994;1(3):207-17.

[28]ISO/TS 17117:2002(E): Health Informatics - Controlled health terminology - Structure and high-level indicators: Technical Commitee ISO/TC 215, *Health Informatics*; 2002.

[29]1995 Documentation Guidelines For Evaluation & Management Services. In: *Centers for Medicare & Medicaid Services*.

[30]Rogers EM. *Diffusion of innovations,* pp 204-251. 4th ed. New York: Free Press; 1995.

[31]Neilson J. *Usability Engineering*, pp 23-48. Boston, MA: Academic Press; 1993.

Address for correspondence

S. Trent Rosenbloom, MD MPH
2209 Garland Avenue, #440
Nashville, TN, USA 37232-8340
trent.rosenbloom@vanderbilt.edu