Original JAMIA Investigations

Research Paper ■

Preliminary Development of the Physician Documentation Quality Instrument

PETER D. STETSON, MD, MA, FRANCES P. MORRISON, MD, MPH, SUZANNE BAKKEN, RN, DNSc, STEPHEN B. JOHNSON, PhD, FOR THE eNOTE RESEARCH TEAM

Abstract **Objectives:** This study sought to design and validate a reliable instrument to assess the quality of physician documentation.

Design: Adjectives describing clinician attitudes about high-quality clinical documentation were gathered through literature review, assessed by clinical experts, and transformed into a semantic differential scale. Using the scale, physicians and nurse practitioners scored the importance of the adjectives for describing quality in three note types: admission, progress, and discharge notes. Psychometric methods including exploratory factor analysis were applied to provide preliminary evidence for the construct validity and internal consistency reliability.

Results: A 22-item Physician Documentation Quality Instrument (PDQI) was developed. Exploratory factor analysis (n = 67 clinician respondents) on three note types resulted in solutions ranging from four (discharge) to six (admission and progress) factors, and explained 65.8% (discharge) to 73% (admission and progress) of the variance. Each factor solution was unique. However, four sets of items consistently factored together across all note types: (1) up-to-date and current; (2) brief, concise, succinct; (3) organized and structured; and (4) correct, comprehensible, consistent. Internal consistency reliabilities were: admission note (factor scales = 0.52–88, overall = 0.86), progress note (factor scales = 0.59–0.84, overall = 0.87), and discharge summary (factor scales = 0.76–0.85, overall = 0.88).

Conclusion: The exploratory factor analyses and reliability analyses provide preliminary evidence for the construct validity and internal consistency reliability of the PDQI. Two novel dimensions of the construct for document quality were developed related to form (Well-formed, Compact). Additional work is needed to assess intrarater and interrater reliability of applying of the proposed instrument and to examine the reproducibility of the factors in other samples.

■ **J Am Med Inform Assoc.** 2008;15:534–541. DOI 10.1197/jamia.M2404.

Introduction

Many health care institutions have begun using and evaluating documentation in electronic health records (EHRs).[1–10] Electronic notes can improve legibility[3,11] and availability.[3] Electronic documentation has the potential to improve the quantity and quality of data available in electronic health records, enabling automated decision support and data analysis. Electronic notes also could enable distribution within an institution's various clinical systems or exchange across institutions to improve clinical communication.[12,13] However, the impact of electronic documentation on the quality of clinical communication is unknown. For the past 5 years, the eNote Research Team at Columbia University Medical Center's Department of Biomedical Informatics has been studying several aspects of electronic documentation.[5,9,14,15] This article reports on our preliminary steps to develop an instrument to measure the quality of clinical notes.

Background

Unintended Consequences of Electronic Documentation

Despite its growing use and great promise, there are certain consequences of electronic documentation that suggest it is not a panacea. It takes physicians longer to write electronic notes than paper notes,[16] which may lead to input errors[17] and increased reliance on time-savers like "cut and paste" or "copy forward."[18] "Cut and paste," although perhaps useful and reasonable in some cases, may result in documentation errors or unreadable notes due to outdated or extraneous information.[4,17–19] One study of the impact of computerized physician documentation by physicians found that it led to an increase in length, redundancy, poor formatting, and more overall "clutter" in clinical notes.[3]

Documentation to Support Transitions

As transitions in care become more common due to the emergence of hospital medicine,[20] the quality of documenta-

Affiliations of the authors: Department of Biomedical Informatics (PDS, FPM, SB, SBJ), School of Nursing (SB), Department of Medicine (PDS), Columbia University, New York, NY.

Supported by National Library of Medicine Training Grant 5T15LM007079-12 (FM), National Library of Medicine grant 1R01 LM 07268-01 (SJ, PDS, SB), National Library of Medicine grant 1K22 LM 8805-01 (PDS), National Institute of Nursing Research grant 1P20 NR007799 (SB).

Correspondence: Dr. Peter D. Stetson, Columbia University Medical Center, Division of General Medicine, 622 West 168th Street, PH9 East, Room 105, New York, NY 10032; e-mail: <peter.stetson@dbmi.columbia.edu>.

Received for review: 02/14/07; accepted for publication: 03/10/08

tion for clinical communication becomes more important.[21] Increased involvement of hospitalists in caring for inpatients, heretofore followed up only by their own primary care givers, introduces discontinuity that could be mitigated by improved communication and documentation, especially with high-quality discharge summaries.[22,23] A tool for assessing the quality of discharge summaries to support transitions might contribute to enhanced patient safety.

Pedagogical Opportunities

Writing high-quality notes is, in part, a training issue. However, residents receive little formal training in how to write good notes for communication, or billing, or legal purposes. Residents self-report that they are not confident in their knowledge about coding for billing purposes.[24] Investigators have promoted the idea of educational interventions early in training to encourage good documentation practices,[4,25,26] in one study leading to successful teaching interventions to enable better coding.[26]

A few studies have evaluated the quality of resident-authored notes. Some investigators have evaluated methods using one-on-one record review: chart audit or chart stimulated recall.[25] Others have utilized standardized patients or objective structured clinical examinations (OSCE) to assess the quality of resident documentation.[27–31] These approaches are important steps in the systematic evaluation of resident-authored notes, but the validity and reliability of the instruments have not been established, or the methods are resource-intensive, requiring the training of standardized patients or development of standardized scenarios.

Template Design in Electronic Health Records

Another challenge that might be aided by an instrument to measure document quality is the development of templates within EHRs. Implementation of commercial EHRs often requires client-side development of documentation templates, and the ability of the client to customize the output of these notes is sometimes limited. Furthermore, electronic documentation could be a platform for delivery of guidelines.[32,33] However, the effect of these "smart" templates on the quality of the produced notes is unknown.

Previous Approaches for Measuring Documentation Quality

Previous reports on clinical documentation quality have evaluated user preference of one format of document entry versus another,[34,35] determining the appropriateness of clinical decisions[6] or treatment,[1] and correctness and completeness—the degree to which specific components of the encounter are accurate and captured.[6,7,27,36–42] However, there are still gaps in knowledge about how best to evaluate clinical documents in a valid and reliable way, and few published instruments.[31] One problem is the way in which previous instruments have been developed. A review of the literature reveals few studies that result in an instrument that demonstrates construct validity and internal consistency reliability. In addition, many of the instruments created have been narrow in scope or were designed to be used in specific clinical situations.[27–29,31,39,43] None were designed to be generalizable for use in evaluating multiple types of clinical documents (e.g., admission, progress, discharge notes). Moreover, past approaches to evaluation of documents have typically placed the focus on content over form. Prior work has emphasized whether the information contained in a document is accurate and complete and whether the actions of the clinician were appropriate, which are all characteristics that principally reflect content. Although content is an important feature, form (in some studies referred to as style)[29,31] is also crucial for a document to convey critical information to other care providers. Consequently, we believe both form and content should be evaluated in a systematic way. We hypothesized that form-related factors would emerge as important components of document quality.

Gaps in Knowledge

Questions that remain unanswered include: What constitutes a good note? Do notes written in an EHR differ in quality from notes written with pen and paper or dictation? Do notes written electronically, but through different entry methods (free text, semistructured, fully structured), differ in quality? Is the effect of "copy and paste" on perceived document quality measurable? Can a simple instrument to measure document quality be useful to determine the effectiveness of educational initiatives for physician documentation? Does embedding guidelines or core measures data elements into electronic templates adversely affect the saliency or readability of notes? To what extent does output format affect the quality of clinical notes? To answer these important questions, we need a simple-to-use, validated, reliable way to measure the quality of physician documentation.

The quality of physician documentation is dependent on what function the documentation is meant to serve. Physician documents are used for many purposes and judged for quality on many different metrics that may not be congruent. For example, a physician may write a note intended to communicate the patient's clinical status for a colleague without regard for whether the note is compliant or would generate a "comprehensive" bill. As such, it might be deemed of excellent quality for the purposes of clinical communication, but poorly compliant and not supportive of a "comprehensive" bill.

This article reports on our preliminary effort to develop and validate an instrument to evaluate the quality of physician documentation to support clinical communication. Our use of established psychometric methods attempts to addresses several methodological issues with previously reported instruments regarding construct validity and internal consistency reliability.

Methods

Scope

In this study, we have limited the scope to the use of physician documentation to support clinical communication with other providers engaged in the care of the patient. We specifically did not predefine high quality in this functional context, as the purpose of this study is to discover key dimensions by assembling many published attributes and compressing them into the simplest instrument that captures most of the important features of high-quality notes.

General Psychometric Approach and Theoretical Basis

Quality is an attribute that is judged by the individual who is reading a note, and this reflects that person's needs, values, and expectations. Because individuals determine subjectively whether notes are "good," and we are developing a measure-

ment instrument based on these attitudes, psychometric approaches are appropriate to use in this context. The semantic differential scale and factor analysis are important psychometric tools used in instrument development.[44] The semantic differential is used to assess attitudes toward a concept, object, or individual. It consists of bipolar scales that are anchored with contrasting adjectives at each end, and respondents numerically rate toward the adjective with which they agree. A factor analysis is then used to group items into a set of factors that are related to the underlying theoretical constructs of interest, in this case identifying the major underlying characteristics that clinicians feel contribute to document quality.

As the theoretical basis for instrument development, we used the four factors recommend by the Institute of Medicine (IOM) for data quality in medical records: legibility, accuracy, completeness, and meaning.[45] These also formed the hypothesized constructs for the factor analysis.

Study Design

We used a descriptive, correlational design to assess the following questions: (1) What attributes do providers believe characterize high-quality physician notes? (2) What are the psychometric properties of the Physician Documentation Quality Instrument (PDQI)?

Definitions

Content validity is defined as the extent to which a measure represents all aspects of a given social construct.[44] For example, an instrument for documentation quality would lack content validity if it only measured completeness of a note, but failed to assess succinctness. Construct validity is defined as the extent to which an instrument measures the hypothetical construct that it purports to measure.[44] It can be used to determine whether a common factor underlies several different items. Face validity is the extent to which an instrument appears to measure what it purports to measure.[46]

Participants

All participants in this study were voluntary. This study was approved by the internal review board at Columbia University Medical Center (CUMC). Seven clinical experts were used to hone the list of possible adjectives derived from our literature review for development of the semantic differential scale component of the study. These clinical experts included one nurse and six physicians (three in internal medicine, one in rehabilitation medicine, one surgeon, and one radiologist). Three of the clinical experts were also study investigators (PS, FM, SB). Attending physicians, fellows, residents, and nurse practitioners in the Department of Medicine at CUMC were invited to participate in the survey for the factor analysis component of the study; 35 residents, 2 fellows, 22 attending physicians, and 4 nurse practitioners responded (3 additional respondents did not specify their training).

Development of the Semantic Differential Scale

A review of the published literature was performed to gather a comprehensive list of adjectives addressing document quality to be incorporated into the semantic differential scale. As no MeSH concept to describe document quality exists, this review consisted of combined keyword searches in Medline, including quality, standards, and clinical with documentation, note, narrative, medical record, and chart. A group of clinical experts (described above) was convened to establish the face validity of the preliminary adjective list. These seven clinicians were asked via an e-mail survey to rate the preliminary list of adjectives. Adjectives were ranked for importance as an attribute of quality in physician documents on a Likert scale from 1 to 5. Suggestions for new adjectives were also solicited from the experts to ensure that important attributes not mentioned in the scientific literature were included. The highest-rated adjectives were included in the semantic differential scale. Antonyms for the adjectives for inclusion in the semantic differential scale were selected by study investigators (FM, SJ, PS) using a thesaurus and group consensus. These activities focused on perceptions of domain completeness; no formal measure of content validity was used.

Construct Validity—Factor Analysis

To establish construct validity, a factor analysis was performed on clinicians' responses to a formal survey based on the semantic differential instrument. Participants in the survey were given a list of the adjectives and their antonyms separated by a 1–7 scale, with alternating position of positive and negative connotation. The following instructions were provided:

> Select the position on the scale that best describes your idea of a high-quality admission note. For example, you would say "A high-quality admission note is _____." The closer the position is toward an adjective, the more strongly you feel it describes a high-quality document.

The three document types—admission note, progress note, and discharge summary—were rated as separate concepts, with a separate set of identical adjective pairs for each type of document.

An invitation for voluntary participation in the survey was e-mailed to all internal medicine residents and general medicine attending physicians at CUMC. The e-mail contained a link to a web-based version of the instrument. Responses from the electronic form were collected automatically into a database. We included nurse practitioners at CUMC because they write the same types of notes as attending physicians and residents. Nurse practitioners were given paper versions of the instrument. Data collected by paper forms were added to the database manually.

After the data were cleaned, descriptive statistics were calculated. Semantic differential responses were then analyzed with factor analysis using SPSS 11.5 for Windows. Separate exploratory factor analyses were performed on discharge summary, admission, and progress notes. Discharge summaries were chosen as the basis for the naming of factors and for comparison with the hypothesized underlying factor structure, i.e., the IOM factors for data quality in medical records: legibility, accuracy, completeness, and meaning. The rationale for this choice was that discharge summaries are the note type most often evaluated for quality in the literature, thereby allowing comparison with other study results. They also contain the clinical reasoning contained in the admission note combined with a summarization of hospital stay reflected in the progress note, and this level of comprehensiveness may allow a higher level of generalizability.

The intercorrelation matrix of responses for each document type was submitted to a principal components factor analysis

Table 1 ■ Semantic Differential Adjectives, Mean Scores, and Standard Deviations

Item	Mean Score (SD)			Reference Number
	Admission	Progress	Discharge	
1. Clear/unclear	6.90 (0.35)	6.85 (0.36)	6.85 (0.44)	11, 29, 51
2. Up-to-date/out-of-date	6.74 (0.56)	6.93 (0.32)	6.64 (0.79)	17, 18
3. Complete/incomplete	6.12 (1.63)	6.06 (1.34)	6.43 (1.09)	6, 7, 17, 25, 27, 36–43, 45
4. Legible/illegible	6.76 (0.80)	6.79 (0.45)	6.82 (0.39)	3, 11, 17, 25, 45
5. Accurate/inaccurate	6.73 (1.07)	6.84 (0.37)	6.87 (0.42)	3, 17, 25–27, 41, 42, 45
6. Thorough/superficial	6.26 (1.06)	5.48 (1.48)	5.96 (1.31)	28
7. Uncluttered/cluttered	6.00 (1.10)	6.39 (0.87)	6.30 (1.29)	3, 18
8. Coherent/incoherent	6.64 (0.87)	6.71 (0.55)	6.68 (0.62)	*
9. Useful/useless	6.80 (0.47)	6.45 (1.22)	6.83 (0.41)	3
10. Correct/incorrect	6.80 (0.50)	6.81 (0.53)	6.91 (0.29)	6, 7, 27, 36–42
11. Brief/lengthy	4.27 (1.50)	5.88 (1.20)	4.79 (1.65)	3, 4, 18, 29, 51
12. Current/outdated	6.64 (0.90)	6.86 (0.35)	6.74 (0.59)	17, 18
13. Organized/disorganized	6.67 (0.59)	6.70 (0.66)	6.73 (0.62)	3, 29
14. Relevant/irrelevant	6.49 (0.82)	6.72 (0.49)	6.61 (0.68)	3, 25, 29
15. Comprehensible/incomprehensible	6.67 (0.56)	6.60 (0.91)	6.65 (0.81)	3
16. Concise/verbose	5.88 (1.35)	6.43 (0.87)	5.93 (1.25)	*
17. Structured/unstructured	6.60 (0.66)	6.54 (0.77)	6.65 (0.64)	*
18. Nonredundant/redundant	6.15 (1.01)	6.34 (0.96)	6.34 (0.93)	3, 18
19. Succinct/long-winded	5.71 (1.35)	6.37 (0.92)	5.70 (1.47)	3
20. Synthesized/unsynthesized	6.34 (1.04)	6.37 (0.91)	6.67 (0.62)	*
21. Focused/unfocused	5.95 (1.12)	6.38 (0.92)	6.25 (1.05)	26, 30
22. Consistent/inconsistent	6.53 (0.77)	6.52 (0.77)	6.51 (1.03)	*

*Adjectives suggested by clinical experts, but not mentioned in published literature.

with Varimax (variance maximizing) rotation and Kaiser normalization.[47] Rotation using this method is an attempt to describe the information in several factors by re-expressing them so that loadings on as few initial variables are as large as possible.[48] This allows differentiation of distinct factors and explication of the fewest factors that explain the largest amount of variance. The Kaiser-Meyer-Olkin index was used to assess sampling adequacy.[49] This is a measure of the degree of intercorrelation among the variables and the appropriateness of factor analysis. Internal consistency reliability estimates were calculated using Cronbach's alpha for each document type and for each factor within a document type. Items that loaded similarly in two or more factors were placed with the factor that maximized the internal consistency reliability of the factors on which they loaded.

Results

Face Validity and Domain Completeness

Twenty-five adjectives were identified through the literature review. Eight of these either scored less well on the survey of experts or were redundant with other retained attributes. The eight that were removed included the following in alphabetical order: bad, cohesive, compelling, dense, good, insightful, long, and salient. Five other adjectives were added at the suggestion of the clinical experts, including: coherent, concise, consistent, structured, and synthesized. These were also scored for importance. The instrument resulting from this process consisted of 22 bipolar pairs of adjectives. The adjectives contained in the preliminary instrument are shown in Table 1, with the mean score, standard deviation, and source (literature citation vs. expert panel) for each item.

Construct Validity

Sample

For the factor analysis, 67 responses were obtained. Of the 67 respondents, 60 were physicians and four were nurse practitioners; three respondents did not identify their training. Of the 60 physicians, there were 35 residents, two fellows, and 22 attending physicians. This resulted in a response rate of 28%. The Kaiser-Meyer-Olkin index was 0.65, 0.71, and 0.71 for admission, progress, and discharge notes respectively,[50] indicating an adequate number of responses for the analyses.

Exploratory Factor Analyses

The data from the three note types were submitted to a principal components factor analysis with Varimax rotation for each note type. Using the Kaiser criterion for including factors with Eigenvalues >1, each analysis resulted in a unique factor solution for each note type.

Admission note. Factor analysis of the admission note type resulted in a six-factor solution. This solution accounted for 73% of total variance (Table 2). Overall internal consistency reliability for admission note items was 0.86, with internal consistency reliability for individual factors ranging from 0.52 to 0.88.

Progress note. The exploratory factor analysis of the progress note type resulted in a six-factor solution (Table 3) that accounted for 73% of total variance. The overall internal consistency reliability was 0.87, with individual factors' internal consistency reliabilities ranging from 0.59 to 0.84.

Discharge summary. The exploratory factor analysis performed on the discharge summary note type resulted in a four-factor solution: factor I (6 items, factor loadings: 0.612–0.805, 37.6% variance), factor II (6 items, factor loadings: 0.566–0.773, 13.7% variance), factor III (6 items, factor loadings: 0.489–0.769, 7.7% variance), and factor IV (4 items, factor loadings: 0.662–0.876, 6.8% variance) (Table 4). This solution accounted for 65.8% of the total variance of the data. Overall internal consistency reliability for discharge notes was 0.88, and internal consistency reliability estimates for the four factors ranged from 0.76 to 0.85.

Table 2 ■ Factor Loadings of Six-factor Solution for Admission Note (n = 67; Overall Internal Consistency Reliability: 0.86)

Item	Factor I	Factor II	Factor III	Factor IV	Factor V	Factor VI
1. Clear	0.687					
2. Up-to-date	0.740					
3. Complete				0.584		
4. Legible	0.866					
5. Accurate						0.840
6. Thorough	0.589					
7. Uncluttered		0.660				
8. Coherent	0.890					
9. Useful				0.704		
10. Correct		0.785				
11. Brief			0.805			
12. Current	0.932					
13. Organized				0.585		
14. Relevant					0.863	
15. Comprehensible		0.665				
16. Concise			0.800			
17. Structured				0.406		
18. Nonredundant			0.463			
19. Succinct			0.803			
20. Synthesized		0.810				
21. Focused					0.643	
22. Consistent		0.730				
Internal consistency reliability estimate (alpha)	0.88	0.73	0.78	0.52	0.70	N/A
Percent variance explained after Varimax rotation	21%	16%	11%	9.9%	8.5%	6.6%

Table 3 ■ Factor Loadings of Six-factor Solution for Progress Note (n = 67; Overall Internal Consistency Reliability: 0.87)

Item	Factor I	Factor II	Factor III	Factor IV	Factor V	Factor VI
1. Clear			0.772			
2. Up-to-date					0.814	
3. Complete	0.743					
4. Legible						0.671
5. Accurate						0.858
6. Thorough	0.750					
7. Uncluttered				0.536		
8. Coherent			0.480			
9. Useful					0.668	
10. Correct				0.879		
11. Brief		0.879				
12. Current					0.524	
13. Organized			0.695			
14. Relevant					0.635	
15. Comprehensible				0.641		
16. Concise		0.791				
17. Structured			0.800			
18. Nonredundant		0.408				
19. Succinct		0.823				
20. Synthesized	0.746					
21. Focused		0.524				
22. Consistent				0.554		
Internal consistency reliability estimate (alpha)	0.74	0.81	0.84	0.74	0.59	0.75
Percent variance explained after Varimax rotation	14%	14%	14%	11%	11%	8.3%

Table 4 ■ Factor Loadings of Four-factor Solution for Discharge Summary (n = 67; Overall Internal Consistency Reliability: 0.88)

Item	Factor I Well-formed	Factor II Comprehensible	Factor III Accurate	Factor IV Compact
Clear	0.654			
Up-to-date			0.769	
Complete			0.726	
Legible		0.625		
Accurate			0.696	
Thorough			0.591	
Uncluttered	0.450			
Coherent		0.745		
Useful		0.566		
Correct		0.773		
Brief				0.778
Current			0.752	
Organized	0.710			
Relevant			0.489	
Comprehensible		0.733		
Concise				0.876
Structured	0.805			
Nonredundant	0.771			
Succinct				0.872
Synthesized	0.612			
Focused				0.662
Consistent		0.618		
Internal consistency reliability estimate (alpha)	0.84	0.85	0.76	0.83
Percent variance explained after Varimax rotation	18%	17%	16%	15%

Comparisons among Exploratory Factor Analyses

Each factor solution was unique; however, four sets of items consistently factored together across all note types: (1) up-to-date and current; (2) brief, concise, succinct; (3) organized and structured; and (4) correct, comprehensible, consistent. Several additional sets of items factored together across two note types: (1) complete and thorough (progress and discharge), (2) uncluttered and consistent (admission and progress), (3) uncluttered and synthesized (admission and discharge), (4) clear and organized (progress and discharge), (5) relevant and current (progress and discharge), and (6) succinct and nonredundant (admission and progress).

Naming of Factors and Comparison with IOM Factors for Data Quality

We named factors based upon the discharge summary note. Factor I, which we called Well-formed, comprised clear, uncluttered, organized, structure, nonredundant, and synthesized. Factor II, Comprehensible, included legible, coherent, useful, correct, comprehensible, and consistent. The six components of Accurate (Factor III) were up-to-date, complete, accurate, thorough, current, and relevant. Compact, the fourth factor, was made up of four items: brief, concise, succinct, and focused. Two of the IOM recommended attributes for data quality, accuracy and completeness, appear to measure the same underlying factor we called accuracy (Table 4). We felt that our factor comprehensible was similar to the IOM attribute meaning, but the IOM attribute legible loads under comprehensible in our analysis. Two of our new factors (compact and well-formed) address form, which was not accounted for in the data quality attributes of the IOM.

Discussion

Using established psychometric methods, we developed and provided preliminary evidence for the construct validity and internal consistency reliability of the PDQI. The fact that the factor structures differed across note types is possibly due to the relatively small sample size. The Kaiser-Meyer-Olkin indexes ranged from 0.65 to 0.71 met the minimum but not optimal criterion for factor analysis. Another potential explanation is that clinician's perceptions regarding note quality vary by note type because each note type serves a particular primary function in the clinical workflow.

The resulting factors differ from those defined by the IOM.[45] This may be due to the original purpose of the IOM terms, which is to describe the quality of data in an EHR, not necessarily the quality of notes. Our factors of Well-formed, Comprehensible, Accurate, and Compact derived from the discharge summary factor analysis have the salutary features of being derived from published attributes, validated by provider attitudes, and inclusive of both form and content of notes. The factor Accurate refers to the extent to which the content of the document reflects the true state of the patient. Well-formed and Compact address how the note is composed (form). Well-formed signifies that the information is being presented in a way that contains a logical structure leading to a reasonable conclusion. Compact indicates that the note has the appropriate density of information within the given content. A fourth factor, Comprehensible, emerged as both content- and form-based, describing whether the transfer of information from the note to the reader occurs with meaning and compre-

hension. One potential reason for the extraction of novel form-related factors that did not appear in the IOM definitions for data quality may be the distinction between data and documents, which are collections of data.

Our findings extend the work of other investigators in three ways. First, development of our documentation quality construct was done using validated psychometric methods for instrument design with measures of validity and reliability, something few studies have done to date. One study that reported on the development of an instrument to evaluate "consult letters" did employ measures of internal consistency reliability, but they observed lower scores (0.21–0.69) than in our study (0.67–0.89).[31] Furthermore, they did not establish content or construct validity. Second, our results suggest that some of the published attributes either measure the same underlying dimension of quality, or do not contribute to the overall construct of document quality. We were able to provide content validation for some of the attributes espoused by other investigators to determine document quality, such as legibility, completeness, correctness, and accuracy.[3,6,7,11,17,25–27,36–43,45] However, our findings suggest that some of these recommended attributes factored together in our solution, suggesting they may measure the same thing. For example, in our exploratory factor analysis of discharge summaries, two attributes proposed in the literature as two separate measures (completeness and accuracy)[7,27,41] factored together (along with up-to-date, current, thorough, and relevant) (Table 4). In contrast, although correctness has been previously reported synonymously with accuracy,[7,27,41] we found it factored separately under the comprehensible factor (Table 4). The third way our results extend previous work is through validation of form as a key constituent of documentation quality. A few previous studies evaluated style,[29,31,51] which is related to form. Coakley et al.[51] determined that the writing style of resident-authored radiology reports was significantly improved after editing by attending physicians. They developed criteria specifically for use with radiology reports, including clarity, brevity, readability, and quality of the impression. Although clarity and brevity were aspects of form reflected in the final instrument we developed, Coakley's instrument did not address accuracy. Furthermore, it was developed for a specific clinical scenario, dictation of computed tomography scan reports by radiology residents. The instrument developed by Myers et al.[29,31] addressed writing style as well as content for resident-authored consultation letters. This 34-item scale broke notes into four components (history, physical examination, impression and plan, and overall writing style). The instrument assessed resident competency for completeness, clarity/organization, and brevity, and had nine items devoted to a formative evaluation of writing style (uses active voice, avoids jargon, avoids repetition, one topic per paragraph, paragraphs with fewer than five sentences, one idea per sentence, use of appropriate headings, and appealing layout). However, the content and construct validity of these items were not established by the authors, and their use may not be generalizable beyond the resident consult letters that they studied. Our analysis discovered, validated, and assessed the reliability of two novel dimensions of the document quality construct related to form (Well-formed and Compact as described above). This is consistent with our prespecified hypothesis that form-related factors are important attributes of document quality.

The primary limitations of our study relate to size and representativeness of study sample and to its implementation in a single clinical setting. In regard to sample size and representativeness, we do not know why our e-mail survey response rate was only 28%, but it is consistent with response rates to e-mail surveys of this population in previous studies conducted by our group. The sample of 67 met minimum but not optimal criteria for factor analysis. Consequently, the extracted factors are unlikely to be stable. In addition, the factors we determined to contribute to note quality almost exclusively reflect the attitudes of physicians (60 of 67 respondents were physicians). Multiple professionals with a diversity of clinical training use documents in the course of giving care to patients, and these individuals may value attributes of quality that are different than those represented here. Third, there was limited variability in mean scores for many of the survey items; this may be the result of the relatively small sample size.

The construct validity of the PDQI requires further investigation in larger samples and for other note types. An assessment of the intrarater and interrater reliability of the use of the PDQI was outside the scope of this study, but will need to be completed to fully test whether the instrument can be reliably applied to score the quality of physician notes. The PDQI, if reliable, could be tested in both resident notes (for training) and attending physician notes (for continuous professional development). As proposed, this instrument could be used to evaluate the output of notes authored in EHRs versus paper, or by different entry modalities within EHRs. Finally, it may be that one or more of the factors are amenable to automated methods of evaluation. For example, the factor Compact could be measured as a ratio of findings within a given note to the length of the note. Natural language processing techniques could be applied to assess this.

Conclusion

Using established psychometric methods, we developed and established preliminary evidence for the construct validity and internal consistency reliability of the PDQI. We found that clinicians believe high-quality clinical notes should be well-formed, comprehensible, accurate, and compact. Two of these factors are novel and related to form (well-formed, compact). These findings resulted in a relatively simple 22-item instrument for measuring document quality. Further work is necessary to assess the applicability of the PDQI to other note types and to determine whether it can be reliably used to score documents.

References ■

1. Buller-Close K, Schriger DL, Baraff LJ. Heterogeneous effect of an emergency department expert charting system. Ann Emerg Med 2003;41:644–52.
2. Clayton PD, Naus SP, Bowes WA 3rd, et al. Physician use of electronic medical records: Issues and successes with direct data entry and physician productivity. AMIA Annu Symp Proc 2005:141–5.
3. Embi PJ, Yackel TR, Logan JR, Bowen JL, Cooney TG, Gorman PN. Impacts of computerized physician documentation in a teaching hospital: Perceptions of faculty and resident physicians. J Am Med Inform Assoc 2004;11:300–9.
4. Hammond KW, Helbig ST, Benson CC, Brathwaite-Sketoe BM. Are electronic medical records trustworthy? Observations on copying, pasting and duplication. AMIA Annu Symp Proc 2003:269–73.
5. Stetson PD, Keselman A, Rappaport D, et al. Electronic discharge summaries. AMIA Annu Symp Proc 2005:1121.

6. Tang PC, LaRosa MP, Gorden SM. Use of computer-based records, completeness of documentation, and appropriateness of documented clinical decisions. J Am Med Inform Assoc 1999;6:245–51.
7. Logan JR, Gorman PN, Middleton B. Measuring the quality of medical records: A method for comparing completeness and correctness of clinical encounter data. Proc AMIA Symp 2001:408–12.
8. Payne T, Perkins M, Kalus R, Reilly D. The transition to electronic documentation on a teaching hospital medical service. Proc AMIA Symp 2006:629–33.
9. Stetson P, Lee P, Shea S. Electronic documentation with eNote enhances physician billing on an academic hospitalist unit. J Hosp Med 2006;1(Suppl 2):27.
10. Rosenbloom ST, Grande J, Geissbuhler A, Miller RA. Experience in implementing inpatient clinical note capture via a provider order entry system. J Am Med Inform Assoc 2004;11:310–5.
11. Thompson A, Jacob K, Fulton J. Anonymized dysgraphia. J R Soc Med 2003;96:51.
12. Dolin RH, Alschuler L, Beebe C, et al. The HL7 Clinical Document Architecture. J Am Med Inform Assoc 2001;8:552–69.
13. Dolin RH, Alschuler L, Boyer S, et al. HL7 Clinical Document Architecture, Release 2. J Am Med Inform Assoc 2006;13:30–9.
14. Haas J, Bakken S, Bright TJ, Melton GB, Stetson P, Johnson SB. Clinicians' perceptions of usability of eNote. AMIA Annu Symp Proc 2005:973.
15. Bright T, Bakken S, Johnson S. Heuristic evaluation of eNote: An electronic notes system. Proc AMIA Symp 2006:864.
16. Poissant L, Pereira J, Tamblyn R, Kawasumi Y. The impact of electronic health records on time efficiency of physicians and nurses: A systematic review. J Am Med Inform Assoc 2005;12:505–16.
17. Weir CR, Hurdle JF, Felgar MA, Hoffman JM, Roth B, Nebeker JR. Direct text entry in electronic progress notes. An evaluation of input errors. Methods Inf Med 2003;42:61–7.
18. Hirschtick RE. A piece of my mind. Copy-and-paste. JAMA 2006;295:2335–6.
19. Thielke S, Hammond K, Helbig S. Copying and pasting of examinations within the electronic medical record. Int J Med Inform 2007;76(suppl 1):S122-S128.
20. Wachter RM. Reflections: The hospitalist movement a decade later. J Hosp Med 2006;1:248–52.
21. JCAHO's 2006 National Patient Safety Goals: Handoffs are biggest challenge. Hosp Peer Rev 2005;30:89–93.
22. Goldman L, Pantilat SZ, Whitcomb WF. Passing the clinical baton: 6 principles to guide the hospitalist. Am J Med 2001;111:36S–39S.
23. Pantilat SZ, Lindenauer PK, Katz PP, Wachter RM. Primary care physician attitudes regarding communication with hospitalists. Am J Med 2001;111:15S–20S.
24. Howell J, Chisholm C, Clark A, Spillane L. Emergency medicine resident documentation: Results of the 1999 American board of emergency medicine in-training examination survey. Acad Emerg Med 2000;7:1135–8.
25. Jennett P, Affleck L. Chart audit and chart stimulated recall as methods of needs assessment in continuing professional health education. J Contin Educ Health Prof 1998;18:163–71.
26. As-Sanie S, Zolnoun D, Wechter ME, Lamvu G, Tu F, Steege J. Teaching residents coding and documentation: Effectiveness of a problem-oriented approach. Am J Obstet Gynecol 2005;193:1790–3.
27. Berner ES, Kasiraman RK, Yu F, Ray MN, Houston TK. Data quality in the outpatient setting: Impact on clinical decision support systems. AMIA Annu Symp Proc 2005:41–5.
28. Deering S, Poggi S, Hodor J, Macedonia C, Satin AJ. Evaluation of residents' delivery notes after a simulated shoulder dystocia. Obstet Gynecol 2004;104:667–70.
29. Keely E, Myers K, Dojeiji S. Can written communication skills be tested in an objective structured clinical examination format? Acad Med 2002;77:82–6.
30. Wilson BE, Performance-based assessment of internal medicine interns: Evaluation of baseline clinical and communication skills. Acad Med 2002;77:1158.
31. Myers KA, Keely EJ, Dojeiji S, Norman GR. Development of a rating scale to evaluate written communication skills of residents. Acad Med 1999;74(Suppl 10):S111–3.
32. Bakken S, Chen E, Choi J, et al. Mobile decision support for advanced practice nurses. Stud Health Technol Inform 2006;122:1002.
33. Linder J, Schnipper J, Palchuk M, Einbinder J, Li Q, Middleton B. improving care for acute and chronic problems with smart forms and quality dashboards. Proc AMIA Symp 2006:1193.
34. Adams DC, Bristol JB, Poskitt KR. Surgical discharge summaries: Improving the record. Ann R Coll Surg Engl 1993;75:96–9.
35. van Walraven C, Duke SM, Weinberg AL, Wells PS. Standardized or narrative discharge summaries. Which do family physicians prefer? Can Fam Physician 1998;44:62–9.
36. Ammenwerth E, Kutscha A, Eichstadter R, Haux R. Systematic evaluation of computer-based nursing documentation. Medinfo 2001;10:1102–6.
37. Aronsky D, Haug PJ. Assessing the quality of clinical data in a computer-based record for calculating the pneumonia severity index. J Am Med Inform Assoc 2000;7:55–65.
38. Dougherty GE, "Conventional" dictated versus database-generated discharge summaries: Timeliness, quality and completeness. CMAJ 1999;160:345–6.
39. Solomon DH, Schaffer JL, Katz JN, et al. Can history and physical examination be used as markers of quality? An analysis of the initial visit note in musculoskeletal care. Med Care 2000;38:383–91.
40. Soto CM, Kleinman KP, Simon SR. Quality and correlates of medical record documentation in the ambulatory care setting. BMC Health Serv Res 2002;2:22.
41. Hogan WR, Wagner MM, Accuracy of data in computer-based patient records. J Am Med Inform Assoc 1997;4:342–55.
42. Wagner MM, Hogan WR. The accuracy of medication data in an outpatient electronic medical record. J Am Med Inform Assoc 1996;3: 234–44.
43. Humphreys T, Shofer FS, Jacobson S, Coutifaris C, Stemhagen A. Preformatted charts improve documentation in the emergency department. Ann Emerg Med 1992;21:534–40.
44. Nunnally J, Bernstein I. Psychometric Theory. 3rd ed. New York, NY: McGraw-Hill, 1994.
45. Institute of Medicine. The Computer-Based Patient Record: An Essential Technology for Health Care, Revised Edition. Washington, DC: National Academies Press, 1997.
46. Polit D, Beck C, Hungler B. Essentials of Nursing Research. 5th ed. Philadelphia, PA: Lippincott, 2001.
47. Kaiser H, The Varimax criterion for analytic rotation in factor analysis. Psychometrika 1958;23:187–200.
48. STATA, STATA 7 Reference Guide. Vol. 1. College Station, TX. STATA Press, 2001.
49. Hair J, Anderson R, Tatham R, Black W. Factor Analysis. 5th ed. Englewood Cliffs, NJ: Prentice Hall, 1998.
50. Strickland OL. Using factor analysis for validity assessment: Practical considerations. J Nurs Meas 2003;11:203–5.
51. Coakley FV, Heinze SB, Shadbolt CL, et al. Routine editing of trainee-generated radiology reports: Effect on style quality. Acad Radiol 2003;10:289–94.