



OPEN ACCESS

QNOTE: an instrument for measuring the quality of EHR clinical notes

Harry B Burke,^{1,2} Albert Hoang,¹ Dorothy Becher,¹ Paul Fontelo,³ Fang Liu,³ Mark Stephens,⁴ Louis N Pangaro,² Laura L Sessums,² Patrick O'Malley,^{1,2} Nancy S Baxi,⁵ Christopher W Bunt,⁴ Vincent F Capaldi,⁵ Julie M Chen,⁵ Barbara A Cooper,⁵ David A Djuric,⁶ Joshua A Hodge,⁶ Shawn Kane,⁵ Charles Magee,² Zizette R Makary,⁵ Renee M Mallory,⁵ Thomas Miller,⁴ Adam Saperstein,⁴ Jessica Servey,⁴ Ronald W Gimbel¹

¹Biomedical Informatics Department, Uniformed Services University of the Health Sciences, Bethesda, Maryland, USA

²Medicine Department, Uniformed Services University of the Health Sciences, Bethesda, Maryland, USA

³National Library of Medicine, National Institutes of Health, Bethesda, Maryland, USA

⁴Family Medicine Department, Uniformed Services University of the Health Sciences, Bethesda, Maryland, USA

⁵Internal Medicine Service, Walter Reed National Military Medical Center, Bethesda, Maryland, USA

⁶Fort Belvoir Community Hospital, Fort Belvoir, Virginia, USA

Correspondence to

Dr Harry B Burke, Biomedical Informatics and Medicine, F Edward Hebert School of Medicine, Uniformed Services University of the Health Sciences, 4301 Jones Bridge Road, Bethesda, MD 20814, USA; harry.burke@usuhs.edu

Received 3 September 2013

Revised 4 December 2013

Accepted 6 December 2013

Published Online First

2 January 2014



CrossMark

To cite: Burke HB, Hoang A, Becher D, *et al.* *J Am Med Inform Assoc* 2014;**21**:910–916.

ABSTRACT

Background and objective The outpatient clinical note documents the clinician's information collection, problem assessment, and patient management, yet there is currently no validated instrument to measure the quality of the electronic clinical note. This study evaluated the validity of the QNOTE instrument, which assesses 12 elements in the clinical note, for measuring the quality of clinical notes. It also compared its performance with a global instrument that assesses the clinical note as a whole.

Materials and methods Retrospective multicenter blinded study of the clinical notes of 100 outpatients with type 2 diabetes mellitus who had been seen in clinic on at least three occasions. The 300 notes were rated by eight general internal medicine and eight family medicine practicing physicians. The QNOTE instrument scored the quality of the note as the sum of a set of 12 note element scores, and its inter-rater agreement was measured by the intraclass correlation coefficient. The Global instrument scored the note in its entirety, and its inter-rater agreement was measured by the Fleiss κ .

Results The overall QNOTE inter-rater agreement was 0.82 (CI 0.80 to 0.84), and its note quality score was 65 (CI 64 to 66). The Global inter-rater agreement was 0.24 (CI 0.19 to 0.29), and its note quality score was 52 (CI 49 to 55). The QNOTE quality scores were consistent, and the overall QNOTE score was significantly higher than the overall Global score ($p=0.04$).

Conclusions We found the QNOTE to be a valid instrument for evaluating the quality of electronic clinical notes, and its performance was superior to that of the Global instrument.

INTRODUCTION

The clinical note documents the clinician's information collection, problem assessment, and patient management. It is a record of the clinician's care of the patient, it assists other clinicians in their management of the patient, and it is a legal document.^{1–2} Outpatient medical records during the first half of the 20th century were typically index cards stored in an envelope.³ During the second half of the 20th century, there was an exponential increase in the complexity of medicine and the recognition that detailed, complete, and accurate medical documentation was necessary for high-quality care.^{4–5} Proposals designed to improve the

clinical note through the use of a problem-based approach were adopted.^{6–8} In addition, starting in the 1960s, there was a great deal of interest in creating computer-based notes.⁹ But until the recent widespread adoption of electronic health records (EHRs), assessing the quality of clinical notes was an arduous task. It required medical records personnel to manually find and pull the patients' paper charts, the physicians' notes had to be legible, raters had to be colocated with the charts, and the raters had to manually review and summarize each of the charts.¹⁰

These difficulties delayed efforts to develop an instrument for assessing clinical note quality. The EHR makes it easier to evaluate clinical notes because they are legible, they can be accessed instantly, and they can be reviewed anywhere. This provides an opportunity for a properly designed instrument that systematically assesses note quality.^{10–12} A global instrument assesses the entire note with one judgment of its quality. This approach has been used to assess inpatient notes.^{13–14} However, there is currently no validated non-global instrument for measuring the quality of a clinical note.

QNOTE is an instrument that assesses a clinical note by judging each of its evaluative elements. The note's quality score is the combination of its element scores. In a previous study, we assessed the external validity of the QNOTE clinical note quality evaluation.¹⁵ In this study, we conducted a validation of the QNOTE instrument and we compared the QNOTE with a global instrument. We hypothesized that, in addition to its prior external validity, the QNOTE instrument would demonstrate high internal validity and be superior to a global measure of note quality.

METHODS

Instruments

QNOTE is an instrument that assesses the quality of the clinical note in terms of 12 clinical elements (figure 1). The external validation of QNOTE has been described.¹⁵ Briefly, the 12 elements and their evaluative components were determined by a group of 61 clinicians privileged to write clinical notes. A set of organizing themes was identified: note characteristics, content, and its relation to the healthcare system. From these themes, a set of note

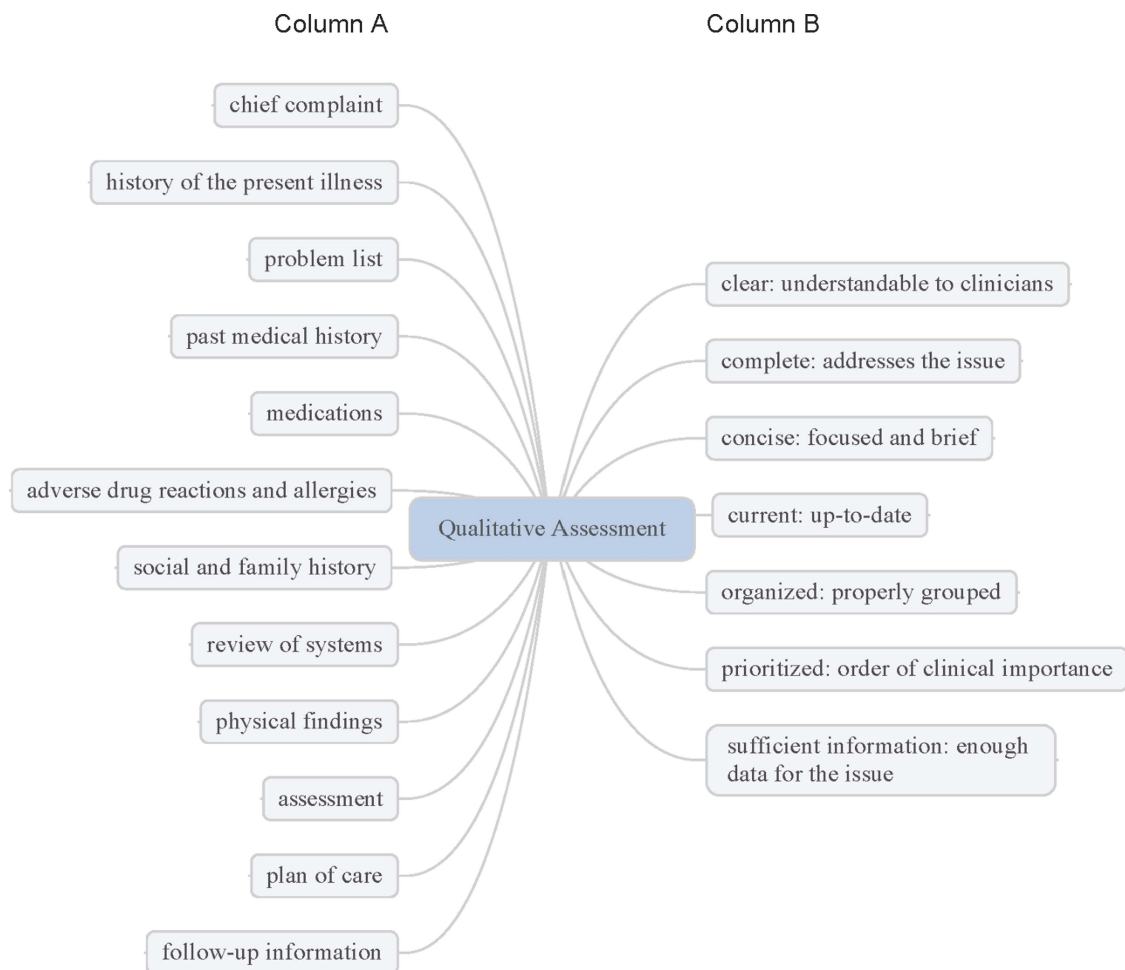


Figure 1 Qualitative assessment of clinical notes. Column A is a list of the evaluative elements, and column B is a list of the components of each element.

elements was developed and the evaluative components of the elements were developed (figure 1). The 12 elements in the clinical note are chief complaint, history of present illness, problem list, past medical history, medications, adverse drug reactions, social and family history, review of systems, physical findings, assessment, plan of care, and follow-up information. The seven evaluative components are clear, complete, concise, current, organized, prioritized, and sufficient information. Not all of the components were used to evaluate all the elements, but every element was evaluated using at least one component. The result was QNOTE (figure 2), an instrument that scores the note based on the evaluative criteria (components) of each element.

Global scores have been used for rating the quality of clinical notes.^{13 14} Essentially, raters review the entire clinical note *de novo* and provide a judgment of the quality of the entire note. We created a global instrument to score the note (figure 3). It provides an overall note score based on the rater's judgment of whether the note supports the diagnosis and the planned work-up or planned treatment and is consistent with the diagnosis.

Clinical notes

Patient records were collected from five US military medical outpatient primary care clinics located in Texas, Mississippi, Nebraska, Nevada, and Virginia populated by predominately civilian patients. An independent contractor blinded to the study was tasked with collecting the charts of at least 100

patients with type 2 diabetes mellitus from each site. Type 2 diabetes was selected because it is the second most common reason for an ambulatory visit (hypertension is the most common)¹⁶ and patients with this condition exhibit multiple comorbid conditions and present with sufficient clinical complexity that they require a proper clinical note. The contractor was instructed to select the notes of outpatients who were treated by a physician, who had a diagnosis of type 2 diabetes mellitus (patients could have comorbid conditions), and who had been seen in clinic on at least three occasions: once approximately 6 months prior to EHR adoption, once approximately 6 months after EHR adoption, and once approximately 5 years after EHR adoption, resulting in three notes per patient. The contractor collected 568 civilian diabetic patients. One-third of the notes were handwritten (pre-EHR) and two-thirds were electronic (post-EHR). The pre-EHR notes were free text, and the post-EHR notes were templates structured by the physicians. An independent contractor photocopied and deidentified the notes prior to their distribution to the investigators. The deidentified notes were reviewed, and 31 patients were removed because at least one of their notes was either incomplete or illegible, resulting in 537 patients. From the pool of 537 patients, 100 study patients were randomly selected, resulting in 300 study notes. Each of the patient's set of three notes was randomly placed in one of four patient groups (figure 4). The result was four patient groups each with 25 patients, with each patient having three clinical notes, resulting in 75 notes per patient group. The presentation

Elements and their components	Scoring – acceptability					
1. CHIEF COMPLAINT(S) A. Sufficient information (enough information to direct HPI; includes pertinent details; includes duration)	<input type="checkbox"/> Missing Fully	<input type="checkbox"/> Partially	<input type="checkbox"/> Unacceptable	C. Clear (understandable to providers and others)	<input type="checkbox"/>	<input type="checkbox"/>
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	D. Concise (focused, brief, not redundant)	<input type="checkbox"/>	<input type="checkbox"/>
				E. Organized (properly grouped, can find important information easily)	<input type="checkbox"/>	<input type="checkbox"/>
2. HISTORY OF PRESENT ILLNESS (HPI)	<input type="checkbox"/> Missing Fully	<input type="checkbox"/> Partially	<input type="checkbox"/> Unacceptable	8. REVIEW OF SYSTEMS	<input type="checkbox"/> Missing Fully	<input type="checkbox"/> Partially Unacceptable
A. Sufficient information (enough information for its purpose; includes pertinent details)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	A. Complete (addresses all pertinent positives and negatives)	<input type="checkbox"/>	<input type="checkbox"/>
B. Concise (focused, brief, not redundant)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	B. Clear (understandable to providers and others)	<input type="checkbox"/>	<input type="checkbox"/>
C. Clear (understandable to providers and others)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	C. Concise (focused, brief, not redundant)	<input type="checkbox"/>	<input type="checkbox"/>
D. Organized (properly grouped, chronological, can find important information easily)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9. PHYSICAL FINDINGS (includes vital signs)	<input type="checkbox"/> Missing Fully	<input type="checkbox"/> Partially Unacceptable
3. PROBLEM (LIST)	<input type="checkbox"/> Missing Fully	<input type="checkbox"/> Not applicable Partially	<input type="checkbox"/> Unacceptable	A. Complete (addresses all pertinent positives and negatives)	<input type="checkbox"/>	<input type="checkbox"/>
A. Current (up-to-date)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	B. Clear (understandable to providers and others)	<input type="checkbox"/>	<input type="checkbox"/>
B. Ordered (from most to least important)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	C. Concise (focused, brief, not redundant)	<input type="checkbox"/>	<input type="checkbox"/>
C. Concise (focused, brief, not redundant)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	10. ASSESSMENT (diagnosis; differential)	<input type="checkbox"/> Missing Fully	<input type="checkbox"/> Partially Unacceptable
D. Complete (addresses all important problems)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	A. Prioritized (displayed in order of importance; signs, symptoms, tests, procedures organized properly, includes care plan)	<input type="checkbox"/>	<input type="checkbox"/>
4. PAST MEDICAL HISTORY	<input type="checkbox"/> Missing Fully	<input type="checkbox"/> Not applicable Partially	<input type="checkbox"/> Unacceptable	B. Sufficient information (enough information for purpose; includes pertinent details)	<input type="checkbox"/>	<input type="checkbox"/>
A. Complete (addresses all important past medical history)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	C. Clear (understandable to providers and others)	<input type="checkbox"/>	<input type="checkbox"/>
B. Concise (focused, brief, not redundant)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	D. Concise (focused, brief, not redundant)	<input type="checkbox"/>	<input type="checkbox"/>
C. Clear (understandable to providers and others)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	11. PLAN OF CARE (with goals and objectives)	<input type="checkbox"/> Missing Fully	<input type="checkbox"/> Partially Unacceptable
D. Organized (properly grouped, can find important information easily)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	A. Prioritized (displayed in order of importance; signs, symptoms, tests, procedures organized properly, includes care plan)	<input type="checkbox"/>	<input type="checkbox"/>
5. MEDICATIONS (LIST)	<input type="checkbox"/> Missing Fully	<input type="checkbox"/> Not applicable Partially	<input type="checkbox"/> Unacceptable	B. Sufficient information (enough information for purpose; includes pertinent details)	<input type="checkbox"/>	<input type="checkbox"/>
A. Current (up to date)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	C. Clear (understandable to providers and others)	<input type="checkbox"/>	<input type="checkbox"/>
B. Complete (contains all the current medications including dosages)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	D. Concise (focused, brief, not redundant)	<input type="checkbox"/>	<input type="checkbox"/>
C. Concise (no non-current medications)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	12. FOLLOW-UP INFORMATION (instructions for the patient; consults; orders; prescriptions)	<input type="checkbox"/> Missing Fully	<input type="checkbox"/> Partially Unacceptable
6. ADVERSE DRUG REACTIONS AND ALLERGIES	<input type="checkbox"/> Missing Fully	<input type="checkbox"/> Not applicable Partially	<input type="checkbox"/> Unacceptable	A. Prioritized (displayed in order of importance)	<input type="checkbox"/>	<input type="checkbox"/>
A. Current (up to date)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	B. Sufficient information (enough information for purpose; includes pertinent details)	<input type="checkbox"/>	<input type="checkbox"/>
B. Sufficient information (enough information for purpose; includes pertinent details)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	C. Clear (understandable to providers and others)	<input type="checkbox"/>	<input type="checkbox"/>
C. Clear (understandable to providers and others)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	D. Concise (focused, brief, not redundant)	<input type="checkbox"/>	<input type="checkbox"/>
7. SOCIAL AND FAMILY HISTORY	<input type="checkbox"/> Missing Fully	<input type="checkbox"/> Partially	<input type="checkbox"/> Unacceptable			
A. Current (up to date)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
B. Sufficient information (enough information for purpose; includes pertinent details)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			

Figure 2 QNOTE, an instrument that scores the note based on the evaluative criteria (components) of each element.

of the clinical notes to the raters was randomized within each of the four patient groups. The Uniformed Services University of the Health Sciences Institutional Review Board approved the study.

Physician raters

We recruited eight general internal medicine and eight family medicine practicing physicians in the national capital medical region (District of Columbia metropolitan area). There were no inclusion or exclusion criteria. We accepted the first eight physicians from each specialty who responded to the announcement. Ten were military physicians and six were civilian physicians, and none had prior experience assessing the quality of clinical notes. Study investigators met the recruited physicians to explain the review process, to describe the Global note and the QNOTE elements and components, and to allow the reviewers to practice on up to three notes. There was no proficiency testing and all the volunteer physicians participated in the study. Within their specialty, the physicians were randomly assigned to one of the four rater groups, resulting in two internal medicine and two family medicine physicians in each of the four-rater groups. Each rater rated two different sets of 75 photocopied deidentified notes. For the QNOTE, raters were asked to score the components of each of the elements of the note as fully

acceptable, partially acceptable, or unacceptable. After the raters completed their component ratings of the elements of the note, we assigned a score to each rating: fully acceptable was scored 100, partially acceptable was scored 50, and unacceptable was scored 0. The average of the component scores for an element was the score for that element. A missing element that was required but not present—that is, an element that was not optional—was assigned a score of zero. Several elements could be rated as not applicable, and they would be omitted from the scoring. For the Global instrument, raters were given the note, asked to read it, and then provide an overall rating of fully acceptable, partially acceptable, or unacceptable. After the raters completed their global rating, we assigned a score to each rating: fully acceptable was scored 100, partially acceptable was scored 50, and unacceptable was scored 0. After the reviewer scoring was completed, it was determined that one reviewer’s scores were significantly different from every one of the other reviewers’ scores and that reviewer was dropped from the study. Reviewers were provided with unique user names and passwords and they scored the notes online using the Global instrument and/or the QNOTE instrument. The results were directly entered into an encrypted database. The reviewers were not contacted during the review process and they were not compensated for participating in the study.

Study design

There were eight study groups (figure 4). Each study group consisted of four raters, each of whom rated two sets of 75 notes. All the study groups used the Global quality instrument to score their notes; four study groups (2, 3, 6, 7) used just the Global

OVERALL NOTE SCORE: the note supports the diagnosis and the planned work-up or planned treatment is consistent with diagnosis
<input type="checkbox"/> Fully acceptable <input type="checkbox"/> Partially acceptable <input type="checkbox"/> Unacceptable

Figure 3 Global quality instrument.

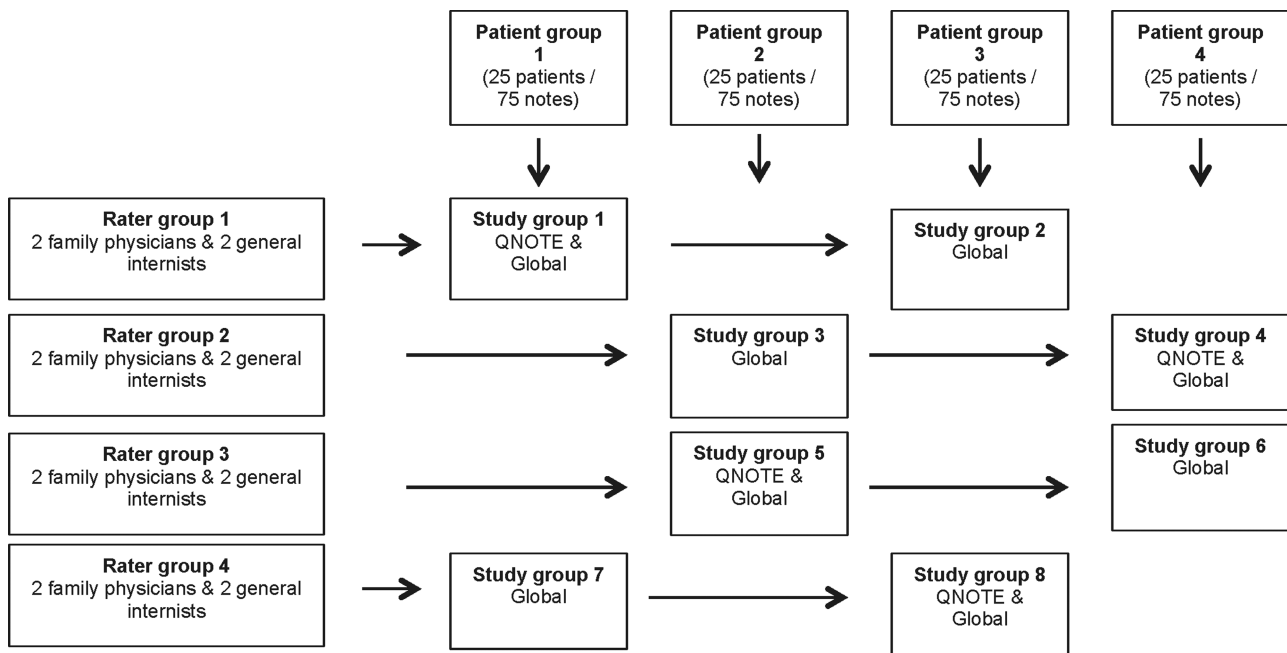


Figure 4 Study design.

quality instrument ('Global 1') and four study groups (1, 4, 5, 8) used the Global quality instrument after using the QNOTE instrument on the same notes ('Global 2').

Statistical analysis

The inter-rater agreement for the QNOTE elements and across elements was assessed by the intraclass correlation coefficient (ICC), and the inter-rater agreement for the Global instrument was assessed by the Fleiss κ score. The Pearson correlation coefficient was used to assess the quality scores across groups. Differences between study groups were tested by Student's t test and the bootstrap method. Quality mean scores and 95% CIs, Fleiss κ mean scores, ICCs, and t tests were calculated using SAS 9.3. The Fleiss κ CI bootstraps and the ICC CIs were calculated using R (R Project, <http://www.r-project.org>).

RESULTS

QNOTE inter-rater agreement results are shown in table 1. The QNOTE mean ICC for each study group was: group 1 score, 0.84 (CI 0.81 to 0.87); group 4 score, 0.80 (CI 0.76 to 0.85); group 5 score, 0.85 (CI 0.83 to 0.88); and group 8 score, 0.78 (CI 0.73 to 0.83). The four study groups were not significantly different. The grand mean QNOTE ICC score was 0.82 (CI 0.80 to 0.84). The QNOTE instrument demonstrated a high level of inter-rater agreement, which suggests that it is a reliable instrument for assessing note quality.

Each of QNOTE's 12 evaluative elements was scored in terms of the mean score of its components. The mean scores for each group, as well as their grand means, are shown in table 2. The QNOTE instrument's mean quality score for each study group was: group 1 score, 61 (CI 60 to 64); group 4 score, 66 (CI 64 to 68); group 5 score, 67 (CI 65 to 69); and group 8 score, 64 (CI 62 to 66). The four study groups were not significantly

Table 1 QNOTE element intraclass correlation coefficients (CI) by rater group

Elements	Group 1	Group 4	Group 5	Group 8	All groups
CC	0.71 (0.59 to 0.81)	0.77 (0.61 to 0.88)	0.62 (0.46 to 0.75)	0.68 (0.49 to 0.80)	0.70 (0.58 to 0.80)
HPI	0.90 (0.85 to 0.93)	0.83 (0.72 to 0.91)	0.81 (0.73 to 0.87)	0.80 (0.69 to 0.88)	0.84 (0.78 to 0.91)
Problems	0.74 (0.60 to 0.83)	0.53 (0.21 to 0.75)	0.84 (0.72 to 0.92)	0.57 (0.25 to 0.75)	0.67 (0.57 to 0.70)
Med history	0.92 (0.89 to 0.94)	0.87 (0.79 to 0.93)	0.93 (0.89 to 0.95)	0.89 (0.82 to 0.93)	0.90 (0.88 to 0.93)
Medications	0.93 (0.87 to 0.94)	0.94 (0.90 to 0.97)	0.94 (0.91 to 0.96)	0.88 (0.81 to 0.93)	0.92 (0.88 to 0.96)
Allergies	0.71 (0.59 to 0.82)	0.75 (0.55 to 0.88)	0.89 (0.84 to 0.92)	0.79 (0.69 to 0.87)	0.79 (0.72 to 0.80)
Soc/family	0.88 (0.83 to 0.92)	0.85 (0.75 to 0.92)	0.84 (0.77 to 0.89)	0.85 (0.76 to 0.91)	0.86 (0.81 to 0.90)
ROS	0.88 (0.83 to 0.92)	0.90 (0.83 to 0.95)	0.93 (0.90 to 0.95)	0.82 (0.72 to 0.89)	0.88 (0.85 to 0.92)
Exam	0.88 (0.83 to 0.92)	0.83 (0.71 to 0.91)	0.94 (0.91 to 0.96)	0.88 (0.80 to 0.92)	0.88 (0.83 to 0.93)
Assess	0.79 (0.69 to 0.86)	0.81 (0.68 to 0.90)	0.84 (0.77 to 0.89)	0.81 (0.70 to 0.88)	0.81 (0.75 to 0.88)
Plan	0.88 (0.82 to 0.92)	0.78 (0.63 to 0.88)	0.85 (0.79 to 0.90)	0.79 (0.67 to 0.87)	0.82 (0.76 to 0.89)
Follow-up	0.84 (0.77 to 0.89)	0.74 (0.57 to 0.86)	0.82 (0.75 to 0.88)	0.63 (0.42 to 0.78)	0.76 (0.68 to 0.84)
Mean	0.84 (0.81 to 0.87)	0.80 (0.76 to 0.85)	0.85 (0.83 to 0.88)	0.78 (0.73 to 0.83)	0.82 (0.80 to 0.84)

CC, chief complaint; HPI, history of the present illness; Problems, problem list; Med history, past medical history; Medication, medication list; Allergies, adverse drug reactions and allergies; Soc/family, social and family history; ROS, review of systems; Exam, physical examination; Assess, assessment; plan, plan of care; Follow-up, follow-up information.

Table 2 QNOTE element quality scores (CI) by rater group

Elements	Group 1	Group 4	Group 5	Group 8	All groups
CC	75 (70 to 81)	80 (75 to 85)	72 (67 to 77)	54 (47 to 61)	71 (68 to 73)
HPI	62 (54 to 69)	74 (67 to 81)	72 (66 to 78)	71 (64 to 77)	70 (66 to 73)
Problems	39 (31 to 46)	40 (34 to 46)	48 (40 to 57)	34 (26 to 42)	39 (37 to 44)
Med history	50 (41 to 60)	53 (44 to 62)	50 (41 to 60)	51 (42 to 60)	51 (47 to 56)
Medications	68 (60 to 76)	71 (63 to 79)	75 (67 to 83)	75 (68 to 83)	72 (68 to 76)
Allergies	78 (72 to 85)	78 (71 to 86)	73 (66 to 80)	71 (64 to 79)	75 (72 to 79)
Soc/family	44 (36 to 55)	42 (35 to 50)	56 (49 to 63)	35 (28 to 43)	45 (41 to 57)
ROS	47 (38 to 55)	58 (49 to 67)	55 (46 to 63)	53 (45 to 62)	53 (49 to 57)
Exam	71 (54 to 75)	77 (71 to 83)	74 (67 to 80)	75 (69 to 81)	74 (71 to 77)
Assess	69 (64 to 75)	75 (70 to 81)	78 (73 to 83)	81 (77 to 86)	76 (73 to 78)
Plan	68 (62 to 74)	76 (72 to 81)	76 (71 to 81)	82 78 to 86)	75 (73 to 78)
Follow-up	68 (62 to 74)	72 (67 to 77)	76 (71 to 82)	78 (72 to 83)	74 (71 to 76)
Mean	61 (60 to 64)	66 (64 to 68)	67 (65 to 69)	64 (62 to 66)	65 (64 to 66)

CC, chief complaint; HPI, history of the present illness; Problems, problem list; Med history, past medical history; Medication, medication list; Allergies, adverse drug reactions and allergies; Soc/family, social and family history; ROS, review of systems; Exam, physical examination; Assess, assessment; plan, plan of care; Follow-up, follow-up information.

different. The grand mean QNOTE quality score was 65 (CI 64 to 66). The mean element scores across the four groups correlated highly ($r=0.89$).

In order to determine whether the QNOTE instrument could reliably assess both handwritten and electronic notes, the ICCs for the two types of notes were assessed. The results are shown in table 3. The QNOTE mean ICC was 0.79 (CI 0.75 to 0.84) for handwritten notes and 0.80 (CI 0.75 to 0.86) for electronic notes, and the two groups were not significantly different. The QNOTE instrument was equally reliable in assessing the quality of handwritten and electronic notes.

The Global instrument assessed the overall quality of the note, where the Global instrument was used either alone (Global 1) or after the QNOTE instrument (Global 2). Each of the four rater groups' global quality mean scores and their inter-rater agreement scores, as well as their grand means, is shown in table 4.

The Global 1 mean Fleiss κ score for each study group was: group 2 score, 0.26 (CI 0.16 to 0.37); group 3 score, 0.25 (CI

0.14 to 0.38); group 6 score, 0.22 (CI 0.13 to 0.34); and group 7 score, 0.22 (CI 0.12 to 0.36). The Global 2 mean Fleiss κ score for each study group was: group 1 score, 0.39 (CI 0.28 to 0.52); group 4 score, 0.23 (CI 0.08 to 0.44); group 5 score, 0.23 (CI 0.13 to 0.35); and group 8 score, 0.10 (CI -0.02 to 0.27). The Global 1 grand mean Fleiss κ score was 0.24 (CI 0.19 to 0.29), the Global 2 grand mean Fleiss κ score was 0.24 (CI 0.17 to 0.31), and the two groups were not significantly different. The Global instrument demonstrated a relatively low level of inter-rater agreement.

The Global 1 mean quality score for each study group was: group 2 score, 50 (CI 43 to 56); group 3 score, 50 (CI 43 to 57); group 6 score, 64 (CI 58 to 69); and group 7 score, 44 (CI 38 to 51). The Global 2 mean quality score for each study group was: group 1 score, 44 (CI 38 to 51); group 4 score, 57 (CI 50 to 60); group 5 score, 54 (CI 47 to 60); and group 8 score, 48 (CI 43 to 54). The Global 1 grand mean quality score was 52 (CI 49 to 55), the Global 2 grand mean quality score was 51 (CI 48 to 54), and the two groups were not significantly different. The Global 1 and 2 group scores correlated highly ($r=0.86$). The fact that there was no difference between the grand means of the two groups and that they correlated highly suggests that the QNOTE scoring did not affect the Global 2 scoring.

The QNOTE quality scores were significantly higher than the Global 1 quality scores ($p=0.04$), and they were significantly higher than the Global 2 quality scores ($p=0.02$). In other words, an examination of the elements leads to higher quality scores than an overall Global assessment.

DISCUSSION

In this large, retrospective multicenter blinded study, we found QNOTE to be a valid instrument for evaluating the quality of clinical notes. It possesses external validity, its elements are congruent with clinical practice,¹⁵ and it possesses internal validity. We found that the systematic examination of the clinical note elements using the QNOTE instrument, compared with a global subjective assessment, resulted in significantly higher quality scores and higher inter-rater agreement. The QNOTE instrument was equally reliable on handwritten and electronic documentation. There was no association between the QNOTE instrument element quality scores and the element inter-rater agreement ($r=0.08$, NS). This, in conjunction with the high

Table 3 Mean intraclass correlation coefficients (CI) for the QNOTE elements for the four groups by type of note

Elements	Handwritten	Electronic
CC	0.63 (0.37 to 0.90)	0.76 (0.55 to 0.97)
HPI	0.78 (0.69 to 0.87)	0.87 (0.78 to 0.96)
Problems	0.63 (0.37 to 0.90)	0.59 (0.35 to 0.83)
Med history	0.90 (0.82 to 0.97)	0.84 (0.75 to 0.94)
Medications	0.93 (0.87 to 0.99)	0.89 (0.75 to 1.00)
Allergies	0.80 (0.68 to 0.92)	0.74 (0.43 to 1.00)
Soc/family	0.77 (0.58 to 0.96)	0.81 (0.66 to 0.96)
ROS	0.81 (0.68 to 0.94)	0.89 (0.82 to 0.96)
Exam	0.86 (0.77 to 0.96)	0.92 (0.83 to 1.00)
Assess	0.81 (0.72 to 0.91)	0.79 (0.62 to 0.96)
Plan	0.86 (0.78 to 0.94)	0.81 (0.65 to 0.96)
Follow-up	0.75 (0.65 to 0.85)	0.74 (0.48 to 1.00)
Mean	0.79 (0.75 to 0.84)	0.80 (0.75 to 0.86)

CC, chief complaint; HPI, history of the present illness; Problems, problem list; Med history, past medical history; Medication, medication list; Allergies, adverse drug reactions and allergies; Soc/family, social and family history; ROS, review of systems; Exam, physical examination; Assess, assessment; plan, plan of care; Follow-up, follow-up information.

Table 4 Global 1 and global 2 mean quality scores (CI) and Fleiss κ inter-rater agreement (CI) by rater group

Global 1					
	Study group 2	Study group 3	Study group 6	Study group 7	Grand mean
Quality score	50 (43 to 56)	50 (43 to 57)	64 (58 to 69)	44 (38 to 51)	52 (49 to 55)
Fleiss κ	0.26 (0.16 to 0.37)	0.25 (0.14 to 0.38)	0.22 (0.13 to 0.34)	0.22 (0.12 to 0.36)	0.24 (0.19 to 0.29)
Global 2					
	Rater group 1	Rater group 4	Rater group 5	Rater group 8	Grand mean
Quality score	44 (38 to 51)	57 (50 to 60)	54 (47 to 60)	48 (43 to 54)	51 (48 to 54)
Fleiss κ	0.39 (0.28 to 0.52)	0.23 (0.08 to 0.44)	0.23 (0.13 to 0.35)	0.10 (−0.02 to 0.27)	0.24 (0.17 to 0.31)

inter-rater agreement, suggests that the QNOTE instrument quality scores reflect true differences in the quality of the notes. The Global ratings that were performed after the QNOTE ratings on the same notes (Global 2) were not affected by the prior QNOTE element-based assessment. This suggests that global assessments and element assessments use two different cognitive evaluation processes.

In terms of the global assessment of notes, Stetson *et al*¹⁴ created the Physician Documentation Quality Instrument (PDQI-9) to measure inpatient note quality. The PDQI-9 assesses the entire note as one global entity. It requires nine global judgments regarding a note. The raters determine, on a 1–5 scale where 1 is ‘not at all’ and 5 is ‘extremely’, whether the overall note is, ‘up-to-date, accurate, thorough, useful, organized, comprehensible, succinct, synthesized, and internally consistent’. The scores for each of the nine judgments are summed to create a total score that can range from 9 to 45. An issue with the global approach is that it does not recognize that the clinical note consists of multiple clinical elements, where each element represents a type of clinical information that can have its own set of quality characteristics and must be evaluated in its own right. It is the combination of these elements that forms the basis for a high-quality note. Further, when physicians create and use a note, they do so in terms of its elements. Clinicians are interested in the element information being clinically useful.¹⁵ Finally, the elements are necessary for medicolegal and billing documentation. A clinical note can be viewed as a set of elements that inform and support the clinician’s reasoning, decision-making, and recommendations to the patient.

This study has several limitations. (1) It was performed using Military Health System records, but it is well established that this health system treats its patients similarly to patients treated in the community.^{17–20} (2) We focused on diabetics because: diabetes is very prevalent (it is the second most common diagnosis in ambulatory medicine); diabetic patients are typical of outpatients with serious diseases; diabetic patients have multiple comorbid conditions so they present with a wide spectrum of diseases. (3) Our results may not be as useful for patients with acute, self-limiting conditions. (4) It is unlikely that another single-item global assessment tool would perform differently from the current instrument because a global assessment is just that, a single overall judgment. This study’s strengths are that it is a large multicenter blinded trial that rigorously evaluated the inter-rater agreement of 16 practicing physicians using QNOTE to assess the quality of electronic clinical notes.

CONCLUSIONS

The documentation of a clinician–patient encounter serves three important purposes: (1) it documents the clinician’s information

collection, problem assessment, and plan for the patient; (2) it creates a complete and accurate record that can be used by other clinicians to care for the patient; and (3) it provides substantiation for what was done for medicolegal reasons and reimbursement. These goals are consistent with the American Medical Informatics Association’s 2011 Health Policy Meeting that stated that the main purpose of documentation is to support patient care and improved outcomes.²¹ A high-quality note contains the necessary and sufficient elements required to achieve these goals. The QNOTE instrument can determine whether a clinical note contains the necessary quality elements and is clinically acceptable.

The QNOTE instrument has been shown to possess external validity and, in this study, it demonstrated high internal validity. It is an important advance because it provides a valid instrument for determining the quality of a clinical encounter, as documented by the clinician’s note. Further, it provides a way for clinicians to assess their electronic notes and to use that knowledge to improve their documentation. Future QNOTE studies will include: (1) changes in note quality before (handwritten notes) and 6 months after EHR adoption; (2) changes in note quality 6 months after EHR adoption and 5 years later; and (3) the relationship between EHR note quality and clinical outcomes. The QNOTE instrument provides a structured, easy to use, transparent, valid and reliable way to assess the quality of electronic clinical notes.

Acknowledgements The authors acknowledge and appreciate the support of: Patrice Waters-Worley, MBA, Michelle Helaire, EdD, and Karla Herres, RN, CPHQ of Lockheed Martin who traveled nationally to the five medical facilities participating in the study, coordinating the collection, deidentification, and indexing of clinical notes. These contributors were compensated for their efforts. Written permission was obtained from each of these contributors.

Contributors All the authors provided substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; they assisted in drafting the article or revising it critically for important intellectual content; and they provided final approval of the version to be published.

Funding Extramural funding was provided by the US Army Medical Research and Materials Command, cooperative agreement W81XWH-08-2-0056. All of the authors are employed by the Uniformed Services University of the Health Sciences or another entity of the US Federal Government.

Disclaimer The views expressed in this manuscript are solely those of the authors and do not necessarily reflect the opinion or position of the Uniformed Services University of the Health Sciences, the US Army Medical Research and Materials Command, the National Library of Medicine, the US Department of Defense, or the US Department of Health and Human Services.

Competing interests None.

Patient consent Obtained.

Ethics approval Uniformed Services University of the Health Sciences.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/3.0/>

REFERENCES

- 1 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.
- 2 Rosenbloom ST, Stead WW, Denny JC, et al. Generating clinical notes for electronic health record systems. *Appl Clin Inform* 2010;1:232–43.
- 3 Kuenssberg EV. Volume and cost of keeping records in a group practice. *Br Med J* 1956;1(Suppl 2681):341–3.
- 4 Dearing WP. Quality of medical care. *Calif Med* 1963;98:331–5.
- 5 Weed LL. The importance of medical records. *Can Fam Physician* 1969;15:23–5.
- 6 Weed LL. Medical records that guide and teach. *N Engl J Med* 1968;278:593–600.
- 7 Weed LL. Medical records that guide and teach. *N Engl J Med* 1968;278:652–7.
- 8 Weed LL. Medical records, medical education and patient care: the problem oriented record as a basic tool. Chicago: Year Book Medical Publishers, Inc, 1969.
- 9 Levy RP, Cammann MR, Smith MJ. Computer handling of ambulatory clinic records. *JAMA* 1964;190:1033–7.
- 10 Embi PJ, Yackel TR, Nowen JL, et al. Impacts of computerized physician documentation in a teaching hospital: perceptions of faculty and resident physicians. *J Am Med Inform Assoc* 2004;11:300–9.
- 11 Berwick DM, Winickoff DE. The truth about doctors' handwriting: a prospective study. *BMJ* 1996;313:1657–8.
- 12 Cimino JJ. Improving the electronic health record—are clinicians getting what they wished for? *JAMA* 2013;309:991–2.
- 13 Stetson PD, Morrison FP, Bakken S, et al. Preliminary development of the physician documentation quality instrument. *J Am Med Inform Assoc* 2008;15:534–41.
- 14 Stetson PD, Bakken S, Wrenn JO, et al. Assessing election note quality using the physician documentation quality instrument (PDQI-9). *Appl Clin Inform* 2012;3:164–74.
- 15 Hanson JL, Stephens MB, Pango LN, et al. Quality of outpatient clinical notes: a stakeholder definition derived through qualitative research. *BMC Health Serv Res* 2012;12:407.
- 16 Centers for Disease Prevention and Control. Selected patient and provider characteristics for ambulatory care visits to physician offices and hospital outpatient and emergency departments: United States, 2009–2010. <http://www.cdc.gov/nchs/fastats/diabetes.htm> (accessed 22 Oct 2013).
- 17 Jackson JL, Strong J, Cheng EY, et al. Patients, diagnoses, and procedures in a military internal medicine clinic: comparison with civilian practices. *Mil Med* 1999;164:194–7.
- 18 Jackson JL, O'Malley PG, Kroenke K. A psychometric comparison of military and civilian medical practices. *Mil Med* 1999;164:112–15.
- 19 Jackson JL, Cheng EY, Jones DL, et al. Comparison of discharge diagnoses and inpatient procedures between military and civilian health care systems. *Mil Med* 1999;164:701–4.
- 20 Gimbel RW, Pangaro L, Barbour G. America's "undiscovered laboratory for health services research. *Med Care* 2010;48:751–6.
- 21 Cusack CM, Hripcsak G, Bloomrosen M, et al. The future state of clinical data capture and documentation: a report from AMIA's 2011 Policy Meeting. *J Am Med Inform Assoc* 2013;20:134–40.