

Assessing the Assessment—Developing and Deploying a Novel Tool for Evaluating Clinical Notes’ Diagnostic Assessment Quality



Mirica Maria, PhD¹, Khazen Maram, PhD², Hussein Sarib, MD³, Ramos Jason, BA⁴, Tewodros Eguale, MD^{1,5}, Linzer Mark, MD⁶, and Schiff D. Gordon, MD^{1,7}

¹Center for Patient Safety Research and Practice, Department of Medicine, Brigham and Women’s Hospital, 3Rd Floor General Medicine, 1620 Tremont St, Boston, MA 02120, USA; ²Haifa University, Haifa, Israel; ³Mass General Brigham Hospital, Boston, MA, USA; ⁴Emory School of Medicine, Atlanta, GA, USA; ⁵Massachusetts College of Pharmacy and Health Sciences (MCPHS), Boston, MA, USA; ⁶Hennepin Healthcare, Minneapolis, MN, USA; ⁷Harvard Medical School, Center for Primary Care, Boston, MA, USA

BACKGROUND: Ambulatory diagnostic errors are increasingly being recognized as an important quality and safety issue, and while measures of diagnostic quality have been sought, tools to evaluate diagnostic assessments in the medical record are lacking.

OBJECTIVE: To develop and test a tool to measure diagnostic assessment note quality in primary care urgent encounters and identify common elements and areas for improvement in diagnostic assessment.

DESIGN: Retrospective chart review of urgent care encounters at an urban academic setting.

PARTICIPANTS: Primary care physicians.

MAIN MEASURES: The Assessing the Assessment (ATA) instrument was evaluated for inter-rater reliability, internal consistency, and findings from its application to EHR notes.

KEY RESULTS: ATA had reasonable performance characteristics (kappa 0.63, overall Cronbach’s alpha 0.76). Variability in diagnostic assessment was seen in several domains. Two components of situational awareness tended to be well-documented (“Don’t miss diagnoses” present in 84% of charts, red flag symptoms in 87%), while Psychosocial context was present only 18% of the time.

CONCLUSIONS: The ATA tool is a promising framework for assessing and identifying areas for improvement in diagnostic assessments documented in clinical encounters.

KEY WORDS: clinical notes; diagnostic assessment; diagnostic error

J Gen Intern Med 38(9):2123–9

DOI: 10.1007/s11606-023-08085-8

© The Author(s), under exclusive licence to Society of General Internal Medicine 2023

BACKGROUND

While ambulatory diagnostic errors are now being increasingly recognized as an important quality and safety issue worldwide,^{1,2} diagnostic safety suffers from a dearth of ways

to measure quality that can be useful and used for catalyzing improvement.³ To make meaningful progress, organizations and clinicians need practical, effective, and sustainable tools that can engage clinicians and patients to help close care and communication gaps in a myriad of interrelated areas.⁴ Akin to an ophthalmoscopic examination, clinical note documentation can provide a window into the quality of the diagnostic process. Notes are not only a recorded representation of the clinician’s diagnostic activities and thinking, but also serve as important communication tools for ensuing visits, other clinicians, and, increasingly, patients accessing their medical records via Open Notes.⁵ Particularly where there is diagnostic uncertainty, clear and thorough documentation is important for transparently capturing clinicians’ assessments and guiding next steps.⁶

Clinical notes remain a challenge with widely voiced issues, especially note quality and time burden for writing notes. Many have expressed concerns that note quality is worsening, with the heavy use of copy/pasted and templated notes.^{5–7} Of particular concern are shortcomings in the clear and accurate conveyance of diagnostic thinking and assessment. For example, notes in diagnostic error malpractice cases often lack meaningful assessments, and in many cases are lacking any differential diagnosis or evidence of meaningful thinking about the patient’s problem.⁸ In many notes, templated check boxes have displaced richer narrative assessments.^{9,10} Understanding note quality both descriptively and prescriptively to improve diagnostic assessment is thus an untapped frontier. As patients increasingly gain access to their notes, the diagnostic content of notes assumes added importance.^{11–13} The lack of standardization, shared best practices, streamlined workflows, and note “assessment” language is so vast, it is safe to say that virtually every physician has a unique style of note writing. This variation can be viewed as another electronic health record (EHR) area in need of more regimentation, or one that can be leveraged as an opportunity for learning.^{14,15} This paper takes the latter view and presents the results of a qualitative content analysis review of urgent care notes to advance our understanding of how well these notes convey differential diagnoses and deal with contextual elements, diagnostic uncertainties, and

Prior Presentations: *Diagnostic Error in Medicine: 14th Annual International Conference, poster.*

Received August 25, 2022

Accepted February 1, 2023

Published online February 28, 2023

contingencies. In addition, since studies highlight the importance of engaging patients by shifting to Open Notes,^{16–18} this study aimed to assess patient and clinician readability.

METHODS

Study Purpose and Design

Objective. The purpose of this study was twofold: (a) to develop a tool assessing diagnostic assessment quality in urgent care settings and (b) to pilot-test this chart assessment tool and validate it on a sample of 100 charts.

Sample. Using a convenience sampling, we recruited 5 primary care physicians (PCPs) (3 females) seeing primarily urgent care patients in an academic primary care clinic who had at least 20 urgent care encounters during the study period. PCPs were invited into the study via email. Physician-informed consent was obtained at the initiation of the study by the Research Assistant (JR). A total of 100 charts of urgent care encounters by the selected 5 PCPs were chosen for analysis via a random number generator.

Study Design. We conducted a retrospective chart review of 100 urgent care encounters which occurred during January–December 2018 at Brigham and Women’s Hospital from 5 PCPs recruited into the study. The review was conducted in three stages. First, the initial tool (see below for description) was tested on a sample of 25 urgent care charts by 6 reviewers who met biweekly for 3 months (two PCPs, a medical student, a research assistant, a communication expert and qualitative researcher, and a health services researcher). In the second stage, three team members (RA, Med Student, Qualitative Research PhD) individually scored 16 training charts, which were taken from the final list of 100 urgent care charts and met 3 times to resolve disagreements. They consulted the principal investigator to resolve scoring discrepancies and reached a 100% agreement. Lastly, one team member (4th-year medical student) scored the 100 charts, using the coding manual and calibrated Likert scales we developed.

Measure Development and Validation

Based on literature review,^{2,19–21} existing EHR evaluation tools, malpractice insurers’ recommendations, and best practices for writing clinical notes, we developed the Assessing the Assessment (ATA) tool to examine the quality of documented diagnostic assessments (Table 1). The tool was aimed at understanding current documentation practices and developing guidelines for optimizing notes’ diagnostic assessments including their organization, content, and readability. It examined a series of diagnostic assessment elements (e.g., addressing chief concern, differential diagnosis, uncertainty, red flags, time frames, contingencies),

Table 1 “Assessment of the Assessment” Tool

- | |
|--|
| 1. Diagnostic assessment |
| a. Chief complaint |
| b. Differential diagnosis |
| c. Psychosocial information |
| d. Possible etiologies |
| e. Degree of certainty |
| 2. Follow-up plan |
| a. Mentions diagnostic tests |
| b. Contingencies |
| c. Time frames |
| d. Rationale |
| 3. Situational awareness/safety nets |
| a. Red flags |
| b. Don’t miss diagnoses |
| c. Pitfalls |
| 4. Other diagnostic factors |
| a. Quality of diagnosis/differential diagnosis |
| b. Adequate tests |
| c. Avoids over-testing |
| d. Succinctness |
| e. Clinician readability |
| f. Patient readability |
| g. Avoids legal liability pejorative red flags |

Scale: 1, absent; 2, minimal/implicit; 3, mentioned but less than good; 4, neutral (obvious mention); 5, adequate/good; 6, very good; 7, excellent; not applicable (N/A)

in addition to evaluating pejorative language, note organization/succinctness, and readability level for clinicians and patients. Although in our past work we have developed a tool to assess errors in history-taking and physical examining shortcomings,^{22,23} here we primarily focused on the recorded diagnostic assessments, given that they are the most tangible representation of the clinician’s diagnostic thinking and decision-making.²⁴

We examined components of the tool’s construct validity, including valid content, response process, internal structure, relations to other variables, and consequences.²⁵ Our analysis showed that we reached good content validity given the sources for the items, response process was straightforward and confirmed by experts in the field, and by a scoring process being carefully documented, with all raters keeping track of their comments and questions, internal consistency/reliability was satisfactory with a reasonable Cronbach’s alpha for internal consistency and kappa for inter-rater reliability (see below), and consequence validity showing favorable ability to categorize charts for diagnostic quality. We could not assess relationship to other variables (external validity), since the ATA is a unique and novel instrument.

The study was approved by the Institutional Review Board as a medical records study.

Analysis

Data were analyzed using content analysis^{26,27} to assess the diagnostic quality of 100 charts by applying the key concepts of the ATA tool, which contained 4 sections (Diagnostic assessment, Follow-up plan, Situational awareness,

and General subjective assessment), each with several sub-items. We iteratively refined the tool based on the pilot reviews including adding several new sub-items to diagnostic domains (e.g., readability for the patient and clinician and utilization of inappropriate copy/paste and templating). Based on team discussions during the first stage of the study, we refined operational definitions for each of the items and scores. For example, we operationally defined our patient readability item examining for the presence (or absence) of medical jargon or abbreviations the patient might not understand, coherence of the prose, conformance to 8th-grade reading level, and general formatting.

Items were scored on a 7-point scale (1, absent; 2, minimal; 3, less than good; 4, neutral; 5, good; 6, very good; 7, excellent) to capture the granularity of the content. We considered an item present if it appeared anywhere in the chart. However, we gave a higher ranking to elements present in the assessment plan. We used $\alpha = 0.05$ for all statistical tests.

RESULTS

The overall kappa for inter-rater reliability during the review of the 16 training charts was 0.63. The highest-scored individual item was “Avoiding legal liability pejorative red flags” (mean 7.0, SD [0.0]). Noting/addressing potential diagnostic pitfalls received the lowest scores (mean 1.5, SD [1.4]). The General diagnostic assessment domain received a slightly higher score (mean 3.4, SD [1.1]) than the domains of Follow-up plan (mean 2.7, SD [1.8]) and Situational awareness (mean 2.7, SD [1.1]). The Other diagnostic factors, a domain containing a variety of descriptive elements, was the highest-scored domain (mean 5.7, SD [1.2]). In general, there was variability in the quality of individual elements comprising the four domains (Table 2). Within the General diagnostic assessment domain, Psychosocial information was absent in 82% of the charts, with a mean score of 1.6 (SD [1.4]). In contrast, a listing of possible etiologies was missing in only 13% of the charts with a mean score of 4.1 (SD [1.5]). Figure 1 shows the percentage of missing elements in the ATA in the charts reviewed.

Situational Awareness

We created a “situational awareness” domain based on recommendations related to key elements recommended by experts and the National Academy of Medicine diagnosis report. Mention of “Don’t miss diagnoses” was absent in 16% of the charts, with a mean score of 3.7 (SD [1.8]). Red flags, a related construct referring to key worrisome symptoms, were absent in only 13% of the charts, whereas pitfalls to avoid scored poorly with 86% of the notes missing this information (mean 3.0, SD [1.4]) (Table 2).

Table 2 Quality of Diagnostic “Assessment of the Assessment” Elements in the Chart: Results of Review of 100 Charts

Domain and individual items	Mean (SD)
Diagnostic assessment	2.0 (0.92)
a. Chief complaint	2.4 (1.87)
b. Differential diagnosis	2.1 (1.66)
c. Psychosocial information	0.6 (1.39)
d. Possible etiologies	3.1 (1.46)
e. Degree of certainty	1.9 (1.64)
Follow-up plan	1.7 (0.83)
a. Mentions diagnostic tests	2.5 (2.06)
b. Contingencies	2.3 (1.83)
c. Time frames	0.9 (1.77)
d. Rationale	1.0 (1.73)
Situational awareness/safety nets	1.7 (1.09)
a. Red Flags:	2.0 (1.36)
b. Don’t miss diagnoses	2.7 (1.75)
c. Pitfalls	0.5 (1.44)
Other diagnostic assessment elements	4.7 (1.20)
a. Quality of diagnosis/differential diagnosis	3.6 (1.32)
b. Adequate tests	4.7 (1.26)
c. Avoids over-testing	5.5 (0.92)
d. Succinctness	4.9 (1.07)
e. Clinician readability	5.7 (0.66)
f. Patient readability	2.9 (0.86)
g. Avoids legal liability pejorative red flags	6.0 (0.00)

Other Key Domains

Within the Follow-up plan, a similar variability was observed. Noting of any planned diagnostic tests was missing in about a third of the charts (mean 3.5, SD [2.1]), while any mention of the clinical time frames was missing from nearly three-quarters of the charts and received the mean score of 1.9 (SD [1.8]).

The Other diagnostic factors domain contained elements which assessed important auxiliary elements of chart writing, such as succinctness, clinician and patient readability, and avoiding pejorative language. The patient readability element received the lowest score in this domain (mean 3.9, SD [0.9]), while avoiding pejorative language was the highest-scored element, with virtually no instances of the use of clearly stigmatizing or pejorative terms found in our study sample (mean 7.0, SD [0]).

Finally, our clinical team made a subjective judgment about the overall quality of diagnosis, whether adequate tests were ordered, and whether providers avoided over-testing. The adequate tests and avoiding over-testing elements both received moderately high scores (means 5.7, SD [1.3] and mean 6.5, SD [0.9] respectively) (Table 2).

A few exemplary high-quality charts were identified, but no trends could be derived from the small sample. Figure 2 illustrates the uneven distribution of the mean scores across the domains.

Internal Consistency

The overall internal consistency between the 19 elements in the ATA was acceptable²⁸ ($\alpha = 0.76$), although

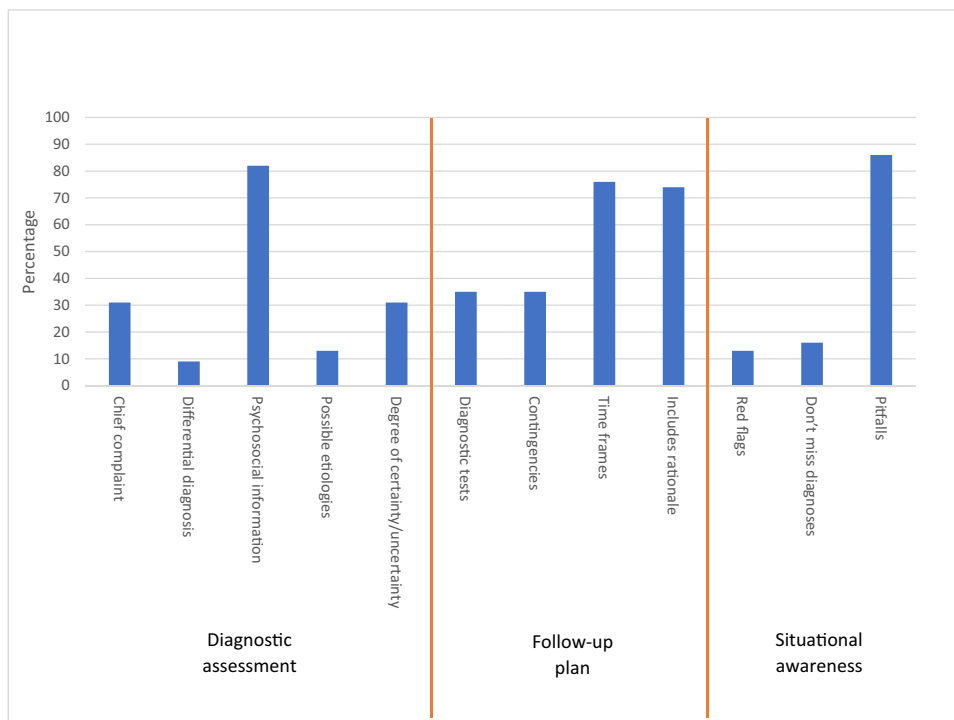


Fig. 1 ATA elements missing in the charts, by domain, %. The y-axis shows percent of charts with a missing element. The elements are grouped by domain.

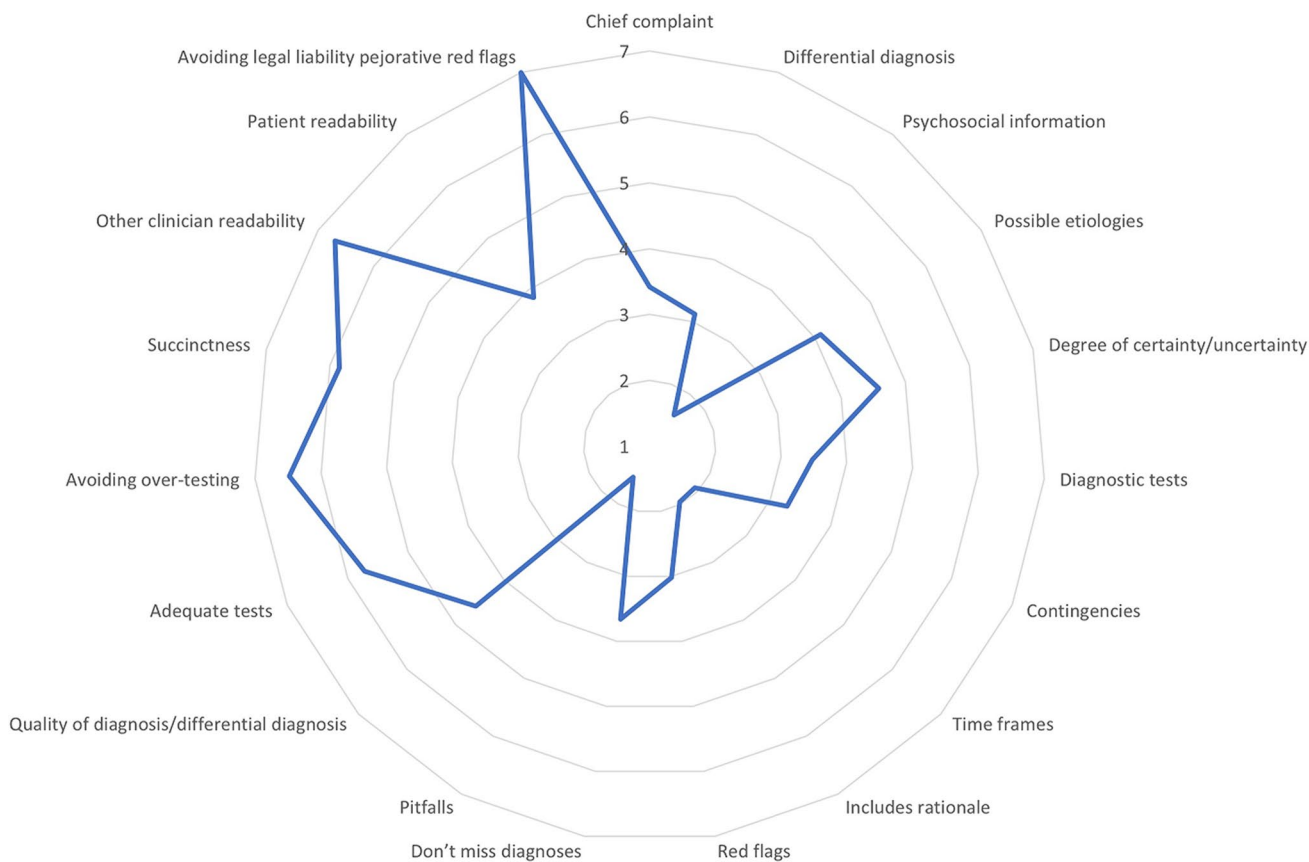


Fig. 2 Mean scores of the ATA elements. All elements of the ATA are included in the chart. The line represents mean score for each of the elements.

some subscales had less favorable consistency. Of the four subscales, Diagnostic assessment had favorable consistency with an alpha of 0.71. Situational awareness had borderline consistency ($\alpha = 0.63$), while the other two scales did not show evidence for internal consistency (Follow-up plan ($\alpha = 0.38$), and General subjective assessment ($\alpha = 0.29$)). There were clinically meaningful ($r > 0.5$) correlations between several of the subscales: the Diagnostic assessment domain was most strongly associated with Situational awareness ($r = 0.60$) and Follow-up plan ($r = 0.57$). The weakest association was observed between Follow-up plan and General subjective assessment ($r = 0.12$).

Within the domain of Diagnostic assessment, the overall internal consistency was acceptable ($\alpha = 0.71$).

The domains of Follow-up plan and General subjective assessment had unacceptable internal consistency ($\alpha = 0.38$ and $\alpha = 0.29$ correspondingly).

DISCUSSION

We developed and deployed a novel tool that was able to be used by a medically trained researcher with reasonable reliability and evidence of many aspects of construct validity in assessing the elements of notes' diagnostic assessment that expert consensus suggests should be present in the documentation of the diagnostic assessment. The quality of the notes varied from both clinician to clinician, encounter to encounter, and, more notably, across different domains of our Assessing the Assessment tool.

We found a somewhat consistent pattern where certain elements of the ATA with the lowest scores were absent from most of the charts (e.g., Psychosocial information and Pitfalls), suggesting clinicians' assessments could benefit from improved documentation in these areas. We also observed a striking imbalance in the quality of documentation of the ATA elements (Fig. 2). While approaching good results in the general writing and readability style, the notes lacked some clinical elements experts recommend good assessment documentation.

Being a novel tool, the internal consistency of the ATA overall and its individual elements has room for improvement. Overall consistency was good, but individual subscales will require further attention with further evolution of the ATA metric.

While we agree that busy clinicians cannot be expected to write (or dictate or have scribed) lengthy narrative assessments, nor is this desirable if the resulting note is not focused or succinct enough to be useful for subsequent readers, certain elements are important and are currently less well documented and possibly less communicated to the patient. To the extent that templated

check boxes have displaced more meaningful narratives, electronic clinical documentation has not fulfilled a role in improving clinicians' notes. It is difficult to defend a diagnostic assessment that does not at least give a meaningful picture of what is going on with the patient including some recognition of uncertainties and probabilities in primary care diagnoses. Thus, some comment on the likely diagnosis and its degree of certainty as well as consideration of other alternatives would seem to be a reasonable expectation. It would also be helpful to have a sense of the clinician's assessment of factors (exposures, etiology, psychosocial) that may be contributing to or impacted by the medical diagnosis the clinician is communicating to the patient—something we found lacking in the majority of notes. Finally, although they rarely were present in the notes we reviewed, consideration of red flags, alternative don't miss diagnoses, and various pitfalls (e.g., limitations of diagnostic test results) could sharpen clinicians' thinking and enhance the quality of their notes, as well provide medical legal support in defending rare but inevitable instances of missed or delayed diagnoses.

Some items, perhaps, stand in conflict with others (i.e., a more succinct note could conflict with producing a more comprehensive note). However, depending on how this metric is used, we can envision it could be helpful for providing a framework for better notes and feedback to a clinician for both the strengths and weaknesses of their own notes. While there are numerous ways notes have been tied to billing and coding, expectations could be changed to allow more emphasis on diagnostic reasoning that would better support the clinical care of the patient. What ultimately defines a good note? Ideally, determining this will need to be tied to various outcomes (patients' health, satisfaction with their care and notes, usability, and usefulness for other clinicians).

How can busy and burnt-out clinicians be better supported in producing better diagnostic assessment documentation?²⁹ It is our hope that outlining the explicit framework embodied in our novel tool can help guide them to write better notes that incorporate these suggested features. With advances in voice recognition³⁰ and potentials for interactive computerized prompts or scribe-facilitated documentation, we can envision ways to both streamline note writing as well as more reliably ensure higher quality notes. This could be useful both educationally, to teach trainees how to write a good assessment, as well as a tool for practicing clinicians to evaluate the quality of their own notes. Given the recent requirement by the 21st Century Cures Act that all notes be made readily available to patients, we envision and would like to help catalyze a transformative sea change in the use and value of this aspect of clinician-patient communication around diagnosis.¹⁸

LIMITATIONS

The assessment tool is based on best practices recommendations from diagnosis safety and malpractice/risk management literature and experts. We recognize that it has not been validated by correlating the ratings with what occurred in the exam room or workloads during the visit, nor with clinical outcomes. However, we chose relatively easily operationally definable measures that did demonstrate good inter-rater reliability and had good face validity for items of importance in a recorded assessment. Whether higher scored notes led to better readers' understanding of the clinician's thinking, enhanced assessments and likelihood of establishing the correct diagnosis with fewer errors or delays, or were worthy investments of added time, requires additional study.

We measured a limited number of physicians who practice in a primary care clinic who primarily see urgent care patients at an academic medical center; this may impede generalizing our findings to more typical primary care encounters more of a mix of chronic disease management and along with an assessment of any acute new symptoms. Furthermore, while this pilot study suggests ATA can be efficiently applied to evaluate clinical notes' assessment features, there is a subjective element to the ratings, which may vary among raters. Nonetheless, we created and iteratively refined standardized operational definitions to anchor the Likert scales for each of the items and reached a reasonably good kappa agreement score on a sample of charts that were blindly reviewed by two reviewers.

Since we did not directly observe the actual encounters for these notes' visits, we were not able to ascertain whether aspects such as red flags, psychosocial assessment, or timeframes were in fact discussed during the visits, rather than simply not documented in the chart. Nonetheless, the goal of this project was to first develop an understanding of how notes are written. Based on this initial project, we have conducted paired evaluation assessment of encounters and their corresponding charts, which is published elsewhere²⁰. Future work will also focus on the reliability of the two subscales with low alphas and attempt to further refine the item inclusion to improve this.

CONCLUSION

We developed a tool to measure the presence and quality of key aspects of documented diagnostic assessments for patients presenting for evaluation of acute symptoms in a primary care clinic. We found significant variation in the completeness and quality across different domains and items, suggesting opportunities for improvement. These

findings assume added importance given the universal patient access to clinical notes mandated in the USA by the 21st Century Cures Act coupled with growing interest in improving diagnostic thinking and processes to ensure more reliable diagnosis and higher quality diagnostic communication with patients.^{31–33}

Corresponding Author: Schiff D. Gordon, MD; Center for Patient Safety Research and Practice, Department of Medicine, Brigham and Women's Hospital, 3Rd Floor General Medicine, 1620 Tremont St, Boston, MA, 02120, USA (e-mail: gschiff@bwh.harvard.edu).

Author Contribution N/a.

Funding This study was funded by CRICO—The Risk Management Foundation of the Harvard Medical Institutions Incorporated. "Tools for Enhancing Primary Care Diagnosis Safety," 2017–2020.

Declarations

Conflict of Interest The authors declare that they do not have a conflict of interest.

Disclosures Dr. Linzer was paid as a consultant for time on this project (funds donated to Hennepin Healthcare Foundation). He is also paid through his employer for his time on burnout reduction projects for ACP, AMA, IHI, ABIM, Optum Office of Provider Advancement, Gillette Children's Hospital, and Essentia Health System, and is supported for other scholarly work and training by NIH and AHRQ.

REFERENCES

1. **Singh H, Schiff GD, Graber ML, Onakpoya I, Thompson MJ.** The global burden of diagnostic errors in primary care. *BMJ Qual Saf.* 2017;26(6):484–94.
2. **Medicine Io, National Academies of Sciences E, Medicine.** Improving Diagnosis in Health Care. Balogh EP, Miller BT, Ball JR, editors. Washington, DC: The National Academies Press; 2015. 472 p.
3. **Singh H, Bradford A, Goeschel C.** Operational measurement of diagnostic safety: state of the science. *Diagnosis (Berlin, Germany).* 2021;8(1):51–65.
4. **Schiff GD.** Diagnosis and diagnostic errors: time for a new paradigm. *BMJ Qual Saf.* 2014;23(1):1–3.
5. **Bell SK, Delbanco T, Elmore JG, Fitzgerald PS, Fossa A, Harcourt K, et al.** Frequency and Types of Patient-Reported Errors in Electronic Health Record Ambulatory Care Notes. *JAMA Netw Open.* 2020;3(6):e205867.
6. **Key Advances in Clinical Informatics.** Aziz Sheikh, Adam Wright, Kathrin Cresswell, editors. 1st ed: Academic Press; 2017 June 28, 2017.
7. **Weiner SJ, Wang S, Kelly B, Sharma G, Schwartz A.** How accurate is the medical record? A comparison of the physician's note with a concealed audio recording in unannounced standardized patient encounters. *Journal of the American Medical Informatics Association : JAMIA.* 2020;27(5):770–5.
8. **Kasick RT, Melvin JE, Perera ST, Perry MF, Black JD, Bode RS, et al.** A diagnostic time-out to improve differential diagnosis in pediatric abdominal pain. *Diagnosis (Berlin, Germany).* 2021;8(2):209–17.
9. **Prater L, Sanchez A, Modan G, Burgess J, Frier K, Richards N, et al.** Electronic Health Record Documentation Patterns of Recorded Primary Care Visits Focused on Complex Communication: A Qualitative Study. *Appl Clin Inform.* 2019;10(2):247–53.
10. **Rajaram A, Patel N, Hickey Z, Wolfrom B, Newbigging J.** Perspectives of undergraduate and graduate medical trainees on documenting clinical notes: Implications for medical education and informatics. *Health Informatics J.* 2022;28(2):14604582221093498.
11. **Gutheil TG, Bursztajn H, Brodsky A.** Malpractice prevention through the sharing of uncertainty. Informed consent and the therapeutic alliance. *The New England Journal of Medicine.* 1984;311(1):49–51.

12. **Kassirer JP.** Our stubborn quest for diagnostic certainty. A cause of excessive testing. *The New England Journal of Medicine.* 1989;320(22):1489-91.
13. **Simpkin AL, Schwartzstein RM.** Tolerating Uncertainty — The Next Medical Revolution? *New England Journal of Medicine.* 2016;375(18):1713-5.
14. **Burke HB, Hoang A, Becher D, Fontelo P, Liu F, Stephens M, et al.** QNOTE: an instrument for measuring the quality of EHR clinical notes. *Journal of the American Medical Informatics Association : JAMIA.* 2014;21(5):910-6.
15. **Hanson JL, Stephens MB, Pangaro LN, Gimbel RW.** Quality of outpatient clinical notes: a stakeholder definition derived through qualitative research. *BMC Health Services Research.* 2012;12(1):407.
16. **Benjamins J, Haveman-Nies A, Gunnink M, Goudkuil A, de Vet E.** How the Use of a Patient-Accessible Health Record Contributes to Patient-Centered Care: Scoping Review. *J Med Internet Res.* 2021;23(1):e17655.
17. **Blease C, Fernandez L, Bell SK, Delbanco T, DesRoches C.** Empowering patients and reducing inequities: is there potential in sharing clinical notes? *BMJ Quality & Safety.* 2020;29(10):1-2.
18. **DesRoches CM, Leveille S, Bell SK, Dong ZJ, Elmore JG, Fernandez L, et al.** The Views and Experiences of Clinicians Sharing Medical Record Notes With Patients. *JAMA Netw Open.* 2020;3(3):e201753.
19. **Olson APJ, Linzer M, Schiff GD.** Measuring and Improving Diagnostic Safety in Primary Care: Addressing the "Twin" Pandemics of Diagnostic Error and Clinician Burnout. *Journal of General Internal Medicine.* 2021;36(5):1404-6.
20. **Khazen M, Sullivan EE, Ramos J, Mirica M, Linzer M, Schiff GD, et al.** Anatomy of diagnosis in a clinical encounter: how clinicians discuss uncertainty with patients. *BMC Primary Care.* 2022;23(1):153.
21. **Stetson PD, Bakken S, Wrenn JO, Siegler EL.** Assessing Electronic Note Quality Using the Physician Documentation Quality Instrument (PDQI-9). *Appl Clin Inform.* 2012;3(2):164-74.
22. **Schiff GD, Kim S, Abrams R, Cosby K, Lambert B, Elstein AS, et al.** Advances in Patient Safety Diagnosing Diagnosis Errors: Lessons from a Multi-institutional Collaborative Project. In: Henriksen K, Battles JB, Marks ES, Lewin DI, editors. *Advances in Patient Safety: From Research to Implementation (Volume 2: Concepts and Methodology).* Rockville (MD): Agency for Healthcare Research and Quality (US); 2005.
23. **Schiff GD, Hasan O, Kim S, Abrams R, Cosby K, Lambert BL, et al.** Diagnostic error in medicine: analysis of 583 physician-reported errors. *Arch Intern Med.* 2009;169(20):1881-7.
24. **Schaye V, Miller L, Kudlowitz D, Chun J, Burk-Rafel J, Cocks P, et al.** Development of a Clinical Reasoning Documentation Assessment Tool for Resident and Fellow Admission Notes: a Shared Mental Model for Feedback. *Journal of General Internal Medicine.* 2022;37(3):507-12.
25. **Cook DA, Beckman TJ.** Current concepts in validity and reliability for psychometric instruments: theory and application. *The American journal of medicine.* 2006;119(2):166.e7-16.
26. **Research design: qualitative, quantitative and mixed methods approaches Research design: qualitative, quantitative and mixed methods approaches J. Creswell Nurse Res 2004 Vol. 12 Issue 1 Pages 82-83 Accession Number: 28718745. <https://doi.org/10.7748/nr.12.1.82.s2>**
27. **Miles M, Huberman, Michael, Saldana, Johnny.** *Qualitative data analysis: A methods sourcebook.* 4 ed. Los Angeles: Sage; 2020.
28. **SPSS for Windows Step by Step: A Simple Guide and Reference.** 17.0 Update D. George and P. Mallery Publisher: Allyn & Bacon 2010
29. **Documentation dynamics: Note composition, burden, and physician efficiency N. C. Apathy, L. Rotenstein, D. W. Bates and A. J. Holmgren Health Services Research 2022 Pages 1-12**
30. **Blackley SV, Huynh J, Wang L, Korach Z, Zhou L.** Speech recognition for clinical documentation from 1990 to 2018: a systematic review. *Journal of the American Medical Informatics Association : JAMIA.* 2019;26(4):324-38.
31. **To accelerate the discovery, development, and delivery of 21st century cures, and for other purposes Congress 2016 Rule Number: 114-255 Issuing Organization: Congress**
32. **Improving the Patient Record IOM In: The Computer-Based Patient Record: Revised Edition: An Essential Technology for Health Care,** edited by R. S. Dick, E. B. Steen and D. E. Detmer National Academies Press (US) 1997
33. **Edwards ST, Neri PM, Volk LA, Schiff GD, Bates DW.** Association of note quality and quality of care: a cross-sectional study. *BMJ Qual Saf.* 2014;23(5):406-13.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.