

BRIEF REPORTS

The Effectiveness of a Bundled Intervention to Improve Resident Progress Notes in an Electronic Health Record

Shannon M. Dean, MD^{1*}, Jens C. Eickhoff, PhD², Leigh Anne Bakel, MD³

¹Department of Pediatrics, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin; ²Department of Biostatistics and Medical Informatics, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin; ³Department of Pediatrics, University of Colorado School of Medicine, Aurora, Colorado.

Providers nationally have observed a decline in the quality of documentation after implementing electronic health records (EHRs). In this pilot study, we examined the effectiveness of an intervention bundle designed to improve resident progress notes written in an EHR and to establish the reliability of an audit tool used to evaluate notes. The bundle consisted of establishing note-writing guidelines, developing an aligned note template, and educating interns about the guidelines and using the template. Twenty-five progress notes written by pediatric interns before and after this intervention were examined using an audit tool. Reliability of the tool was evaluated using the intraclass correlation coefficient (ICC).

The total score of the audit tool was summarized in terms of means and standard deviation. Individual item responses were summarized using percentages and compared between the pre- and postintervention assessment using the Fisher exact test. The ICC for the audit tool was 0.96 (95% confidence interval: 0.91–0.98). A significant improvement in the total note score and in questions related to note clutter was seen. No significant improvement was seen for questions related to copy-paste. The study suggests that an intervention bundle can lead to some improvements in note writing. *Journal of Hospital Medicine* 2015;10:104–107. © 2014 Society of Hospital Medicine

There are described advantages to documenting in an electronic health record (EHR).^{1–5} There has been, however, an unanticipated decline in certain aspects of documentation quality after implementing EHRs,^{6–8} for example, the overinclusion of data (note clutter) and inappropriate use of copy-paste.^{6–17}

The objectives of this pilot study were to examine the effectiveness of an intervention bundle designed to improve resident progress notes written in an EHR (Epic Systems Corp., Verona, WI) and to establish the reliability of an audit tool used to assess the notes. Prior to this intervention, we provided no formal education for our residents about documentation in the EHR and had no policy governing format or content. The institutional review board at the University of Wisconsin approved this study.

METHODS

The Intervention Bundle

A multidisciplinary task force developed a set of Best Practice Guidelines for Writing Progress Notes in the EHR (see Supporting Information, Appendix 1, in the online version of this article). They were designed to

promote cognitive review of data, reduce note clutter, promote synthesis of data, and discourage copy-paste. For example, the guidelines recommended either the phrase, “Vital signs from the last 24 hours have been reviewed and are pertinent for...” or a link that included minimum/maximum values rather than including multiple sets of data. We next developed a note template aligned with these guidelines (see Supporting Information, Appendix 2, in the online version of this article) using features and links that already existed within the EHR. Interns received classroom teaching about the best practices and instruction in use of the template.

Study Design

The study was a retrospective pre-/postintervention. An audit tool designed to assess compliance with the guidelines was used to score 25 progress notes written by pediatric interns in August 2010 and August 2011 during the pre- and postintervention periods, respectively (see Supporting Information, Appendix 3, in the online version of this article).

Progress notes were eligible based on the following criteria: (1) written on any day subsequent to the admission date, (2) written by a pediatric intern, and (3) progress note from the previous day available for comparison. It was not required that 2 consecutive notes be written by the same resident. Eligible notes were identified using a computer-generated report, reviewed by a study member to ensure eligibility, and assigned a number.

Notes were scored on a scale of 0 to 17, with each question having a range of possible scores from 0 to 2.

*Address for correspondence and reprint requests: Shannon M. Dean, MD, 600 Highland Avenue H4/410 (MC 4108), Madison, WI 53792; Telephone: 608-265-5545; Fax: 608-265-8074; E-mail: sdean@uwhealth.org

Additional Supporting Information may be found in the online version of this article.

Received: June 13, 2014; Revised: October 10, 2014; Accepted: October 28, 2014

2014 Society of Hospital Medicine DOI 10.1002/jhm.2283

Published online in Wiley Online Library (Wileyonlinelibrary.com).

TABLE 1. Comparison of Percentage of “Yes” Responses Between Pre- and Postintervention for Each Question

Question	Preintervention, N = 25*	Postintervention, N = 25	P Value†
1. Does the note header include the name of the service, author, and training level of the author?	0%	68%	<0.0001
2. Does it appear that the subjective/interval history section of the note was newly written? (ie, not copied in its entirety from the previous note)	100%	96%	0.9999
3. Is the vital sign section noncluttered?	4%	84%	<0.0001
4. Is the entire medication list included in the note?	96%	96%	0.9999
5. Is the intake/output section noncluttered?	0%	16%	0.3076
6. Does it appear that the physical exam was newly written? (ie, not copied in its entirety from the previous note)	80%	68%	0.9103
7. Is the lab section noncluttered?	64%	44%	0.5125
8. Is the imaging section noncluttered?	100%	100%	0.9999
9. Does it appear that the assessment was newly written?	48%	28%	0.5121
10. Does it appear that the plan was newly written or partially copied with new information added?	48% partial	52% partial	0.9999
11. If the assessment includes abnormal lab values, is there also an accompanying diagnosis? (eg, inclusion of “patient has hemoglobin of 6.2, also includes diagnosis of anemia”)	88%	96%	0.9477
12. Is additional visual clutter prevented by excluding other objective data found elsewhere in the chart?	96%	96%	0.9999
13. Is the author’s name and contact information (pager, cell) included at the bottom of the note?	0%	28%	0.0035
	0%	72%	<0.0001

NOTE: *Percentages calculated from the first rater. †Adjusted P value (for evaluating multiple items) using the Benjamini-Hochberg false discovery rate method.

Some questions related to inappropriate copy-paste (questions 2, 9, 10) and a question related to discrete diagnostic language for abnormal labs (question 11) were weighted more heavily in the tool, as compliance with these components of the guideline was felt to be of greater importance. Several questions within the audit tool refer to “clutter.” We defined clutter as any additional data not endorsed by the guidelines or not explicitly stated as relevant to the patient’s care for that day.

Raters were trained to score notes through practice sessions, during which they all scored the same note and compared findings. To rectify inter-rater scoring discrepancies identified during these sessions, a reference manual was created to assist raters in scoring notes (see Supporting Information, Appendix 4, in the online version of this article). Each preintervention note was then systematically assigned to 2 raters, comprised of a physician and 3 staff from health information management. Each rater scored the note individually without discussion. The inter-rater reliability was determined to be excellent, with kappa indices ranging from 88% to 100% for the 13 questions; each note in the postintervention period was therefore assigned to only 1 rater. Total and individual questions’ scores were sent to the statistician for analysis.

Statistical Analysis

Inter-rater reliability of the audit tool was evaluated by calculating the intraclass correlation (ICC) coefficient using a multilevel random intercept model to account for the rater effect.¹⁸ The study was powered to detect an anticipated ICC of at least 0.75 at the 1-sided 0.05 significance level, assuming a null hypothesis that the ICC is 0.4 or less. The total score was

summarized in terms of means and standard deviation. Individual item responses were summarized using percentages and compared between the pre- and post-intervention assessment using the Fisher exact test. The analysis of response patterns for individual item scores was considered exploratory. The Benjamini-Hochberg false discovery rate method was utilized to control the false-positive rate when comparing individual item scores.¹⁹ All P values were 2-sided and considered statistically significant at <0.05. Statistical analyses were conducted using SAS software version 9.2 (SAS Institute Inc., Cary, NC).

RESULTS

The ICC was 0.96 (95% confidence interval: 0.91-0.98), indicating an excellent level of inter-rater reliability. There was a significant improvement in the total score (see Supporting Information, Appendix 5, in the online version of this article) between the preintervention (mean 9.72, standard deviation [SD] \pm 1.52) and postintervention (mean 11.72, SD \pm 1.62) periods ($P < 0.0001$).

Table 1 shows the percentage of “yes” responses to each individual item in the pre- and postintervention periods. Our intervention had a significant impact on reducing vital sign clutter (4% preintervention, 84% postintervention, $P < 0.0001$) and other visual clutter within the note (0% preintervention, 28% postintervention, $P = 0.0035$). We did not observe a significant impact on the reduction of input/output or lab clutter. There was no significant difference observed in the inclusion of the medication list. No significant improvements were seen in questions related to copy-paste. The intervention had no significant impact on areas with an already high baseline performance: newly written interval histories, newly written

physical exams, newly written plans, and the inclusion of discrete diagnostic language for abnormal labs.

DISCUSSION

Principal Findings

Improvements in electronic note writing, particularly in reducing note clutter, were achieved after the implementation of a bundled intervention. Because the intervention is a bundle, we cannot definitively identify which component had the greatest impact. Given the improvements seen in some areas with very low baseline performance, we hypothesize that these are most attributable to the creation of a compliant note template that (1) guided authors in using data links that were less cluttered and (2) eliminated the use of unnecessary links (eg, pain scores and daily weights). The lack of similar improvements in reducing input/output and lab clutter may be due to the fact that even with changes to the template suggesting a more narrative approach to these components, residents still felt compelled to use data links. Because our EHR does not easily allow for the inclusion of individual data elements, such as specific drain output or hemoglobin as opposed to a complete blood count, residents continued to use links that included more data than necessary. Although not significant findings, there was an observed decline in the proportion of notes containing a physical exam not entirely copied from the previous day and containing an assessment that was entirely new. These findings may be attributable to having a small sample of authors, a few of whom in the postintervention period were particularly prone to using copy-paste.

Relationship to Other Evidence

The observed decline in quality of provider documentation after implementation of the EHR has led to a robust discussion in the literature about what really constitutes a quality provider note.^{7-10,20} The absence of a defined gold standard makes research in this area challenging. It is our observation that when physicians refer to a decline in quality documentation in the EHR, they are frequently referring to the fact that electronically generated notes are often unattractive, difficult to read, and seem to lack clinical narrative.

Several publications have attempted to define note quality. Payne et al. described physical characteristics of electronically generated notes that were deemed more attractive to a reader, including a large proportion of narrative free text.¹⁵ Hanson performed a qualitative study to describe outpatient clinical notes from the perspective of multiple stakeholders, resulting in a description of the characteristics of a quality note.²¹ This formed the basis for the QNOTE, a validated tool to measure the quality of outpatient notes.²² Similar work has not been done to rigorously define quality for inpatient documentation. Stetson did develop an instrument, the Physician Documenta-

tion Quality Instrument (PDQI-9) to assess inpatient notes across 9 attributes; however, the validation method relied on a gold standard of a general impression score of 7 physician leaders.^{23,24}

Although these tools aim to address overall note quality, an advantage provided by our audit tool is that it directly addresses the problems most attributable to documenting in an EHR, namely note clutter and copy-paste. A second advantage is that clinicians and nonclinicians can score notes objectively. The QNOTE and PDQI-9 still rely on subjective assessment and require that the evaluator be a clinician.

There has also been little published about how to achieve notes of high quality. In 2013, Shoolin et al. did publish a consensus statement from the Association of Medical Directors of Information Systems outlining some guidelines for inpatient EHR documentation.²⁵ Optimal strategies for implementing such guidelines, however, and the overall impact such an implementation would have on improving note writing has not previously been studied. This study, therefore, adds to the existing body of literature by providing an example of an intervention that may lead to improvements in note writing.

Limitations

Our study has several limitations. The sample size of notes and authors was small. The short duration of the study and the assessment of notes soon after the intervention prevented an assessment of whether improvements were sustained over time.

Unfortunately, we were not evaluating the same group of interns in the pre- and postintervention periods. Interns were chosen as subjects as there was an existing opportunity to do large group training during new intern orientation. Furthermore, we were concerned that more note-writing experience alone would influence the outcome if we examined the same interns later in the year.

The audit tool was also a first attempt at measuring compliance with the guidelines. Determination of an optimal score/weight for each item requires further investigation as part of a larger scale validation study. In addition, the cognitive review and synthesis of data encouraged in our guideline were more difficult to measure using the audit tool, as they require some clinical knowledge about the patient and an assessment of the author's medical decision making. We do not assert, therefore, that compliance with the guidelines or a higher total score necessarily translates into overall note quality, as we recognize these limitations of the tool.

Future Directions

In conclusion, this report is a first effort to improve the quality of note writing in the EHR. Much more work is necessary, particularly in improving the clinical narrative and inappropriate copy-paste. The

examination of other interventions, such as the impact of structured feedback to the note author, whether by way of a validated scoring tool and/or narrative comments, is a logical next step for investigation.

Acknowledgements

The authors acknowledge and appreciate the support of Joel Buchanan, MD, Ellen Wald, MD, and Ann Boyer, MD, for their contributions to this study and manuscript preparation. We also acknowledge the members of the auditing team: Linda Brickert, Jane Duckert, and Jeannine Strunk.

Disclosure: Nothing to report.

References

1. 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(3):245–251.
2. Amarasingham R, Moore BJ, Tabak YP, et al. An automated model to identify heart failure patients at risk for 30-day readmission or death using electronic medical record data. *Med Care.* 2010;48(11):981–988.
3. Amarasingham R, Plantinga L, Diener-West M, Gaskin DJ, Powe NR. Clinical information technologies and inpatient outcomes: a multiple hospital study. *Arch Intern Med.* 2009;169(2):108–114.
4. Makam AN, Nguyen OK, Moore B, Ma Y, Amarasingham R. Identifying patients with diabetes and the earliest date of diagnosis in real time: an electronic health record case-finding algorithm. *BMC Med Inform Decis Mak.* 2013;13:81.
5. Poon EG, Wright A, Simon SR, et al. Relationship between use of electronic health record features and health care quality: results of a statewide survey. *Med Care.* 2010;48(3):203–209.
6. 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(4):300–309.
7. Hartzband P, Groopman J. Off the record—avoiding the pitfalls of going electronic. *N Engl J Med.* 2008;358(16):1656–1658.
8. Hirschtick RE. A piece of my mind. Copy-and-paste. *JAMA.* 2006;295(20):2335–2336.
9. Siegler EL, Adelman R. Copy and paste: a remediable hazard of electronic health records. *Am J Med.* 2009;122(6):495–496.
10. O'Donnell HC, Kaushal R, Barron Y, Callahan MA, Adelman RD, Siegler EL. Physicians' attitudes towards copy and pasting in electronic note writing. *J Gen Intern Med.* 2009;24(1):63–68.
11. Cimino JJ. Improving the electronic health record—are clinicians getting what they wished for? *JAMA.* 2013;309(10):991–992.
12. 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.
13. Siegler EL. The evolving medical record. *Ann Intern Med.* 2010;153(10):671–677.
14. 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(1):61–67.
15. Payne TH, Patel R, Beahan S, Zehner J. The physical attractiveness of electronic physician notes. *AMIA Annu Symp Proc.* 2010;2010:622–626.
16. Yackel TR, Embi PJ. Copy-and-paste-and-paste. *JAMA.* 2006;296(19):2315; author reply 2315–2316.
17. Hammond KW, Helbig ST, Benson CC, Brathwaite-Sketoe BM. Are electronic medical records trustworthy? Observations on copying, pasting and duplication. *AMIA Annu Symp Proc.* 2003:269–273.
18. Raudenbush S, Bruk, AS. *Hierarchical Linear Models: Applications and Data Analysis Methods.* 2nd ed. Thousand Oaks, CA: Sage; 2002.
19. Benjamini Y, Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach for multiple testing. *J R Stat Soc Series B Stat Methodol* 1995;57(1):289–300.
20. Sheehy AM, Weissburg DJ, Dean SM. The role of copy-and-paste in the hospital electronic health record. *JAMA Intern Med.* 2014;174(8):1217–1218.
21. Hanson JL, Stephens MB, Pangaro LN, Gimbel RW. Quality of outpatient clinical notes: a stakeholder definition derived through qualitative research. *BMC Health Serv Res.* 2012;12:407.
22. Burke HB, Hoang A, Becher D, et al. QNOTE: an instrument for measuring the quality of EHR clinical notes. *J Am Med Inform Assoc.* 2014;21(5):910–916.
23. 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):164–174.
24. Stetson PD, Morrison FP, Bakken S, Johnson SB. Preliminary development of the physician documentation quality instrument. *J Am Med Inform Assoc.* 2008;15(4):534–541.
25. Shoolin J, Ozeran L, Hamann C, Bria W II. Association of Medical Directors of Information Systems consensus on inpatient electronic health record documentation. *Appl Clin Inform.* 2013;4(2):293–303.