ORIGINAL RESEARCH

Check for updates

Validating the Physician Documentation Quality Instrument for Intensive Care Unit–Ward Transfer Notes

Patrick G. Lyons[1,2], Juan C. Rojas[3], Alice F. Bewley[4], Sara M. Malone[5], and Lekshmi Santhosh[6]

[1]Division of Pulmonary, Allergy, and Critical Care Medicine, Department of Medicine, and [2]Department of Medical Informatics and Clinical Epidemiology, Oregon Health and Science University, Portland, Oregon; [3]Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, Rush University, Chicago, Illinois; [4]Division of Infectious Diseases, Department of Medicine, and [5]Division of Public Health Sciences, Department of Surgery, Washington University School of Medicine, St. Louis, Missouri; and [6]Division of Pulmonary and Critical Care Medicine, Department of Medicine, University of California, San Francisco, San Francisco, California

ORCID IDs: 0000–0002–1557–2787 (P.G.L.); 0000–0002–9897–3462 (L.S.)

ABSTRACT

Background: Physician communication failures during transfers of patients from the intensive care unit (ICU) to the general ward are common and can lead to adverse events. Efforts to improve written handoffs during these transfers are increasingly prominent, but no instruments have been developed to assess the quality of physician ICU–ward transfer notes.

Objective: To collect validity evidence for the modified nine-item Physician Documentation Quality Instrument (mPDQI-9) for assessing ICU–ward transfer note usefulness across several hospitals.

Methods: Twenty-four physician raters independently used the mPDQI-9 to grade 12 notes collected from three academic hospitals. *A priori*, we excluded the "up-to-date" and "accurate" domains, because these could not be assessed without giving the rater access to the complete patient chart. Assessments therefore used the domains "thorough," "useful," "organized," "comprehensible," "succinct," "synthesized," and "consistent." Raters scored each domain on a Likert scale ranging from 1 (*low*) to 5 (*high*). The total mPDQI-9 was the sum of these domain scores. The primary outcome was the raters' perceived clinical utility of the notes, and the primary measures of interest were criterion validity (Spearman's ρ) and interrater reliability (intraclass correlation [ICC]).

Results: Mean mPDQI-9 scores by note ranged from 19 (SD = 5.5) to 30 (SD = 4.2). Mean note ratings did not systematically differ by rater expertise (for interaction, $P = 0.15$). The proportion of raters perceiving each note as independently sufficient for patient care (the primary outcome) ranged from 33% to 100% across the set of notes. We found a moderately positive correlation between mPDQI-9 ratings and raters' overall assessments of each note's clinical utility ($\rho = 0.48$, $P < 0.001$). Interrater reliability was strong; the overall ICC was 0.89 (95% confidence interval [CI], 0.80–0.85), and ICCs were similar among reviewer groups. Finally, Cronbach's α was 0.87 (95% CI, 0.84–0.89), indicating good internal consistency.

Conclusions: We report moderate validity evidence for the mPDQI-9 to assess the usefulness of ICU–ward transfer notes written by internal medicine residents.

Keywords:
ICU-ward transfers, handoffs, quality and safety, clinical informatics, electronic health records

Physician communication failures during patient transfers from the intensive care unit (ICU) to the general ward (ICU–ward) are common and associated with adverse events, near-misses, and worsened patient and family experiences (1–7). Such handoff miscommunication may take the form of omitted information or incorrect information (1, 2), both of which may be propagated by electronic health record (EHR) issues such as copy/paste (8–10).

Efforts to improve ICU–ward physician handoffs are nascent and have, to date, focused on education (11–15) and standardization, including the development of structured tools analogous to those proven beneficial for other handoff situations (16). For instance, the structured illness severity, patient summary, action list, situational awareness, and synthesis by receiver (or, I-PASS) program has been associated with reductions in medical errors and in preventable adverse events when implemented in pediatric settings (16, 17). In critical care settings, implementation of a standardized electronic handoff tool led to earlier and more detailed completion of ICU transfer notes in patient charts (14).

Despite these efforts, a persistent gap in this area is that no instruments have been developed to assess the quality of physician ICU-ward transfer notes. Thus, it is unknown whether, for example, notes that are deemed thorough or complete would also be accurate or helpful (14). Moreover, this gap has direct relevance to postgraduate educational curricula on

quality and safety, physician and interpersonal communication, documentation, and clinical reasoning. It is, therefore, essential for us to evaluate the central question underlying these assumptions empirically: Are the notes that we require (and train) our house staff and faculty to write actually useful to other clinicians?

Although the use of a standard template might achieve some degree of consistency within and among note writers, an assessment tool is needed to measure consistency and quality empirically (18). One such tool, the nine-item Physician Documentation Quality Instrument (PDQI-9) has validity for assessing EHR daily progress notes for hospitalized patients (19, 20). However, no analogous tools currently exist for ICU–ward transfer notes, which differ fundamentally from other note types in purpose and content (summarizing and transferring content/ responsibility vs. documenting evolving clinical information and reasoning). Thus, we aimed to provide validity evidence for extending this instrument to the assessment of ICU–ward transfer notes by testing it across several hospitals.

Some of the results of these studies have been previously reported in the form of an abstract (21).

METHODS
Study Design, Setting, and Data

We conducted a retrospective observational study of ICU–ward transfer notes from three academic hospitals: Barnes Jewish Hospital (the teaching hospital of Washington University School of Medicine); the University of Chicago Medicine; and the University of California, San Francisco. Each hospital's institutional review board deemed this work as not human subjects research and, therefore, exempt from review.

We retrospectively collected and deidentified ICU-to-ward transfer notes (selected randomly from daily patient transfer lists) from medical ICU stays occurring between July 1 and September 31, 2019. During deidentification, we also removed any references to the hospital, including local acronyms and hospital-specific jargon (i.e., raters could not necessarily identify the hospital from which each note had originated).

At all three sites, medical ICU-to-ward transfer notes are routinely written by internal medicine residents and are entered into the EHR at or around the time of transfer to the wards. These transfer notes are not used for billing and are not routinely edited or cosigned by ICU attending physicians; at each institution, transfer notes were entered as separate entities from the daily ICU progress note. During the study period, no templates were used for these notes, and no specific training or guidance was provided to residents on note format or content.

Raters

We recruited physician raters from each study site and pooled all raters for the primary analysis. All raters reviewed all notes. To determine whether particular educational or administrative expertise (beyond subject matter familiarity) would be relevant to raters' ability to use the instrument (which would have implications for programmatic use and scale-up), we assigned raters to one of two groups *post hoc*. Faculty members who served in internal medicine residency program leadership roles or who had methods or subject matter expertise in patient safety, quality improvement, and/or handoffs were labeled "Group A." The other raters (hereinafter termed "Group B") included ward attending

Table 1. Domains from the mPDQI–9

mPDQI–9 Domain	Definition
Thorough	Complete and documents all issues of importance to the patient
Useful	Extremely relevant, providing useful information and/or analysis
Organized	Well formed and structured in a way that helps the reader understand the patient's clinical course
Comprehensible	Clear, without ambiguity or sections that are difficult to understand
Succinct	Brief, to the point, and without redundancy
Synthesized	Reflects the author's understanding of the patient's status and ability to develop a plan of care
Internally consistent	No part of the note ignores or contradicts any other part

Definition of abbreviation: mPDQI–9 = modified nine–item Physician Documentation Quality Instrument. Each item is rated on a 5–point Likert scale ranging from 1 (*not at all*) and 5 (*extremely*).

physicians, critical care fellows, and internal medicine chief residents.

The PDQI–9 Instrument

In REDCap (22), we electronically presented each rater with individual transfer notes, a PDQI-9 rubric, and a rating sheet. The PDQI-9 has been validated for assessing the quality of progress notes and discharge summaries. This tool assesses note quality in nine domains (Table 1): up-to-date, accurate, thorough, useful, organized, comprehensible, succinct, synthesized, and consistent. Each domain is rated on a 5-point Likert scale ranging from 1 (*not at all*) to 5 (*extremely*). *A priori*, we excluded the "up-to-date" and "accurate" domains, because raters could not assess these items without access to the patients' charts. Thus, the total score for the modified PDQI (mPDQI-9) could range from 7 (indicating the worst possible note quality) to 35 (indicating the best possible quality).

Outcomes, Measures, and Sample Size Estimation

Modern theory posits that "validity" is essentially a hypothesis that can be accepted or rejected on the basis of validity evidence (i.e., qualitative or quantitative data informing our overall evaluative judgment of the adequacy and appropriateness of inferences drawn from an assessment's results). Within Messick's framework (23), we sought content and internal structure validity evidence for using the mPDQI-9 to assess ICU–ward transfer note quality. Here, "content" refers to the relationship between test items and the construct they are intended to measure (in terms of themes, wording, and item format), and "internal structure" refers to the extent to which each item fits the instrument's underlying constructs (generally through quantitative measures).

The primary outcome was each note's perceived clinical utility, which we measured as the binary "yes" or "no" response to the general question "Without any additional information, could you use this note to manage this patient if called for help?"

With general clinical utility as the intended construct, we identified, *a priori*, two primary measures of interest: Spearman's rank correlation coefficient (ρ) between the total modified PDQI-9 score and the general utility question (evidence

of content validity, related to the previously used concept of criterion validity) (24) and interrater reliability (the extent to which any two independent raters would agree, evidence of internal structure validity) for the total mPDQI-9 score. We measured interrater reliability with averaged two-way random effects intraclass correlations (ICCs) (25, 26). As further internal structure validity evidence, a secondary measure of interest was internal consistency reliability (Cronbach's α).

On the basis of our expected interrater reliability (ICC, ~0.67, near the lower boundary of PDQI-9 ICCs for other types of notes) (19) and identification of 24 raters, we estimated that 12 transfer notes would yield precision ±0.2 and therefore captured four notes per clinical site (27).

Exploratory Analyses

We performed several prespecified exploratory analyses. First, because differential experience, skill, and expectations might influence how receiving clinicians perceive ICU–ward transfer notes, their quality, and their clinical utility, we prespecified an exploratory subgroup analysis by area of expertise in handoffs and patient safety (i.e., Group A vs. Group B) (19). We used an analysis of variance to test whether mPDQI-9 scoring varied by expertise, with plans for *post hoc* testing if interaction term P values were <0.01 (to minimize risk of false discovery) (28).

Second, to generate hypotheses regarding which note qualities might be most essential for ICU-ward transfer documentation, we measured Spearman correlation coefficients between each individual mPDQI-9 component and the general utility question.

Finally, longer ICU–ward transfer notes are associated with negative patient outcomes (29), likely because of multiple mechanisms including patient severity and complexity,

diagnostic certainty, and challenges in gleaning information from unnecessarily long notes. It is important to note that the original 22-item PDQI contained four factors directly related to length ("brief," "concise," "succinct," and "focused") as well as multiple items that are likely to interact with brevity (e.g., shorter notes may be more likely to be "nonredundant," whereas "complete" and "thorough" notes may require additional length) (20). Thus, in an exploratory analysis, we also calculated each note's word count and compared these with mPDQI-9 ratings and the overall utility assessment with Pearson and Spearman correlation coefficients, respectively.

Statistical Analyses

We used frequencies (with percentage) for categorical data and mean (with standard deviation [SD]) to describe rater characteristics and mPDQI-9 rating distributions. We used R 4.1 (The R Project for Statistical Computing) and the *tidyverse* for all analyses with the exception of ICCs (for which we used the *psych* package), and Cronbach's α (for which we used *ltm*) (30–32).

RESULTS

We recruited 24 raters for 12 notes; each rater independently rated all notes. Raters in Group A had generally been in practice longer and held more senior academic ranks than the remainder of the raters (*see* Table E1 in the data supplement).

Among the 12 notes reviewed (Figure 1 and Table 2), total mPDQI-9 scores ranged from 8 to 35, with mean scores by note ranging from 19 (SD = 5.5) to 30 (SD = 4.2). Mean note ratings did not systematically differ by rater expertise (for interaction, $P = 0.15$). The proportion of raters perceiving each note as independently sufficient for patient care (the primary outcome) ranged from 33% to 100% across the set of notes.
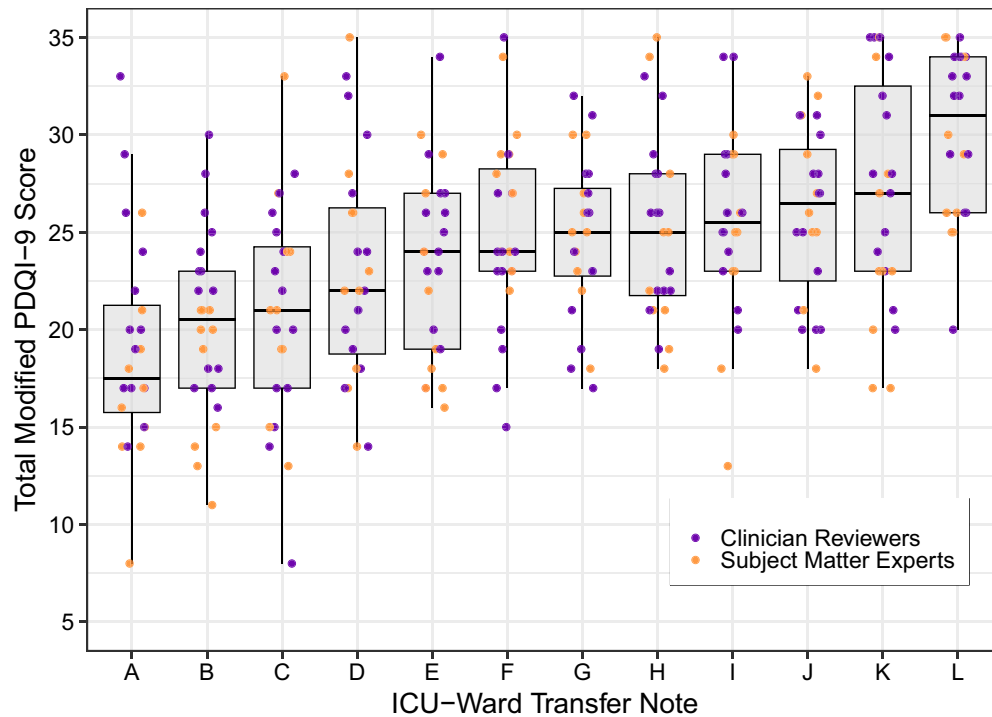
Figure 1. Modified nine-item Physician Documentation Quality Instrument (mPDQI-9) score distributions varied across individual notes. For each note (*x*-axis), scattered dots represent individual raters' mPDQI scores (*y*-axis), and the boxplots indicate median scores with interquartile ranges. Whiskers on each boxplot represent the 95% confidence interval for each note's scores.

We found a moderately positive correlation between mPDQI-9 ratings and raters' overall assessments of each note's clinical utility ($\rho = 0.49$, $P < 0.001$; Table 3 and Figure E1). Correlations were similar between rater groups (Group A, $\rho = 0.53$, $P < 0.001$; Group B, $\rho = 0.45$, $P < 0.001$). Interrater reliability was strong; the overall ICC was 0.89 (95% confidence interval [CI], 0.78–0.96), and ICCs were similar among Group A reviewers (0.81, 95% CI, 0.61–0.94) and Group B reviewers (0.82; 95% CI, 0.64–0.94). Finally, Cronbach's α was 0.87 (95% CI, 0.84–0.89), indicating good internal consistency.

Individual mPDQI-9 components generally differed between helpful and unhelpful transfer notes (ρ range = 0.25–0.51, $P < 0.001$; Figure 2), with the exception of succinctness negative rho value ($\rho = 0.-0.2$, $P = 0.76$). Note length was only modestly correlated

with mPDQI score (Pearson coefficient, 0.14, $P = 0.02$) and the primary outcomes ($\rho = 0.16$, $P = 0.008$).

DISCUSSION

We tested the modified PDQI-9's ability to assess ICU–ward transfer notes, providing moderate validity evidence in terms of content and internal structure. Although interrater reliability and internal consistency were high, the correlation between mPDQI-9 scores and raters' overall assessments was only modest. The latter measurement might indicate the mPDQI-9's lack or underdevelopment of one or more content areas within the overall concept of utility, such as being up-to-date or accurate (the domains omitted from this study). Conversely, the relatively low correlation could also indicate that one or more items

Table 2. Characteristics of patient transfer note artifacts

Characteristic	Note A, Site 1	Note B, Site 2	Note C, Site 1	Note D, Site 3	Note E, Site 3	Note F, Site 3	Note G, Site 2	Note H, Site 1	Note I, Site 2	Note J, Site 2	Note K, Site 1	Note L, Site 3
Word count, no.	637	1,106	148	1,333	2,111	262	943	435	370	2,133	626	1,054
mPDQI-9 score, mean (SD)	19.1 (5.5)	20.2 (4.7)	20.9 (5.6)	23 (5.7)	23.5 (4.8)	24.7 (4.2)	25 (4.8)	25.2 (4.9)	25.5 (4.7)	26 (4.3)	26.9 (5.9)	30.1 (4.2)
mPDQI-9 scores, range	8–33	11–30	8–33	14–35	16–34	17–32	15–35	18–35	13–34	18–33	17–35	20–35
Proportion of "useful" ratings on the general utility question	33.3	58.3	33.3	83.3	87.5	87.5	58.3	91.7	83.3	83.3	79.2	100

Definition of abbreviations: mPDQI-9 = modified nine-item Physician Documentation Quality Instrument; SD = standard deviation.

have an unexpected relationship with utility. In particular, the perceived importance of brevity and succinctness may vary on a case-by-case basis; for example, transfer notes for patients with prolonged complex ICU stays may not be helpful if they are too succinct.

Overall, these results suggest that this instrument may be useful for the objective assessment of transfer notes when patients move from the ICU to the wards. To our knowledge, our work appears to be the first attempt to collect validity evidence for a standardized instrument to assess ICU–ward transfer note quality. Further-more, our findings have several additional important implications. First, the mPDQI-9 could serve as a process measure in the evaluation of educational and/or quality and safety interventions in the ICU. For example, measuring mPDQI-9 scores before and after implementing a standard-ized ICU–ward transfer note template would provide valuable objective data beyond clinician perspectives (33–35) and might offer important mechanistic information to aid interpretation of other implementation, process, and patient outcomes.

From the educational standpoint, a scoring tool supported by validity evidence offers the potential to enhance standardized objective assessment of learner performance in the domains of communication skills, practice-based learning and improvement, and systems-based practice (36). In the most recent set of Accreditation Council for Graduate Medical Education Milestones, transitions of care ("System Navigation for Patient-Centered Care") are similarly emphasized (37). Objective measurements of trainees' transfer notes might have substan-tial value as cross-sectional or longitudinal adjuncts to direct observation and narrative assessments in these areas. As the

Table 3. Validity evidence of mPDQI-9 performance in assessing notes on patient transfers from the ICU to the general ward

Measure and Raters	Value	95% Confidence Interval	P Value
Spearman's ρ			
All	0.49	NA	<0.001
Group A	0.53	NA	<0.001
Group B	0.45	NA	<0.001
Intraclass coefficient			
All	0.89	0.78–0.96	NA
Group A	0.81	0.61–0.94	NA
Group B	0.82	0.64–0.94	NA
Cronbach's α			
All	0.87	0.84–0.89	NA
Group A	0.88	0.84–0.91	NA
Group B	0.86	0.83–0.89	NA

Definition of abbreviations: ICU = intensive care unit; mPDQI-9 = modified Physician Documentation Quality Instrument; NA = not applicable.
Group A consisted of raters who were faculty members serving in internal medicine residency program leadership roles or who had methods or subject matter expertise in patient safety, quality improvement, and/or handoffs. Group B consisted of raters including ward attending physicians, critical care fellows, and internal medicine chief residents.

educational landscape moves away from summative assessments to more high-quality feedback and formative assessments such as structured entrustable professional activities (38), trainees might receive high-quality feedback on a transfer note by means of a tool such as the PDQI, thus improving their communication and systems-based practice skills. Specifically, we envision that the PDQI could be used to educate interns and residents in writing effective transfer summaries. For example, performing a successful ICU–ward transfer could even serve as an entrustable professional activity that faculty could observe and sign off on.

Finally, a tool such as the mPDQI-9 might also help trainees answer the central question of whether the time they spend on documentation actually translates to other clinicians understanding their critical thinking and patient care plans. At the individual level, reflecting on this question could encourage self-directed learning and continuous improvement. More important, using such a tool could aid introspection in our profession on how we might compose briefer, more comprehensible, and higher yield notes, rather than contributing to ever-persistent "note bloat" (39–41). Future directions might also explore using the PDQI to train artificial intelligence algorithms to autosuggest improvements to transfer summaries with predictive text technology.

An important limitation of this work is that mPDQI-9 assessments require time from clinician raters who are familiar with ICU syndromes and care processes, which
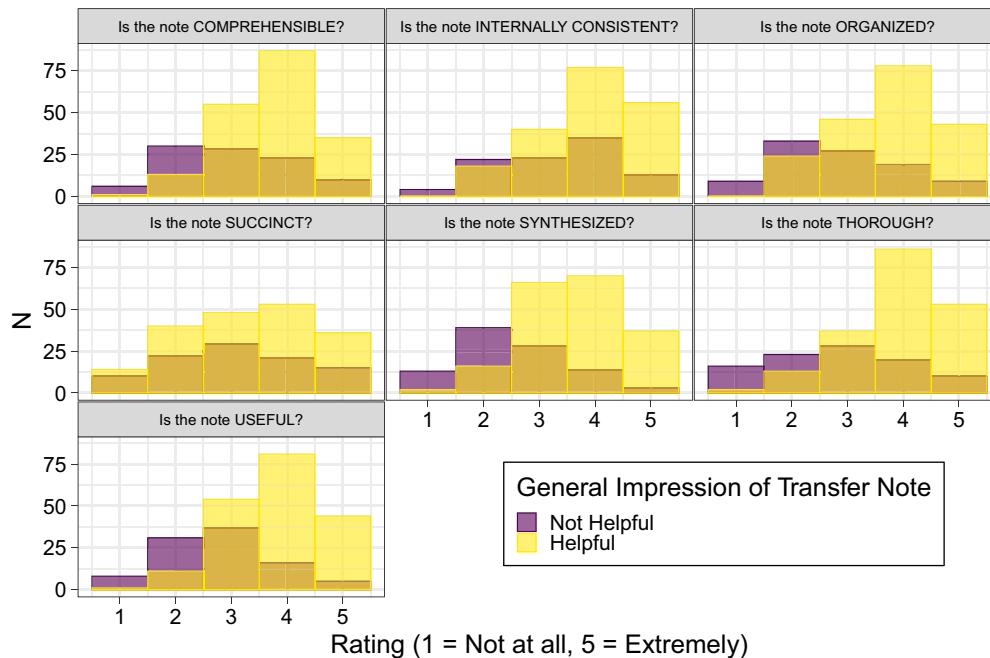
Figure 2. Subscores on the Modified nine-item Physician Documentation Quality Instrument generally differentiated helpful from unhelpful notes on patient transfers from the intensive care unit to the general ward. The *x*-axis indicates the Likert score of each item (averaged over all notes), and the *y*-axis indicates the number of note ratings with each score value. Histograms are shaded by raters' responses to the overall utility question (purple indicates unhelpful and yellow indicates helpful).

may limit immediate utility to smaller scale assessment activities such as focused educational initiatives or research studies with dedicated effort toward process measurement. It is important to note that rater expertise did not have much bearing on mPDQI-9 performance, suggesting that particular educational or administrative expertise (beyond subject matter familiarity) may not be required for use. Moreover, this limitation may be overcome with help from rapidly evolving artificial intelligence tools such as large language models (42); we speculate that, in the near future, a model based on medical lexicon training data may be able to provide near-human performance on discrete tasks such as mPDQI-9 ratings, allowing assessments to be collected on a large scale without substantial rater time. A second limitation is the unavailability of chart review necessitating modification

of the PDQI-9 instrument (omitting two domains that could not be assessed without full access to patient charts). Despite this limitation, raters did not report difficulty rating notes' overall utility or the remaining PDQI domains, which appear to have value for our purpose. Finally, this study provides only moderate validity evidence and does not examine other aspects of validity (e.g., consequences). Future work should explore these areas and test the extent to which including the omitted domains yields stronger validity evidence.

In conclusion, we found moderate validity evidence for the modified PDQI-9 to assess ICU–ward transfer notes written by internal medicine residents. This instrument may have value for educational, quality, and informatics activities in and around the ICU.

REFERENCES

1. Santhosh L, Lyons PG, Rojas JC, Ciesielski TM, Beach S, Farnan JM, et al. Characterising ICU-ward handoffs at three academic medical centres: process and perceptions. *BMJ Qual Saf* 2019;28:627–634.

2. Lyons PG, Arora VM, Farnan JM. Adverse events and near-misses relating to intensive care unit-ward transfer: a qualitative analysis of resident perceptions. *Ann Am Thorac Soc* 2016;13:570–572.

3. Li P, Stelfox HT, Ghali WA. A prospective observational study of physician handoff for intensive-care-unit-to-ward patient transfers. *Am J Med* 2011;124:860–867.

4. Sauro KM, Soo A, de Grood C, Yang MMH, Wierstra B, Benoit L, et al. Adverse events after transition from ICU to hospital ward: a multicenter cohort study. *Crit Care Med* 2020;48:946–953.

5. Li P, Boyd JM, Ghali WA, Stelfox HT. Stakeholder views regarding patient discharge from intensive care: suboptimal quality and opportunities for improvement. *Can Respir J* 2015;22:109–118.

6. Stelfox HT, Lane D, Boyd JM, Taylor S, Perrier L, Straus S, et al. A scoping review of patient discharge from intensive care: opportunities and tools to improve care. *Chest* 2015;147:317–327.

7. Pilcher DV, Duke GJ, George C, Bailey MJ, Hart G. After-hours discharge from intensive care increases the risk of readmission and death. *Anaesth Intensive Care* 2007;35:477–485.

8. 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;23:406–413.

9. Weis JM, Levy PC. Copy, paste, and cloned notes in electronic health records. *Chest* 2014;145:632–638.

10. Steinkamp J, Kantrowitz JJ, Airan-Javia S. Prevalence and sources of duplicate information in the electronic medical record. *JAMA Netw Open* 2022;5:e2233348.

11. Alali H, Antar M, AlShehri A, AlHamouieh O, Al-Surimi K, Kazzaz Y. Improving physician handover documentation process for patient transfer from paediatric intensive care unit to general ward. *BMJ Open Qual* 2020;9:e001020.

12. Herbst LA, Desai S, Benscoter D, Jerardi K, Meier KA, Statile AM, et al. Going back to the ward-transitioning care back to the ward team. *Transl Pediatr* 2018;7:314–325.

13. Correia PC, Gomes de Macedo P, Santos JFG, Moreira Júnior JR, de Oliveira C, Malbouisson LMS. Impact of customised ICU handover protocol on the quality of ICU discharge reports. *BMJ Open Qual* 2022;11:e001647.

14. Stelfox HT, Brundin-Mather R, Soo A, Whalen-Browne L, Kashyap D, Sauro KM, et al. A multiple baseline trial of an electronic ICU discharge summary tool for improving quality of care. *Crit Care Med* 2022;50:1566–1576.

15. Parsons Leigh J, Brundin-Mather R, Whalen-Browne L, Kashyap D, Sauro K, Soo A, et al. Effectiveness of an electronic communication tool on transitions in care from the intensive care unit: protocol for a cluster-specific pre-post trial. *JMIR Res Protoc* 2021;10:e18675.

16. Starmer AJ, Spector ND, Srivastava R, West DC, Rosenbluth G, Allen AD, et al.; I-PASS Study Group. Changes in medical errors after implementation of a handoff program. *N Engl J Med* 2014; 371:1803–1812.

17. Starmer AJ, Sectish TC, Simon DW, Keohane C, McSweeney ME, Chung EY, et al. Rates of medical errors and preventable adverse events among hospitalized children following implementation of a resident handoff bundle. *JAMA* 2013;310:2262–2270.

18. Pastores SM, O'Connor MF, Kleinpell RM, Napolitano L, Ward N, Bailey H, et al. The Accreditation Council for Graduate Medical Education resident duty hour new standards: history, changes, and impact on staffing of intensive care units. *Crit Care Med* 2011;39:2540–2549.

19. Stetson PD, Bakken S, Wrenn JO, Siegler EL. Assessing electronic note quality using the Physician Documentation Quality Instrument (PDQI-9). *Appl Clin Inform* 2012;3:164–174.

20. Stetson PD, Morrison FP, Bakken S, Johnson SB; eNote Research Team. Preliminary development of the physician documentation quality instrument. *J Am Med Inform Assoc* 2008;15:534–541.

21. Lyons P, Rojas JC, Baty J, Santhosh L. How reliably can we grade intensive care unit-ward transfer notes? Validating the Physician Documentation Quality Instrument in assessment of ICU-ward transfer notes [abstract]. *Am J Respir Crit Care Med* 2020;201:A1398.

22. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377–381.

23. Messick S. Validity. In: Linn RL, editor. Educational measurement. 3rd ed. Phoenix, Ariz: Oryx Press; 1993.

24. Spearman C. The proof and measurement of association between two things. *Am J Psychol* 1904;15: 72–101.

25. Bartko JJ. The intraclass correlation coefficient as a measure of reliability. *Psychol Rep* 1966;19:3–11.

26. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;15:155–163.

27. Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. *Stat Med* 2002;21:1331–1335.

28. Chen SY, Feng Z, Yi X. A general introduction to adjustment for multiple comparisons. *J Thorac Dis* 2017;9:1725–1729.

29. Orenstein EW, Ferro DF, Bonafide CP, Landrigan CP, Gillespie S, Muthu N. Hidden health IT hazards: a qualitative analysis of clinically meaningful documentation discrepancies at transfer out of the pediatric intensive care unit. *JAMIA Open* 2019;2:392–398.

30. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, Francois R, et al. Welcome to the tidyverse. *J Open Source Softw* 2019;4:1686.

31. Revelle W. psych: procedures for psychological, psychometric, and personality research [R package psych version 2.3.6]. 2023 [accessed 2023 Jun 27]. Available from: https://CRAN.R-project.org/package=psych.

32. Rizopoulos D. ltm: an R package for latent variable modeling and item response analysis. *J Stat Softw* 2007;17:1–25.

33. Fukui EM, Lyons PG, Harris E, McCune EK, Rojas JC, Santhosh L. Improving communication in intensive care unit to ward transitions: protocol for multisite national implementation of the ICU-PAUSE handoff tool. *JMIR Res Protoc* 2023;12:e40918.

34. Santhosh L, Rojas JC, Garcia B, Thomashow M, Lyons PG. Cocreating the ICU-PAUSE tool for intensive care unit–ward transitions. *ATS Scholar* 2022;3:312–323.

35. Cornell EG, Harris E, McCune E, Fukui E, Lyons PG, Rojas JC, et al. Scaling up a diagnostic pause at the ICU-to-ward transition: an exploration of barriers and facilitators to implementation of the ICU-PAUSE handoff tool. *Diagnosis (Berl)* 2023;10:417–423.

36. Swing SR, Beeson MS, Carraccio C, Coburn M, Iobst W, Selden NR, et al. Educational milestone development in the first 7 specialties to enter the next accreditation system. *J Grad Med Educ* 2013; 5:98–106.

37. Edgar L, Roberts S, Holmboe E. Milestones 2.0: a step forward. *J Grad Med Educ* 2018;10: 367–369.

38. Fessler HE, Addrizzo-Harris D, Beck JM, Buckley JD, Pastores SM, Piquette CA, *et al.* Entrustable professional activities and curricular milestones for fellowship training in pulmonary and critical care medicine: report of a multisociety working group. *Chest* 2014;146:813–834.

39. Kahn D, Stewart E, Duncan M, Lee E, Simon W, Lee C, *et al.* A prescription for note bloat: an effective progress note template. *J Hosp Med* 2018;13:378–382.

40. Rule A, Bedrick S, Chiang MF, Hribar MR. Length and redundancy of outpatient progress notes across a decade at an academic medical center. *JAMA Netw Open* 2021;4:e2115334.

41. 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: 293–303.

42. Yilmaz Y, Jurado Nunez A, Ariaeinejad A, Lee M, Sherbino J, Chan TM. Harnessing natural language processing to support decisions around workplace-based assessment: machine learning study of competency-based medical education. *JMIR Med Educ* 2022;8:e30537.