

# Structured Chart Review: Assessment of a Structured Chart Review Methodology

Ashley Siems, MD,<sup>a</sup> Russell Banks, MS,<sup>b</sup> Richard Holubkov, PhD,<sup>b</sup> Kathleen L. Meert, MD,<sup>c</sup> Christian Bauerfeld, MD,<sup>c</sup> David Beyda, MD,<sup>d</sup> Robert A. Berg, MD,<sup>e</sup> Yonca Bulut, MD,<sup>f</sup> Randall S. Burd, MD, PhD,<sup>a</sup> Joseph Carcillo, MD,<sup>g</sup> J. Michael Dean, MD,<sup>b</sup> Eleanor Gradidge, MD,<sup>d</sup> Mark W. Hall, MD,<sup>h</sup> Patrick S. McQuillen, MD,<sup>i</sup> Peter M. Mourani, MD,<sup>j</sup> Christopher J.L. Newth, MD, FRCPC,<sup>k</sup> Daniel A. Notterman, MD,<sup>l</sup> Margaret A. Priestley, MD,<sup>e</sup> Anil Sapru, MD,<sup>f</sup> David L. Wessel, MD,<sup>a</sup> Andrew R. Yates, MD,<sup>h</sup> Athena F. Zuppa, MD,<sup>e</sup> Murray M. Pollack, MD<sup>a</sup>

## ABSTRACT

**BACKGROUND AND OBJECTIVES:** Chart reviews are frequently used for research, care assessments, and quality improvement activities despite an absence of data on reliability and validity. We aim to describe a structured chart review methodology and to establish its validity and reliability.

**METHODS:** A generalizable structured chart review methodology was designed to evaluate causes of morbidity or mortality and to identify potential therapeutic advances. The review process consisted of a 2-tiered approach with a primary review completed by a site physician and a short secondary review completed by a central physician. A total of 327 randomly selected cases of known mortality or new morbidities were reviewed. Validity was assessed by using postreview surveys with a Likert scale. Reliability was assessed by percent agreement and interrater reliability.

**RESULTS:** The primary reviewers agreed or strongly agreed in 94.9% of reviews that the information to form a conclusion about pathophysiological processes and therapeutic advances could be adequately found. They agreed or strongly agreed in 93.2% of the reviews that conclusions were easy to make, and confidence in the process was 94.2%. Secondary reviewers made modifications to 36.6% of cases. Duplicate reviews ( $n = 41$ ) revealed excellent percent agreement for the causes (80.5%–100%) and therapeutic advances (68.3%–100%).  $\kappa$  statistics were strong for the pathophysiological categories but weaker for the therapeutic categories.

**CONCLUSIONS:** A structured chart review by knowledgeable primary reviewers, followed by a brief secondary review, can be valid and reliable.



<sup>a</sup>School of Medicine and Health Sciences, The George Washington University and Children's National Health System, Washington, District of Columbia; <sup>b</sup>School of Medicine, University of Utah, Salt Lake City, Utah; <sup>c</sup>Wayne State University and Children's Hospital of Michigan, Detroit, Michigan; <sup>d</sup>College of Medicine—Phoenix, University of Arizona and Phoenix Children's Hospital, Phoenix, Arizona; <sup>e</sup>Children's Hospital of Philadelphia, Philadelphia, Pennsylvania; <sup>f</sup>University of California, Los Angeles and University of California, Los Angeles Mattel Children's Hospital, California; <sup>g</sup>University of Pittsburgh Medical Center Children's Hospital of Pittsburgh, Pittsburgh, Pennsylvania; <sup>h</sup>Nationwide Children's Hospital, Columbus, Ohio; <sup>i</sup>University of California, San Francisco and University of California, San Francisco Benioff Children's Hospital, San Francisco, California; <sup>j</sup>University of Colorado and Children's Hospital of Colorado, Denver, Colorado; <sup>k</sup>Keck School of Medicine, University of Southern California and Children's Hospital Los Angeles, Los Angeles, California; and <sup>l</sup>Princeton University, Princeton, New Jersey

www.hospitalpediatrics.org

DOI: <https://doi.org/10.1542/hpeds.2019-0225>

Copyright © 2020 by the American Academy of Pediatrics

Address correspondence to Murray M. Pollack, MD, Children's National Medical Center, 111 Michigan Ave NW, Washington, DC 20010.

E-mail: [mpollack@childrensnational.org](mailto:mpollack@childrensnational.org)

HOSPITAL PEDIATRICS (ISSN Numbers: Print, 2154-1663; Online, 2154-1671).

**FINANCIAL DISCLOSURE:** The authors have indicated they have no financial relationships relevant to this article to disclose.

**FUNDING:** Supported, in part, by the following cooperative agreements from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Department of Health and Human Services: U10HD050096, U10HD049981, U10HD049983, U10HD050012, U10HD063108, U10HD063114 and U01HD049934. UG1HD083171, UG1HD083166, and UG1HD083170. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Funded by the National Institutes of Health (NIH).

**POTENTIAL CONFLICT OF INTEREST:** The authors have indicated they have no potential conflicts of interest to disclose.

Drs Siems, Meert, and Pollack conceptualized and designed the study, drafted the initial manuscript, participated in data collection, chart reviews, and the data analysis, and reviewed and revised the manuscript; Mr Banks conceptualized and designed the study, coordinated and supervised data collection, and critically reviewed the manuscript for important intellectual content; Drs Beyda, Berg, Burd, Carcillo, Dean, Hall, Newth, Notterman, Sapru, Wessel, and Zuppa helped design the study, assisted with data collection, reviewed the data analysis, and reviewed and revised the manuscript; Drs Bauerfeld, Bulut, Gradidge, McQuillen, Mourani, Priestley, and Yates helped with data collection, data analysis, and revisions of the manuscript; and all authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

Chart review is a methodology frequently used to perform retrospective research and care assessments and conduct quality improvement activities, including morbidity and mortality conferences, critical event reviews, and root cause analyses.<sup>1</sup> This methodology facilitates epidemiological investigations, quality assessments, education, and clinical research.<sup>2-4</sup> Retrospective chart reviews are used in 15% to 25% of scientific articles in emergency medical journals.<sup>5,6</sup> Additionally, >90% of PICUs use this approach to inform their morbidity and mortality conferences.<sup>7</sup> Despite their widespread use and potential impact, chart reviews are time consuming, often expensive, and often yield substantial variability among reviewers, a finding not unexpected when assessing complex data and forming subjective assessments.<sup>8,9</sup> Chart review is seldom taught in medical school, and medical students report having limited experience with this process.<sup>10</sup>

The reliability and validity of chart review are dependent on features of the medical record as well as the subjectivity of the review elements. Chart reviews may include appraising the medical record's physician and nursing notes, outpatient and emergency department reports, consultations, admission and discharge documentation, and diagnostics.<sup>11</sup> Challenging features of the medical record include incomplete documentation, variation in the location of information, discrepancy in information between locations, reliance on jargon or handwriting interpretation in older records, unrecoverable information, and dependency on accurate documentation. Challenging features of chart reviews include training reviewers, creating an operational procedure for review and assessment, resolving discrepancies among multiple reviewers, and assessing validity and reliability.<sup>12</sup> Objective review criteria with defined, measurable content are likely to have high reliability and validity, whereas subjective review criteria may have comparatively reduced agreement, validity, and reproducibility. In randomized controlled trials, subjective outcomes when compared with objective outcomes are associated with bias and overestimation of results.<sup>13,14</sup>

Subjective variables have poor interrater reliability and agreement when compared with items that have clear criteria.<sup>15,16</sup> Regardless of these limitations, standardized chart reviews have high face validity and are commonly used. For example, standardized medical record review has been used to classify areas of potential harm and to detect adverse events and negligence.<sup>3,17-19</sup>

Because of the limitations of previous chart review methodologies, our goal was the development of an improved and generalizable structured chart review methodology as part of a larger initiative of the Collaborative Pediatric Critical Care Research Network (CPCCRN), Informing the Research Agenda. CPCCRN's goal was to better understand the pathophysiology associated with significant new morbidity and mortality and to identify therapeutic advances that might have prevented or ameliorated these adverse outcomes at the individual patient level. Subsequent publications in which this methodology is used are in preparation. Determining these elements is subjective in the context of complex critical illness. Our aims were (1) the development of a structured, reproducible, time-efficient, and generalizable method of chart review and (2) the assessment of this methodology for validity and reliability.

## METHOD

### Patient Data

Patient lists were generated from the Trichotomous Outcome Prediction in Critical Care (TOPICC) data set, a database used to investigate morbidity and mortality in the PICU population.<sup>20</sup> Eligible patients included those who had sustained a significant new morbidity at hospital discharge, defined as a Functional Status Scale (FSS) score increase of  $\geq 2$  in a single FSS domain, or died with an admission mortality risk of  $< 80\%$  (ie, not "dead on arrival"). The FSS is a validated granular assessment of function based on the principles of adaptive behavior in 6 domains (feeding, respiratory, motor, communication, mental, and sensory), with higher scores indicating more significant morbidity.<sup>21,22</sup> For this study, eligible patients from the TOPICC data set at each site were

randomly sorted. Sites were provided a list of the study identifiers linked to these patients, along with dates of admission and discharge, the FSS scores at admission and hospital discharge, and survival status at hospital discharge. The individual sites then linked the TOPICC identification numbers to the patient's medical record number and provided this information to the primary reviewer. The study was approved by the CPCCRN centralized Institutional Review Board.

### Structured Chart Review

To develop the structured chart review process, we followed a common-sense multistep approach adapted from the trigger tool methodology (detailed in Table 1).<sup>23,24</sup> First, we developed goals and objectives for the project. The goals and objectives of the project were to identify pathophysiological processes resulting in new morbidities and mortalities and to identify needed therapeutic additions or advances that could potentially reduce morbidity and mortality. The classification schemes developed by the CPCCRN steering committee using an expert panel employing qualitative research methods are shown in the data collection form (Supplemental Fig 2).<sup>25-27</sup> More than 1 pathophysiological process or potential therapeutic addition or advance could be relevant to each case.

We also developed performance parameters, including time requirements for the reviews and expertise requirements for the reviewers. These criteria were intended to ensure that the project was feasible and could be completed in a timely fashion and to guarantee that the person performing the chart review was qualified. We standardized the instructional process to allow the study to be conducted in multiple institutions. This process included supervision of the initial reviews by a study co-principal investigator. Finally, we included a process for oversight and peer checking with the central reviewer as the final phase of the chart review. This central review confirmed the conclusions of the primary reviews and helped ensure uniformity and consistency in the classifications.

In Fig 1, the chart review process is outlined with steps for the primary reviewer to

**TABLE 1** Steps Taken to Develop the Chart Review Process

Step	Examples From Our Process
Establish goals and objectives of the chart review	Identify pathophysiology and potential therapeutic advances in ICU morbidities and mortalities
Define the parameters that achieve the goals and objectives of the chart review	Trigger: new morbidity or mortality (by using the FSS)
Establish a standardized data collection tool	Causes and interventions: CPCCRN expert panel
Identify possible locations within the EHR to find data	See Supplemental Fig 2
Establish time limits for each phase of the chart review	Admission note, discharge note, consultant notes
Identify criteria for primary reviewers	5 min for confirmation of trigger; 15 min for identification of the cause; 10 min for the interventions; 5 min for the central review
Develop training program for primary reviewers	Minimum of 2 y of critical care training
	Online module with supervision of the first few chart reviews

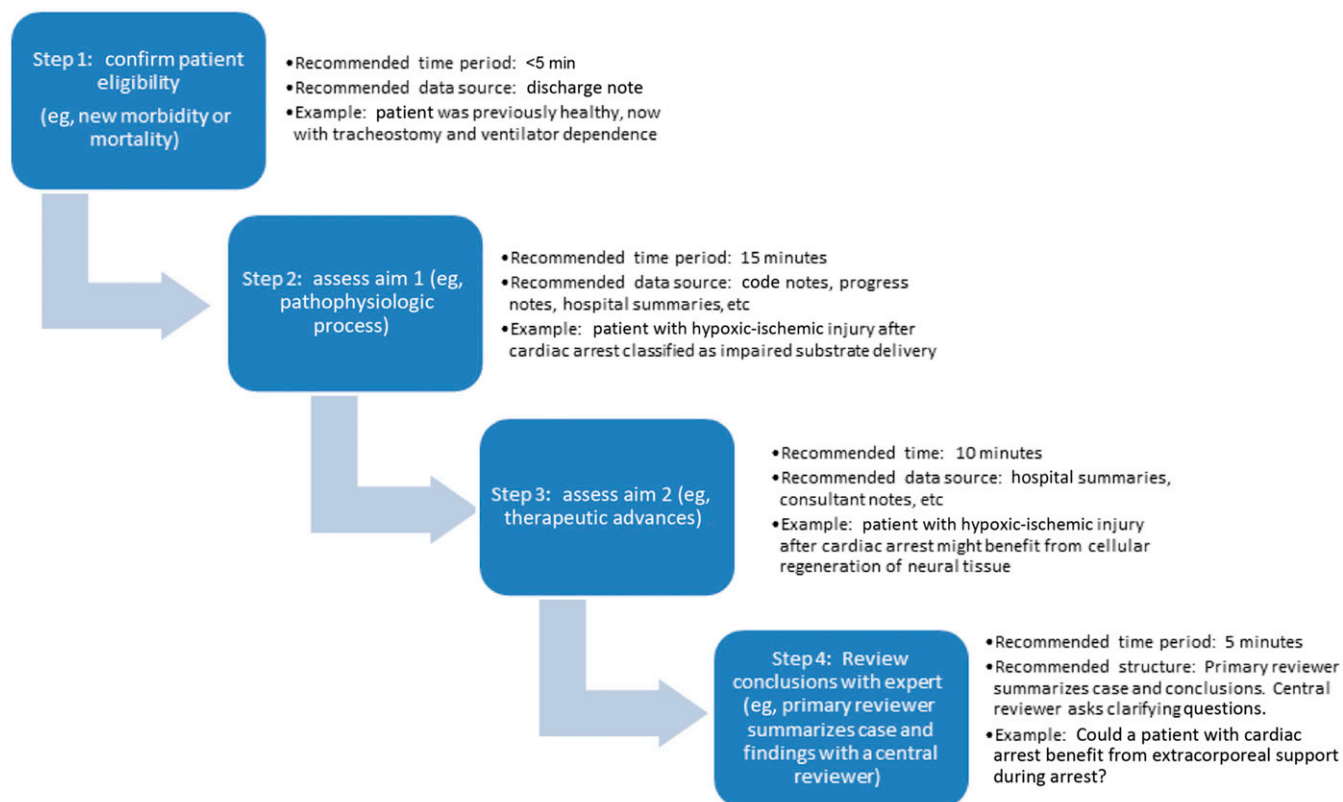
EHR, electronic health record.

follow, recommended time frames for each step, and details on where to find information relevant to this project with examples of how it was applied to this particular project. This process was adapted from the methods developed for the assessment of safety and quality of health care and was most recently used by the Institute for Healthcare

Improvement.<sup>28–30</sup> In developing this process, we made broad recommendations for where to find the information (eg, a rehabilitation medicine consultation note for the functional status of a patient), acknowledging that the reviewer would have better insight into their own electronic health records and health system. When generalizing this process, the phases of the

chart review in Fig 1 can be defined as a trigger (in our study, confirming a new morbidity or mortality), a cause for that trigger (the pathophysiology of the trigger), and a potential intervention (therapeutic advance) followed by a confirmatory process (central review).

The primary reviewers were familiar with the features of their local medical record



**FIGURE 1** Process for the structured chart review: step by step process for the chart review with the recommended time period, source data, and general and project-specific information.

systems and determined the most efficient review sequence for their site. The initial chart review step was to confirm patient eligibility by confirming the dates of illness and the FSS data, which established a new morbidity. The time allotted for this was ~5 minutes. Primary reviewers evaluated 327 patients and could not confirm new morbidity in 35 patients, resulting in a final sample size of 292 chart reviews. Because the FSS from the original TOPICC database was obtained from communications with caregivers, it was expected that some new morbidities would not be confirmed via chart review. Step 2 was to determine the pathophysiological processes responsible for the observed morbidity or mortality. More than 1 pathophysiological process may have been selected. An example would be a patient who experienced a cardiac arrest and has new morbidity from hypoxic-ischemic injury to the brain. This pathophysiology might be impaired substrate delivery. If the cardiac arrest was due to traumatic injury, then direct tissue injury may also be a selected pathophysiology. The time allotted for this step was ~15 minutes. Step 3 involved determining the therapies or interventions that, if these had existed or if existing therapies had been more effective, might have mitigated the new morbidity or mortality. For example, hypoxic-ischemic brain injury might be treated with therapies to promote neuronal cellular regeneration. The suggested time allotment for this step was 10 minutes. Step 4 was a review conducted by a central reviewer (co-principal investigator M.M.P. or K.L.M.) with the primary reviewer. The intent of this step was to assess the primary reviewer's conclusions and to ensure consistency of classification among the sites. The central reviewer was considered an expert in both the chart review methodology and the classification system. For example, the primary reviewer might discuss a patient with severe hypoxic-ischemic injury and recommend cellular regeneration as a potential therapeutic advance, and the central reviewer might suggest that extracorporeal life support advances could also benefit this patient. The expected duration of step 4 was ~5 minutes. The

expected duration of the total process was ~30 minutes per review.

The minimum qualifications for the primary reviewers were expertise and experience in critical care pathophysiology and therapies and were operationalized as completion of at least 2 years of a critical care fellowship, current status as a physician at the institution, and familiarity with the site's medical record. The primary reviewers underwent a standardized Web-based teaching session that included the central reviewers, the site's project manager, the site's primary reviewer, and the site's CPCCRN principal investigator. In this training session, the protocol was reviewed, including pertinent definitions, each pathophysiological process, each therapeutic advance, and potential data locations for the information. The training session also included a review of documents in which information could be

found and a review of the data collection tool and allowed time for the review steps. After this orientation, a central reviewer participated in at least 2 reviews at each site by teleconference until the primary reviewer demonstrated competence performing the methodology alone. Other issues were resolved with the central reviewer.

## Statistics

The structured chart review methodology is reported for all initial chart reviews ( $N = 327$ ) by summarizing responses found on both the primary and the central review evaluation forms. Continuous variables, such as the time needed to complete the primary reviewer's medical record review, are reported by using means with SDs and medians with interquartile ranges (IQRs), whereas categorical variables are summarized as counts and percentages.

**TABLE 2** Primary Reviewer Questionnaires ( $N = 327$ )

Questionnaire Items	Results
Time to complete the medical record review, min	
Mean (SD)	30.2 (16.7)
Median (Q1, Q3)	25.0 (20.0, 35.0)
Did you find the information you felt was needed to make the requested conclusions? $n$ (%)	
Strongly agree	221 (67.6)
Agree	89 (27.2)
Probably	15 (4.6)
Disagree	0 (0)
Strongly disagree	1 (0.3)
Unknown	1 (0.3)
It was easy to form a conclusion regarding this case, $n$ (%)	
Strongly agree	214 (65.4)
Agree	92 (28.1)
Probably	17 (5.2)
Disagree	2 (0.6)
Strongly disagree	1 (0.3)
Unknown	1 (0.3)
I am confident in the conclusion I came to, $n$ (%)	
Strongly agree	200 (61.2)
Agree	109 (33.3)
Probably	16 (4.9)
Disagree	0 (0)
Strongly disagree	1 (0.3)
Unknown	1 (0.3)

Q1, first quartile; Q3, third quartile.

Validity of the structured chart review methodology was assessed by questionnaires completed by the primary and central reviewers after the chart review. The primary reviewer questionnaire was used to assess the ease of finding information and forming conclusions and to assess confidence in the process (see Results, Table 2). The central reviewer questionnaire included similar items regarding the ability of the process to identify important issues, modifications made to the conclusions of the primary reviewer, and confidence in the process (see Results, Table 3). These subjective assessments were measured by using Likert response scales. The primary reviewer recorded the duration required for each review, and the central reviewer recorded the time needed for each final review. Confidence in the process was used to assess face validity, whereas the number of modifications reflected content validity.

A duplicative review of 41 charts was used to assess the reliability of the chart review process. Each central reviewer completed an independent review of a minimum of

20 charts at their home site using the same methodology as the primary reviewers but without the secondary review. The central reviewer had not reviewed these charts as part of the described process. The interrater reliability for the primary pathophysiological processes and therapeutic advances is reported as percent agreement and  $\kappa$  statistics with corresponding 95% confidence intervals (CIs). Categories infrequently selected will have wider CIs, and thus they will have less reliable point estimates. Because multiple categories could be selected for each case, a  $\kappa$  statistic and percent agreement were calculated for each category. There is a lack of general guidelines for acceptable  $\kappa$  statistics in the medical literature. As a common standard,  $\kappa$  statistics between 0.41 and 0.6 are assigned as moderate reliability,  $\kappa$  statistics between 0.61 and 0.80 as good reliability, and  $\kappa$  statistics between 0.81 and 1.00 as very good reliability.<sup>31</sup> Authors of previous chart review studies have used a minimum  $\kappa$  statistic ranging from 0.41 to 0.6 as acceptable agreement.<sup>15,32,33</sup>

## RESULTS

There were 8 primary reviewers from the 7 institutions with experience ranging from third-year fellow to professor. Most primary reviewers (62.5%) had previous chart review experience mostly with previous research or associated morbidity and mortality conferences. The median time spent working in the PICU at the time of the chart review was 92 months, with a median of 18 months in a dedicated cardiac ICU. A majority (75%) worked in a medical PICU, whereas the remainder worked both in a medical PICU and a cardiac ICU.

The primary reviewers conducted 327 reviews (Table 2) and confirmed new morbidity or mortality in 292 cases. The mean time required for the chart reviews was 30.2 (SD 16.7) minutes per case. The primary reviewers agreed or strongly agreed in 94.9% (277 of the 292) of reviews that the information to form a conclusion about pathophysiological processes and therapeutic advances could be adequately found during the chart review process. The reviewers agreed or strongly agreed in 93.2% (272 of 292) of the reviews that conclusions were easy to make. Reviewers also had confidence in the review process for 94.2% (275 of 292) of reviews, indicating agreement or strong agreement that they were confident in their conclusions. The secondary reviews of the central reviewers with the primary reviewers (Table 3) took 4.6 (SD 1.9) minutes per case. The central reviewer concluded that the primary reviewer had identified all or most major issues in all but one review and were confident or very confident in the classification in 93.8% (274 of 292) of reviews. Central reviewers made modifications to 36.6% of the classifications (107 of 292); 71.0% (76 of 107) involved the therapeutic advances, and 47.7% (51 of 107) involved the pathophysiological processes.

Forty-one reviews underwent 2 independent reviews (Table 4). Percent agreement for the pathophysiological processes for these reviews ranged from 80.5% to 100% and was >85% for 9 of 11 pathophysiological process categories. The  $\kappa$  statistics for the pathophysiological processes ranged from

**TABLE 3** Review Process Evaluation by the Central Reviewer

	Overall (N = 292)
Minutes to complete record review with the primary reviewer	
Mean (SD)	4.6 (1.90)
Median (Q1, Q3)	4.0 (3.0, 5.5)
Did the primary reviewer identify the important issues? n (%)	
The reviewer seemed to identify the major issues required for the classifications.	254 (87.0)
The reviewer seemed to identify most of the major issues required for the classifications.	37 (12.7)
The major issues were insufficiently identified for the classifications. This seemed to be because the issues were complex, and more time or expertise was needed.	1 (0.3)
The major issues were insufficiently identified for the classifications. This seemed to be because the reviewer did not adequately assess the record.	0 (0)
Confident in the classification n (%)	
Very confident	170 (58.2)
Confident	104 (35.6)
Neutral	17 (5.8)
Not very confident	0 (0)
Not confident	1 (0.3)
Modifications to the reviewer conclusions, n (%)	107 (36.6)
Modification changed the pathophysiological classification <sup>a</sup>	51 (47.7)
Modification changed the therapeutic options classification <sup>a</sup>	76 (71.0)

Q1, first quartile; Q3, third quartile.

<sup>a</sup> Percentages are out of reviews with a modification made by a central reviewer.



**TABLE 4** Reviewer Agreement for Primary Pathophysiological Processes and Therapeutic Advances (*N* = 41)

	Standard Review/Independent Review				Agreement, <i>n/N</i> (%)	$\kappa$ (95% CI)
	Yes/Yes	No/No	Yes/No	No/Yes		
Pathophysiological processes						
Impaired substrate delivery	14	19	4	4	33/41 (80.5)	0.60 (0.36 to 0.85)
Coagulation dysfunction	4	33	2	2	37/41 (90.2)	0.61 (0.26 to 0.96)
Inflammation	11	25	5	0	36/41 (87.8)	0.73 (0.51 to 0.94)
Immune dysfunction	4	33	3	1	37/41 (90.2)	0.61 (0.27 to 0.95)
Toxicities	5	33	1	2	38/41 (92.7)	0.73 (0.43 to 1.00)
Tissue injury (direct)	4	33	3	1	37/41 (90.2)	0.61 (0.27 to 0.95)
Malnutrition	3	34	1	3	37/41 (90.2)	0.55 (0.16 to 0.94)
Electrical signaling dysfunction	9	28	3	1	37/41 (90.2)	0.75 (0.52 to 0.98)
Abnormal growth or abnormal cell cycle	2	31	5	3	33/41 (80.5)	0.22 (−0.16 to 0.60)
Capillary or vascular dysfunction	1	36	1	3	37/41 (90.2)	0.29 (−0.21 to 0.79)
Mitochondrial dysfunction	1	40	0	0	41/41 (100.0)	1.00 (1.00 to 1.00)
Therapeutic needs						
Mechanical respiratory support	4	29	7	1	33/41 (80.5)	0.40 (0.08 to 0.72)
Inhaled respiratory treatments	0	39	1	1	39/41 (95.1)	−0.03 (−0.06 to 0.01)
Renal replacement therapy and plasmapheresis	0	40	0	1	40/41 (97.6)	—
Extracorporeal support and artificial organs	6	32	2	1	38/41 (92.7)	0.76 (0.49 to 1.00)
Organ transplant	3	27	3	8	30/41 (73.2)	0.20 (−0.12 to 0.53)
Blood and blood products	0	38	1	2	38/41 (92.7)	−0.03 (−0.08 to 0.01)
Drugs	20	14	5	2	34/41 (82.9)	0.65 (0.42 to 0.88)
Drug delivery	0	41	0	0	41/41 (100.0)	—
Immune and inflammatory modulation	6	30	4	1	36/41 (87.8)	0.63 (0.34 to 0.92)
Nutritional support	0	35	4	2	35/41 (85.4)	−0.07 (−0.14 to −0.00)
Therapeutic devices	1	34	5	1	35/41 (85.4)	0.19 (−0.20 to 0.59)
Monitoring devices	0	37	1	3	37/41 (90.2)	−0.04 (−0.10 to 0.02)
Cellular regeneration	5	23	9	4	28/41 (68.3)	0.23 (−0.08 to 0.53)
Suspended animation	0	41	0	0	41/41 (100.0)	—
Mitochondrial support	1	40	0	0	41/41 (100.0)	1.00 (1.00 to 1.00)

A total of 41 cases underwent duplicate reviews at 2 sites. Agreement is indicated by “yes/yes” or “no/no” by the 2 reviewers. Disagreement is indicated by “yes/no” or “no/yes.” —, not applicable.

0.22 to 1.00. Nine of the 11 categories had  $\kappa$  statistics  $\geq 0.4$ , and 8 had  $\kappa$  statistics  $\geq 0.6$ . The highest  $\kappa$  statistics occurred in the assessments that were the most objective: electrical signaling dysfunction (ie, cardiac arrhythmias and seizures), toxicities, and inflammation. The reliability measures for therapeutic needs were not as good as those for the pathophysiological processes. The percent agreement for therapeutic needs ranged from 68.3% to 100%, with 11 of 15  $> 85\%$ . The  $\kappa$  statistics for therapeutic advances ranged from −0.07 to 1.00, with only 5 of 15  $\geq 0.4$ . Three of the 15 therapeutic categories were never selected and could not be evaluated by using the  $\kappa$  statistic. The 95% CIs for the

$\kappa$  statistics were often wide, especially for the items infrequently selected.

## DISCUSSION

The structured chart review process we developed was intended to be generalizable, and our results support the potential use of this process to fulfill purposes beyond the scope of this specific project. Studies of medical malpractice litigation and liability insurance issues highlighted the use of structured chart reviews. To investigate malpractice litigation, the relevant data, especially related to hospital injuries, required that the Harvard Medical Malpractice Study develop a standardized approach to chart review.<sup>34,35</sup> This endeavor

resulted in insights about adverse event causation and lead to the initiation of root cause analysis.<sup>3,28,36</sup> The trigger tool methodology was developed where medical records were used to identify potential or actual adverse events during patient care, a process that is now largely automated.<sup>37–40</sup>

Our approach was a common-sense framework that was consistent with the original Harvard Medical Malpractice Study principles.<sup>34</sup> The critical steps that we took to develop the chart review process include clearly outlining goals and objectives, creating a classification scheme and tool to achieve those goals and objectives, reviewing times and source data that make

sense to the project, setting criteria for who can conduct the chart reviews, and including a final step of peer review with a person who has core knowledge on the goals and objectives of the project. These steps all help standardize the chart review process and contributed to the agreement we saw in our data analysis. Training of reviewers also clearly plays a critical role in making sure the chart review is standardized across reviewers and ensuring that the reviewer understands the process. Structured chart review methodology has been described previously, most notably by the Institute of Healthcare Improvement's Global Trigger Tool to identify adverse events in patients who are hospitalized, but the validity and reliability have been variable.<sup>23,41</sup> In our study, the survey results of the primary and central reviewers support the content and face validity of the process. Reviewers were generally confident in their ability to find information needed for reaching conclusions, they felt that conclusions were easy to form, and they were generally confident in their conclusions. The primary reviewers were all highly trained and had a range of critical care experience from third-year fellow to senior faculty. A second-level review by a central reviewer provided classification consistency. The central reviewers were confident that the primary reviewers identified all or most important issues in all but one review. The process was concluded within the expected time period of ~30 minutes for the primary review and 5 minutes for the secondary review.

The reliability of previous medical chart reviews for assessing adverse events ranged from poor to moderate.<sup>42,43</sup> In this study, the percent agreement was high for both the pathophysiological and therapeutic categories. The  $\kappa$  statistic revealed generally strong reliability for the pathophysiological processes but less strong performance for the therapeutic categories. The reviewers were more familiar with identifying pathophysiological processes because it is part of their training and care experiences, so it is expected that there would be greater reliability in the pathophysiological

categories. In contrast, advances in therapies included hypothetical therapies and was not a core expertise of the reviewers.

There are several important limitations to this study. First, we developed this chart review methodology for a purpose that required advanced experience and content expertise for both site and central reviewers. These highly trained and experienced reviewers might have contributed to the strong reliability and validity data. We did not assess applicability to projects with different reviewer requirements. Second, the interrater and agreement studies were done in only 2 sites because of issues involving protected health information. The positive reliability and validity results would be more robust with a larger number of institutions.

We developed an efficient, structured chart review methodology that, when applied to individual patients in critical care, had excellent validity and percent agreement, with sufficient interrater reliability. Despite the relatively complex assessments, the case reviews averaged 30 minutes per case. Structured chart reviews with clear objectives by knowledgeable primary reviewers followed by a brief secondary review by an appropriate expert can be valid and reliable and applied to research, quality assessments, and process improvement activities.

## REFERENCES

- de Vries EN, Eikens-Jansen MP, Hamersma AM, Smorenburg SM, Gouma DJ, Boermeester MA. Prevention of surgical malpractice claims by use of a surgical safety checklist. *Ann Surg.* 2011; 253(3):624–628
- Klasco RS, Wolfe RE, Lee T, et al. Can medical record reviewers reliably identify errors and adverse events in the ED? *Am J Emerg Med.* 2016;34(6): 1043–1048
- Brennan TA, Localio AR, Leape LL, et al. Identification of adverse events occurring during hospitalization. A cross-sectional study of litigation, quality assurance, and medical records at two teaching hospitals. *Ann Intern Med.* 1990;112(3):221–226
- Allison JJ, Wall TC, Spettell CM, et al. The art and science of chart review. *Jt Comm J Qual Improv.* 2000;26(3):115–136
- Gilbert EH, Lowenstein SR, Koziol-McLain J, Barta DC, Steiner J. Chart reviews in emergency medicine research: where are the methods? *Ann Emerg Med.* 1996; 27(3):305–308
- Worster A, Bledsoe RD, Cleve P, Fernandes CM, Upadhye S, Eva K. Reassessing the methods of medical record review studies in emergency medicine research. *Ann Emerg Med.* 2005;45(4):448–451
- Cifra CL, Bembea MM, Fackler JC, Miller MR. The morbidity and mortality conference in PICUs in the United States: a national survey. *Crit Care Med.* 2014; 42(10):2252–2257
- Resar RK, Rozich JD, Classen D. Methodology and rationale for the measurement of harm with trigger tools. *Qual Saf Heal Care.* 2003;12(suppl 2):ii39–ii45
- Horwitz RI, Yu EC. Assessing the reliability of epidemiologic data obtained from medical records. *J Chronic Dis.* 1984;37(11):825–831
- Holderried F, Heine D, Wagner R, et al. Problem-based training improves recognition of patient hazards by advanced medical students during chart review: a randomized controlled crossover study. *PLoS One.* 2014;9(2): e89198
- Gearing RE, Mian IA, Barber J, Ickowicz A. A methodology for conducting retrospective chart review research in child and adolescent psychiatry. *J Can Acad Child Adolesc Psychiatry.* 2006; 15(3):126–134
- Aaronson LS, Burman ME. Use of health records in research: reliability and validity issues. *Res Nurs Health.* 1994; 17(1):67–73
- Savović J, Jones HE, Altman DG, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled

- trials. *Ann Intern Med.* 2012;157(6): 429–438
14. Moustgaard H, Bello S, Miller FG, Hróbjartsson A. Subjective and objective outcomes in randomized clinical trials: definitions differed in methods publications and were often absent from trial reports. *J Clin Epidemiol.* 2014;67(12):1327–1334
  15. Saczynski JS, Kosar CM, Xu G, et al. A tale of two methods: chart and interview methods for identifying delirium. *J Am Geriatr Soc.* 2014;62(3):518–524
  16. Lehmann LS, Puopolo AL, Shaykevich S, Brennan TA. Iatrogenic events resulting in intensive care admission: frequency, cause, and disclosure to patients and institutions. *Am J Med.* 2005;118(4): 409–413
  17. Larsen GY, Donaldson AE, Parker HB, Grant MJ. Preventable harm occurring to critically ill children. *Pediatr Crit Care Med.* 2007;8(4):331–336
  18. Takata GS, Mason W, Taketomo C, Logsdon T, Sharek PJ. Development, testing, and findings of a pediatric-focused trigger tool to identify medication-related harm in US children's hospitals. *Pediatrics.* 2008;121(4). Available at: [www.pediatrics.org/cgi/content/full/121/4/e927](http://www.pediatrics.org/cgi/content/full/121/4/e927)
  19. Agarwal S, Classen D, Larsen G, et al. Prevalence of adverse events in pediatric intensive care units in the United States. *Pediatr Crit Care Med.* 2010;11(5):568–578
  20. Pollack MM, Holubkov R, Funai T, et al; Eunice Kennedy Shriver National Institute of Child Health and Human Development Collaborative Pediatric Critical Care Research Network. Simultaneous prediction of new morbidity, mortality, and survival without new morbidity from pediatric intensive care: a new paradigm for outcomes assessment. *Crit Care Med.* 2015;43(8):1699–1709
  21. Pollack MM, Holubkov R, Glass P, et al; Eunice Kennedy Shriver National Institute of Child Health and Human Development Collaborative Pediatric Critical Care Research Network. Functional Status Scale: new pediatric outcome measure. *Pediatrics.* 2009; 124(1). Available at: [www.pediatrics.org/cgi/content/full/124/1/e18](http://www.pediatrics.org/cgi/content/full/124/1/e18)
  22. Pollack MM, Holubkov R, Funai T, et al. Relationship between the functional status scale and the pediatric overall performance category and pediatric cerebral performance category scales. *JAMA Pediatr.* 2014;168(7): 671–676
  23. Resar RK, Rozich JD, Simmonds T, Haraden GR. A trigger tool to identify adverse events in the intensive care unit. *Jt Comm J Qual Patient Saf.* 2006;32(10): 585–590
  24. Lau H, Litman KC. Saving lives by studying deaths: using standardized mortality reviews to improve inpatient safety. *Jt Comm J Qual Patient Saf.* 2011; 37(9):400–408
  25. Weiss W, Bolton P. *Training in Qualitative Research Methods for PVOs & NGOs (& Counterparts): A Trainers Guide to Strengthen Program Planning and Evaluation.* Baltimore, MD: Center for Refugee and Disaster Studies, Johns Hopkins University School of Public Health; 2000
  26. Regis T, Steiner MJ, Ford CA, Byerley JS. Professionalism expectations seen through the eyes of resident physicians and patient families. *Pediatrics.* 2011; 127(2):317–324
  27. Rollins N, Chanza H, Chimbwandira F, et al. Prioritizing the PMTCT implementation research agenda in 3 African countries: Integrating and Scaling up PMTCT through Implementation REsearch (INSPIRE). *J Acquir Immune Defic Syndr.* 2014; 67(suppl 2):S108–S113
  28. Brennan TA, Leape LL, Laird NM, et al. Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. *N Engl J Med.* 1991;324(6): 370–376
  29. Griffin FA, Resar RK. *IHI Global Trigger Tool for Measuring Adverse Events.* 2nd ed. Cambridge, MA: Institute for Healthcare Improvement; 2009
  30. Naessens JM, O'Byrne TJ, Johnson MG, Vansuch MB, McGlone CM, Huddleston JM. Measuring hospital adverse events: assessing inter-rater reliability and trigger performance of the Global Trigger Tool. *Int J Qual Health Care.* 2010; 22(4):266–274
  31. Flight L, Julious SA. The disagreeable behaviour of the kappa statistic. *Pharm Stat.* 2015;14(1):74–78
  32. Stavem K, Hoel H, Skjaker SA, Haagensen R. Charlson comorbidity index derived from chart review or administrative data: agreement and prediction of mortality in intensive care patients. *Clin Epidemiol.* 2017;9:311–320
  33. Hogan H, Healey F, Neale G, Thomson R, Black N, Vincent C. Learning from preventable deaths: exploring case record reviewers' narratives using change analysis. *J R Soc Med.* 2014; 107(9):365–375
  34. Hiatt HH, Barnes BA, Brennan TA, et al. A study of medical injury and medical malpractice. *N Engl J Med.* 1989;321(7): 480–484
  35. Haukland EC, Mevik K, von Plessen C, Nieder C, Vonon B. Contribution of adverse events to death of hospitalised patients. *BMJ Open Qual.* 2019;8(1): e000377
  36. Leape LL, Brennan TA, Laird N, et al. The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II. *N Engl J Med.* 1991; 324(6):377–384
  37. Landrigan CP, Parry GJ, Bones CB, Hackbarth AD, Goldmann DA, Sharek PJ. Temporal trends in rates of patient harm resulting from medical care [published correction appears in *N Engl J Med.* 2010;363(26):2573]. *N Engl J Med.* 2010; 363(22):2124–2134
  38. de Almeida SM, Romualdo A, de Abreu Ferraresi A, Zelezoglo GR, Marra AR, Edmond MB. Use of a trigger tool to detect adverse drug reactions in an emergency department. *BMC Pharmacol Toxicol.* 2017;18(1):71



39. Maaskant JM, Smeulers M, Bosman D, et al. The trigger tool as a method to measure harmful medication errors in children. *J Patient Saf.* 2018;14(2):95–100
40. Stockwell DC, Pollack MM, Turenne WM, Slonim AD. Leadership and management training of pediatric intensivists: how do we gain our skills? *Pediatr Crit Care Med.* 2005;6(6):665–670
41. Classen DC, Resar R, Griffin F, et al. 'Global trigger tool' shows that adverse events in hospitals may be ten times greater than previously measured. *Health Aff (Millwood).* 2011;30(4):581–589
42. Thomas EJ, Lipsitz SR, Studdert DM, Brennan TA. The reliability of medical record review for estimating adverse event rates [published correction appears in *Ann Intern Med* 2002;137(2):147]. *Ann Intern Med.* 2002;136(11):812–816
43. Brennan TA, Localio RJ, Laird NL. Reliability and validity of judgements concerning adverse events suffered by hospitalized patients. *Med Care.* 1989;27(12):1148–1158