

Feasibility and Reliability Testing of Manual Electronic Health Record Reviews as a Tool for Timely Identification of Diagnostic Error in Patients at Risk

Jalal Soleimani¹ Yuliya Pinevich¹ Amelia K. Barwise² Chanyan Huang^{1,3} Yue Dong¹
Vitaly Herasevich¹ Ognjen Gajic² Brian W. Pickering¹

¹Department of Anesthesiology and Perioperative Medicine, Mayo Clinic, Rochester, Minnesota, United States

²Division of Pulmonary and Critical Care Medicine, Mayo Clinic, Rochester, Minnesota, United States

³Department of Anesthesiology, the First Affiliated Hospital, Sun Yat-Sen University, Guangzhou, Guangdong, China

Address for correspondence Jalal Soleimani, MD, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, United States
(e-mail: Soleimani.jalal@mayo.edu).

Appl Clin Inform 2020;11:474–482.

Abstract

Background Although diagnostic error (DE) is a significant problem, it remains challenging for clinicians to identify it reliably and to recognize its contribution to the clinical trajectory of their patients. The purpose of this work was to evaluate the reliability of real-time electronic health record (EHR) reviews using a search strategy for the identification of DE as a contributor to the rapid response team (RRT) activation.

Objectives Early and accurate recognition of critical illness is of paramount importance. The objective of this study was to test the feasibility and reliability of prospective, real-time EHR reviews as a means of identification of DE.

Methods We conducted this prospective observational study in June 2019 and included consecutive adult patients experiencing their first RRT activation. An EHR search strategy and a standard operating procedure were refined based on the literature and expert clinician inputs. Two physician-investigators independently reviewed eligible patient EHRs for the evidence of DE within 24 hours after RRT activation. In cases of disagreement, a secondary review of the EHR using a taxonomy approach was applied. The reviewers categorized patient experience of DE as Yes/No/Uncertain.

Results We reviewed 112 patient records. DE was identified in 15% of cases by both reviewers. Kappa agreement with the initial review was 0.23 and with the secondary review 0.65. No evidence of DE was detected in 60% of patients. In 25% of cases, the reviewers could not determine whether DE was present or absent.

Conclusion EHR review is of limited value in the real-time identification of DE in hospitalized patients. Alternative approaches are needed for research and quality improvement efforts in this field.

Keywords

- ▶ testing
- ▶ evaluation
- ▶ electronic health record
- ▶ communication
- ▶ new diagnosis
- ▶ diagnostic error

received
March 24, 2020
accepted
May 24, 2020

© 2020 Georg Thieme Verlag KG
Stuttgart · New York

DOI <https://doi.org/10.1055/s-0040-1713750>.
ISSN 1869-0327.

Background and Significance

In the United States, one in 20 adult patients experiences a diagnostic error (DE) every year.¹ Autopsy studies suggest that DEs occur in 10 to 20% of deaths.^{2,3} It has been estimated from postmortem studies that of the 540,000 deaths annually in the intensive care unit (ICU), 34,000 may die due to a severe DE.⁴ Real-time and accurate communication is essential for the provision of safe and effective care.⁵ In 2015, the Institute of Medicine, now the National Academy of Medicine, defined DE as (1) the failure to establish an accurate explanation of the patient's health problem or (2) the failure to communicate that explanation to the patient and within the electronic health record (EHR).¹ In many cases, the causes of DE may be complex and multifactorial.^{6–8} The multiple factors that cause DEs range from cognitive failures on the part of individuals, including diagnosing clinicians, system failures for tracking and managing test results, and poor teamwork and communication.^{9–12}

Despite the importance of DE, the process of reporting errors and near misses remains underdeveloped and lacks standardized measurement tools.^{13,14} DEs more often result in death and are the leading cause of claims-associated death and disability; indeed, until a decade ago, much of what was known about DE was learned through the evaluation of data from malpractice claims.^{15,16} In recent years, investigators have employed different approaches, including autopsy studies, case reviews, surveys of patients and physicians, voluntary reporting systems, use of standardized patients, second reviews, diagnostic testing audits, and closed claims reviews.¹⁷ Each of these methods has limitations, and none is well suited to establishing the presence of DE in real time when used alone. In most situations, the described methods have been applied only after the patient hospitalization episode and are not suited to real-time reporting or identification of DE.

Identifying and measuring DE through a detailed review of the EHR is particularly challenging and, expert reviewers often disagree about whether an error has happened.¹⁸ The lack of reliable measurement methods and difficulties in detecting DE in hospitalized patients continue to present significant barriers to designing solutions to reduce DE.¹⁹

Rapid response teams (RRTs) are a well-established patient safety strategy and are frequently implemented to manage deteriorating patients early by providing appropriate care quickly.²⁰ Up to 31% of clinical deteriorations requiring RRT activations have been attributed to medical errors, 68% of which were related to DEs.²¹ Therefore, for the purpose of this study, RRT activations were used to identify a population of hospitalized patients with increased risk for DE. In addition, RRT is often a prelude to ICU admission or escalation of care. Acutely ill patients are patients who become suddenly unwell or are physiologically deteriorating, and patients with a critical illness are patients who need ICU-level care as they are physiologically unstable or deteriorating. These patient groups may overlap. Both of these types of patients may have an RRT activation for stabilization and consideration for transfer to the ICU. Given the importance of early recognition and treatment of critical illness to reduce morbidity and cost, robust mechanisms to identify DE are vital.^{22,23}

The specific aim of this study was to reliably determine which patients experienced a DE within 24 hours after an RRT event. Our study used the kappa value as a criterion for applying a secondary review to resolve disagreements between independent reviewers.

Objectives

The objective of this work was to test the feasibility and reliability of prospective, real-time EHR reviews as a means to identify the presence of DE. Furthermore, we wanted to describe the process of developing, applying, and refining standard operating procedures (SOPs) using an acute-care learning laboratory to improve the identification of DE and inform future approaches.

Methods

Setting and Study Design

This single-center prospective observational study was conducted at Mayo Clinic, Rochester, Minnesota, United States, from June 1 through June 30, 2019. The study protocol was approved as a minimal risk study by the Institutional Review Board. Mayo Clinic is a quaternary care academic center with over 2,000 inpatient beds. The number of hospital inpatient admissions is approximately 62,000 per year. There are 3,800 RRT activations annually.²⁴ An RRT event is defined as an activation of the RRT to review a patient. The RRT team generally consists of a critical care fellow, a respiratory therapist, and an ICU nurse. A staff intensivist supervises the team, which provides ICU-level care responding to hospitalized patients with signs of deterioration in non-ICU settings. The RRT call can be activated by any member of the primary care team or other team members involved with the patient's care. Daily RRT reports and the pager system were used to identify patients' RRT activations in real time. The criteria for RRT activation are described in **Appendix Table 1**. RRT activations were used as an enriched population for identifying DE, because in acute situations like this, more evaluations increase the possibility of new diagnosis and approaches that were missed before.

Study Participants

A physician-investigator screened all patients experiencing an RRT activation. Inclusion criteria were age ≥ 18 years old, documented research authorization, and first RRT activation of that hospitalization. We excluded those patients who were experiencing an RRT activation within 6 hours of hospital admission, an RRT activation that occurred during an admission that was within 30 days of the last hospital discharge, and repeated RRT activations.

Data Collection

We abstracted data, including demographic information such as age, gender, ethnicity, race, reason for hospital admission, and admission time, from the EHR. We also abstracted data about the RRT activations such as the reason for the RRT call, time of RRT call, assessment at RRT activation, treatment

Table 1 SOP of the initial review

Diagnostic criteria	Data to be collected
New diagnostic label within 24 hours after RRT	<ul style="list-style-type: none"> Is the reason for RRT adequately explained by the current diagnosis? Is there any new diagnosis in the past 24 hours? Does this new diagnosis adequately explain the reason for RRT?
Time: features >6 hours before initial presentation of new diagnosis	<ul style="list-style-type: none"> Review Symptoms Signs Consults (problem list) Diagnostics (laboratories, images) Interventions (surgery, procedures)

Abbreviations: RRT, rapid response team; SOP, standard operating procedure.

during RRT activation, location of RRT activation, the current location of the patient, and whether the patient was transferred to the ICU post-RRT activation.

Search Strategy

Developing Standard Operating Procedures

An SOP was developed to review the EHR to identify DE systematically (→ **Table 1**).

Applying Standard Operating Procedures

During the initial iterative search strategy approach, two observers independently reviewed the EHR of eligible patients for the presence of a new diagnosis within 24 hours of RRT activation. The reviewers assessed the EHR, including documentation about the RRT activation, admission notes, and other clinical notes recorded prior to the RRT activation and up to 24 hours after RRT activation.

When a new diagnosis was detected, the physician-investigators reviewed the EHR to determine whether any features, indicative of that diagnosis, were present for greater than 6 hours prior to the first documentation of that new diagnosis. The window of 6 hours was chosen as a reasonable one in which patients at risk of critical illness, presenting to a quaternary care academic center, should be expected to have an accurate differential diagnosis communicated in the EHR. Features included physiological features such as symptoms and signs, as well as consults, diagnostics test results (laboratories, images), and interventions (surgery, procedures). In making their judgments, the reviewers were asked to classify each patient as either “Yes,” “No,” or “Uncertain” for DE.

Refining Standard Operating Procedures

In cases of disagreement about DE from the initial review, a secondary EHR chart review strategy was applied. For the second strategy, we used a taxonomy-based approach,

Table 2 SOP of the secondary review

Components of collected data	Data to be collected
History	<ul style="list-style-type: none"> Failure/delay in eliciting critical piece of history data Inaccurate/misinterpretation of history Suboptimal weighing of the piece of history Failure/delay to follow up of critical piece of history
Physical exam	<ul style="list-style-type: none"> Failure/delay in eliciting a critical physical-exam finding Inaccurate/misinterpreted critical physical-exam finding Suboptimal weighing of critical exam finding Failure/delay to follow up on critical exam finding
Tests (laboratories/radiology)	<ul style="list-style-type: none"> Ordering: <ul style="list-style-type: none"> Failure/delay in ordering needed test(s) Failure/delay in performing ordered test(s) Suboptimal test sequencing Ordering of wrong test(s) Performance: <ul style="list-style-type: none"> Sample mix-up/mislabeled Technical errors/poor processing of specimen/test Erroneous laboratory/radiology reading of tests Failed/delayed transmission or result to clinician Clinician processing: <ul style="list-style-type: none"> Failed/delayed follow-up action on test result Erroneous clinician interpretation of test
Assessment	<ul style="list-style-type: none"> Hypothesis generation: <ul style="list-style-type: none"> Failure/delay in considering the correct diagnosis Suboptimal weighing/prioritization: <ul style="list-style-type: none"> Too much weight to low(er) probability/priority diagnosis Too little consideration of high(er) probability/priority dx Too much weight on competing diagnosis Recognizing urgency/complications: <ul style="list-style-type: none"> Failure to appreciate urgency/acuity of illness Failure/delay in recognizing complication(s)
Referral/consultation	<ul style="list-style-type: none"> Failure/delay in ordering needing referral Inappropriate/unneeded referral Suboptimal consultation diagnostic performance Failed/delayed communication/follow-up of consultation
Follow-up	<ul style="list-style-type: none"> Failure to refer to a setting for close monitoring Failure/delay in real-time follow-up/rechecking of patient

Abbreviations: dx, diagnosis; SOP, standard operating procedure.

initially described by “Schiff et al,”⁸ and then replicated by “Jayaprakash et al” in our institution.²⁵ Judgments about the presence of DE were made based on answers to specific questions in the six categories outlined in → **Table 2**. The same reviewers in the initial review answered questions independently after reviewing the EHR and patients were classified as having a DE if the reviewer responded “Yes” to any of these questions. The reviewers, again, categorized the cases as Yes/No/Uncertain as they had during initial screening.

Final Review

The final step involved combining the agreements of the initial review with the total results of the secondary review (→ **Fig. 1**).

Kappa Statistic

Interobserver variation can be measured with a kappa coefficient in any situation in which two or more independent observers are evaluating the same thing. A kappa of 1

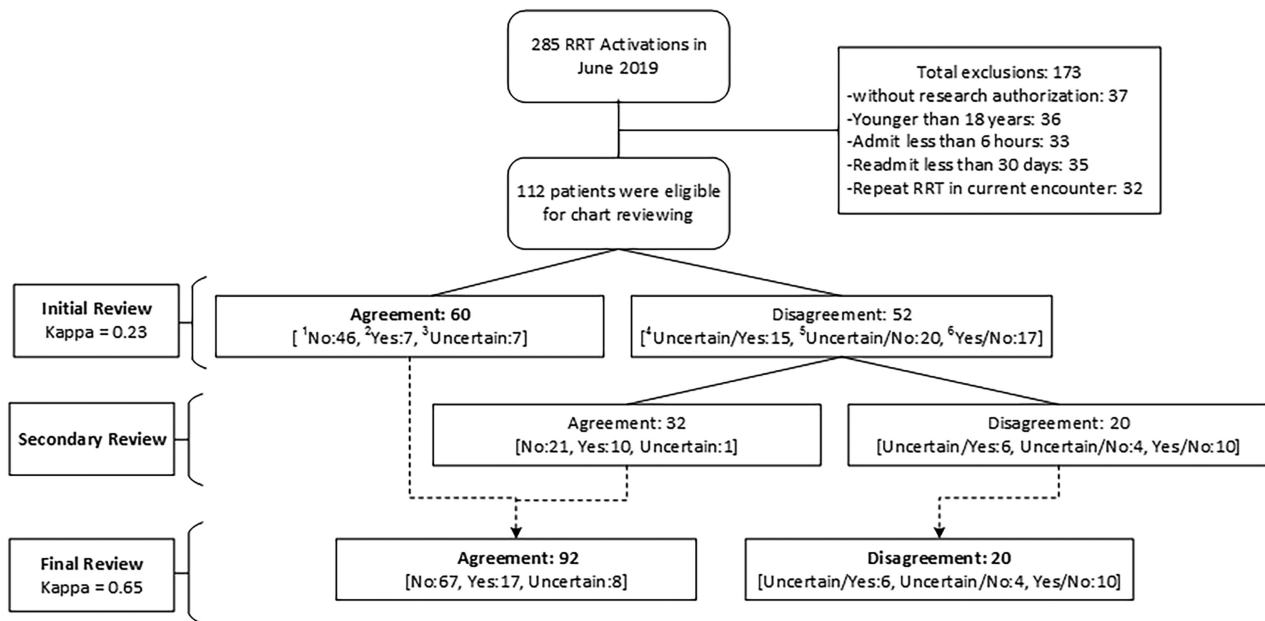


Fig. 1 A diagram of reviews. RRT, rapid response team. ¹No: two reviewers agree on “No” error; ²Yes: two reviewers agree on “Yes” error; ³Uncertain: two reviewers agree on “Uncertain” error; ⁴Uncertain/Yes: two reviewers disagree on “Yes or Uncertain” error; ⁵Uncertain/No: two reviewers disagree on “No or Uncertain” error; ⁶Yes/No: two reviewers disagree on “Yes or No” error.

indicates perfect agreement, whereas a kappa of 0 indicates agreement equivalent to chance.²⁶ In this study, our target range of kappa values from the initial review was moderate or 0.41 to 0.60 (→ **Appendix Table 2**).

We performed a Cohen’s kappa statistic score measurement with three distinct categories (Yes, No, and Uncertain) and six levels of agreement (Yes/Yes, No/No, Uncertain/Uncertain, Yes/No, Yes/Uncertain, and No/Uncertain). Kappa was used to assess the reliability of chart review and the level of agreement between both reviewers for DE.

General Statistic

We used JMP statistical software (version Pro 14) for statistical analysis. All continuous variables were reported as means with standard deviation (SD). All categorical variables were reported as counts with percentages.

Results

Cohort Identification

A total of 285 RRT activations occurred during June 2019, of which 173 cases were excluded. Reasons for exclusion were as follows: 36/173 (21%) ≤18 years; 37/173 (21.5%)—no research authorization; 33/173 (19%)—admission less than 6 hours before RRT activation; 35/173 (20%)—readmission less than 30 days after the most recent previous encounter; and 32/173 (18.5%)—repeated RRT in current admission. The total number of eligible patients was 112 (→ **Fig. 1**).

Demographics

The mean age of patients was 68 (SD: 16.3) years, and 46/112 (41%) were females. The number of patients admitted to the ICU after RRT activation was 55/112 (49%). The most common documented reason for RRT activation was respiratory

failure, 35/112 (31%), followed by altered mental status, 21/112 (19%), and hypotension, 19/112 (17%). Other documented reasons included cardiac arrhythmia, chest pain, stroke symptoms, and hyperthermia or fever (→ **Table 3**).

Primary Outcomes—Agreement

The data presented in → **Table 4** indicate results after use of the SOP of the initial review, the agreement between the two reviewers was 60/112 (54%) and the kappa value was 0.23 (fair) when both reviewers agreed on “Yes,” “No,” or “Uncertain.” After deploying the secondary review, the agreement increased to 92/112 (82%), and the kappa to 0.65 (substantial).

Secondary Outcomes—Proportions of DE

Using the initial review, DE was identified in 7/112 (6%) of eligible cases, no evidence of DE in 46/112 (41%) of eligible cases, and 7/112 (6%) of eligible cases were classified as uncertain by both reviewers. In 52/112 (46%) of eligible cases, there was no agreement between the reviewers. After combining the agreements in the initial review with the secondary review, the results changed to 17/112 (15%) agreement that DE had occurred, 67/112 (60%) no evidence of DE, 8/112 (7%) uncertainty, and 20/112 (18%) disagreement about whether DE had occurred (→ **Fig. 1** and → **Table 4**).

Discussion

Most evidence regarding DE in acutely ill patients comes from retrospective chart review or postmortem studies.^{27–29} In this article, we explore the reliability of chart review strategies to support real-time recognition of DE risk in acutely ill patients.

Over the years, DE has been recognized as a common cause of preventable harm to patients and negatively

Table 3 Characteristics of patients with RRT activation

Characteristics	Number (%) (N = 112)
Age in years, mean (SD)	67.9 (64.88–70.91)
Sex (female)	46 (41%)
Race	
White	94 (84%)
Unknown	8 (7%)
African American	5 (4%)
Asian	4 (3%)
American Indian	1 (<1%)
Transfer to ICU after RRT	55 (49%)
Patient status (alive)	103 (92%)
Reason for admission	
Cardiac diseases	17 (15%)
Infectious diseases	15 (13%)
Orthopedic diseases	15 (13%)
GI diseases	14 (15%)
Respiratory diseases	13 (12%)
Hemato/oncology diseases	13 (12%)
Neurologic diseases	11 (10%)
Other	8 (7%)
Metabolic/renal	6 (5.5%)
Reason for RRT activation	
Respiratory failure	35 (31%)
Altered mental status	21 (19%)
Hypotension	19 (17%)
Cardiac arrhythmia	13 (12%)
Other	10 (9%)
Chest pain	7 (6%)
Stroke symptoms	4 (3%)
Hyperthermia or fever	3 (2%)

Abbreviations: GI, gastrointestinal; ICU, intensive care unit; RRT, rapid response team; SD, standard deviation; SOP, standard operating procedure.

impacting the quality of health care.^{1,7,8} Given the positive impact of real-time identification of life-threatening diagnoses and treatment on patient outcomes, it was determined that effective DE recognition strategies for critical illness would need to be applicable in real time. For this reason, it was decided to develop a chart review strategy that would allow reviewers to reliably identify the presence of DE in a patient within 24 hours of an RRT call. We set a window of 24 hours as a reasonable one in which the most likely contributors to acute illness should be determinable in an academic center. Our initial focus was on testing the feasibility of a structured chart review that might form the basis of an automated search strategy that could be applied within 24 hours of a tracer event such as an RRT call.

Despite our best efforts, the agreement between reviewers in our study was elusive. The initial screening strategy outlined in the Methods section above resulted in a poor agreement between observers (kappa: 0.23). The agreement improved with secondary screening (kappa: 0.65) using a taxonomy approach, but the reviewers were unable to agree or were uncertain about the DE classification in nearly 25% of cases. There are several possible explanations for this.

In studies of DE, which utilize retrospective chart reviews, the discharge summary and subsequent coding reviews have been used as key data points. Our emphasis on real-time identification of DE meant that data documented after 24 hours post-RRT were not available to our reviewers to inform their judgment. The lack of agreement raises the question of the reliability of the medical record as a means of communicating diagnosis in a real-time fashion. There is evidence from other sources that the electronic medical record is not optimized for the communication of diagnostic reasoning or certainty.^{17,18,30} This limitation is particularly problematic in the care of critically ill patients where the urgency of treatment takes priority over documentation,³¹ and real-time diagnosis is impeded by information overload leading to errors of cognition.³²

The desire to satisfy regulatory and billing requirements cannot be ignored as a driver of commitment to specific International Classification of Diseases-9 or -10 diagnostic codes. This has implications for the validity of these data for the purposes of DE research or quality improvement. In addition, these requirements leave no room for uncertainty in diagnosis, which may be at odds with the messy reality of the acute care diagnostic process we encountered when reviewing the records of patients in real time after a critical RRT event.

One of the key objectives of our acute care learning laboratory is to develop automated search strategies that can reliably identify DE using data available in real time in the EHR.^{33,34} We were particularly interested, therefore, in the reliability of real-time chart review performed by trained observers. For this reason, we maintained independence between reviewers and refrained from using a third “super”-reviewer to arbitrate decisions as others have done. We tested the reproducibility of the search strategy as a proof of concept for an automated alert. The addition of a super-reviewer would have defeated that purpose as he/she would have eliminated any disagreement. This undoubtedly contributed to the resulting kappa score and is in contrast to retrospective chart review studies of DE, which rely on arbitration to achieve consensus agreement and where accuracy rather than inter-rater reliability is the goal.¹⁴

Our independent reviewers agreed that 15% of patients in the cohort were exposed to DE. This proportion is similar to that found in our previous study, which demonstrated a proportion of almost 18%. In the previous study, the entire EHR was utilized, and a consensus super-reviewer helped arrive at a final decision about DE.²⁵ In our study, the findings were based on the review of data available within 24 hours after the RRT call, and it is the first study we are aware of to describe the identification of DE this proximal to the RRT event. Our proportions of DE are in keeping with those of other DE

Table 4 Agreement and disagreement between two reviewers

		Initial review (N = 112)		Final review (N = 112)	
Agreement	Error	Count (%)	CI	Count (%)	CI
	No ^a	46 (41)	0.32–0.50	67 (60)	0.50–0.67
	Yes ^b	7 (6)	0.03–0.12	17 (15)	0.10–0.23
	Uncertain ^c	7 (6)	0.03–0.12	8 (7)	0.04–0.15
No agreement	Uncertain/Yes ^d	15 (13)	0.08–0.21	6 (5)	0.02–0.11
	Uncertain/No ^e	20 (18)	0.12–0.26	4 (4)	0.01–0.09
	Yes/No ^f	17 (15)	0.10–0.23	10 (9)	0.05–0.16
Kappa		0.23		0.65	

Abbreviation: CI, confidence interval.

^aNo: two reviewers agree on “No” error.

^bYes: two reviewers agree on “Yes” error.

^cUncertain: two reviewers agree on “Uncertain” error.

^dUncertain/Yes: two reviewers disagree on “Yes or Uncertain” error.

^eUncertain/No: two reviewers disagree on “No or Uncertain” error.

^fYes/No: two reviewers disagree on “Yes or No” error

methods and approaches. According to the report of the National Academy of Medicine, the general proportion of DE is 13.8%.¹ In a new study of unscheduled return visits to the emergency department with ICU admission, approximately 14% represented failures in the initial diagnostic pathway.³⁵ The proportion of DE can vary for different populations and settings, and depends on study methodology.^{4,27–29}

Each review was estimated to take approximately 1 hour per reviewer per patient. This has obvious implications for the scalability and use of this tool outside of our acute care learning laboratory environment. As discussed above, the intention was to use the refined search strategy as a starting point for the development of an automated EHR search instrument that could be exported to other settings to support scalability. The study findings have prompted us to reevaluate the feasibility of using the data available in the EHR as a trigger for real-time identification of vulnerability to DE.

Given the high proportions of DE in RRT patients, their vulnerability to deterioration, and the challenges associated with identifying DE in individual patients from the chart review alone, it may be time to consider an intervention that avoids harm from DE, and that may be applied to any patient who experiences an RRT call. At a minimum, RRT events should be considered as triggers for a diagnostic review by the responding team, which consists of a critical care fellow, a critical care respiratory therapist, and an ICU nurse.³⁶ This has implications for the composition of the team and speaks to the need to involve a diagnostician as part of that team. Indeed, our trained physician reviewers failed to establish an agreement (18%) or were uncertain about the presence of DE (7%) in this group of patients. This suggests that situations will arise in which an RRT cannot establish a diagnosis and should consider escalating the care of the patient to a panel of diagnostic experts.

Limitations

There are several limitations to this study. The design of this study is subject to weaknesses such as a narrow patient

population of RRT cases, a lack of previous prospective research studies about DEs, inability to generalize the results, being a single-center study, and potential false-positive results because of the inaccuracy of documentation. The single-center nature of the study has implications for generalizability. The population, processes, and EHR documentation standards found in the study setting may not be reflected in other settings. The decision to exclude readmitted patients or patients with recurrent RRT activations has implications for the reported proportions of DE. The inclusion of these groups may have had a significant impact on the proportions reported in this study and should be taken into consideration when interpreting the findings. In determining the proportion of DE, reviewers relied solely on EHR documentation that fails to account for other forms of communication, including verbal. As the record is not optimized to support real-time communication through accurate documentation, the emerging diagnosis, or differential diagnosis of critically ill patients, there is a possibility that reviewers underestimated the amount of communication that occurred.

Strengths

The work described in this article is part of a broader learning laboratory approach to better understand and prevent DE in acutely ill patients. This study is the first of its kind to use a prospective methodology to assess the reliability of EHR review to assess DE. By applying an SOP and doing blinded independent EHR reviews as well as deploying kappa statistics to test agreement, we can establish whether our strategy is genuinely reliable and has the potential to be useful in the future. The EHR review was done by physician-investigators with sound clinical knowledge and an understanding of the clinical practice.

Future Direction

We believe our acute care learning laboratory findings support inquiry and further study in at least two directions. First, the

question of how best to supplement or redesign the EHR must be tackled if we are to enable real-time detection of DE. This has implications for the development of advanced analytics approaches to a prediction that may be impaired, as our expert reviewers were, by an over-reliance on data derived solely from the EHR. This is particularly important in critically ill patients who will likely benefit most from early diagnosis and treatment.^{22,23} Real-time recognition of DE in this group could have a significant impact on outcomes and costs. Second, we believe the results of this study support a discussion about the composition and role of acute care outreach teams such as RRTs. Clearly, this population of patients is vulnerable to harm from unrecognized DE, and the benefits of the introduction of a diagnostic time-out should be debated.

Conclusion

Detecting the presence of DE within 24 hours of RRT activation, using available EHR data, is challenging. A moderate agreement can be reached, but consensus cannot be achieved in nearly 25% of cases. An EHR search strategy, such as the one described here, is time consuming and not easily scalable outside of a study environment. Alternative approaches to chart review for the real-time identification of DE are needed. A diagnostic time-out following RRT may be a useful intervention to prevent harm from unrecognized DE in this patient population.

Clinical Relevance Statement

It is difficult to reliably establish whether a DE has occurred using prospective medical record review strategies. The medical record, as a means of communicating diagnosis, requires improvements. Alternative approaches to medical record review for the real-time identification of diagnostic errors are needed.

Multiple Choice Questions

1. Kappa is a useful statistical measure to do which of the following:
 - a. Measure differences in continuous outcomes.
 - b. Measure differences in categorical outcomes.
 - c. Measure intraobserver variability two or more independent observers are evaluating the same thing.
 - d. Measure when consensus is reached.

Correct Answer: The correct answer is option c. Studies that measure the agreement between two or more observers should include a statistic that takes into account the fact that observers will sometimes agree or disagree simply by chance. The kappa statistic (or kappa coefficient) is the most commonly used statistic for this purpose.
2. Which of the below measurements has been used to identify diagnostic error?
 - a. Autopsy studies
 - b. Voluntary reporting systems

- c. Chart review
- d. All of the above

Correct Answer: The correct answer is option d. Investigators have employed different approaches, including autopsy studies, case reviews, surveys of patients and physicians, voluntary reporting systems, use of standardized patients, second reviews, diagnostic testing audits, and closed claims reviews.

3. Which method is ideally suited for identifying diagnostic error in real time?
 - a. Retrospective reviewing of the electronic medical record.
 - b. Using medical record as a means of communicating diagnosis.
 - c. The evaluation of data from malpractice claims.
 - d. All of the methods have limitations for identifying diagnostic error in real time.

Correct Answer: The correct answer is option d. Each of the methods has limitations, and none is well suited to establishing the presence of misdiagnosis in real time when used alone. In most situations, these methods are applied only after the patient hospitalization episode and are not suited to real-time reporting or identification of misdiagnosis.
4. What was the proportion of diagnostic error in critically ill patients identified by both reviewers?
 - a. Diagnostic error proportion in critically ill patients was 35%.
 - b. Diagnostic error proportion in critically ill patients was 15%.
 - c. Diagnostic error proportion in critically ill patients was 20%.
 - d. Diagnostic error proportion in critically ill patients was 10%.

Correct Answer: The correct answer is option b. After combining the agreements in the initial review with the secondary review, the final result was 17/112 (15%) agreement that DE had occurred, 67/112 (60%) no evidence of DE, 8/112 (7%) uncertainty, and 20/112 (18%) disagreement about whether DE had occurred.

Protection of Human and Animal Subjects

The study was performed in compliance with the World Medical Association Declaration of Helsinki on Ethical Principles for Medical Research Involving Human Subjects and was reviewed by Mayo Clinic Institutional Review Board.

Note

This work was performed at Mayo Clinic, Rochester, Minnesota, United States. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality or the Society of Critical Care Medicine.

Funding

This work was supported by the Agency for Healthcare Research and Quality (grant number R18HS026609,

received by Dr. Brian W. Pickering) and the Society of Critical Care Medicine 2019 SCCM Discovery Grant Award (received by Dr. Amelia K. Barwise).

Conflict of Interest

B.W.P. reports other from Ambient Clinical Analytics, outside the submitted work. In addition, B.W.P. has a patent Presentation of Critical Patient Data with royalties paid to Ambient Clinical Analytics. J.S. reports grants from the Agency for Healthcare Research and Quality and grants from the Society of Critical Care Medicine 2019 SCCM Discovery Grant Award, during the conduct of the study. Y.D. reports grants from AHRQ, during the conduct of the study.

Acknowledgments

We would like to thank Dr. Alice Gallo De Moraes, the RRT oversight committee members, RRT team members, our Mayo Clinic, and patients for their support in this study.

References

- National Academies of Sciences, Engineering, and Medicine. Improving Diagnosis in Health Care. Washington, DC: The National Academies Press; 2015
- Goldman L, Sayson R, Robbins S, Cohn LH, Bettmann M, Weisberg M. The value of the autopsy in three medical eras. *N Engl J Med* 1983;308(17):1000–1005
- Aalten CM, Samson MM, Jansen PA. Diagnostic errors; the need to have autopsies. *Neth J Med* 2006;64(06):186–190
- Winters B, Custer J, Galvagno SM Jr, et al. Diagnostic errors in the intensive care unit: a systematic review of autopsy studies. *BMJ Qual Saf* 2012;21(11):894–902
- Murphy DR, Singh H, Berlin L. Communication breakdowns and diagnostic errors: a radiology perspective. *Diagnosis (Berl)* 2014; 1(04):253–261
- Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. *Acad Med* 2003;78(08):775–780
- Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med* 2005;165(13):1493–1499
- Schiff GD, Kim S, Abrams Ret al. Diagnosing diagnosis errors: lessons from a multi-institutional collaborative project. In: Henriksen K, Battles JB, Marks ES, et al., eds. *Advances in Patient Safety: From Research to Implementation. Vol 2: Concepts and Methodology*. Rockville, MD: Agency For Healthcare Research and Quality; 2005:255–278
- Schattner A. Researching and preventing diagnostic errors: chasing patient safety from a different angle. *QJM* 2016;109(05):293–294
- Reilly JB, Myers JS, Salvador D, Trowbridge RL. Use of a novel, modified fishbone diagram to analyze diagnostic errors. *Diagnosis (Berl)* 2014;1(02):167–171
- Tobin MJ. Why physiology is critical to the practice of medicine: a 40-year personal perspective. *Clin Chest Med* 2019;40(02): 243–257
- Weingart SN, Yaghi O, Barnhart L, et al. Preventing diagnostic errors in ambulatory care: an electronic notification tool for incomplete radiology tests. *Appl Clin Inform* 2020;11(02):276–285
- Sabov M, Jayaprakash N, Chae J, Samavedam S, Gajic O, Pickering B. 1259: Reliability of identifying diagnostic error and delay in critically ill patients. *Crit Care Med* 2018;46(01):613
- Bergl PA, Nanchal RS, Singh H. Diagnostic error in the critically ill: defining the problem and exploring next steps to advance intensive care unit safety. *Ann Am Thorac Soc* 2018;15(08):903–907
- Saber Tehrani AS, Lee H, Mathews SC, et al. 25-Year summary of US malpractice claims for diagnostic errors 1986–2010: an analysis from the National Practitioner Data Bank. *BMJ Qual Saf* 2013; 22(08):672–680
- Kachalia A, Gandhi TK, Puopolo AL, et al. Missed and delayed diagnoses in the emergency department: a study of closed malpractice claims from 4 liability insurers. *Ann Emerg Med* 2007;49(02):196–205
- Graber ML. The incidence of diagnostic error in medicine. *BMJ Qual Saf* 2013;22(Suppl 2):ii21–ii27
- Zwaan L, Singh H. The challenges in defining and measuring diagnostic error. *Diagnosis (Berl)* 2015;2(02):97–103
- Graber ML, Kissam S, Payne VL, et al. Cognitive interventions to reduce diagnostic error: a narrative review. *BMJ Qual Saf* 2012;21 (07):535–557
- Chalwin R, Giles L, Salter A, Eaton V, Kapitola K, Karnon J. Reasons for repeat rapid response team calls, and associations with in-hospital mortality. *Jt Comm J Qual Patient Saf* 2019;45(04):268–275
- Braithwaite RS, DeVita MA, Mahidhara R, Simmons RL, Stuart S, Foraida M; Medical Emergency Response Improvement Team (MERIT). Use of medical emergency team (MET) responses to detect medical errors. *Qual Saf Health Care* 2004;13(04):255–259
- Kim HI, Park S. Sepsis: early recognition and optimized treatment. *Tuberc Respir Dis (Seoul)* 2019;82(01):6–14
- Diaz JV, Riviello ED, Papali A, Adhikari NKJ, Ferreira JC. Global critical care: moving forward in resource-limited settings. *Ann Glob Health* 2019;85(01):3
- Mayo Clinic Intranet. Quality data and analysis. Available at: <http://intranet.mayo.edu/charlie/quality/quality-data-and-analysis/>. Accessed August 22, 2019
- Jayaprakash N, Chae J, Sabov M, Samavedam S, Gajic O, Pickering BW. Improving diagnostic fidelity: an approach to standardizing the process in patients with emerging critical illness. *Mayo Clin Proc Innov Qual Outcomes* 2019;3(03):327–334
- Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med* 2005;37(05):360–363
- Dubosh NM, Edlow JA, Lefton M, Pope JV. Types of diagnostic errors in neurological emergencies in the emergency department. *Diagnosis (Berl)* 2015;2(01):21–28
- Chellis M, Olson J, Augustine J, Hamilton G. Evaluation of missed diagnoses for patients admitted from the emergency department. *Acad Emerg Med* 2001;8(02):125–130
- O'Connor PM, Dowe KE, Bell PM, Irwin ST, Dearden CH. Unnecessary delays in accident and emergency departments: do medical and surgical senior house officers need to vet admissions? *J Accid Emerg Med* 1995;12(04):251–254
- Fowles JB, Fowler EJ, Craft C. Validation of claims diagnoses and self-reported conditions compared with medical records for selected chronic diseases. *J Ambul Care Manage* 1998;21(01):24–34
- Kaye W, Mancini ME, Truitt TL. When minutes count—the fallacy of accurate time documentation during in-hospital resuscitation. *Resuscitation* 2005;65(03):285–290
- Pickering BW, Dong Y, Ahmed A, et al. The implementation of clinician designed, human-centered electronic medical record viewer in the intensive care unit: a pilot step-wedge cluster randomized trial. *Int J Med Inform* 2015;84(05):299–307
- Hudspeth J, El-Kareh R, Schiff G. Use of an expedited review tool to screen for prior diagnostic error in emergency department patients. *Appl Clin Inform* 2015;6(04):619–628
- Smith M, Murphy D, Laxmisan A, et al. Developing software to “track and catch” missed follow-up of abnormal test results in a complex sociotechnical environment. *Appl Clin Inform* 2013;4 (03):359–375
- Aaronson E, Jansson P, Wittbold K, Flavin S, Borczuk P. Unscheduled return visits to the emergency department with ICU admission: A trigger tool for diagnostic error. *Am J Emerg Med* 2019; S0735-6757(19)30579-0. Doi: 10.1016/j.ajem.2019.158430
- Barwise A, Thongprayoon C, Gajic O, Jensen J, Herasevich V, Pickering BW. Delayed rapid response team activation is associated with increased hospital mortality, morbidity, and length of stay in a tertiary care institution. *Crit Care Med* 2016;44(01):54–63

Appendix Table 1 Rapid response team criteria at Mayo Clinic Rochester

• A staff member is worried about the patient
• Acute and persistent declining oxygen saturations <90%
• Acute and persistent change in HR: <40 or >130
• Acute and persistent change in systolic BP: <90
• Acute and persistent change in RR: <10 or >28
• Acute chest pain suggestive of ischemia
• Acute and persistent change in the conscious state (including agitated delirium)
• New onset of symptoms suggestive of stroke

Abbreviations: BP, blood pressure; HR, heart rate; RR, respiratory rate.

Appendix Table 2 Interpretation of kappa^a

Kappa value	Agreement
0	Poor
0.01–0.20	Slight
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Substantial
0.81–1.00	Almost perfect

^aViera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med* 2005;37:360–363