Review

Predicting Patient Deterioration: A Review of Tools in the Digital Hospital Setting

Kay D Mann[1], MMathStat, PhD; Norm M Good[1], MSc; Farhad Fatehi[2,3], MD, PhD, FIAHSI; Sankalp Khanna[1], MInfTech, PhD, FAIDH; Victoria Campbell[4,5,6], MBBS, FRACP, FCICM; Roger Conway[4], RN, PGDip; Clair Sullivan[2,7], MBBS, MD, FRACP, FAIDH; Andrew Staib[5,8,9], MBBS, PhD, FACEM, FAIDH; Christopher Joyce[8,9], MB, ChB, PhD, FANZCA, FCICM; David Cook[8,9,10], MBBS, PhD, FANZCA, FCICM

[1]The Australian e-Health Research Centre, Commonwealth Scientific and Industrial Research Organisation, Brisbane, Australia
[2]Centre for Health Services Research, Faculty of Medicine, The University of Queensland, Brisbane, Australia
[3]School of Psychological Sciences, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Australia
[4]Sunshine Coast University Hospital, Sunshine Coast Hospital and Health Service, Birtinya, Australia
[5]Clinical Excellence Queensland, Queensland Health, Queensland, Australia
[6]School of Medicine, Griffith University, Nathan Campas, Australia
[7]Metro North Hospital and Health Service, Brisbane, Australia
[8]Princess Alexandra Hospital, Metro South Hospital and Health Service, Brisbane, Australia
[9]Faculty of Medicine, The University of Queensland, Brisbane, Australia
[10]School of Computer Science, Faculty of Science, Queensland University of Technology, Brisbane, Australia

Corresponding Author:
Norm M Good, MSc
The Australian e-Health Research Centre
Commonwealth Scientific and Industrial Research Organisation
Level 7, Surgical Treatment and Rehabilitation Service
Herston
Brisbane, 4029
Australia
Phone: 61 732533640
Email: norm.good@csiro.au

Abstract

Background: Early warning tools identify patients at risk of deterioration in hospitals. Electronic medical records in hospitals offer real-time data and the opportunity to automate early warning tools and provide real-time, dynamic risk estimates.

Objective: This review describes published studies on the development, validation, and implementation of tools for predicting patient deterioration in general wards in hospitals.

Methods: An electronic database search of peer reviewed journal papers from 2008-2020 identified studies reporting the use of tools and algorithms for predicting patient deterioration, defined by unplanned transfer to the intensive care unit, cardiac arrest, or death. Studies conducted solely in intensive care units, emergency departments, or single diagnosis patient groups were excluded.

Results: A total of 46 publications were eligible for inclusion. These publications were heterogeneous in design, setting, and outcome measures. Most studies were retrospective studies using cohort data to develop, validate, or statistically evaluate prediction tools. The tools consisted of early warning, screening, or scoring systems based on physiologic data, as well as more complex algorithms developed to better represent real-time data, deal with complexities of longitudinal data, and warn of deterioration risk earlier. Only a few studies detailed the results of the implementation of deterioration warning tools.

Conclusions: Despite relative progress in the development of algorithms to predict patient deterioration, the literature has not shown that the deployment or implementation of such algorithms is reproducibly associated with improvements in patient outcomes. Further work is needed to realize the potential of automated predictions and update dynamic risk estimates as part of an operational early warning system for inpatient deterioration.

(J Med Internet Res 2021;23(9):e28209) doi: 10.2196/28209

KEYWORDS

patient deterioration; early warning scores; digital tools; vital signs; electronic medical record

Introduction

Background

In recent years, proactive clinical processes have been developed to target timely and appropriate care for deteriorating or high-risk patients. Emergency responses such as *Rapid Response Systems* have been implemented with the aim to intervene and avoid preventable death, cardiac arrest, or transfer to an intensive care unit (ICU) in adult [1-3] and pediatric [4,5] patients. These systems have evolved to consist of a recognition (afferent) limb, commonly known as early warning scores (EWSs), and a response (efferent) limb (escalation and intervention). The responders rely on an accurate recognition limb, which in turn relies on a combination of empirical rules, statistical models, and clinical judgment to recognize deterioration. Initial EWSs were limited to vital signs, as these were the only routinely collected physiological data available for analysis in real time. The EWSs are currently available as paper or digital tools and vary significantly in their modeling, design, and escalation guidance [3,6-10].

The growth of rich, detailed, and dynamic clinical digital documentation in electronic medical records (EMRs) raises the possibility of using a broader range of clinical data, including pathology and diagnoses. An EMR collects the detailed phenotype of the patient in real time. Data collected, such as previous history, comorbidities, and demographic descriptors, are static during an admission. Observations of dynamic processes such vital signs, clinical measurements, imaging, and laboratory results that document biological and pathological processes are also recorded and continuously updated. Further evolving diagnoses, events, and interventions (eg, operations and drugs that are administered) capture the changing status of a patient. Finally, a rich source of dynamic information lies in the metadata: the timing, frequency, and location of actions and observations that occur to the patient and patient movements in the system. These data enable refined predictive models and more effective, patient-specific treatments. To create clinical decision support, these data must be analyzed and risk interpreted, and then critically, the clinical decision support communicating this real-time risk needs to be engineered back into the routine clinical workflows of the clinicians caring for the patient.

There is a diversity of models for predicting patient deterioration. Some risk estimates are static systems that identify high-risk patients at the time of diagnosis and allow triage of patients to a higher intensity care destination. Other approaches use vital sign observations to maintain an up-to-date risk evaluation. Typically, these dynamic systems either identify an extreme singular derangement or use weighted sums of a few vital signs and their variation from normal values. In both situations, the likelihood of deterioration and poor outcome anticipated with worsening values is increased. In some cases, these models have been developed and validated only for specific and narrow patient groups [11-13], whereas other general models are applicable to wider adult or pediatric ward patients [3-5,10,14].

Objectives

The aim of this review is to identify studies conducted within a general hospital setting that have attempted to develop prediction algorithms for detecting a deteriorating ward patient in real time, based primarily on routinely collected EMR data. This review includes model statistical validation and, where available, the results of digital hospital implementation of new or existing prediction models or rule-based systems for predicting patient deterioration. A secondary aim is to review those successful examples for common data elements, approaches, and statistical or machine learning techniques that were associated with successful clinical use.

Methods

This review followed the recommendations of the Center for Reviews and Dissemination [15] and PRISMA (Preferred Reporting in Systematic Reviews and Meta-Analyses) [16].

Data Sources and Search Strategy

A systematic search of PubMed, Scopus, Web of Science, IEEE Xplore, and ACM Digital Library using a combination of controlled vocabulary (eg, MeSH [Medical Subject Headings] terms) and free text keywords was conducted. The search strategy was first developed for PubMed, guided by the recommendations of Hausner et al [17] and Fatehi et al [18], and transposed to other databases. Across databases, free text keywords remained the same, but controlled vocabulary was mapped where possible (eg, from MeSH to Index Terms). The search was limited by language (English), date of publication (January 1, 2008, to June 30, 2020) and type of publication (original papers).

Screening and Study Selection

Reports that have developed prediction algorithms for detection of clinical patient deterioration in real time, primarily based on routinely collected patient data, were identified. The included studies had to (1) use any kind of (electronic) patient record, (2) use an early warning tool for patient deterioration, (3) use of a system that was dynamic, or observations of a patient over time, and (4) document the model statistical accuracy and performance. The studies were peer reviewed and published in journals or conference proceedings. The focus of this review was on tools in a general hospital ward and in a real time setting. Excluded studies were those conducted solely in a critical care or an emergency department setting, those limited to a single diagnosis or organ dysfunction patient cohort, those that used a static time point or an observation to assess risk of deterioration, and those that were qualitative with no quantitative assessments.

The results of electronic database searches were exported into an EndNote [19] library, and duplicates were removed. Title and abstract screening was coordinated on Rayyan web application (Qatar Computing Research Institute) [20]. Two independent reviewers screened titles and abstracts for relevance, and unresolved differences were adjudicated by a third reviewer. Full texts of remaining papers were assessed by 2 independent reviewers against the inclusion and exclusion criteria. Unresolved differences were adjudicated by a third reviewer. Where more than one publication was found for the same project, the papers were grouped, and the publication with the most comprehensive findings was included.

Data Extraction and Synthesis

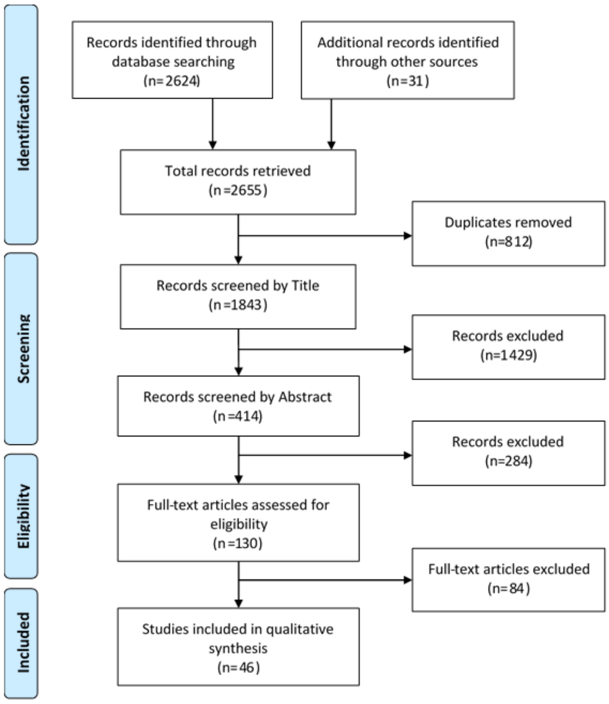
An electronic web form was developed according to the aims of this review and was used for data extraction from full-text papers. Data extraction was conducted by one reviewer and checked by a second reviewer. Owing to the heterogeneity of the included studies in terms of study designs, aims, and outcome measures, it was not possible to conduct a meta-analysis. Thus, a narrative approach was adopted for data synthesis. An assessment of study quality, including risks of bias, was conducted using PROBAST (Prediction Model Risk of Bias Assessment Tool; UMC Utrecht) [21] and ROBINS-I (Risk of Bias in Non-Randomised Intervention Studies; Cochrane) tool [22].

Results

Overview

The initial electronic search of five databases yielded 2624 records, and an additional 31 records were identified from other sources (Figure 1). After removing duplicates, 1843 unique records remained for eligibility assessment. Following the screening of titles and abstracts, 130 papers were deemed relevant, and their full texts were obtained. After a full-text inspection, 46 papers met the eligibility criteria and were included in this review.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of study selection.



Characteristics of the Studies

Of the 46 papers in this review, the majority (37/46, 80%) were retrospective studies that attempted to develop and/or evaluate the performance of a prediction model or compare the performance of a number of predictive models using historical patient data [23-59]. One study reported both the retrospective development of a prediction model and a prospective observational study of the developed model [60]. The remaining studies included one randomized controlled trial [61], 4 before-after (implementation) studies [62-65], and 3 prospective observational studies [66-68].

Overview of Studies

Studies from the same setting and institution were grouped and ordered according to the chronological model development, evaluation, and implementation, if available.

Of the identified studies, 6 described a body of research undertaken by researchers from Kaiser Permanente in California, as well as the University of Chicago [34,35,40,44,62,63] (Table 1). Kaiser Permanente is a large integrated managed care organization in the United States and began the deployment of EMRs in its hospitals in 2006. Escobar et al [35] reported the development of the early detection of impending physiological deterioration algorithm, which was further developed and subsequently named the Advanced Alert Monitor [44] and the early detection of impending physiological deterioration version 2 algorithm [40]. These models were developed to predict the risk of unplanned ICU transfer or mortality, using historical data of hundreds of thousands of patients retrieved from the Epic system, have also been retrospectively tested in a simulation study for feasibility [40], piloted in 2 hospitals [62], and more recently implemented as the Advanced Alert Monitor program and evaluated across the remaining Kaiser Permanente Hospitals [63,69,70].

Table 1. Studies undertaken on Kaiser Permanente Hospitals.

Author and year	Study design	Settings	Study aim	Model type or used	Outcome measure	Key findings
Escobar et al, 2012 [35]	RS[a]: tool development	102,422 hospitalizations; 14 KP[b] hospitals; 2006-2009	To develop a model for the prediction of unplanned ICU[c] transfer using EMR[d] data	Pooled logistic regression models	Transfer to ICU or death	EMR-based detection of impending deterioration outside the ICU is feasible. The overall performance of the model incorporating physiology, diagnosis, and longitudinal data was superior to MEWS[e]. Model *c* statistic 0.85, validation model 0.78
Escobar et al, 2013 [34]	RS: tool development	391,584 hospitalizations; 248,383 patients; 21 KP hospitals; 2008-2011	To develop a risk adjustment methodology applicable to all hospitalized patients	Pooled logistic regression models	Death	Risk adjustment of hospital mortality using EMR is feasible. Incorporation of physiological data increased model discrimination and explanatory power. Model *c* statistic 0.80, validation model 0.88
Kipnis et al, 2016 [44]	RS: tool development and evaluation	649,418 hospitalizations; 374,838 patients; 21 KP hospitals; 2010-2013	To describe the development and performance of an automated EWS[f] based on EMR data: The AAM[g]	Discrete time logistic regression	Transfer to ICU or death	The AAM had better performance compared with NEWS[h] and eCART[i] in all metrics and prediction intervals (AUC[j] 0.82). Around half the alerts triggered occurred within 12 hours of the event and almost two-thirds within 24 hours.
Escobar et al, 2016 [62]	Prospective study: tool implementation	2 KP hospitals	To detail technical and operational challenges of deploying an EWS	AAM	Pre- and postimplementation metrics	The pilot implementation in 2 hospitals was successful and further deployment to other hospitals will go ahead.
Hu et al, 2018 [40]	RS: tool simulation and feasibility	174,632 hospitalizations; 21 KP hospitals	To evaluate the impact of proactive transfer to ICU based on EDIP2[k] score, on mortality rate and LOS[l]	EDIP2	Transfer to ICU or death	Proactively transferring of the most severe patients could reduce mortality rates without sacrificing other patient outcomes.
Escobar et al, 2020 [63]	RS: tool implementation and evaluation	548,838 hospitalizations; 326,816 patients; 21 KP hospitals	To evaluate a staggered deployment of an automated predictive model; identifying patients at high risk for clinical deterioration.	AAM	Pre- and postimplementation outcomes: transfer to ICU, 30-day mortality, LOS, favorable status	30-day mortality after an alert was lower in the intervention compared with control (three deaths avoided per 1000 eligible patients). The intervention was also associated with lower incidence of ICU admission, higher percentage of patients with favorable status 30 days after alert, shorter LOS, and longer survival.

[a]RS: retrospective study.

[b]KP: Kaiser Permanente.

[c]ICU: intensive care unit.

[d]EMR: electronic medical record.

[e]MEWS: Modified Early Warning Score.

[f]EWS: early warning score.

[g]AAM: Advanced Alert Monitor.

[h]NEWS: National Early Warning Score.

[i]eCART: electronic Cardiac Arrest Risk Triage.

[j]AUC: area under the receiver operating characteristic curve.

[k]EDIP2: early detection of impending deterioration version 2.

[l]LOS: length of stay.

Another group of papers from the University of Chicago program was based on experiments using the electronic patient records from a number of hospitals in the Chicago area [29-33,37,66] (Table 2). Churpek et al [29] in several papers reported the development and validation of the electronic Cardiac Arrest Risk Triage score and other tools for predicting cardiac arrest and ICU transfer in Chicago (Table 2). They also showed that mortality and cardiac arrest were easier to predict than ICU transfer [29], and machine learning methods were more accurate than logistic regression for predicting patient deterioration [32].

Table 2. Studies conducted in Chicago hospitals.

Author and year	Study design	Setting	Study aim	Model type or used	Prediction event	Key findings
Churpek et al, 2012 [31]	RS[a]: tool development and evaluation	47,427 patients, 1 hospital, 2008-2011	To develop a CART[b] score and compare with the MEWS[c]	Logistic regression	CA[d] or transfer to ICU[e]	The CART score more accurately predicted cardiac arrest than the MEWS. Model AUC[f] 0.84
Churpek et al, 2013 [29]	RS: tool development	59,643 patients, 1 hospital, 2008-2011	To assess the impact of outcome selection on the performance of prediction algorithms	Logistic regression (4 models)	CA, transfer to ICU, death, all combined	Mortality is the easiest outcome to predict (AUC range 0.73-0.82), and ICU transfer is the most difficult.
Churpek et al, 2014 [30]	RS: tool development	59,301 patients, 1 hospital, 2008-2011	To derive and validate a prediction model for CA	Logistic regression	CA and transfer to ICU	The model can simultaneously predict the risk of CA and ICU transfer and was more accurate than ViEWS[g]. Model AUC 0.88 for CA, 0.77 for ICU
Churpek et al, 2014 [33]	RS: tool development	269,999 admissions, 5 hospitals, 2008-2011	To develop and validate eCART[h] score using commonly collected EMR[i] data	Survival analysis	CA, transfer to ICU, or death	eCART score was more accurate than MEWS for detecting CA, ICU transfer, or death. Model AUC 0.83 for CA, 0.75 for ICU transfer, 0.93 for death and 0.77 all combined
Somanchi et al, 2015 [56]	RS: tool development	133,000 patients, 4 hospitals, 2006-2011	To develop a prediction model for Code Blue, using EMR data, and compare with MEWS	SVM[j] and logistic regression	Code blue event in the next x hours	The model was able to predict Code Blue with ~80% recall and 20% false positive rate 4 hours ahead of the event. It out-performed MEWS.
Churpek et al, 2016 [32]	RS: tool development and evaluation	269,999 patients, 5 hospitals, 2008-2013	To compare the accuracy of different techniques for detecting clinical deterioration on the wards	Logistic, decision trees, SVM, K-NN[k], neural net, MEWS	CA, transfer to ICU, or death	This multicenter study showed that several machine learning methods can more accurately predict clinical deterioration than logistic regression.
Kang et al, 2016 [66]	Prospective study: feasibility study	3889 admissions, 3 wards, 2013-2014	To assess the feasibility of a real-time risk stratification tool	eCART	CA, transfer to ICU, or death	eCART score identified more CA and ICU transfers, many hours in advance, compared with standard RRT[l] activation.
Green et al, 2018 [37]	RS: tool evaluation	107,868 admissions, 5 hospitals, 2008-2013	To compare the BTF[m] calling criteria to MEWS, NEWS[n] and eCART score	BTF, NEWS, MEWS, eCART	CA, transfer to ICU, or death (24 hours)	eCART was more accurate than BTF, MEWS, NEWS for predicting the composite outcome of CA, ICU transfer and death. eCART AUC 0.80 (0.79-0.80)
Bartkowiak et al, 2019 [25]	RS: tool evaluation	32,537 admissions, 1 hospital, 2008-2016	To assess the accuracy of three EWS[o] postoperatively	NEWS, MEWS, eCART	CA, ICU transfer or ward, or death	The eCART score was the most accurate followed by MEWS. Maximum respiratory rate was the most predictive vital sign.
Mayampurath et al, 2019 [48]	RS: tool development	115,825 admissions, 1 hospital, 2008-2016	To develop a model from visual timelines to predict mortality	Convolutional neural network	Death	The model was more accurate than MEWS and SOFA[p], validation model AUC 0.91, and visual timelines enabled interpretation of a deep neural network.

[a]RS: retrospective study.

[b]CART: Cardiac Arrest Risk Triage.

[c]MEWS: Modified Early Warning Score.

[d]CA: cardiac arrest.

[e]ICU: intensive care unit.

[f]AUC: area under the receiver operating characteristic curve.

[g]ViEWS: VitalPAC Early Warning Score.

[h]eCART: electronic Cardiac Arrest Risk Triage.

[i]EMR: electronic medical record.
[j]SVM: support vector machine.
[k]K-NN: K-nearest neighbors.
[l]RRT: rapid response team.
[m]BTF: Between the Flags.
[n]NEWS: National Early Warning Score.
[o]EWS: early warning score.
[p]SOFA: Sequential Organ Failure Assessment.

Several other studies have evaluated screening tools such as the National Early Warning Score, Modified Early Warning Score, Rothman Index, and Sequential Organ Failure Assessment, for the early warning and detection of patient deterioration in hospital settings across different countries [26,39,41,50,52,54,57,59,64,65,67] (Table 3). The evaluation included applying tools retrospectively on historical clinical data to assess the feasibility of future use as well as assessing tools prospectively alongside the standard clinical systems. The studies of screening tools reported differing results of both good and poor predictive accuracy and usefulness for escalation of care through medical emergency team (MET) or rapid response team (Table 3).

Of the studies detailing the implementation of scoring tools across institutions [50-52,61,62,64,65], high-risk patients were appropriately identified as aiding in clinical response (Table 4). However, when comparing intervention and control patient cohorts, differing results were seen with either significant reductions or no impact on the assessed deterioration events.

Within the reviewed papers, deterioration prediction methodologies included single and multiparameter scoring tools, such as the National Early Warning Score and Modified Early Warning Score, as well as statistical and machine learning methods. Single and multiparameter scores were derived from a set of vital sign threshold derangements (Tables 3 and 4). The statistical and machine learning methods included logistic regression, survival models, Cox regression, Gaussian process regression, Markov models, decision trees (random forest and gradient boosted trees), K-nearest neighbor, support vector machine, and neural networks (Tables 1 and 2, and Multimedia Appendix 1 [23,24,27,28,36,38,42,43,45-47,49,51,53,55,58,60,68]). Most of these models attempted to account for changes in physiologic measures over time using novel model frameworks, for example, taking a sliding time window looking forward or backward in time to predict outcomes.

Across all the studies, models of greater complexity and statistical and machine learning methods were shown to have superior performance in predicting deterioration than scoring tools (Multimedia Appendix 1). For example, there was timelier detection and earlier warning of high deterioration risk in more complex models. Furthermore, the discrimination of patient deterioration using statistical and machine learning methods (as assessed with the area under the receiver operating characteristic curve or *c* statistic) outperformed conventional tools.

All patient deterioration prediction tools used vital sign measures, most commonly blood pressure, heart rate, respiratory rate, temperature, oxygen saturation, and a level of consciousness measure (Multimedia Appendix 2, Tables S2 and S3 [23-39,41-45,47-61,63-65,67,68]). In addition, most of the models included basic patient demographic data, such as age and gender, as well as administrative measures, such as admission status, time since admission, length of stay, and patient location. Many of the models attempted to incorporate various pathology or laboratory test results where available, noting that they would experience some level of time delay. Composite indices or scores for severity of illness, longitudinal comorbidities, and combined laboratory results were often included in models where available. The more complex models considered higher-level features of their data such as the frequency, change, minimum, maximum, moving average, and patterns of physiological measures over time [23,24,38,42,45,56,58].

Of the studies that reported missing data, the majority were filled by propagating a previous value carried forward if a current value within a set time window was not measured. If no prior value was available, values were imputed with a representative value such as a population-based estimate or normal value. In models that consider dynamic irregularly sampled physiologic time series data, Gaussian process models were used to deal with sparsity in the data [23,24,55].

Overall, the quality of the studies included in this review was high, with low and unclear risks of bias and concern for the applicability of prediction models in addressing our review questions (Multimedia Appendix 2, Table S2). The majority of prediction model studies appropriately selected model data for inclusion, assessed model predictors and outcomes, approached model development, and adequately tested models. Studies assessed as unclear were because of ambiguous details in reporting the number of participants or samples per derivation and validation data sets, handling of missing data, and lack of detail in model validation. The study quality of the implementation studies we identified was also high, with low to moderate risk of bias identified (Multimedia Appendix 2, Table S4). Moderate concerns were because of the lack of adjustment for potential confounding, participant selection, and lack of detail for handling of missing data.

Table 3. Retrospective studies evaluating scoring tools.

Author, year, and country	Settings	Study aim	Scoring tools	Prediction event	Key findings
Lighthall et al, 2009 [67], United States	1089 patients, 1 hospital, 2006	To evaluate vital signs and association with critical events	MET[a] call criteria	CA[b], ICU[c] transfer or death	Even a single recording of an abnormal vital sign increases the risk of critical events in hospitalized patients.
Huh et al 2014, [41], Korea	3030 events, 1 hospital, 2008-2010	To evaluate the efficacy of screening triggered alerts for MET management	Medical alert system criteria	MET activation	The automatic alert system triggers, along with a skilled intervention team, were successful in managing the MET
Romeo-Brufau et al, 2014 [54], United States	34,898 patients, 2 hospitals, 2011	Comparative analysis of the performance of common EWS[d] methods and how they would function if automated	MEWS[e], SEWS[f], GMEWS[g], Worthing, ViEWS[h], NEWS[i]	Resuscitation call, RRS[j] activation or ICU transfer	The evaluated scores did not offer good predictive capabilities for an automated alarm system. Positive predictive values ranged from <0.01-0.21, and sensitivity ranged from 0.07-0.75.
Yu et al, 2014 [59], United States	328 cases, 328 controls, 1 hospital, 2009-2010	To compare the ability of 9 risk prediction scores in detecting clinical deterioration among non-ICU ward patients	SOFA[k], PIRO[l], ViEWS, SCS[m], MEDS[n], MEWS, SAPS II[o], REMS[p], APACHE II[q]	Critical care consult, ICU transfer or death	Prediction scores can be used to estimate a ward patient's risk of clinical deterioration, with good discriminatory ability comparable with that of existing track-and-trigger systems. 0-12 hours before clinical deterioration, 7 of 9 scores performed with acceptable discrimination (AUC[r]>0.70).
Wengerter et al, 2018 [57], United States	217 cases, 868 controls, 1 hospital, 2013-2015	To evaluate whether Rothman Index variability can predict RRT[s] activation in surgical patients	Rothman Index	RRT activation, mortality	Rothman Index variability predicted likelihood of RRT activation.
Bedoya et al, 2019 [26], United States	85,322 patients, 2 hospitals, 2014-2016	To determine the effectiveness of NEWS implementation on predicting and preventing patient deterioration	NEWS	ICU transfer or death	No change after implementing NEWS. At both academic and community hospitals, NEWS had poor performance characteristics and was generally ignored by nursing staff.
Heller et al [39], 2020, Germany	3827 patients, 2 wards, 2016-2017	To develop a prediction model for Code Blue, using EMR[t] data, and compare with MEWS	MEWS with paging functionality	CA or ICU transfer	The rate of CA and ICU transfers significantly decreased after implementing MEWS with paging functionality.

[a]MET: medical emergency team.
[b]CA: cardiac arrest.
[c]ICU: intensive care unit.
[d]EWS: early warning score.
[e]MEWS: Modified Early Warning Score.
[f]SEWS: Standardized Early Warning Score.
[g]GMEWS: Global Modified Early Warning Score.
[h]ViEWS: VitalPAC Early Warning Score.
[i]NEWS: National Early Warning Score.
[j]RRS: rapid response system.
[k]SOFA: Sequential Organ Failure Assessment.
[l]PIRO: Predisposition, Infection, Response, Organ, Dysfunction Score.
[m]SCS: simple clinical score.
[n]MEDS: Mortality in Emergency Department Sepsis.
[o]SAPS II: Simple Acute Physiology Score II.
[p]REMS: Rapid Emergency Medicine Score.
[q]APACHE II: Acute Physiology and Chronic Health Evaluation Score II.
[r]AUC: area under the receiver operating characteristic curve.
[s]RRT: rapid response team.
[t]EMR: electronic medical record.

Table 4. Studies of scoring tool implementation.

Author, year, and country	Study design	Setting	Study aim	Scoring tool	Intervention assessment	Implementation outcome	Key findings or conclusions
Bailey et al, 2013 [61], United States	PS[a]: RCT[b]	19,116 patients, 1 hospital, 2007-2011	To validate the EWS[c] in general medical wards and trial real time alerting.	Two-tiered EWS [38]	Comparison of alerts between intervention and control patients.	Among patients identified by EWS, there were no differences in proportion transferred to ICU[d] or died in the intervention group compared with control.	Alerts generated for patients meeting the threshold were highly specific for ICU transfer and death. However, sending real time alerts to the nurse manager did not improve event outcomes.
Evans et al, 2015 [64], United States	PS: observational study	6289 patients, 1 hospital, 2012-2013	To develop and evaluate a detection and alert system for monitoring patients every 5 min	MEWS[e] with pager alerts	Comparison of events from patients in 2 wards pre and post intervention	Ward A patients had more ICU transfers, MET[f] calls and greater LOS[g] but fewer deaths during intervention compared with preintervention. No differences were seen in ward B.	Implementation of the predictive model increased appropriate MET calls. Mortality decreased in the ward with older patients and multiple comorbidities, but not in the other ward.
Subbe et al, 2017, [65], United Kingdom	PS: before and after study	4402 patients, 1 hospital, 2014-2015	To assess the effect of a vital sign monitoring and alert system on outcomes	Vital sign monitoring system	Comparison of serious events between control and intervention periods	Deaths, CAs[h] and, for patients transferred to ICU, severity of illness scores were lower in the intervention compared with control.	Deploying an EWS based on vital signs increased RRT[i] calls and decreased mortality and CAs.
Oh et al, 2018 [52], Korea	RS[j]: before and after study	207,054 surgeries, 1 hospital, 2008-2016	To evaluate whether a RRS[k] reduces incidence of postoperative CA	RRS with thresholds and calling criteria	Change in cardiopulmonary arrest rate in patients before and after intervention	Cardiopulmonary arrest relative risk (pre vs post intervention) was 0.56 during RRS operational hours but was unchanged during nonoperational hours. These associations remained after comorbidity adjustment.	Implementation of the RRS reduced postoperative cardiopulmonary arrest incidence but only during RRS operational hours.
Morgan et al, 2020 [50], United States	PS: quality improvement study	30,292 patients, 1 hospital, 2017-2018	To evaluate the implementation of a continuous cloud-based EWS to activate an RRT.	Cloud-based modified NEWS[l] with RRT call threshold	Comparison of intervention with control patients for time to first lactate order, ICU[d] transfer and mortality.	The intervention group had improved the time to the first lactate order within 24 hours of modified NEWS ≥7. There was no significant improvement in time to ICU transfer, ICU length of stay, or hospital mortality.	The study provides preliminary evidence for a pragmatic integration of cloud-based, automated monitoring with standardized and timely RRT intervention.

[a]PS: prospective study.

[b]RCT: randomized controlled trial.

[c]EWS: early warning score.

[d]ICU: intensive care unit.

[e]MEWS: Modified Early Warning Score.

[f]MET: medical emergency team.

[g]LOS: length of stay.

[h]CA: cardiac arrest.

[i]RRT: rapid response team.

[j]RS: retrospective study.
[k]RRS, rapid response system.
[l]NEWS: National Early Warning Score.

Discussion

Principal Findings

This review examines the use of data sets collected in EMRs to develop and implement decision support for clinicians to predict and prevent inpatient deterioration. The current literature confirms that it is possible for routinely collected EMR data to be used to anticipate patient deterioration. However, there are few reports on the performance and efficacy of these systems when used in clinical settings. Despite the wide and increasing adoption of EMRs, the successful implementation of EMR-based early warning systems or their impact on patient-centered outcomes is not commonly reported.

The studies that met the eligibility criteria and the variability in methods created a narrative review rather than a quantitative review. There was considerable variation in the models developed, methodological approach, and data collected. The common study designs (retrospective, small cohorts, and before and after studies) pose a risk of bias. The institutions where models were developed and implemented predominately from the United States, which may pose additional risks of bias because of the nature of the health care system and environment. The heterogeneity in the methodology of the developed models makes comparison of their statistical accuracy and performance difficult. The area under the receiver operating characteristic curve statistic has been used as a comparison metric, but it is important to note the differences in model specification, complexity, and outcome when interpreting model output (eg, scores) and their usefulness for implementation in a real-time clinical setting. In addition, the reporting of superior performance in statistics from models of greater complexity may be the result of model overfitting if this has not been appropriately addressed in model development. Consideration should be made such that data used to develop early warning tools reflect previous cohorts of patients and may not be relevant for future cohorts. It is reasonable that retrospective patterns of deterioration will repeat themselves in the future; however, with advances in health care technology and the potential for new and emerging health disease trends, adapting or updating tools will be required to maintain relevance. We conclude that the effectiveness of current EMR-based digital early warning tools remains promising but has not been reproducibly demonstrated.

We propose that there are multiple factors that make a patient-centered outcome evaluation of EMR-based deterioration prediction difficult. Demonstration of improved patient outcomes following early warning tool implementation relies on the successful performance of all four components of implementation, recognition, escalation, and response. Where there has been successful statistical validation for the EWSs, validation of the escalation and response (ie, implementation) has been elusive. Several factors may contribute, which are not unique to digital warning implementation. The current metrics for outcome measures are problematic. Death, an outcome that is easy to count, does not always account for treatment limitations. Cardiac arrest events are rare; therefore, they are insensitive measures. ICU transfer can reflect MET resourcing deficiencies, ICU bed availability, and local admission practices or end-of-life planning rather than preventable deterioration. Emergency callout rates may reflect individual clinician sentiment rather than true risk or deterioration, or alternatively, not be called when they should be. There is inconsistency in efferent limb performance (assessment, intervention, transfer, monitoring, and follow-up). The time required for implementation and translation into improved outcomes will frequently be confounded by other system improvements, changes within an individual health service, or the maturity of the rapid response system itself. Therefore, the relationship between implementation and patient outcomes remains unstudied, rather than unproven by the current studies.

These issues will continue to challenge the empirical validation of a complete rapid response system. The lifecycle of these technologies has yet to mature to the point where evaluation and assessment of a possible effective intervention can be performed. One example of practical limitations is to recognize which data are contemporaneously available. Data can only inform a ward-based, real-time prediction model once the information is available in the EMR for analysis. This is relevant to model development and evaluation. For a pathology result, for example, there are times of sample collection, arrival at the laboratory or analysis, and the time the result is available for integration into the clinical picture. The results of the test are only available for integration once they are available in the clinical space. However, the metadata confirming that the sample was actually ordered or collected, where it was done, and the number of samples, is immediately available. The result of a positive blood culture can take hours to days, and a definitive negative result can only be finalized 2-5 days after collection. Providing actionable decision support with such dynamic data sets is a challenge, and it could be argued that the performance of the model on an experimental data set is irrelevant unless, once deployed to the EMR, the relevant decision support is actionable by clinicians in time to avoid an adverse outcome.

As maturity grows in the development cycle within EMRs, there will be potential opportunities to improve EMR-based deterioration tools and assess their impact on patient care. The development of algorithms can help monitor in real time rich clinical data from our patients, including vital signs, investigation results, drug prescription, and provide useful decision support to improve care trajectories. These clinical data can then be used at the patient level to provide visibility of deteriorating or at-risk patients to individual clinicians, wards, clinical teams, and the MET responders, and place decision support back into the EMR to support clinicians in accurately predicting and preventing inpatient deterioration. At the system level, the data could provide feedback to guide the optimization of the digital or clinician interaction and maximize response efficiency. Successful implementation should improve patient-centered outcomes, reduce suffering, incapacity, and mortality, as well as reduce length of stay, increase hospital capacity, improve efficiency, and increase care delivery to meet ever-increasing demands.

One final, but highly important, factor for consideration in the future implementation of EMR-based deterioration tools is the aspect of medical device regulation. For example, with the recent introduction of regulations for software based medical devices in Australia by the Therapeutic Goods Administration, clinical decision support software such as algorithms to predict patient deterioration that meet the definition of a medical device would be subject to regulation and possibly inclusion in the Australian Register of Therapeutic Goods. The legislation also allows for clinical decision support software to be considered as an exempt medical device subject to how it is intended to be used. Either way, it is expected that regulation will lead to increased clinical acceptance and uptake of such algorithms.

Conclusions

The development and accuracy of digital EWSs is increasing, facilitated by the growing availability of digital data sets. However, despite the relative performance of algorithms that can predict patient deterioration, the current literature shows that limited deployment of such algorithms into clinical practice is associated with improvement in patient outcomes. There is a paucity of quality studies in this area, and further work is needed to explore potential clinical benefits, including optimizing of the digital or clinician interaction, consideration of limitations in implementation, such as the requirement for real-time data availability, and use of standardized measures.

Acknowledgments

The work presented in this manuscript was supported by Clinical Excellence Queensland, Queensland Health, and by the National Health and Medical Research Council Centre for Research Excellence in Digital Health (1134919). However, these sponsors had no influence on the study design; on the collection, analysis, and interpretation of data, on the writing of the manuscript; or decision to submit the manuscript for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Studies of more complex patient deterioration models.
[DOCX File , 22 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Full details of the papers included in this review, including information on prediction model input variables, performance metrics, and quality assessment.
[DOCX File , 49 KB-Multimedia Appendix 2]

References

1. Gao H, McDonnell A, Harrison DA, Moore T, Adam S, Daly K, et al. Systematic review and evaluation of physiological track and trigger warning systems for identifying at-risk patients on the ward. Intensive Care Med 2007 Apr;33(4):667-679. [doi: 10.1007/s00134-007-0532-3] [Medline: 17318499]
2. McGaughey J, Alderdice F, Fowler R, Kapila A, Mayhew A, Moutray M. Outreach and Early Warning Systems (EWS) for the prevention of intensive care admission and death of critically ill adult patients on general hospital wards. Cochrane Database Syst Rev 2007 Jul 18(3):CD005529. [doi: 10.1002/14651858.CD005529.pub2] [Medline: 17636805]
3. Smith MB, Chiovaro JC, O'Neil M, Kansagara D, Quiñones AR, Freeman M, et al. Early warning system scores for clinical deterioration in hospitalized patients: a systematic review. Ann Am Thorac Soc 2014 Nov;11(9):1454-1465. [doi: 10.1513/AnnalsATS.201403-102OC] [Medline: 25296111]
4. Chapman SM, Wray J, Oulton K, Peters MJ. Systematic review of paediatric track and trigger systems for hospitalised children. Resuscitation 2016 Dec;109:87-109. [doi: 10.1016/j.resuscitation.2016.07.230] [Medline: 27496259]
5. Jacob N, Moriarty Y, Lloyd A, Mann M, Tume LN, Sefton G, et al. Optimising paediatric afferent component early warning systems: a hermeneutic systematic literature review and model development. BMJ Open 2019 Nov 14;9(11):e028796 [FREE Full text] [doi: 10.1136/bmjopen-2018-028796] [Medline: 31727645]
6. O'Connell A, Flabouris A, Thompson CH. Optimising the response to acute clinical deterioration: the role of observation and response charts. Intern Med J 2020 Jul;50(7):790-797. [doi: 10.1111/imj.14444] [Medline: 31389119]
7. Varghese J, Kleine M, Gessner SI, Sandmann S, Dugas M. Effects of computerized decision support system implementations on patient outcomes in inpatient care: a systematic review. J Am Med Inform Assoc 2018 May 01;25(5):593-602 [FREE Full text] [doi: 10.1093/jamia/ocx100] [Medline: 29036406]

8. Alam N, Hobbelink EL, van Tienhoven AJ, van de Ven PM, Jansma EP, Nanayakkara PW. The impact of the use of the Early Warning Score (EWS) on patient outcomes: a systematic review. Resuscitation 2014 May;85(5):587-594. [doi: 10.1016/j.resuscitation.2014.01.013] [Medline: 24467882]
9. Fang AH, Lim WT, Balakrishnan T. Early warning score validation methodologies and performance metrics: a systematic review. BMC Med Inform Decis Mak 2020 Jun 18;20(1):111 [FREE Full text] [doi: 10.1186/s12911-020-01144-8] [Medline: 32552702]
10. Gerry S, Bonnici T, Birks J, Kirtley S, Virdee PS, Watkinson PJ, et al. Early warning scores for detecting deterioration in adult hospital patients: systematic review and critical appraisal of methodology. Br Med J 2020 May 20;369:m1501 [FREE Full text] [doi: 10.1136/bmj.m1501] [Medline: 32434791]
11. Amland RC, Sutariya BB. Quick Sequential [Sepsis-Related] Organ Failure Assessment (qSOFA) and St. John sepsis surveillance agent to detect patients at risk of sepsis: an observational cohort study. Am J Med Qual 2018;33(1):50-57 [FREE Full text] [doi: 10.1177/1062860617692034] [Medline: 28693336]
12. Li L, Walter S, Rathnayake K, Westbrook J. Evaluation and Optimisation of Risk Identification Tools for the Early Detection of Sepsis in Adult Inpatients. Sydney: Australian Institute of Health Innovation, Macquarie University; 2018.
13. Delahanty RJ, Kaufman D, Jones SS. Development and evaluation of an automated machine learning algorithm for in-hospital mortality risk adjustment among critical care patients. Crit Care Med 2018 Jun;46(6):e481-e488. [doi: 10.1097/CCM.0000000000003011] [Medline: 29419557]
14. Malycha J, Bonnici T, Clifton DA, Ludbrook G, Young JD, Watkinson PJ. Patient centred variables with univariate associations with unplanned ICU admission: a systematic review. BMC Med Inform Decis Mak 2019 May 15;19(1):98 [FREE Full text] [doi: 10.1186/s12911-019-0820-1] [Medline: 31092256]
15. Centre for Reviews and Dissemination. Systematic Reviews: CRD's Guidance for Undertaking Reviews in Healthcare. York, England: University of York NHS Centre for Reviews & Dissemination; Dec 01, 2008.
16. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-269 [FREE Full text] [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
17. Hausner E, Waffenschmidt S, Kaiser T, Simon M. Routine development of objectively derived search strategies. Syst Rev 2012 Feb 29;1:19 [FREE Full text] [doi: 10.1186/2046-4053-1-19] [Medline: 22587829]
18. Fatehi F, Gray LC, Wootton R. How to improve your PubMed/MEDLINE searches: 3. advanced searching, MeSH and My NCBI. J Telemed Telecare 2014 Mar;20(2):102-112. [doi: 10.1177/1357633X13519036] [Medline: 24614997]
19. EndNote: EndNote X9.: Clarivate URL: https://clarivate.libguides.com/endnote_training/users/enx9 [accessed 2021-08-10]
20. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
21. Wolff RF, Moons KG, Riley RD, Whiting PF, Westwood M, Collins GS, PROBAST Group†. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 2019 Jan 01;170(1):51-58 [FREE Full text] [doi: 10.7326/M18-1376] [Medline: 30596875]
22. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. Br Med J 2016 Oct 12;355:i4919 [FREE Full text] [doi: 10.1136/bmj.i4919] [Medline: 27733354]
23. Alaa A, Van Der Schaar M. A hidden absorbing semi-Markov model for informatively censored temporal data: learning and inference. J Mach Learn Res 2018 Jan 1;19(1):108-169. [doi: 10.5555/3291125.3291129]
24. Alaa AM, Yoon J, Hu S, van der Schaar M. Personalized risk scoring for critical care prognosis using mixtures of gaussian processes. IEEE Trans Biomed Eng 2018 Jan;65(1):207-218. [doi: 10.1109/TBME.2017.2698602] [Medline: 28463183]
25. Bartkowiak B, Snyder AM, Benjamin A, Schneider A, Twu NM, Churpek MM, et al. Validating the Electronic Cardiac Arrest Risk Triage (eCART) score for risk stratification of surgical inpatients in the postoperative setting: retrospective cohort study. Ann Surg 2019 Jun;269(6):1059-1063 [FREE Full text] [doi: 10.1097/SLA.0000000000002665] [Medline: 31082902]
26. Bedoya AD, Clement ME, Phelan M, Steorts RC, O'Brien C, Goldstein BA. Minimal impact of implemented early warning score and best practice alert for patient deterioration. Crit Care Med 2019 Jan;47(1):49-55 [FREE Full text] [doi: 10.1097/CCM.0000000000003439] [Medline: 30247239]
27. Capan M, Ivy JS, Rohleder T, Hickman J, Huddleston JM. Individualizing and optimizing the use of early warning scores in acute medical care for deteriorating hospitalized patients. Resuscitation 2015 Aug;93:107-112 [FREE Full text] [doi: 10.1016/j.resuscitation.2014.12.032] [Medline: 25597507]
28. Cho KJ, Kwon O, Kwon JM, Lee Y, Park H, Jeon KH, et al. Detecting patient deterioration using artificial intelligence in a rapid response system. Crit Care Med 2020 Apr;48(4):285-289. [doi: 10.1097/CCM.0000000000004236] [Medline: 32205618]
29. Churpek MM, Yuen TC, Edelson DP. Predicting clinical deterioration in the hospital: the impact of outcome selection. Resuscitation 2013 May;84(5):564-568 [FREE Full text] [doi: 10.1016/j.resuscitation.2012.09.024] [Medline: 23022075]

30. Churpek MM, Yuen TC, Park SY, Gibbons R, Edelson DP. Using electronic health record data to develop and validate a prediction model for adverse outcomes in the wards*. Crit Care Med 2014 Apr;42(4):841-848 [FREE Full text] [doi: 10.1097/CCM.0000000000000038] [Medline: 24247472]
31. Churpek MM, Yuen TC, Park SY, Meltzer DO, Hall JB, Edelson DP. Derivation of a cardiac arrest prediction model using ward vital signs*. Crit Care Med 2012 Jul;40(7):2102-2108 [FREE Full text] [doi: 10.1097/CCM.0b013e318250aa5a] [Medline: 22584764]
32. Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. Crit Care Med 2016 Feb;44(2):368-374 [FREE Full text] [doi: 10.1097/CCM.0000000000001571] [Medline: 26771782]
33. Churpek MM, Yuen TC, Winslow C, Robicsek AA, Meltzer DO, Gibbons RD, et al. Multicenter development and validation of a risk stratification tool for ward patients. Am J Respir Crit Care Med 2014 Sep 15;190(6):649-655 [FREE Full text] [doi: 10.1164/rccm.201406-1022OC] [Medline: 25089847]
34. Escobar GJ, Gardner MN, Greene JD, Draper D, Kipnis P. Risk-adjusting hospital mortality using a comprehensive electronic record in an integrated health care delivery system. Med Care 2013 May;51(5):446-453. [doi: 10.1097/MLR.0b013e3182881c8e] [Medline: 23579354]
35. Escobar GJ, LaGuardia JC, Turk BJ, Ragins A, Kipnis P, Draper D. Early detection of impending physiologic deterioration among patients who are not in intensive care: development of predictive models using data from an automated electronic medical record. J Hosp Med 2012;7(5):388-395. [doi: 10.1002/jhm.1929] [Medline: 22447632]
36. Fejza A, Geneves P, Layaida N, Bosson J. Scalable and interpretable predictive models for electronic health records. In: Proceedings of the IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA). 2018 Presented at: 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA); Oct 1-3, 2018; Turin, Italy URL: https://hal.inria.fr/hal-03124966 [doi: 10.1109/DSAA.2018.00045]
37. Green M, Lander H, Snyder A, Hudson P, Churpek M, Edelson D. Comparison of the between the flags calling criteria to the MEWS, NEWS and the electronic Cardiac Arrest Risk Triage (eCART) score for the identification of deteriorating ward patients. Resuscitation 2018 Feb;123:86-91 [FREE Full text] [doi: 10.1016/j.resuscitation.2017.10.028] [Medline: 29169912]
38. Hackmann G, Chen M, Chipara O, Lu C, Chen Y, Kollef M, et al. Toward a two-tier clinical warning system for hospitalized patients. AMIA Annu Symp Proc 2011;2011:511-519 [FREE Full text] [Medline: 22195105]
39. Heller AR, Mees ST, Lauterwald B, Reeps C, Koch T, Weitz J. Detection of deteriorating patients on surgical wards outside the ICU by an automated MEWS-based early warning system with paging functionality. Ann Surg 2020 Jan;271(1):100-105. [doi: 10.1097/SLA.0000000000002830] [Medline: 29771722]
40. Hu W, Chan CW, Zubizarreta JR, Escobar GJ. An examination of early transfers to the ICU based on a physiologic risk score. Manuf Serv Oper Manag 2018 Jul;20(3):531-549. [doi: 10.1287/msom.2017.0658]
41. Huh JW, Lim C, Koh Y, Lee J, Jung Y, Seo H, et al. Activation of a medical emergency team using an electronic medical recording-based screening system*. Crit Care Med 2014 Apr;42(4):801-808. [doi: 10.1097/CCM.0000000000000031] [Medline: 24335439]
42. Keim-Malpass J, Clark MT, Lake DE, Moorman JR. Towards development of alert thresholds for clinical deterioration using continuous predictive analytics monitoring. J Clin Monit Comput 2020 Aug;34(4):797-804. [doi: 10.1007/s10877-019-00361-5] [Medline: 31327101]
43. Kia A, Timsina P, Joshi HN, Klang E, Gupta RR, Freeman RM, et al. MEWS++: enhancing the prediction of clinical deterioration in admitted patients through a machine learning model. J Clin Med 2020 Jan 27;9(2):343 [FREE Full text] [doi: 10.3390/jcm9020343] [Medline: 32012659]
44. Kipnis P, Turk BJ, Wulf DA, LaGuardia JC, Liu V, Churpek MM, et al. Development and validation of an electronic medical record-based alert score for detection of inpatient deterioration outside the ICU. J Biomed Inform 2016 Dec;64:10-19 [FREE Full text] [doi: 10.1016/j.jbi.2016.09.013] [Medline: 27658885]
45. Kirkland LL, Malinchoc M, O'Byrne M, Benson JT, Kashiwagi DT, Burton MC, et al. A clinical deterioration prediction tool for internal medicine patients. Am J Med Qual 2013;28(2):135-142. [doi: 10.1177/1062860612450459] [Medline: 22822159]
46. Kwon J, Lee Y, Lee Y, Lee S, Park J. An algorithm based on deep learning for predicting in-hospital cardiac arrest. J Am Heart Assoc 2018 Jun 26;7(13):e008678 [FREE Full text] [doi: 10.1161/JAHA.118.008678] [Medline: 29945914]
47. Loekito E, Bailey J, Bellomo R, Hart GK, Hegarty C, Davey P, et al. Common laboratory tests predict imminent death in ward patients. Resuscitation 2013 Mar;84(3):280-285. [doi: 10.1016/j.resuscitation.2012.07.025] [Medline: 22863543]
48. Mayampurath A, Sanchez-Pinto LN, Carey KA, Venable L, Churpek M. Combining patient visual timelines with deep learning to predict mortality. PLoS One 2019 Jul 31;14(7):e0220640 [FREE Full text] [doi: 10.1371/journal.pone.0220640] [Medline: 31365580]
49. Mohamadlou H, Panchavati S, Calvert J, Lynn-Palevsky A, Le S, Allen A, et al. Multicenter validation of a machine-learning algorithm for 48-h all-cause mortality prediction. Health Informatics J 2020 Sep;26(3):1912-1925 [FREE Full text] [doi: 10.1177/1460458219894494] [Medline: 31884847]

50. Morgan CK, Amspoker AB, Howard C, Razjouyan J, Siddique M, D'Avignon S, et al. Continuous cloud-based early warning score surveillance to improve the safety of acutely ill hospitalized patients. J Healthc Qual 2021;43(1):59-66. [doi: 10.1097/JHQ.0000000000000272] [Medline: 32604130]
51. O'Brien C, Goldstein BA, Shen Y, Phelan M, Lambert C, Bedoya AD, et al. Development, implementation, and evaluation of an in-hospital optimized early warning score for patient deterioration. MDM Policy Pract 2020 Jan 10;5(1):2381468319899663 [FREE Full text] [doi: 10.1177/2381468319899663] [Medline: 31976373]
52. Oh TK, Kim S, Lee DS, Min H, Choi YY, Lee EY, et al. A rapid response system reduces the incidence of in-hospital postoperative cardiopulmonary arrest: a retrospective study. Can J Anaesth 2018 Dec;65(12):1303-1313. [doi: 10.1007/s12630-018-1200-5] [Medline: 30076577]
53. Redfern OC, Pimentel MA, Prytherch D, Meredith P, Clifton DA, Tarassenko L, et al. Predicting in-hospital mortality and unanticipated admissions to the intensive care unit using routinely collected blood tests and vital signs: development and validation of a multivariable model. Resuscitation 2018 Dec;133:75-81 [FREE Full text] [doi: 10.1016/j.resuscitation.2018.09.021] [Medline: 30253229]
54. Romero-Brufau S, Huddleston JM, Naessens JM, Johnson MG, Hickman J, Morlan BW, et al. Widely used track and trigger scores: are they ready for automation in practice? Resuscitation 2014 Apr;85(4):549-552 [FREE Full text] [doi: 10.1016/j.resuscitation.2013.12.017] [Medline: 24412159]
55. Shamout FE, Zhu T, Sharma P, Watkinson PJ, Clifton DA. Deep interpretable early warning system for the detection of clinical deterioration. IEEE J Biomed Health Inform 2020 Feb;24(2):437-446. [doi: 10.1109/JBHI.2019.2937803] [Medline: 31545746]
56. Somanchi S, Adhikari S, Lin A, Eneva E, Ghani R. Early prediction of cardiac arrest (code blue) using electronic medical records. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2015 Aug 10 Presented at: KDD '15: 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 10 - 13, 2015; Sydney NSW Australia p. 2119-2126. [doi: 10.1145/2783258.2788588]
57. Wengerter BC, Pei KY, Asuzu D, Davis KA. Rothman Index variability predicts clinical deterioration and rapid response activation. Am J Surg 2018 Jan;215(1):37-41. [doi: 10.1016/j.amjsurg.2017.07.031] [Medline: 28818297]
58. Wong J, Taljaard M, Forster AJ, Escobar GJ, van Walraven C. Addition of time-dependent covariates to a survival model significantly improved predictions for daily risk of hospital death. J Eval Clin Pract 2013 Apr;19(2):351-357. [doi: 10.1111/j.1365-2753.2012.01832.x] [Medline: 22409151]
59. Yu S, Leung S, Heo M, Soto GJ, Shah RT, Gunda S, et al. Comparison of risk prediction scoring systems for ward patients: a retrospective nested case-control study. Crit Care 2014 Jun 26;18(3):R132 [FREE Full text] [doi: 10.1186/cc13947] [Medline: 24970344]
60. Ye C, Wang O, Liu M, Zheng L, Xia M, Hao S, et al. A real-time early warning system for monitoring inpatient mortality risk: prospective study using electronic medical record data. J Med Internet Res 2019 Jul 05;21(7):e13719 [FREE Full text] [doi: 10.2196/13719] [Medline: 31278734]
61. Bailey TC, Chen Y, Mao Y, Lu C, Hackmann G, Micek ST, et al. A trial of a real-time alert for clinical deterioration in patients hospitalized on general medical wards. J Hosp Med 2013 May;8(5):236-242. [doi: 10.1002/jhm.2009] [Medline: 23440923]
62. Escobar GJ, Turk BJ, Ragins A, Ha J, Hoberman B, LeVine SM, et al. Piloting electronic medical record-based early detection of inpatient deterioration in community hospitals. J Hosp Med 2016 Nov;11 Suppl 1(Suppl 1):18-24 [FREE Full text] [doi: 10.1002/jhm.2652] [Medline: 27805795]
63. Escobar GJ, Liu VX, Schuler A, Lawson B, Greene JD, Kipnis P. Automated identification of adults at risk for in-hospital clinical deterioration. N Engl J Med 2020 Nov 12;383:1951-1960. [doi: 10.1056/nejmsa2001090]
64. Evans RS, Kuttler KG, Simpson KJ, Howe S, Crossno PF, Johnson KV, et al. Automated detection of physiologic deterioration in hospitalized patients. J Am Med Inform Assoc 2015 Mar;22(2):350-360 [FREE Full text] [doi: 10.1136/amiajnl-2014-002816] [Medline: 25164256]
65. Subbe CP, Duller B, Bellomo R. Effect of an automated notification system for deteriorating ward patients on clinical outcomes. Crit Care 2017 Mar 14;21(1):52 [FREE Full text] [doi: 10.1186/s13054-017-1635-z] [Medline: 28288655]
66. Kang MA, Churpek MM, Zadravecz FJ, Adhikari R, Twu NM, Edelson DP. Real-time risk prediction on the wards: a feasibility study. Crit Care Med 2016 Aug;44(8):1468-1473 [FREE Full text] [doi: 10.1097/CCM.0000000000001716] [Medline: 27075140]
67. Lighthall GK, Markar S, Hsiung R. Abnormal vital signs are associated with an increased risk for critical events in US veteran inpatients. Resuscitation 2009 Nov;80(11):1264-1269. [doi: 10.1016/j.resuscitation.2009.08.012] [Medline: 19744762]
68. Arnold J, Davis A, Fischhoff B, Yecies E, Grace J, Klobuka A, et al. Comparing the predictive ability of a commercial artificial intelligence early warning system with physician judgement for clinical deterioration in hospitalised general internal medicine patients: a prospective observational study. BMJ Open 2019 Oct 10;9(10):e032187 [FREE Full text] [doi: 10.1136/bmjopen-2019-032187] [Medline: 31601602]

69. Dummett BA, Adams C, Scruth E, Liu V, Guo M, Escobar GJ. Incorporating an early detection system into routine clinical practice in two community hospitals. J Hosp Med 2016 Nov;11 Suppl 1(Suppl 1):25-31 [FREE Full text] [doi: 10.1002/jhm.2661] [Medline: 27805798]
70. Paulson SS, Dummett BA, Green J, Scruth E, Reyes V, Escobar GJ. What do we do after the pilot is done? Implementation of a hospital early warning system at scale. Jt Comm J Qual Patient Saf 2020 Apr;46(4):207-216 [FREE Full text] [doi: 10.1016/j.jcjq.2020.01.003] [Medline: 32085952]

Abbreviations

EMR: electronic medical record
EWS: early warning score
ICU: intensive care unit
MeSH: Medical Subject Headings
MET: medical emergency team
PRISMA: Preferred Reporting in Systematic Reviews and Meta-Analyses
PROBAST: Prediction Model Risk of Bias Assessment Tool
ROBINS-I: Risk of Bias in Non-Randomised Intervention Studies

Edited by R Kukafka; submitted 24.02.21; peer-reviewed by V Liu, M Dugas, E Lee; comments to author 05.04.21; revised version received 14.07.21; accepted 27.07.21; published 30.09.21

Please cite as:
Mann KD, Good NM, Fatehi F, Khanna S, Campbell V, Conway R, Sullivan C, Staib A, Joyce C, Cook D
Predicting Patient Deterioration: A Review of Tools in the Digital Hospital Setting
J Med Internet Res 2021;23(9):e28209
URL: https://www.jmir.org/2021/9/e28209
doi: 10.2196/28209
PMID: