

Algorithms for Prediction of Clinical Deterioration on the General Wards: A Scoping Review

Roel V Peelen, MD^{1*}, Yassin Eddahchouri, MD², Mats Koeneman, MSc³, Tom H van de Belt, MSc, PhD³, Harry van Goor, MD, PhD², Sebastian JH Bredie, MD, PhD^{1,3}

¹Radboud University Medical Center, Department of Internal Medicine, Nijmegen, The Netherlands; ²Radboud University Medical Center, Department of Surgery, Nijmegen, The Netherlands; ³Radboud University Medical Center, RShape and Innovation Center, Nijmegen, The Netherlands.

OBJECTIVE: The primary objective of this scoping review was to identify and describe state-of-the-art models that use vital sign monitoring to predict clinical deterioration on the general ward. The secondary objective was to identify facilitators, barriers, and effects of implementing these models.

DATA SOURCES: PubMed, Embase, and CINAHL databases until November 2020.

STUDY SELECTION: We selected studies that compared vital signs–based automated real-time predictive algorithms to current track-and-trace protocols in regard to the outcome of clinical deterioration in a general ward population.

DATA EXTRACTION: Study characteristics, predictive characteristics and barriers, facilitators, and effects.

RESULTS: We identified 1,741 publications, 21 of which were included in our review. Two of these were clinical trials, 2 were prospective observational studies,

and the remaining 17 were retrospective studies. All of the studies focused on hospitalized adult patients. The reported area under the receiver operating characteristic curves ranged between 0.65 and 0.95 for the outcome of clinical deterioration. Positive predictive value and sensitivity ranged between 0.223 and 0.773 and 7.2% to 84.0%, respectively. Input variables differed widely, and predicted endpoints were inconsistently defined. We identified 57 facilitators and 48 barriers to the implementation of these models. We found 68 reported effects, 57 of which were positive.

CONCLUSION: Predictive algorithms can detect clinical deterioration on the general ward earlier and more accurately than conventional protocols, which in one recent study led to lower mortality. Consensus is needed on input variables, predictive time horizons, and definitions of endpoints to better facilitate comparative research. *Journal of Hospital Medicine* 2021;16:XXX-XXX. © 2021 Society of Hospital Medicine

The early identification of clinical deterioration among adult hospitalized patients remains a challenge.¹ Delayed identification is associated with increased morbidity and mortality, unplanned intensive care unit (ICU) admissions, prolonged hospitalization, and higher costs.^{2,3} Earlier detection of deterioration using predictive algorithms of vital sign monitoring might avoid these negative outcomes.⁴ In this scoping review, we summarize current algorithms and their evidence.

Vital signs provide the backbone for detecting clinical deterioration. Early warning scores (EWS) and outreach protocols were developed to bring structure to the assessment of vital signs. Most EWS claim to predict clinical endpoints such as unplanned ICU admission up to 24 hours in advance.^{5,6} Reviews of EWS showed a positive trend toward reduced length of stay and mortality. However, conclusions about general efficacy could not be generated because of case heterogeneity and methodologic shortcomings.^{4,7}

Continuous automated vital sign monitoring of patients on the general ward can now be accomplished with wearable devices.⁸ The first reports on continuous monitoring showed earlier detection of deterioration but not improved clinical endpoints.^{4,9} Since then, different reports on continuous monitoring have shown positive effects but concluded that unprocessed monitoring data per se falls short of generating actionable alarms.^{4,10,11}

Predictive algorithms, which often use artificial intelligence (AI), are increasingly employed to recognize complex patterns or abnormalities and support predictions of events in big data sets.^{12,13} Especially when combined with continuous vital sign monitoring, predictive algorithms have the potential to expedite detection of clinical deterioration and improve patient outcomes. Predictive algorithms using vital signs in the ICU have shown promising results.¹⁴ The impact of predictive algorithms on the general wards, however, is unclear.

The aims of our scoping review were to explore the extent and range of and evidence for predictive vital signs–based algorithms on the adult general ward; to describe the variety of these algorithms; and to categorize effects, facilitators, and barriers of their implementation.¹⁵

MATERIALS AND METHODS

We performed a scoping review to create a summary of the

*Corresponding Author: Roel V Peelen, MD; Email: roelpeelen@gmail.com; Telephone: +31 6 3043 8698; Twitter: @PeelenRoel.

Find additional supporting information in the online version of this article.

Received: August 12, 2020; Revised: March 17, 2021; Accepted: March 30, 2021

© 2021 Society of Hospital Medicine DOI 10.12788/jhm.3630

TABLE 1. Descriptive Characteristics of the Selected Studies

Characteristics	Non-AI algorithms, No. (%)	AI algorithms, No. (%)	Total, No. (%)
Total number of studies	12 (100)	9 (100)	21 (100)
Conference abstracts	1 (8)	1 (11)	2 (10)
Study design			
Retrospective cohort	8 (67)	9 (100)	17 (81)
Prospective cohort	2 (17)	0 (0)	2 (10)
Intervention study	2 (17)	0 (0)	2 (10)
State of development			
Developed	7 (58)	9 (100)	16 (76)
Technically implemented	2 (17)	0 (0)	2 (10)
Clinically implemented	3 (25)	0 (0)	3 (14)
Population			
General ward	11 (92)	4 (44)	15 (71)
General ward and intensive care unit	1 (8)	5 (56)	6 (29)
No. of admissions			
<2,000 or unknown	1 (8)	2 (22)	3 (14)
2,000-30,000	6 (50)	2 (22)	8 (38)
> 30,000	5 (42)	5 (56)	10 (48)
Input			
Vitals only	1 (8)	4 (44)	5 (24)
Vitals and lab results	7 (58)	3 (33)	10 (48)
Vitals and patient characteristics	9 (75)	4 (44)	13 (62)
Input variables not specified	2 (17)	0 (0)	2 (10)
Vitals input			
Intermittent	6 (50)	4 (44)	10 (48)
Continuous vitals	1 (8)	2 (22)	3 (14)
All registered	5 (42)	3 (33)	8 (38)
Horizon of prediction			
<2 hours	1 (8)	1 (11)	2 (10)
2-12 hours	3 (25)	4 (44)	7 (33)
>12 hours	8 (67)	4 (44)	12 (57)
Country			
United States	11 (92)	7 (78)	18 (86)
Other	1 (8)	2 (22)	3 (14)
Center			
Single center	8 (67)	4 (44)	12 (57)
Multicenter	4 (33)	5 (56)	9 (43)

Abbreviation: AI, artificial intelligence.

current state of research. We used the five-step method of Levac and followed the Preferred Reporting Items for Systematic reviews and Meta-Analyses Extension for Scoping Reviews guidelines.^{16,17} (Appendix 1).

PubMed, Embase, and CINAHL databases were searched for English-language articles written between January 1, 2010, and November 20, 2020. We developed the search queries with an experienced information scientist, and we used database-specific terms and strategies for input, clinical outcome, method, predictive capability, and population (Appendix 2).

Additionally, we searched the references of the selected articles, as well as publications citing these articles.

All studies identified were screened by title and abstract by two researchers (RP and YE). The selected studies were read in their entirety and checked for eligibility using the following inclusion criteria: automated algorithm; vital signs-based; real-time prediction; of clinical deterioration; in an adult, general ward population. In cases where there were successive publications with the same algorithm and population, we selected the most recent study.

For screening and selection, we used the Rayyan QCRI online tool (Qatar Computing Research Institute) and Endnote X9 (Clarivate Analytics). We extracted information using a data extraction form and organized it into descriptive characteristics of the selected studies (Table 1): an input data table showing number of admissions, intermittent or continuous measurements, vital signs measured, laboratory results (Appendix Table 1), a table summarizing study designs and settings (Appendix Table 2), and a prediction performance table (Table 2). We report characteristics of the populations and algorithms, prediction specifications such as area under the receiver operating curve (AUROC), and predictive values. Predictive values are affected by prevalence, which may differ among populations. To compare the algorithms, we calculated an indexed positive predictive value (PPV) and a number needed to evaluate (NNE) using a weighted average prevalence of clinical deterioration of 3.0%.

We defined clinical deterioration as endpoints, including rapid response team activation, cardiopulmonary resuscitation, transfer to an ICU, or death. Real-time was defined by the ability to automatically update predictions as new measurements are added. Predictions were defined as data-derived warnings for events in the near future. Prediction horizon was defined as the period for which a prediction is made. Special interest was given to algorithms that involved AI, which we defined as any form of machine learning or other nonclassical statistical algorithm.

Effects, facilitators, and barriers were identified and categorized using ATLAS.ti 8 software (ATLAS.ti) and evaluated by three researchers (RP, MK, and THvdB). These were categorized using the adapted frameworks of Gagnon et al¹⁸ for the barriers and facilitators and of Donabedian¹⁹ for the effects (Appendix 3).

The Gagnon et al framework was adapted by changing two of four domains—that is, “Individual” was changed to “Professional” and “Human” to “Physiology.” The domains of “Technology” and “Organization” remained unchanged. The Donabedian domains of “Outcome,” “Process,” and “Structure” also remained unchanged (Table 3).

We divided the studies into two groups: studies on predictive algorithms with and without AI when reporting on characteristics and performance. For the secondary aim of exploring implementation impact, we reported facilitators and barriers in a narrative way, highlighting the most frequent and notable findings.

RESULTS

As shown in the Figure, we found 1,741 publications, of which we read the full-text of 109. There were 1,632 publications that did not meet the inclusion criteria. The publications by Churpek et al,^{20,21} Bartkiowak et al,²² Edelson et al,²³ Escobar et al,^{24,25} and Kipnis et al²⁶ reported on the same algorithms or databases but had significantly different approaches. For multiple publications using the same algorithm and population, the most recent was named with inclusion of the earlier findings.^{20,21,27–29} The resulting 21 papers are included in this review.

TABLE 2. Table With Prediction Characteristics

Publication	Horizon, h	Prediction specifications ^a			
		AUC	PPV @ 3.0% prevalence	NNE @ 3.0% prevalence	Sensitivity, %
Non-AI algorithms					
Alvarez ⁴²	24	0.85	0.130	7.7	42.2
Kollef ²⁷	4	-	0.109	9.2	41.1
Bartkiowak ²²	24	0.79	0.172	5.8	47.0
Churpek ²⁰	24	0.77	0.236	4.2	20.0
Correia ³¹	24	0.65	-	-	-
Escobar ²⁴	12	0.78	-	-	-
Evans ³⁵	-	-	-	-	-
Ghosh ³⁶	24	0.77	0.527	1.9	7.2
Kang ³⁷	24	0.80	0.124	8.1	52.5
Escobar ²⁵	12	0.82	0.112	8.9	49.0
Kirkland ⁵¹	12	0.71	-	-	-
Rothman ³⁹	24	0.93	-	-	-
AI algorithms					
Alaa ⁴⁰	12	0.81	-	-	60
Bai ³³	12	-	-	-	-
Churpek ²¹	8	0.80	-	-	-
Edelson ²³	24	0.76	-	-	-
Hu ³⁴	2	-	0.336	3.0	81.8
Hu ³⁸	12	0.92	0.565	1.8	84.0
Kwon ³⁰	48	0.95	0.429	2.3	80.0
Mohamadlou ⁴¹	24	0.88	0.166	6.0	72.9
Shamout ³²	24	0.85	0.419	2.4	37.3

Abbreviations: AUC, area under the receiver operator curve; PPV, positive predictive value; NNE, number needed to evaluate.

^a If data were insufficient to calculate the relevant metric, a dash is displayed.

Descriptive characteristics of the studies are summarized in Table 1. Nineteen of the publications were full papers and two were conference abstracts. Most of the studies (n = 18) were from the United States; there was one study from South Korea,³⁰ one study from Portugal,³¹ and one study from the United Kingdom.³² In 15 of the studies, there was a strict focus on general or specific wards; 6 studies also included the ICU and/or emergency departments.

Two of the studies were clinical trials, 2 were prospective observational studies, and 17 were retrospective studies. Five studies reported on an active predictive model during admission. Of these, 3 reported that the model was clinically implemented, using the predictions in their clinical workflow. None of the implemented studies used AI.

TABLE 3. Effects and Facilitators of, and Barriers to Implementation of Predictive Algorithms

Effect	Positive effects	Reference	Negative effects	Reference
Outcomes	28		6	
AUROC-characteristics and predictive values	17	20-23, 25, 27*, 30, 32, 34, 36, 37, 38*, 39-41	1	31
Detection of deterioration	4	20, 25, 26, 42		
Personalized/subtypes	3	32, 40, 41		
Patient related outcomes	4	25*, 26, 35	1	37
Validity and Limitations			4	27, 33, 39, 42
Process	26		4	
Smoothing	2	24, 40		
Predictive alarming	7	22, 27, 30, 36, 38, 40*		
Follow up	3	35*, 42	1	27
Workflow	6	24, 27, 35*, 37, 42		
Alarm fatigue	8	20, 22, 26, 30, 32, 33, 35, 40	3	27, 42*
Structure	3		1	
Resource management	1	40		
Training and education			1	34
Information Infrastructure	2	24, 37		
Facilitator or barrier	Facilitators	Reference	Barriers	Reference
Factors related to technology	18		20	
Prediction (horizon)	2	41, 42	2	39, 42
Implementation (variables, accessibility)	5	26*, 32, 37, 41	4	22, 24, 40, 42
Interpretability	1	32	7	30, 32*, 33, 35, 39, 41
Development	3	25, 26, 42	1	26
External influences on prediction	1	42	3	26*, 41
Excess alarms/alarm fatigue	1	36		
Patient groups	1	41		
Data overfitting			2	21, 38
Robustness			1	33
Electronic medical record integration	3	25, 27, 36		
Professionals	6		5	
Attitude	1	35	1	39
Interpretation and acting upon	5	21, 26, 27, 32, 39	2	21, 38
Discriminating variables			1	26
End-of-life decisions			1	24
Physiology	8		9	
Prevalence			2	24, 42
Vital related	3	21, 34, 51		
Preventability and consequences of interventions	4	21, 40*, 41	2	24, 38
Heterogeneity of patient groups	1	40	5	24, 25, 30, 38, 41
Organization	25		14	
Time and workload/efficiency	5	21, 23, 27, 35, 38	1	25
Workflow/alarm fatigue	6	25*, 31, 33, 38, 51		
Use of resources	3	21, 27, 41	1	25
Implementation	3	25, 30, 41	2	25, 47
Transferable to other hospital	3	25, 30, 39		
Quality of data	1	21	3	24, 25, 32
Quantity of data	3	24, 30, 33	2	24, 32
Route of access to data			3	21, 22, 33
Provability of usefulness	1	21	2	24, 37

* Multiple effects, facilitators, or barriers of this category in this reference.

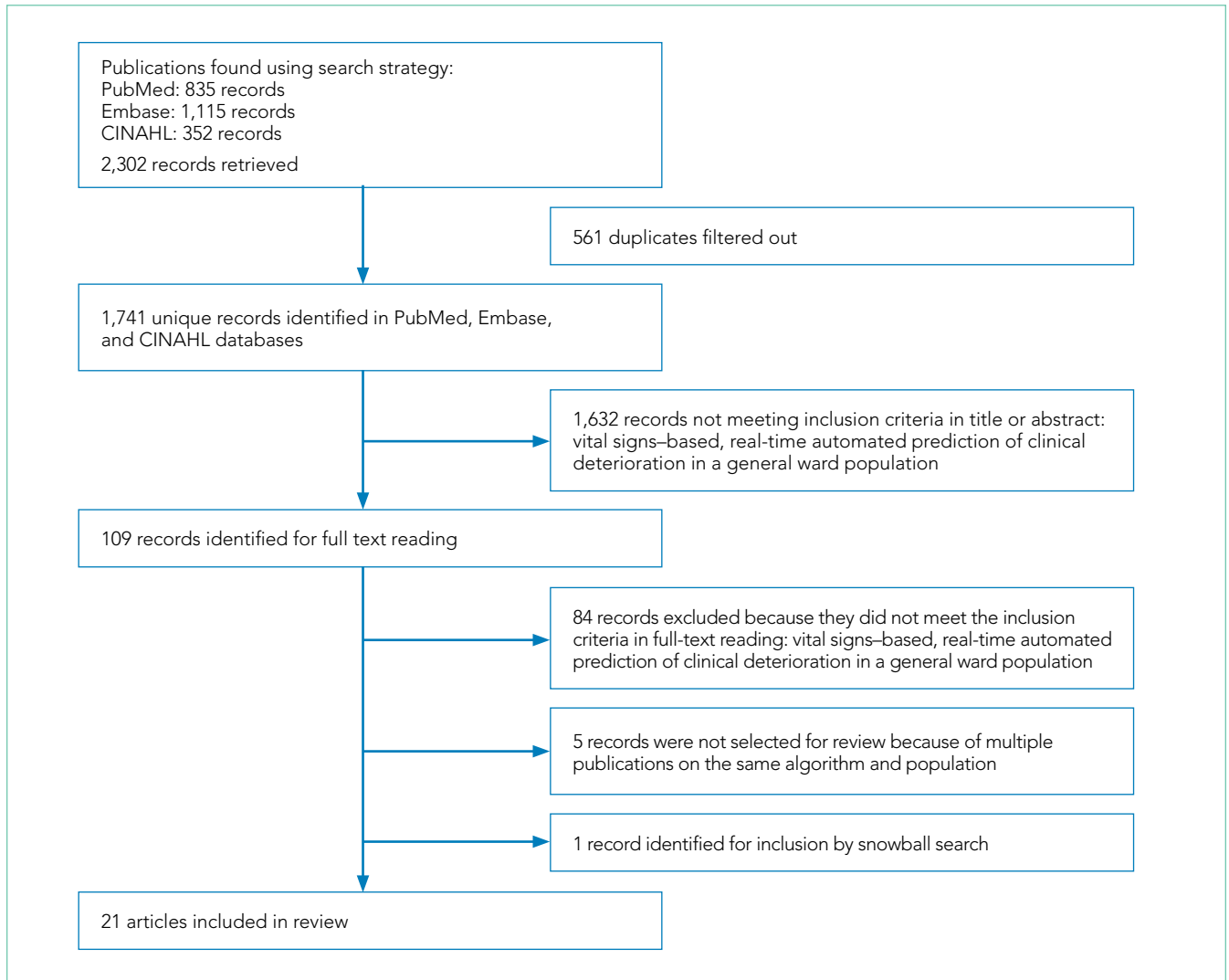


FIG. Study Screening and Selection Flow Diagram

All input variables are presented in Appendix Table 1. In 10 of the studies, vital signs were combined with laboratory results; in 13 of the studies, vital signs were combined with patient characteristics. All of the studies used data derived from electronic medical records (EMRs), except for Bai et al³³ and Hu et al,³⁴ who used single-source waveforms directly from the bedside monitor. Three studies focused on continuous vital sign measurements.^{27,33,34}

Most authors reported an AUROC to describe the predictive value of their algorithms. As shown in Table 2, AUROCs varied from 0.65 to 0.95, with indexed PPVs between 0.24 and 0.75. Sensitivity ranged from 7.2% to 52.5% in non-AI models and up to 82.4% in AI models. Prediction definitions, horizons, and the reported metrics differed too much to directly compare studies.

The non-AI algorithm prediction horizons ranged from 4 to 24 hours, with a median of 24 hours (interquartile range [IQR], 12-24 hours). The AI algorithms ranged from 2 to 48 hours and had a median horizon of 14 hours (IQR, 12-24 hours).

We found three studies reporting patient outcomes. The most recent of these was a large multicenter implementation study by Escobar et al²⁵ that included an extensive follow-up response. This study reported a significantly decreased 30-day mortality in the intervention cohort. A smaller randomized controlled trial reported no significant differences in patient outcomes with earlier warning alarms.²⁷ A third study reported more appropriate rapid response team deployment and decreased mortality in a subgroup analysis.³⁵

Effects, Facilitators, and Barriers

As shown in the Appendix Figure and further detailed in Table 3, the described effects were predominantly positive—57 positive effects vs 11 negative effects. These positive effects sorted primarily into the outcome and process domains.

All of the studies that compared their proposed model with one of various warning systems (eg, EWS, National Early Warning Score [NEWS], Modified Early Warning Score [MEWS]) showed superior performance (based on AUROC

and reported predictive values). In 17 studies, the authors reported their model as more useful or superior to the EWS.^{20,23,26-28,34,36-41} Four studies reported real-time detection of deterioration before regular EWS,^{20,26,42} and three studies reported positive effects on patient-related outcomes.^{26,35} Four negative effects were noted on the controllability, validity, and potential limitations.^{27,42}

There were 26 positive effects on the clinical process mentioned, 7 of which pointed out the effects of earlier, predictive alarming. Algorithms with higher PPVs reported greater rates of actionable alarms, less alarm fatigue, and improved workflow.^{21,22,24-27,30,32,33,35-38,40} Potential alarm fatigue was named as a barrier.^{27,42} Smoother scoring instead of binned categories was mentioned positively.^{24,26}

In the infrastructure domain, very few items were reported. The increased need for education on the used techniques was reported once as a negative effect.³⁴ One of the positive infrastructural effects noted was more efficient planning and use of resources.^{24,37,40}

We identified 57 facilitators and 48 barriers for the clinical implementation and use of real-time predictive analytics (Appendix Figure). In the Technology domain, there were 18 facilitators and 20 barriers cited, and in the Organization domain, 25 and 14, respectively. They were equally present in the Professional and Physiology domains (6 vs 5, 8 vs 9).

Of the 38 remarks in the Technology domain, difficulty with implementation in daily practice was a commonly cited barrier.^{22,24,40,42} Difficulties included creating real-time data feeds out of the EMR, though there were mentions of some successful examples.^{25,27,36} Difficulty in the interpretability of AI was also considered a potential barrier.^{30,32,33,35,39,41} There were remarks as to the applicability of the prolonged prediction horizon because of the associated decoupling from the clinical view.^{39,42}

Conservative attitudes toward new technologies and inadequate knowledge were mentioned as barriers.³⁹ Repeated remarks were made on the difficulty of interpreting and responding to a predicted escalation, as the clinical pattern might not be recognizable at such an early stage. On the other hand, it is expected that less invasive countermeasures would be adequate to avert further escalation. Earlier recognition of possible escalations also raised potential ethical questions, for instance, when to discuss palliative care.²⁴

The heterogeneity of the general ward population and the relatively low prevalence of deterioration were mentioned as barriers.^{24,30,38,41} There were also concerns that not all escalations are preventable and that some patient outcomes may not be modifiable.^{24,38}

Many investigators expected reductions in false alarms and associated alarm fatigue (reflected as higher PPVs). Furthermore, they expected workflow to improve and workload to decrease.^{21,23,27,31,33,35,38,41} Despite the capacity of modern EMRs to store large amounts of patient data, some investigators felt improvements to real-time access, data quality and validity, and data density are needed to ensure valid associated predictions.^{21,22,24,32,37}

DISCUSSION

As the complexity and comorbidity of hospitalized adults grow, predicting clinical deterioration is becoming more important. With an ever-increasing amount of available patient data, real-time algorithms can predict the patient's clinical course with increasing accuracy, positively affecting outcomes.^{4,21,25,43} The studies identified in this scoping review, as measured by higher AUROC scores and improved PPVs, show that predictive algorithms can outperform more conventional EWS, enable earlier and more efficient alarming, and be successfully implemented on the general wards. However, formal meta-analysis was made infeasible by differences in populations, use of different endpoint definitions, cut-off points, time-horizons to prediction, and other methodologic heterogeneity.

There are several important limitations across these studies. In a clinical setting, these models would function as a screening test. Almost all studies report an AUROC; however, sensitivity and PPV or NNE (defined as 1/PPV) may be more useful than AUROC when predicting low-frequency events with high-potential clinical impact.⁴⁴ Assessing the NNE is especially relevant because of its relation to alarm fatigue and responsiveness of clinicians.⁴³ Alarm fatigue and lack of adequate response to alarms were repeatedly cited as potential barriers for application of automated scores. A more useful metric might be NNE over a certain timeframe and across a specified number of patients to more clearly reflect associated workload. Future studies should include these metrics as indicators of the usability and clinical impact of predictive models. This review could not assess PPV or NNE systematically due to inconsistencies in the reporting of these metrics.

Although the results of our scoping review are promising, there are limited data on clinical outcomes using these algorithms. Only three of five algorithms were used to guide clinical decision-making.^{25,27,35} Kollef et al²⁷ showed shorter hospitalizations and Evans et al³⁵ found decreased mortality rates in a multimorbid subgroup. Escobar et al²⁵ found an overall and consistent decrease in mortality in a large, heterogenic population of inpatients across 21 hospitals. While Escobar et al's findings provide strong evidence that predictive algorithms and structured follow-up on alarms can improve patient outcomes, it recognizes that not all facilities will have the resources to implement them.²⁵ Dedicated round-the-clock follow-up of alarms has yet to be proven feasible for smaller institutions, and leaner solutions must be explored. The example set by Escobar et al²⁵ should be translated into various settings to prove its reproducibility and to substantiate the clinical impact of predictive models and structured follow-up.

According to expert opinion, the use of high-frequency or continuous monitoring at low-acuity wards and AI algorithms to detect trends and patterns will reduce failure-to-rescue rates.^{4,9,43} However, most studies in our review focused on periodic spot-checked vital signs, and none of the AI algorithms were implemented in clinical care (Appendix Table 1). A significant barrier to implementation was uncertainty surrounding how to react to generated alarms.^{9,45} As algorithms become

more complex and predict earlier, interpretability and causality in general can diminish, and the response to this type of alarm will be different from that of an acute warning from an EWS.

The assessment of predictive algorithm protocols must include their impact on clinical workflow, workload, and resource utilization. Earlier detection of deterioration can potentially allow coordinated alarm follow-up and lead to more efficient use of resources.^{20,21,31,43,46,47}

Greater numbers of variables do not always improve the quality of monitoring. For example, in one study, an algorithm combining only heart rate, respiration rate, and age outperformed an EWS that tracked six vital sign measures.²³ Algorithms using fewer variables may facilitate more frequent and less complex error-sensitive monitoring. Leaner measurements may also lead to higher patient and clinician acceptance.^{43,45}

The end goal of implementing predictive algorithms on the general ward is to provide timely, reliable, and actionable clinical decision support.⁴³ As shown in a recent study by Blackwell et al,⁴⁸ multiple prediction models for specific clinical events may increase interpretability and performance. Disease-specific algorithms may complement general algorithms for clinical deterioration and enhance overall performance.

STRENGTHS AND LIMITATIONS

We performed a comprehensive review of the current literature using a clear and reproducible methodology to minimize the risk of missing relevant publications. The identified research is mainly limited to large US centers and consists of mostly retrospective studies. Heterogeneity among inputs, endpoints, time horizons, and evaluation metrics make comparisons challenging. Comments on facilitators, barriers, and effects were limited. Positive publication bias may have led to overrepresentation of models showing clinical benefit.

RECOMMENDATIONS FOR FUTURE RESEARCH

Artificial intelligence and the use of continuous monitoring hold great promise in creating optimal predictive algorithms. Future studies should directly compare AI- and non-AI-based algorithms using continuous monitoring to determine predictive accuracy, feasibility, costs, and outcomes. A consensus on endpoint definitions, input variables, methodology, and reporting is needed to enhance reproducibility, comparability, and generalizability of future research. The current research is limited to a few research groups, predominantly in the United States. More international research could enhance validity and increase applicability across varying populations and settings. Greater collaboration would accelerate research and enhance short-cycled continuous improvements. Sharing databases with different populations, variables, and outcomes, such as the Medical Information Mart for Intensive Care database,⁴⁹ could help develop, test, and compare models and contribute to consensus in data standardization and consistent reporting of results. Studies should be designed to determine clinical, societal, and economic effects in accordance with the Quadruple Aim principle.⁵⁰ Successful implementation will depend not only on improved patient outcomes but

also on cost-effectiveness, robust statistics, and end-user acceptance. Follow-up protocols and workflows also should be studied and optimized.

CONCLUSION

Predictive analytics based on vital sign monitoring can identify clinical deterioration at an earlier stage and can do so more accurately than conventional EWS. Implementation of such monitoring can simultaneously decrease alarm-related workload and enhance the efficiency of follow-up. While there is also emerging evidence of associated mortality reduction, it may be too soon to know how representative these findings are. The current literature is limited by heterogeneity across populations studied, monitoring frequency, definitions of deterioration, and clinical outcomes. Consensus is therefore needed to better compare tools and harmonize results. Artificial intelligence and continuous monitoring show great promise in advancing the field; however, additional studies to assess cost, actionability of results, and end-user acceptance are required.

Disclosures: The authors have no conflicts of interest to disclose.

Funding: This work was funded by internal funds of the Radboud University Medical Center.

References

- van Galen LS, Struik PW, Driesen BEJM, et al. Delayed recognition of deterioration of patients in general wards is mostly caused by human related monitoring failures: a root cause analysis of unplanned ICU admissions. *PLoS One*. 2016;11(8):e0161393. <https://doi.org/10.1371/journal.pone.0161393>
- Mardini L, Lipes J, Jayaraman D. Adverse outcomes associated with delayed intensive care consultation in medical and surgical inpatients. *J Crit Care*. 2012;27(6):688-693. <https://doi.org/10.1016/j.jccr.2012.04.011>
- Young MP, Gooder VJ, McBride K, James B, Fisher ES. Inpatient transfers to the intensive care unit: delays are associated with increased mortality and morbidity. *J Gen Intern Med*. 2003;18(2):77-83. <https://doi.org/10.1046/j.1525-1497.2003.20441.x>
- Khanna AK, Hoppe P, Saugel B. Automated continuous noninvasive ward monitoring: future directions and challenges. *Crit Care*. 2019;23(1):194. <https://doi.org/10.1186/s13054-019-2485-7>
- Ludikhuijze J, Hamming A, de Jonge E, Fikkers BG. Rapid response systems in The Netherlands. *Jt Comm J Qual Patient Saf*. 2011;37(3):138-197. [https://doi.org/10.1016/s1553-7250\(11\)37017-1](https://doi.org/10.1016/s1553-7250(11)37017-1)
- Cuthbertson BH, Boroujerdi M, McKie L, Aucott L, Prescott G. Can physiological variables and early warning scoring systems allow early recognition of the deteriorating surgical patient? *Crit Care Med*. 2007;35(2):402-409. <https://doi.org/10.1097/01.ccm.0000254826.10520.87>
- Alam N, Hobbelenk EL, van Tienhoven AJ, van de Ven PM, Jansma EP, Nanayakkara PWB. The impact of the use of the Early Warning Score (EWS) on patient outcomes: a systematic review. *Resuscitation*. 2014;85(5):587-594. <https://doi.org/10.1016/j.resuscitation.2014.01.013>
- Weenk M, Koeneman M, van de Belt TH, Engelen LJLPG, van Goor H, Bredie SJH. Wireless and continuous monitoring of vital signs in patients at the general ward. *Resuscitation*. 2019;136:47-53. <https://doi.org/10.1016/j.resuscitation.2019.01.017>
- Cardona-Morrell M, Prgomet M, Turner RM, Nicholson M, Hillman K. Effectiveness of continuous or intermittent vital signs monitoring in preventing adverse events on general wards: a systematic review and meta-analysis. *Int J Clin Pract*. 2016;70(10):806-824. <https://doi.org/10.1111/ijcp.12846>
- Brown H, Terrence J, Vasquez P, Bates DW, Zimlichman E. Continuous monitoring in an inpatient medical-surgical unit: a controlled clinical trial. *Am J Med*. 2014;127(3):226-232. <https://doi.org/10.1016/j.amjmed.2013.12.004>
- Mestrom E, De Bie A, van de Steeg M, Driessen M, Atallah L, Bezemer R. Implementation of an automated early warning scoring system in a

- surgical ward: practical use and effects on patient outcomes. *PLoS One*. 2019;14(5):e0213402. <https://doi.org/10.1371/journal.pone.0213402>
12. Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol*. 2017;2(4):230-243. <https://doi.org/10.1136/svn-2017-000101>
 13. Iwashyna TJ, Liu V. What's so different about big data? A primer for clinicians trained to think epidemiologically. *Ann Am Thorac Soc*. 2014;11(7):1130-1135. <https://doi.org/10.1513/annalsats.201405-185as>
 14. Jalali A, Bender D, Rehman M, Nadkanri V, Nataraj C. Advanced analytics for outcome prediction in intensive care units. *Conf Proc IEEE Eng Med Biol Soc*. 2016;2016:2520-2524. <https://doi.org/10.1109/embc.2016.7591243>
 15. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med Res Methodol*. 2018;18(1):143. <https://doi.org/10.1186/s12874-018-0611-x>
 16. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol*. 2005;8(1):19-32. <https://doi.org/10.1080/1364557032000119616>
 17. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med*. 2018;169(7):467-473. <https://doi.org/10.7326/m18-0850>
 18. Gagnon MP, Desmartis M, Gagnon J, et al. Framework for user involvement in health technology assessment at the local level: views of health managers, user representatives, and clinicians. *Int J Technol Assess Health Care*. 2015;31(1-2):68-77. <https://doi.org/10.1017/s0266462315000070>
 19. Donabedian A. The quality of care. How can it be assessed? *JAMA*. 1988;260(12):1743-1748. <https://doi.org/10.1001/jama.260.12.1743>
 20. Churpek MM, Yuen TC, Winslow C, et al. Multicenter development and validation of a risk stratification tool for ward patients. *Am J Respir Crit Care Med*. 2014;190(6):649-655. <https://doi.org/10.1164/rccm.201406-1022oc>
 21. 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;44(2):368-374. <https://doi.org/10.1097/ccm.0000000000001571>
 22. Bartkowiak B, Snyder AM, Benjamin A, 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;269(6):1059-1063. <https://doi.org/10.1097/sla.0000000000002665>
 23. Edelson DP, Carey K, Winslow CJ, Churpek MM. Less is more: detecting clinical deterioration in the hospital with machine learning using only age, heart rate and respiratory rate. Abstract presented at: American Thoracic Society International Conference; May 22, 2018; San Diego, California. *Am J Resp Crit Care Med*. 2018;197:A4444.
 24. 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. <https://doi.org/10.1002/jhm.1929>
 25. 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;383(20):1951-1960. <https://doi.org/10.1056/nejmsa2001090>
 26. Kipnis P, Turk BJ, Wulf DA, 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;64:10-19. <https://doi.org/10.1016/j.jbi.2016.09.013>
 27. Kolfel MH, Chen Y, Heard K, et al. A randomized trial of real-time automated clinical deterioration alerts sent to a rapid response team. *J Hosp Med*. 2014;9(7):424-429. <https://doi.org/10.1002/jhm.2193>
 28. Hackmann G, Chen M, Chipara O, et al. Toward a two-tier clinical warning system for hospitalized patients. *AMIA Annu Symp Proc*. 2011;2011:511-519.
 29. Bailey TC, Chen Y, Mao Y, Lu, C, Hackmann G, Micek ST. A trial of a real-time alert for clinical deterioration in patients hospitalized on general medical wards. *J Hosp Med*. 2013;8(5):236-242. <https://doi.org/10.1002/jhm.2009>
 30. Kwon JM, 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;7(13):e008678. <https://doi.org/10.1161/jaha.118.008678>
 31. Correia S, Gomes A, Shahriari S, Almeida JP, Severo M, Azevedo A. Performance of the early warning system vital to predict unanticipated higher-level of care admission and in-hospital death of ward patients. *Value Health*. 2018;21(5):S360. <https://doi.org/10.1016/j.jval.2018.09.2152>
 32. 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;24(2):437-446. <https://doi.org/10.1109/jbhi.2019.2937803>
 33. Bai Y, Do DH, Harris PRE, et al. Integrating monitor alarms with laboratory test results to enhance patient deterioration prediction. *J Biomed Inform*. 2015;53:81-92. <https://doi.org/10.1016/j.jbi.2014.09.006>
 34. Hu X, Sapo M, Nenov V, et al. Predictive combinations of monitor alarms preceding in-hospital code blue events. *J Biomed Inform*. 2012;45(5):913-921. <https://doi.org/10.1016/j.jbi.2012.03.001>
 35. Evans RS, Kuttler KG, Simpson KJ, et al. Automated detection of physiologic deterioration in hospitalized patients. *J Am Med Inform Assoc*. 2015;22(2):350-360. <https://doi.org/10.1136/amiajnl-2014-002816>
 36. Ghosh E, Eshelman L, Yang L, Carlson E, Lord B. Early deterioration indicator: data-driven approach to detecting deterioration in general ward. *Resuscitation*. 2018;122:99-105. <https://doi.org/10.1016/j.resuscitation.2017.10.026>
 37. Kang MA, Churpek MM, Zdravec FJ, Adhikari R, Twu NM, Edelson DP. Real-time risk prediction on the wards: a feasibility study. *Crit Care Med*. 2016;44(8):1468-1473. <https://doi.org/10.1097/ccm.0000000000001716>
 38. Hu SB, Wong DJL, Correa A, Li N, Deng JC. Prediction of clinical deterioration in hospitalized adult patients with hematologic malignancies using a neural network model. *PLoS One*. 2016;11(8):e0161401. <https://doi.org/10.1371/journal.pone.0161401>
 39. Rothman MJ, Rothman SI, Beals J 4th. Development and validation of a continuous measure of patient condition using the electronic medical record. *J Biomed Inform*. 2013;46(5):837-848. <https://doi.org/10.1016/j.jbi.2013.06.011>
 40. 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;65(1):207-218. <https://doi.org/10.1109/tbme.2017.2698602>
 41. Mohamadlou H, Panchavati S, Calvert J, et al. Multicenter validation of a machine-learning algorithm for 48-h all-cause mortality prediction. *Health Informatics J*. 2020;26(3):1912-1925. <https://doi.org/10.1177/1460458219894494>
 42. Alvarez CA, Clark CA, Zhang S, et al. Predicting out of intensive care unit cardiopulmonary arrest or death using electronic medical record data. *BMC Med Inform Decis Mak*. 2013;13:28. <https://doi.org/10.1186/1472-6947-13-28>
 43. Vincent JL, Einav S, Pearce R, et al. Improving detection of patient deterioration in the general hospital ward environment. *Eur J Anaesthesiol*. 2018;35(5):325-333. <https://doi.org/10.1097/eja.0000000000000798>
 44. Romero-Brufau S, Huddleston JM, Escobar GJ, Liebow M. Why the C-statistic is not informative to evaluate early warning scores and what metrics to use. *Crit Care*. 2015;19(1):285. <https://doi.org/10.1186/s13054-015-0999-1>
 45. Weenk M, Bredie SJ, Koeneman M, Hesselink G, van Goor H, van de Belt TH. Continuous monitoring of the vital signs in the general ward using wearable devices: randomized controlled trial. *J Med Internet Res*. 2020;22(6):e15471. <https://doi.org/10.2196/15471>
 46. Wellner B, Grand J, Canzone E, et al. Predicting unplanned transfers to the intensive care unit: a machine learning approach leveraging diverse clinical elements. *JMIR Med Inform*. 2017;5(4):e45. <https://doi.org/10.2196/medinform.8680>
 47. Elliott M, Baird J. Pulse oximetry and the enduring neglect of respiratory rate assessment: a commentary on patient surveillance. *Br J Nurs*. 2019;28(19):1256-1259. <https://doi.org/10.12968/bjon.2019.28.19.1256>
 48. Blackwell JN, Keim-Malpass J, Clark MT, et al. Early detection of inpatient deterioration: one prediction model does not fit all. *Crit Care Explor*. 2020;2(5):e0116. <https://doi.org/10.1097/cce.0000000000000116>
 49. Johnson AEW, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. *Sci Data*. 2016;3:160035. <https://doi.org/10.1038/sdata.2016.35>
 50. Bodenheimer T, Sinsky C. From triple to quadruple aim: care of the patient requires care of the provider. *Ann Fam Med*. 2014;12(6):573-576. <https://doi.org/10.1370/afm.1713>
 51. Kirkland LL, Malinchoc M, O'Byrne M, et al. A clinical deterioration prediction tool for internal medicine patients. *Am J Med Qual*. 2013;28(2):135-142. <https://doi.org/10.1177/1062860612450459>