
Research and Applications

Implementation approaches and barriers for rule-based and machine learning-based sepsis risk prediction tools: a qualitative study

Mugdha Joshi ¹, Keizra Mecklai², Ronen Rozenblum^{2,3}, and Lipika Samal^{2,3}

¹Department of Medicine, Stanford University, Stanford, California, USA, ²Harvard Medical School, Boston, Massachusetts, USA, and ³Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital, Boston, Massachusetts, USA

Corresponding Author: Mugdha Joshi, 300 Pasteur Drive, Lane Building Room 154, Stanford, CA 94305, USA; mugdhaj@stanford.edu

Received 12 October 2020; Revised 16 January 2022; Editorial Decision 23 March 2022; Accepted 30 March 2022

ABSTRACT

Objective: Many options are currently available for sepsis surveillance clinical decision support (CDS) from electronic medical record (EMR) vendors, third party, and homegrown models drawing on rule-based (RB) and machine learning (ML) algorithms. This study explores sepsis CDS implementation from the perspective of implementation leads by describing the motivations, tool choices, and implementation experiences of a diverse group of implementers.

Materials and Methods: Semi-structured interviews were conducted with and a questionnaire was administered to 21 hospital leaders overseeing CDS implementation at 15 US medical centers. Participants were recruited via convenience sampling. Responses were coded by 2 coders with consensus approach and inductively analyzed for themes.

Results: Use of sepsis CDS is motivated in part by quality metrics for sepsis patients. Choice of tool is driven by ease of integration, customization capability, and perceived predictive potential. Implementation processes for these CDS tools are complex, time-consuming, interdisciplinary undertakings resulting in heterogeneous choice of tools and workflow integration. To improve clinician acceptance, implementers addressed both optimization of the alerts as well as clinician understanding and buy in. More distrust and confusion was reported for ML models, as compared to RB models. Respondents described a variety of approaches to overcome implementation barriers; these approaches related to alert firing, content, integration, and buy-in.

Discussion: While there are shared socio-technical challenges of implementing CDS for both RB and ML models, attention to user education, support, expectation management, and dissemination of effective practices may improve feasibility and effectiveness of ML models in quality improvement efforts.

Conclusion: Further implementation science research is needed to determine real world efficacy of these tools. Clinician acceptance is a significant barrier to sepsis CDS implementation. Successful implementation of less clinically intuitive ML models may require additional attention to user confusion and distrust.

Key words: sepsis, predictive analytics, machine learning, implementation, clinical decision support

Lay Summary

Sepsis is a life-threatening illness. Improving sepsis care is a growing priority for many hospitals. Patients at risk of developing sepsis can be identified before they get very sick using tools that analyze data from computerized medical records systems. A variety of options are available from different sources. Some tools are programmed using established sepsis screening criteria used in clinical practice. Others rely on machine learning, where computer algorithms identify patterns in the available data without being pre-programmed by a human being. In this study, we interviewed 21 individuals at 15 US medical centers who oversaw hospital level implementations of these tools. Teams were motivated by wanting to improve quality of care for patients with sepsis. One major challenge was making the tools identify as many patients truly at risk for sepsis as possible while limiting false identification of patients not actually at risk. Many interviewees also described lack of trust in the tools from the nurses and doctors using the tools. There was more distrust and confusion reported by implementers of tools that relied on machine learning than tools that programmed human logic. Strategies emphasizing user education, user support, and expectation management were reported to be helpful.

INTRODUCTION

Sepsis is a life-threatening illness and an expensive cause for hospitalization affecting more than 1.7 million American adults annually.¹ Since early resuscitation and antibiotic administration can reduce mortality, sepsis recognition and care has become a nationwide priority.² In 2015, Centers for Medicare and Medicaid implemented a sepsis bundled payment (SEP-1), mandating placement of specific orders within 3 h of sepsis onset or hospital presentation.³ The bundle has led many hospitals to invest money, time, and energy into measuring quality of sepsis care.⁴

Clinical decision support (CDS) tools have been an essential part of hospital efforts to comply with bundle requirements. These tools have traditionally been rule-based (RB), meaning that they rely on logic that represents the opinion of clinical experts and clinical guidelines and therefore only identify situations that have been previously identified and programmed into their logic. Traditional early warning systems using Systemic Inflammatory Response System (SIRS) criteria relying on vital sign and lab abnormalities have been criticized for producing excessive false positive alerts.^{2,5,6} Existing data are inconsistent about whether these tools actually change the likelihood of bundle interventions or patient outcomes.^{2,5,7-15} Some prospective studies have shown significant reduction in time to blood culture and antibiotics without significant reductions in mortality.^{9,12,16} While others have largely been negative, including 2 randomized controlled trials.^{2,8,10-12}

In contrast to traditional RB models, in machine learning (ML) algorithms, the associations between patient variables and clinical outcomes such as deterioration, death, and confirmed infection are derived by the computer rather than being pre-programmed. These algorithms often incorporate many more variables than RB models, ranging from vital signs to lab values to demographics and billing codes. ML models have demonstrated improved diagnostic accuracy and reduced false positives compared to clinical assessment tools like SIRS, SOFA, and MEWS scores.^{7,17-20} Furthermore, there are data suggesting that such algorithms may enable earlier intervention and improve outcomes such as length of stay and mortality.^{14,21} However, ML algorithms are new to clinical practice and raise practical, ethical, and face validity concerns.

Given the challenge of the clinical diagnosis, the policy incentives, and the potential for ML to add value to the status quo, the market for CDS options is diverse and growing. Leaders looking to CDS for improvement in their sepsis outcomes can choose from third party ML applications, commercial vendor produced ML models and RB best practice advisory toolkits, and homegrown ML and RB models.

OBJECTIVE

The purpose of this study is to explore sepsis CDS implementation from the perspective of implementation leads by describing the motivations, tool choices, and implementation experiences of a diverse group of implementers.

METHODS

Study design and sample

This descriptive study includes semi-structured interviews and a questionnaire ([Supplementary Material S1](#)). This approach was chosen as qualitative methods can provide “rich descriptions of complex phenomena” typical of healthcare organization and IT implementation.²⁰ A convenience sample was developed of hospital leaders from various hospitals nationwide with criteria intended to include implementors of CDS tools from 3 sources: third party, commercial vendor, and homegrown. Initial sampling involved directly contacting hospitals who had implemented third party tools identified via literature review and web search of third party tools. All identified hospitals were contacted (18) of which 4 agreed to participate. Additional sites were recruited via informatics professional group email lists increasing sample size to 11 institutions. In order to attain representation of the 3 major tool categories and reach saturation,²² 4 additional hospitals were recruited through professional networks, for a total of 15. The selected institutions self-identified individuals overseeing the implementation of the tool for interview who independently agreed to participate in the study. Five out of 13 interviews were simultaneously conducted with more than 1 individual for a total of 21 participants.

Data collection and procedures

All participants were asked to complete a Qualtrics questionnaire about organizational characteristics and a semi-structured telephone interview. The interview guide (see [Supplementary Material S1](#)) was developed based on literature review and expert opinion, piloted with clinicians and qualitative experts, and modified accordingly to ensure credibility and dependability of the data collection methodology. The interview guide was developed to specifically elicit perspective of implementation leaders, not clinician end users. Interviews consisted of 30–60 minute telephone calls conducted by a single team member (MJ or LS) between June 2019 and January 2020. Participants were informed about the confidentiality of their responses, received no compensation, and provided verbal consent. This study

was considered IRB exempt by the Harvard Longwood Institutional Review Board.

Qualitative data analysis

Transcribed interviews were deidentified, imported into Excel, with each independent thought assigned its own row (Microsoft Windows 7). Data were analyzed using a thematic content analysis approach, guided by expertise of team members with similar qualitative methods.^{23–28} Two investigators (MJ and KM) independently assigned concepts for each complete thought in the first 5 transcripts and recurrent concepts were formalized into codes with operational definitions through a 2-person consensus approach. Codes were independently assigned to complete thoughts for all interviews by both coders. To ensure credibility and dependability of the analysis, the team used multiple methods including a process of debriefing among researchers, engagement with the raw data and codes, the use of reflective notes, and iterative reconciliation and adjustment of operational definitions of codes between the 2 coders. Finally, codes were grouped into emergent themes and relationships after iterative reading and discussion with the 2 other authors (LS and RR) with clinical, informatics, and qualitative expertise.

RESULTS

Characterizing the study sample

The study captured a set of institutions heterogeneous in size, location, patient population, EMR vendor, and current sepsis prediction tool (Table 1). Seven community and 8 academic hospitals were included. Three had less than 300 beds, 4 had between 300 and 500 beds, and 8 had greater than 500 beds. All but 2 served adult populations. Nine were located in the Northeast, 1 on the west coast, 5 in the Midwest, and 1 in the Southeast. Most used Epic as their EMR vendor. Other vendors included MEDITECH, VistA, and Cerner. At time of interview, 7 employed a ML tool and 8 employed a RB tool. Homegrown tools were most common (9 hospitals), followed by EMR vendor tools (4 hospitals), and third party tools (2 hospitals). The individuals interviewed as the primary leader for the implementation effort carried a wide variety of titles: 5 carried an informatics leadership title, 10 named a clinical leader, and 6 identified an individual with an executive position.

Motivation for using CDS tools to target sepsis

All interviewees cited quality improvement for sepsis patients as the primary driver for their sepsis prediction tool initiative. Several interviewees mentioned policy changes involving public reporting of data at the state and national level as well as compliance with the CMS SEP-1 bundle to be important drivers for their institutional emphasis on early detection of sepsis. A few interviewees specifically describe being motivated by inferior sepsis outcomes compared to other regional hospitals. For example, “we don’t do all that great compared to the big academic centers... trying to figure out why there’s such a difference.”

CDS was described as a “logical progression” to a “multi-pronged approach” to sepsis quality improvement. Efforts started with manual sepsis screening and reporting requirements which were subsequently automated. Participants attributed success to the alerts in concert with manual screenings, order sets, and efforts to increase awareness and communication amongst clinicians.

Implementation process

While there was significant diversity in the title of the “primary implementation leader” (Table 1), all individuals were part of large interdisciplinary teams. All sites reported teams including executive leadership, clinicians, educators, quality improvement, risk management, Information Technology (IT), and informatics. Clinician leadership with informatician support was most common, but several institutions had informatician-led teams. The 3 teams that did not have informatics involvement were community hospitals. They emphasized the importance of a connection between clinicians and IT and noted feeling challenged without specialized resources to facilitate tool implementation. Many specifically noted the importance of having strong informatics leadership. The majority also noted the importance of support from the executive leadership level.

Implementation timelines varied from 3 months to 3 years. Almost all participants noted that implementation took longer than anticipated. Those reporting implementation times extending beyond 2 years tended to have homegrown models, while most vendor supplied or third party tools had implementation times shorter than 2 years.

Most interviewees were dissatisfied with their current tool operation. A minority reported being satisfied with how the tool operates. Stakeholders that reported satisfaction with the status quo tended to occupy executive roles (Chief Medical Officer, Chief Medical Informatics Officer [CMIO]) as opposed to clinical roles.

Choosing a tool

As demonstrated in Table 1, a broad mix of tools are in use. In addition to RB or ML, the tools can be further categorized as homegrown, EMR vendor provided, or third party. Of the 15 institutions, 7 attempted an ML approach with intention for use in clinical workflow, of which 5 are using the ML tool in practice. One ran an EMR vendor supplied tool in the background and was not able to progress to clinical integration at the time of the interview. The other abandoned their integrated third party ML tool, returning to a vendor provided RB tool.

Most participants used a homegrown tool or relied on a vendor provided tool. Only 4 hospitals tried a third party tool with 2 returning to an EMR vendor provided tool and 2 continuing their third party solution. In choosing a tool, interviewees valued ease of integration, customization capability, and predictive potential. They avoided tools with added contracting, cost, and distrust (Table 2).

Implementers were mixed about their impression on whether ML tools were actually predictive with clinically meaningful specificity compared to existing SIRS-based alerts. One noted a reduction of alerts by 66% after switching from an RB approach to an ML approach suggesting improved specificity. In contrast, another noted disappointment about predictive potential:

“The tool... was supposedly predictive, but we discovered... it wasn’t predictive... it was really telling providers that they’ve met the criteria for severe sepsis which... is not really predictive because they’ve already met it. It wasn’t that you were getting it before it happened so even though they were selling it as a predictive model I’m not so convinced it was predictive”

There is a lack of consensus about whether the clinical problem of sepsis is appropriate for a CDS solution due the “nebulous” or “continuous” nature of sepsis. As one participant noted:

“Sepsis is a continuum. There is no line for sepsis. If you put a line you are not successful”

Table 1. Characteristics of participating sites

	Size	Type	Region	Adult/pediatric	EMR vendor	Tool type	Tool source	Interviewee title	Method of identification
1	<300	Community	W	Adult	Vista	ML	3rd party	Chief Medical Officer (CMO)	Tool website
2	<300	Community	NE	Adult	MEDITECH	RB	EMR	Chief Nursing Officer (CNO)	Tool website
3	>500	Community	MW	Adult	Epic	ML	EMR	1. System Epidemiologist 2. Business Intelligence Manager	Tool website
4	300–500	Academic	NE	Adult	Epic	RB	Home Grown	Associate Chief Medical Informatics Officer (CMIO)	Email list
5	>500	Community	MW	Adult	Epic	ML	EMR	Chief Medical Informatics Officer (CMIO)	Email list
6	>500	Academic	MW	Adult	Epic	ML	3rd party	Executive Director of Clinical Operations	Email list
7	>500	Academic	NE	Adult	Epic	ML	Home Grown	Senior Director Clinical Operations	Email list
8	>500	Academic	MW	Adult	Epic	ML	EMR	1. Director of Clinical Effectiveness and Informatics 2. Clinical Informatics Lead	Tool website
9	>500	Community	SE	Adult	Epic	RB	Home grown	Chief Medical Informatics Officer (CMIO)	Email list
10	300–500	Academic	NE	Pediatric	Cerner	RB	Home Grown	1. Chief Medical Informatics Officer (CMIO) 2. Director of Critical Care	Email list
11	300–500	Academic	NE	Pediatric	Epic	ML	Home Grown	Emergency Department Director of Clinical Care	Email list
12	>500	Academic	NE	Adult	Epic	RB	Home Grown	Medical Director Intensive Care Unit (ICU) ^a	Professional network
13	300–500	Community	NE	Adult	Epic	RB	Home Grown	1. Pulmonary/Critical Care Medicine Chief 2. Emergency Department Educator ^a	Professional network
14	>500	Academic	NE	Adult	Epic	RB	Home Grown	1. Emergency Physician, Clinical Content Lead for EM, Clinical Content Lead for CDS 2. Co-Director of Sepsis Task Force ^a	Professional network
15	>300	Community	NE	Adult	Epic	RB	Home Grown	Intensive Care Unit Director ^a	Professional network

Note: Size: number of beds.

Regional Abbreviations: W: West; SE: Southeast; NE: Northeast; MW: Midwest.

^aSingle Data Analyst from Clinical Informatics Team representing all 4 hospitals additionally participated.

Table 2. Factors in choosing a tool

	Factors in choice	Representative quotes
Favorable factors	Ease of integration	<p>“We implement a lot of Epic functionality it’s tightly integrated. . .there’s a little bit more work to do when we introduce. . .third-party models”</p> <p>“We had some trouble getting the information across to them. . .it was taking upwards of 10–15 seconds which in a clinical workflow is really just not okay”</p>
	Customization capability	<p>“Why did we decide to build it ourselves versus go with what’s in the EMR? The problem was it still would have taken quite a bit of lifting and we still wouldn’t have had much control over the parameters”</p> <p>“The other major problem. . .is the one size fits all nature because it is designed to be implemented by multiple different organizations they had to dumb it down. Had to normalize, smooth out the curve, sacrifice accuracy for being able to universally implement. . .So there are probably some features . . .we could have used, but they excluded because they didn’t feel confident that all Epic organizations would have that data available”</p>
	Predictive potential	<p>“Every study we saw said to identify patients sooner in order to have better outcomes because. . . earlier our ability to intervene, the better outcomes. . . So wanting to know sooner was inherent in identifying those patients at all”</p> <p>“I would say that we tried [SIRS criteria rule based surveillance] at first. . .and realized that it would fire way too frequently. It would have a huge false positive rate. In fact it fired for somewhere between 20 to 30% of all patients that were admitted to the hospital”</p>
Avoided factors	Contracting	<p>“We generally like to do things as much as possible within our EMR without involving third-party vendors”</p> <p>“You know there’s always contracting issues and a lot of components like that which are often out of scope of the clinical team to manage. . . having to get legal involvement adds steps to things. . .it was not as easy as using your own EMR”</p>
	Cost	<p>“We looked at outside solutions but we didn’t purchase. The cost was too high”</p> <p>“And because we have Epic, because there was no additional cost to implement their method, this is in all honesty, it was determined that that could be where we could start.”</p>
	Distrust	<p>“Either you purchase a program through your EMR vendor, or you try to build it yourself, or you purchase a third-party solution and hope that they are not lying to you. Or you know putting lipstick on the pig. Or you know just making it sound better. . .”</p> <p>“I think it was. . . the external one because they really were pushing artificial intelligence and the predictive model. People didn’t understand that as much and because they’re not your employees and they’re still people you’re always skeptical about what people are telling you and selling to you is very different”</p>

The problem is additionally complicated by a difficulty establishing true positives and negatives. A participant described the challenge as:

“The definition of gold standard for sepsis makes it hard. . .to apply any standard, but machine learning in particular because it is much easier. . . if you have true positives and true negatives. And then the ambiguous cases can be used for learning but often nobody really knows what to do with [the ambiguous cases]”

The challenge of establishing true positives and negatives was also noted for RB models by another participant as:

“How do you account for when the BPA fired, everyone did the right thing, and prevented this bad outcome which is the gold standard for measuring if that alert worked or not. There’s a flaw in the overall methodology that I don’t know if there is a good way to account for it”

Despite these challenges, one participant, who had successfully implemented an ML model, favored an ML approach and said that rule based heuristics are too simplistic to capture the clinical complexity of sepsis.

Workflow integration of the tool

There is considerable heterogeneity in how the tools are integrated into the workflow. Most alerts were integrated within the EMR. Both alerts appearing outside the EMR were ML. The most common clinical setting for these alerts was the floors. Many had the alert

fire in multiple locations. The most common combination was floor and ED with a few cases (4) of ED, floor, and ICU.

Most alerts targeted both nurses and physicians, while some alerted only nurses. In a minority of cases, a team external to the primary care team was informed. In one case, the external team notified the primary team, in the other case, a rapid response team was alerted in addition to the primary team. None alerted only physicians. A minority of those alerting both nurses and physicians used differential thresholds for the 2 groups.

Alerts were more commonly linked to actions than not. Most alerts not linked to action were ML models and correspondingly either third party or vendor provided. Most alerts did not incorporate a hard stop requiring clinicians to engage with the alert.

Two hospitals had double-layered logic with a sensitive alert followed by a more specific alert. Both hospitals were pediatric hospitals. About half of the hospitals had systems that included logic to suppress alerts to decrease redundant alert volume. All these systems were homegrown with the majority being RB alerts in the ED and one being ML on the floors. Most alerts were accompanied by an explanation of why the alert fired, among those that did not, most were ML.

Implementation barriers

Almost all interviewees expressed that the process for implementation was more difficult than anticipated and encouraged others to be persistent. In one interviewee’s words:

“Don’t give up. You got to just keep chugging. Sometimes it’s a lot of little steps that sometimes feel like you’re climbing a mountain that doesn’t end”

Barriers were defined as experiences that impeded, slowed, or made implementations difficult in some way. Those that emerged from the interviews clustered to technical build, optimization of alerts, workflow integration, tool validation, implementation time, working with external vendors, and clinician acceptance. These were shared by RB and ML models alike. Promoting clinician acceptance was the dominant challenge for all implementation leaders (Table 3).

Optimizing the alert to appropriately identify patients and trigger clinician response was the first major barrier to generating clinician acceptance. Optimizing the alert consisted of fine-tuning thresholds, content, and integration of the alert into the workflow. Themes focused on lack of consensus, tension between alert placement and disruption of workflow, and burden of optimization falling to individual institutions (Table 3).

The second major barrier was generating clinician buy in. The most frequently cited concern for clinician buy in for both RB and ML models was avoiding alert fatigue, or over-alerting users. A few remarked that this concern was present with all their decision support tools. Users of ML models tended to be less concerned with alert fatigue. One implementer of an ML model stated the model was chosen specifically to reduce alert fatigue and did not feel over alerted. Another felt that use of ML to combine multiple data points in the EMR reduced overall alert fatigue. Other themes focused on concerns about clinical relevance, difficulty explaining why models fire, confusion with understanding what an alert means, and distrust stemming from mismatched expectations (Table 3). Confusion was reported more by implementers of ML models than RB models. All mentions of distrust pertained to ML models.

Approaches to overcome identified implementation barriers

Hospitals employed a variety of approaches to challenges with alert firing, alert content, workflow integration, and promoting buy-in among clinicians (Table 4). Approaches were defined as ideas to overcome identified barriers and processes that were described to be going well or without problems. Most teams worked to minimize alerts through manipulating thresholds for the alerts to fire. A few employed heuristics to minimize redundant notifications, used differential thresholds for different provider types, or had a two-phase alert system with a sensitive alert followed by one more specific. Concise content with explanations for firing was well-received. A few noted that this was not possible or was more confusing, but most felt inclusion of explanations was helpful. From a workflow perspective, most systems did not incorporate hard stops. The only one that did reported a negative experience. A few reported using differential workflows to use the tool in different settings and noted that actionability specifically enabling placement of orders that have not yet been placed would be helpful.

To improve clinician buy in, implementers reported garnering support by showing outcomes data to clinicians, providing direct feedback to teams, offering support at the point of care to interpret alerts, emphasizing ongoing multimodal user education, using relatable metaphors to make tool outputs more intuitive for users, incorporating front line practitioners onto implementation teams, and managing expectations (Table 4).

DISCUSSION

In response to national and institutional quality improvement priorities, institutions ranging from small community hospitals to large academic healthcare systems nationwide have turned to CDS to improve care for patients with sepsis. Our study describes the perspective of implementation leaders for CDS tools that were either RB or ML-based. In choosing tools, implementers tried to maximize ease of integration, customization capability, and predictive potential while minimizing contracting, cost, and distrust. Implementation efforts were found to be large and heterogeneous undertakings requiring significant activation energy and sustained commitment from interdisciplinary teams, and almost universally reveal significant implementation barriers and dissatisfaction with CDS tools.

Clinician acceptance of CDS tools was difficult to achieve for both ML and RB models. Both faced barriers with alert optimization and clinician buy in. Barriers to alert optimization included lack of consensus for what optimal means, a tension between alerting appropriately and disrupting workflow, and having to reinvent the wheel of optimization decisions on the institutional level. With regards to buy in, for RB models, barriers centered mostly on minimizing alert fatigue. In addition to alert fatigue, ML models carried additional challenges around clinical relevance, difficult explanations, confusing outputs, and expectation management. To improve clinician buy in, interviewees worked to garner support with outcomes data, feedback to teams, accessible in-person alert interpretation support, ongoing education, and managing user expectations.

Existing studies have not clearly demonstrated improvement in outcomes with these tools.^{2,8–12,14,21} Some randomized studies support reduction in length of stay and in hospital mortality,^{21,29} while others have shown no significant change in clinical outcomes.^{2,8} There is however prospective data to suggest that these alerts may lead to faster interventions such as time to blood culture or antibiotics.^{9–12,14} If compliance with sepsis quality metrics is the key driver of these implementations, there may be data to support efficacy; however, strong data supporting change in patient outcomes are yet lacking. Of note, presence or absence of supportive data did not seem to prominently factor into decision making about implementation by implementation leaders in our study.

It is perhaps not surprising that clinicians are hesitant to accept these tools, as the data support an experience of too many alerts without significant value. Consistent with our findings, existing literature specifically surveying clinician end users describes that alerts do not change perception of patient’s risk and alerts vary in their ability to alter management.^{30,31} Furthermore, a recent large validation cohort study focusing on characterizing alert fatigue around the Epic sepsis detection algorithm reported low sensitivity and many missed cases of sepsis despite generation of a large number of alerts.³²

However, the argument can be made that further evaluation is needed and that variability in results may be related to variability between hospitals and differential workflow creation.³³ One study, particularly highlighting significant change management effort and delivery of alerts via a mobile application in addition to implementation of an algorithm, did demonstrate 53% decrease in mortality.¹³ Another study, conducted at 21 hospitals involving 374 838 patients, demonstrated lower mortality, a lower incidence of ICU admission, and a shorter length of hospital stay using a full rapid response intervention program built around a validated deterioration prediction model.³⁴ This study emphasized an approach using remote monitoring by a dedicated nursing team specifically designed

Table 3. Barriers to promoting clinician acceptance

Major barrier	Theme	Representative quotes
Optimizing the alert	<u>No consensus for what optimal means:</u> There is no clear consensus for setting thresholds, what to include in the alert, or how to tie alerts to actions.	“It was just a gut decision made by our sepsis team on like how many patients are we comfortable being correct on and incorrect on”
	<u>Drawing attention without being disruptive:</u> There is tension between placing the alert in the workflow such that it prompts action but is not disruptive to the workflow.	“I don’t think any of them are totally plug and play. That play is going to depend on a lot of other factors” “One is. . .the tension that people have to respond to it but also is it isn’t invasive enough that it disrupts people’s workflows”
	<u>Reinventing the wheel:</u> Sites spend considerable effort on the institutional level optimizing these features.	“One of the things that we struggled with is that, you can’t really close the chart if you have a BPA fired that’s open. And that was a real nuisance to a lot of people” “They do not have any model builds that says you should do this. . . and here’s the alerts you can build. We’ve determined all of that. . .there were not recommendations from Epic in that regard. Those were all decided at an institutional level” “You can then surface that information up anyway you want, displaying information or a column on a patient list or as an alert”
Clinician buy in	<u>Alert Fatigue:</u> Trying to avoid overalerting clinician users	“It’s a challenge if you overwhelm providers with warnings then they’ll ignore them all. So many alerts are false positives but you don’t want a lot of misses so we’re trying to find the correct balance right now”
	<u>Concerns about clinical relevance:</u> Clinical endpoints are important, endpoints are limited by what data are available and people are skeptical of billing based codes	“if you are going to design a screening tool, basically by definition you are going to get a lot of false positive alerts. So we were concerned that that could lead to alert fatigue and that you know it would be driving everyone crazy by having them run around for false positive alerts” “Doctors drop codes at any time during the admission. . .the four hours before someone drops a code, I don’t know if that’s going to help me. . .even if I bought the model, and I agreed with it, I’m not sure how you implement it clinically”
	<u>Difficult to explain:</u> ML models are confusing for clinicians because they are difficult, sometimes impossible, to explain why the system fired.	“People often had clinical ideas for. . .what would be helpful in terms of detection but translating that to actual numbers or data points that can be interpreted was a big challenge” “Knowing that there’s 127+ rules that contribute, it’s not as easy to say these are the things and so we’ve made some changes to try to make that a little more visible in our alerts”
	<u>Confusing to understand:</u> The outputs of ML models are not clinically intuitive.	“The third party vendor would never actually identify to the provider what they saw in the record that made the patient be warned for severe sepsis. You couldn’t give any clinical information. They would just say the third-party vendors review the record and the patient is at risk for severe sepsis. There was no other information that they would give us nor did we have the algorithm they were working on” “I think the hardest part about a predictive model is not specific to sepsis, but understanding that are predictive model is really a forecast” “A lot of people get confused. . .so say you get 25, when the patient’s really sick and then the number goes to twenty, does that mean the patients getting better? What do all of the subsequent numbers mean? If it goes up to 30, is the patient getting worse? So a lot of clinicians who looked at this model thought that that number is some kind of measure of patient clinical status and in fact it has nothing to do with that and the model completely breaks down after that first time you get the score because you can get new data points that come in and I don’t even know what the

(continued)

Table 3. continued

Major barrier	Theme	Representative quotes
		score means after that first time. So that's another major issue with the model. we don't know what the numbers mean in a longitudinal fashion"
	<u>Mismatched expectations</u> : Sites are challenged with losing trust and buy-in for the tool when it does not match the clinician's expectations	<p>"When you bring a bunch of doctors in the room and explain to them the model, they start interpreting the model in the way they want it to work, rather than the way it actually works. You can explain it until you're blue in the face but that's not how the model was built the model can do this, you know it can do A but it can't do B, C and D. They still, they're stuck in the way they want it to work"</p> <p>"The clinician has a high expectation that this alert is going off for patients who have sepsis, and that is just not the case. It is going off for patients who are at risk for sepsis, many of whom will not have sepsis. An alarm went off for a patient that is clearly not septic, that has a GI bleed so to get clinicians to buy into that concept of being alerted for patients who are at risk didn't really seem to work"</p>

to shield practicing clinicians from alert fatigue. Together, these studies suggest that effective integration, change management, and strategies to improve acceptance, buy in, and trust may be the factors limiting a demonstration of benefit of these tools. While there are not many studies emphasizing workflow design, it is compelling that these are the studies that demonstrate improvement in mortality as an outcome. More such studies, especially with prospective controlled designs, are needed to understand whether effective change management that improves buy in and trust could make these tools effective for clinically meaningful outcomes.

"Meaningful decision support" and "explainability" have been described as 2 key implementation science barriers specific to ML models, which carried additional buy in challenges in our study.³⁵ In order for decision support to be "meaningful," users need to first trust individual predictions enough to act on them.³⁶ Our study suggests that over-alerting and excessive false positives may detract from trust and clinical meaningfulness of both RB and ML models. However, our study also shows greater distrust of ML tools because they come from third parties and rely on billing based rather than clinical inputs. Furthermore, because they do not draw on criteria that are taught in clinical training like their RB counterparts, they are confusing and non-intuitive to the clinical user. Strategies focusing on user education, accessible alert interpretation support, and management of expectations may help promote clinician acceptance because they attempt to address this distrust and confusion.

With regards to explainability, it is well established in the literature that "black boxes" of ML in clinical medicine are difficult for clinicians to accept and that demystifying the "black box" is critical for establishing trust in the model.^{36,37} RB models had more straightforward explanations since the rules are derived from expert clinicians. However, for ML models, efforts to provide explanations sometimes resulted in greater confusion and sometimes were not possible at all. A growing body of work in ML literature explores the components of effective explanations of models so that they can be understood more easily by users who do not have in-depth ML backgrounds.^{36,38-41} Further work is needed in this area to understand and disseminate effective practices to help implementers explain models.

Considerations for implementation

- Alerts should be credible and not excessive. ML models offer the advantage of improved specificity if they can be implemented effectively.
- Effective implementation requires clinically relevant models whose outputs are easily understood by clinical users.
- User education and support need to be prioritized with emphasis on how to interpret the output of a predictive model, building trust, and user expectation management.
- Burden of re-creating effective solutions would be reduced with guidance from tool developers on how to integrate tools into workflows and how to educate and support users.
- Implementing these tools requiring significant time, energy, and as a result, cost to institutions. Outcomes of implementations should be studied to understand whether these significantly resource intensive implementations are worthwhile and to establish comparisons for successful implementation approaches.

This study has several limitations. Its generalizability is potentially limited by selection bias from limited sample size and convenience sampling. The surveyed group is diverse but not clearly representative of American hospitals. While the heterogeneity of the sampled institutions does not overcome possible selection bias, it provides insight that predictive analytics are being employed for quality improvement even in small community hospitals without academic informatics departments. Furthermore, intentionally representing each tool type allowed exploration of tool choice and difference in ease of implementation based on tool type.

The initial group had selection bias given initial response rate of 22%. By augmenting recruitment through professional networks, the remainder of the sample was biased to institutions with informatician involvement. Institutions contacted via third party application websites introduced community perspective to the discussion. While all perspectives that may have been excluded due to non-response and selection bias cannot be known, the themes and examples reported, especially with regards to alert fatigue and clinician buy in, were consistent throughout the interviews, and likely reflective of the obstacles faced by many institutions even if not generalizable to

Table 4. Approaches to overcome identified implementation barriers

Implementation barrier	Approach	Representative quote
Minimize alert firing	Threshold optimization	“We did a lot of testing to see at what threshold could we have the minimum number of alerts. . . we are very sensitive to alert fatigue”
	Heuristics to reduce redundant alerts	“If we alert the rapid response doctors, we won’t alert them again for the next 8 hours. Because we don’t want to be continuously sending the same alert”
	Different thresholds for different provider types	“We have an upper and lower kind of threshold, and at a lower threshold we alert the frontline team. So that would be the front line nurse and the front line provider. And at the upper threshold the plan we actually text, the platform will actually text our rapid response providers”
	Two-phase alerts	“We came up with a model that incorporates vital signs, past medical history, certain high risk factors, high risk neurological conditions, presence of a central line, sickle cell, some other things to develop an initial screening alert that’s targeting the inpatient nurse that is largely vital sign driven and then based on follow up assessments that they document and also presence or absence of some of those high risk conditions, a secondary alert would appear to the entire team”
Alert content	Concise alert messages	“Keep it as simple as possible. . . doctors and nurses are inundated by alerts all the time. If you expect them to read it, it is not going to happen. The alert needs to be very straightforward and specific”
	De-emphasize wordsmithing of alert	“The more clear you can be with that message the better but like changing the tense of a verb here or doing this or doing that doesn’t make any bit of a difference. I mean we’ve looked at the amount of time people spend in these alerts and it’s like a fraction of a second so it is not long enough to even notice a typo”
	Include explanations when possible	“I think it’s important for users to know why this alert went off. Now when we get an alert and it says. . . some indication for why this alert went off. I think that actually reduced the amount of negative feedback that we were getting”
Workflow integration	Avoid hard stops	“I think that having some acknowledgement reason [that] captures whether you agree or disagree with the alert is a bad thing”
	Ability to place orders that have not been placed	“I think at the time one of the draws was the ability to place orders. . . as a follow-up so if I was missing something [the tool] could say hey you are missing you know a second lactate and here is the order to place”
	Use different alerts for different locations	“Many hospitals decide to take two workflows. One for the ED and one for inpatients. This model requires that data is in place in order to make the prediction. Like, you know lab results, flowsheet values, medications. . . if there are no lab tests or medications you know for that patient, it’s not going to predict very well. So you know talking to Epic, they stated that many hospitals chose to take a two branch approach to the prediction”
Clinician buy-in	Garner support with data	“Just showing people data of how often it fires, who it fires for, where the false positives are, and giving them visual patterns of how is succeeding or failing is a powerful tool”
	Direct feedback to teams	“We have demonstrated that direct feedback to the clinicians certainly results in higher compliance with antibiotics and bundle elements”
	Point of care clinical support	“We created a resource through the virtual care team that allowed nursing staff, provider staff to call anytime 24/7. . . you tell them this is my number, what does that mean? And we would say it is just a number, let’s look at everything that went into it, let’s talk about it and then let’s talk about what that means for what we need to do for our patient”
	Emphasis on ongoing multimodal user education	“I think you need to approach education from a couple of angles, because there’s different folks who learn in different ways. You need a video, you need a PowerPoint, it needs to be referenceable, there needs to be frontline people who go out and support units”
	Use of metaphors and analogies to address intuitiveness of tool output	“[We created a video] comparing predictive models to a weather forecast. It doesn’t mean you’re going to put the rainboots on now because it’s not raining right now”

(continued)

Table 4. continued

Implementation barrier	Approach	Representative quote
Incorporating frontline practitioners onto implementation teams		“I think the fact that we as the clinical effectiveness team are clinicians, I think really helps”
Managing expectations		“[We] have to manage expectations that we are not yet at a point where these rules are going to be able to define sepsis without help from humans. . .”

institutions who may not have responded because of differences in satisfaction or frustration with their implementations.

Our study notably focuses on the perspective of implementation leaders, making it particularly relevant to leaders such as CMIOs at medium to large scale hospitals considering CDS software as a component of their hospital's sepsis strategy. However, we learned hospitals with smaller IT infrastructure are also implementing both types of tools. They are struggling with the same issues of clinician buy-in, alert optimization, distrust, and confusion with even wider gaps between clinician and IT leaders without always having informaticians to serve as a bridge. Better dissemination of helpful approaches to address these common issues is of perhaps even greater importance to these smaller institutions. Implementation leaders at these settings are motivated by the same quality metrics as larger, more resourced settings, but may be less equipped to generate institution level solutions to the numerous barriers we have described.

There is much work to be done to facilitate implementations and share successful strategies. While several leaders interviewed in this study also had clinical roles, they were interviewed for their perspective as implementation leaders. What is perceived as lack of clinician acceptance and trust may be rooted in reasonable concerns about benefit to patient care which can be further characterized with future study from front line user perspective. Additional study should also elicit vendor perspectives about improving user experience, support, and education. Other next steps are to propose a framework characterizing elements contributing to confusion, distrust, and expectation mismatch that institutions can use to develop user support and education.

CONCLUSION

In this small but diverse set of hospitals, we find broad heterogeneity in institutional application of CDS to improve sepsis outcomes. Implementation of all tools was time consuming and complicated, with the job of making tools clinically useful falling largely to individual institutions. While both RB and ML models posed significant challenges to optimization of the alert and integration into the workflow, ML models posed additional barriers to clinical meaningfulness and acceptance due to issues of confusion, distrust, and expectation mismatch. Attention to user education, alert interpretation support, and expectation management and dissemination of effective practices related to these areas may improve feasibility and effectiveness of ML models being used in quality improvement efforts.

FUNDING

LS is funded under R01 DK116898 from National Institutes of Health (NIH).

AUTHOR CONTRIBUTIONS

LS and MJ were responsible for study conception, design, acquiring approvals, and recruitment. Data were collected by MJ and LS, transcribed by MJ, and analyzed by KM and MJ with guidance and oversight from RR and LS. All authors were involved in interpretation of data, drafting and revising of final manuscript, and approving the final version to be published. All authors agree to be accountable to all aspects of the work in ensuring that questions related to accuracy and integrity are appropriately investigated and resolved.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *JAMIA Open* online.

CONFLICT OF INTEREST STATEMENT

None declared.

DATA AVAILABILITY

The data underlying this article cannot be shared publicly for the privacy of the individuals who participated in the study. Data may be shared upon reasonable request to the corresponding author.

REFERENCES

- Rhee C, Dantes R, Epstein L, *et al.*; CDC Prevention Epicenter Program. Incidence and trends of sepsis in US hospitals using clinical vs claims data, 2009–2014. *JAMA* 2017; 318 (13): 1241–9.
- Semler MW, Weavind L, Hooper MH, *et al.* An electronic tool for the evaluation and treatment of sepsis in the ICU: a randomized controlled trial. *Crit Care Med* 2015; 43 (8): 1595–602.
- Klompas M, Rhee C. The CMS sepsis mandate: right disease, wrong measure. *Ann Intern Med* 2016; 165 (7): 517–9.
- Walkey AJ, Lindenauer PK. Keeping it simple in sepsis measures. *J Hosp Med* 2017; 12 (12): 1019–20.
- Nguyen SQ, Mwakalindile E, Booth JS, *et al.* Automated electronic medical record sepsis detection in the emergency department. *PeerJ* 2014; 2: e343.
- Makam AN, Nguyen OK, Auerbach AD. Diagnostic accuracy and effectiveness of automated electronic sepsis alert systems: a systematic review. *J Hosp Med* 2015; 10 (6): 396–402.
- Taylor RA, Pare JR, Venkatesh AK, *et al.* Prediction of in-hospital mortality in emergency department patients with sepsis: a local big data-driven, machine learning approach. *Acad Emerg Med* 2016; 23 (3): 269–78.
- Downing NL, Rolnick J, Poole SF, *et al.* Electronic health record-based clinical decision support alert for severe sepsis: a randomised evaluation. *BMJ Qual Saf* 2019; 28 (9): 762–8.
- Nelson JL, Smith BL, Jared JD, Younger JG. Prospective trial of real-time electronic surveillance to expedite early care of severe sepsis. *Ann Emerg Med* 2011; 57 (5): 500–4.

10. Sawyer AM, Deal EN, Labelle AJ, *et al.* Implementation of a real-time computerized sepsis alert in nonintensive care unit patients. *Crit Care Med* 2011; 39 (3): 469–73.
11. Umscheid CA, Betesh J, VanZandbergen C, *et al.* Development, implementation, and impact of an automated early warning and response system for sepsis. *J Hosp Med* 2015; 10 (1): 26–31.
12. Narayanan N, Gross AK, Pintens M, Fee C, MacDougall C. Effect of an electronic medical record alert for severe sepsis among ED patients. *Am J Emerg Med* 2016; 34 (2): 185–8.
13. Manaktala S, Claypool SR. Evaluating the impact of a computerized surveillance algorithm and decision support system on sepsis mortality. *J Am Med Inform Assoc* 2017; 24 (1): 88–95.
14. Giannini HM, Ginestra JC, Chivers C, *et al.* A machine learning algorithm to predict severe sepsis and septic shock: development, implementation, and impact on clinical practice. *Crit Care Med* 2019; 47 (11): 1485–92.
15. Paranjape K, Schinkel M, Nannan Panday R, Car J, Nanayakkara P. Introducing artificial intelligence training in medical education. *JMIR Med Educ* 2019; 5 (2): e16048.
16. Seymour CW, Liu VX, Iwashyna TJ, *et al.* Assessment of clinical criteria for sepsis for the third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA* 2016; 315 (8): 762–74.
17. Islam MM, Nasrin T, Walther BA, Wu CC, Yang HC, Li YC. Prediction of sepsis patients using machine learning approach: a meta-analysis. *Comput Methods Programs Biomed* 2019; 170: 1–9.
18. Desautels T, Calvert J, Hoffman J, *et al.* Prediction of sepsis in the intensive care unit with minimal electronic health record data: a machine learning approach. *JMIR Med Inform* 2016; 4 (3): e28.
19. Schinkel M, Paranjape K, Nannan Panday RS, Skyttberg N, Nanayakkara PWB. Clinical applications of artificial intelligence in sepsis: a narrative review. *Comput Biol Med* 2019; 115: 103488.
20. Fleuren LM, Klausch TLT, Zwager CL, *et al.* Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive Care Med* 2020; 46 (3): 383–400.
21. Shimabukuro DW, Barton CW, Feldman MD, Mataraso SJ, Das R. Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: a randomised clinical trial. *BMJ Open Respir Res* 2017; 4 (1): e000234.
22. Saunders B, Sim J, Kingstone T, *et al.* Saturation in qualitative research: exploring its conceptualization and operationalization. *Qual Quant* 2018; 52 (4): 1893–907.
23. Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis: striving to meet the trustworthiness criteria. *Int J Qual Methods* 2017; 16 (1): 160940691773384.
24. Samal L, Dykes PC, Greenberg J, *et al.* The current capabilities of health information technology to support care transitions. *AMIA Annu Symp Proc* 2013; 2013: 1231.
25. Dykes PC, Samal L, Donahue M, *et al.* A patient-centered longitudinal care plan: vision versus reality. *J Am Med Inform Assoc* 2014; 21 (6): 1082–90.
26. Samal L, Dykes PC, Greenberg JO, *et al.* Care coordination gaps due to lack of interoperability in the United States: a qualitative study and literature review. *BMC Health Serv Res* 2016; 16 (1): 1–8.
27. Rozenblum R, Jang Y, Zimlichman E, *et al.* A qualitative study of Canada's experience with the implementation of electronic health information technology. *CMAJ* 2011; 183 (5): 281–8.
28. Wells S, Rozenblum R, Park A, Dunn M, Bates DW. Organizational strategies for promoting patient and provider uptake of personal health records. *J Am Med Inform Assoc* 2015; 22 (1): 213–22.
29. Kollef 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): 424–9.
30. Guidi JL, Clark K, Upton MT, *et al.* Clinician perception of the effectiveness of an automated early warning and response system for sepsis in an academic medical center. *Ann Am Thorac Soc* 2015; 12 (10): 1514–9.
31. Ginestra JC, Giannini HM, Schweickert WD, *et al.* Clinician perception of a machine learning-based early warning system designed to predict severe sepsis and septic shock. *Crit Care Med* 2019; 47 (11): 1477–84.
32. Wong A, Otles E, Donnelly JP, *et al.* External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Intern Med* 2021; 181 (8): 1065–70.
33. Habib AR, Lin AL, Grant RW. The epic sepsis model falls short—the importance of external validation. *JAMA Intern Med* 2021; 181 (8): 1040–1.
34. 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): 1951–60.
35. Shaw J, Rudzicz F, Jamieson T, Goldfarb A. Artificial intelligence and the implementation challenge. *J Med Internet Res* 2019; 21 (7): e13659.
36. Tulio Ribeiro M, Singh S, Guestrin C. “Why Should I Trust You?”: explaining the predictions of any classifier. *arXiv e-prints*. February 2016:arXiv:1602.04938. <https://ui.adsabs.harvard.edu/abs/2016arXiv160204938T>.
37. Braunstein ML. *Health Informatics on FHIR: How HL7's New API is Transforming Healthcare*. Springer; 2018. doi: 10.1007/978-3-319-93414-3.
38. Lundberg SM, Nair B, Vavilala MS, *et al.* Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng* 2018; 2 (10): 749–60.
39. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: Guyon I, Luxburg UV, Bengio S, *et al.*, eds. *Advances in Neural Information Processing Systems 30*. Long Beach, CA: Curran Associates, Inc.; 2017: 4765–74. <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf>.
40. Tajgardo M, Samayamuthu MJ, Calzoni L, Visweswaran S. Patient-specific explanations for predictions of clinical outcomes. *ACI Open* 2019; 03 (02): e88–e97.
41. Caruana R, Kangerloo H, Dionisio JD, Sinha U, Johnson D. Case-based explanation of non-case-based learning methods. *Proceedings AMIA Symp* 1999; 212–5.