Title
Alert Driven Rescue: Do In-Hospital Sepsis Interventions Following an Advanced Early Warning System Alert Differ Substantially Between Decedents and Survivors?

Author
Linnen, Daniel Theodore

Publication Date
2018

Peer reviewed|Thesis/dissertation

Alert Driven Rescue: Do In-Hospital Sepsis Interventions Following an Advanced Early Warning System Alert Differ Substantially Between Decedents and Survivors?

by

Daniel T. Linnen, PhD(c), MS, RN-BC

DISSERTATION

Submitted in partial satisfaction of the requirements for degree of

DOCTOR OF PHILOSOPHY

in

Nursing

in the

[redacted]

Dedication

To Katherine Louise Ruff, a strong nurse with a big heart.

† September 27, 2017

Acknowledgements

I climbed Mount PhD with the help and encouragement of many. I thank my wife, Karin Louise Linnen, and our children, Helena Sofia Linnen and Hendrik Lukas Linnen, for their love. Throughout this journey, they reminded me what matters in life. I thank Kenneth and Katherine Ruff for being family.

At UCSF, many brilliant and highly accomplished scientists supported my training and development and generously shared their knowledge and expertise. I am completing this final milestone with deep gratitude to my faculty advisor, mentor, and dissertation committee chair Dr. Caroline Stephens, PhD, RN, GNP; my qualifying exam chair, Dr. Xiao Hu, PhD; the former and current Deans of the School of Nursing, Dr. David Vlahov, PhD, RN, FAAN and Dr. Catherine Gilliss, PhD, RN, FAAN; the faculty of the School of Nursing, particularly Dr. Linda Franck, PhD, RN for her mentorship and encouragement; and the faculty of the Training in Clinical Research Program in the School of Medicine's Department of Epidemiology & Biostatistics. Substantial funding by the School of Nursing's Maribelle & Stephen Leavitt Scholarship and the Jonas Scholars program allowed me to complete my degree in three years.

As Predoctoral Research Fellow, I also received generous funding support from the Kaiser Permanente Nurse Scholars Academy and research and administrative support from the Kaiser Permanente Division of Research in Northern California. My sincere thanks go to my mentor and teacher Dr. Gabriel Escobar, MD, Director of the Systems Research Initiative at the KP Division of Research; Dr. Marilyn Chow, PhD, RN, former Vice President of National Patient Care Services; Dr. Theresa Brodrick, PhD, RN, Regional Chief Nurse Executive and Vice President; Dr. Jim D'Alfonso, DNP,

RN, Regional Executive Director of the Nurse Scholars Academy; Dr. Tracy Lieu, MD, Director of the Division of Research; Dr. Lisa Herrinton, PhD, Senior Research Scientist at the Division of Research; Dr. Patricia Kipnis, PhD, Principal Statistician at the Division of Research; John Greene, MA, Senior Programmer at the Division of Research; Ms. Jill Pope, Senior Technical Writer and Editor at the Center for Health Research in Portland, Oregon; Dr. Vinnie Liu, MD, Research Scientist at the Division of Research and Director of Hospital Advanced Analytics in Northern California; Catherine Dower, JD, National Director of Nursing Research and Policy; Dr. Dan Weberg, PhD, RN, National Senior Director of Innovation and Leadership; Ms. Linda Leavell, National Executive Director of Patient Care Services; Dr. Gretchen Summer, PhD, RN, Regional Director of Nursing Research; Dr. Elizabeth Scruth, PhD, RN, Clinical Practice Consultant, Quality & Safety in Northern California; Ms. Julie Vilardi, MS, RN, former National Executive Director of Informatics and Innovation; and Mr. Hunter Burgoon, MS, RN, former National Director of National Biomedical Informatics.

" Zwei Dinge sind zu unserer Arbeit nötig: Unermüdliche Ausdauer und die Bereitschaft, etwas, in das man viel Zeit und Arbeit gesteckt hat, wieder wegzuwerfen."

- Albert Einstein

Abstract

Significance: Patients with worsening sepsis on general hospital wards are at high risk for clinical deterioration, unplanned admission to the intensive care unit, and death. Many hospitals have begun to employ early warning systems (EWSs) to alert nurses and providers that a patient is predicted to deteriorate. Multiple papers in the rapid response team (RRT) and sepsis literature describe typical problems leading to an EWS alert (e.g., systemic inflammatory response syndrome, low blood pressure) as well recommended interventions such as intravenous fluid bolus, antibiotics, and transfer to the intensive care unit (ICU). Despite the evidence base, it remains unclear which actions following an EWS alert might improve 30-day survival.

Methods: 1) We performed a systematic review of the evidence of advanced early warning systems detecting patient deterioration risk using multivariate regression or machine learning vs. point-score systems. We then systematically quantified results of model performance (e.g., area under the curve, sensitivity, PPV) and alerts generated per positive case. 2) We conducted two rounds of clinical chart reviews evaluating patient characteristics (e.g., severity of illness, comorbidities), clinical notes and process markers of early sepsis care following a clinical deterioration alert including Do Not Resuscitate (DNR) order time; fluid bolus therapy; new antibiotics; and transfer to the intensive care unit. 3) Using a retrospective matched pair cohort design, we evaluated the impact of sepsis interventions following a clinical deterioration alert on sepsis survival in patients who were admitted in stable condition to general medical wards of Kaiser Permanente hospitals with an advanced EWS. Using a pool of hospitalized patients, we investigated whether specific fluid bolus processes (Time from EWS alert

to fluid bolus administration and total 24h fluid bolus volume) occurred more frequently in survivors.

Results: 1) Advanced EWSs using multivariate regression or machine learning had better prognostic accuracy than point-score EWSs and decreased the RRT and hospitalist evaluation workload substantially. 2) The advanced EWS alert frequently occurred within hours after hospital admission, requiring exclusion of the time period with therapeutic overlap with the initial bundle of sepsis interventions. DNR order change occurred frequently before death, making it an unsuitable exclusion criterion for "expected death" in hospital populations. 3) More sepsis survivors received additional antibiotics, and often before the alert. Decedents received more than twice as much fluid bolus therapy following the alert; had more vital sign documentations and laboratory orders following the alert; more transfers to ICU; and more DNR or comfort care orders following the alert. Some proportion of decedents may have been on a fixed end-of-life trajectory.

Discussion: This dissertation offers a novel approach to characterizing and measuring the impact of fluid bolus therapy on sepsis survival and has the potential to improve outcomes among sepsis patients outside the ICU. Early additional antibiotic coverage may aid survival. Fluid bolus therapy does not appear to aid survival. Measuring expected vs. unexpected mortality in future research may offer additional insights into the treatments effects of sepsis interventions relative to the alert.

Table of Contents

Table of Contents (continued)

List of Tables

List of Tables (continued)

List of Figures

List of Figures (continued)

Chapter 5: Conclusion

Chapter 1

An Introduction to Alert-Driven Rescue

Statement of the Problem

Hospitalized patients with worsening sepsis on general wards are at high risk for clinical deterioration, unplanned admission to the intensive care unit, and death (Angus et al., 2001; Rivers et al., 2001a; Rivers et al., 2001b; Martin, Mannino, Eaton, & Moss, 2003; Dombrovskiy, Martin, Sunderram, & Paz, 2007; Gaieski et al., 2010; O'Neill, Morales, & Jule, 2012; Pastores, Dakwar, & Halpern, 2012; Liu, Morehouse, Soule, Whippy, & Escobar, 2013; Institute of Medicine, 2015; Park et al., 2017; Prasad et al., 2017). An estimated 900,000 to 3 million patients are diagnosed with sepsis in the United States annually (Gaieski, Edwards, Kallan, & Carr, 2013), and an estimated 180,000 to 750,000 patients die of sepsis each year (Gaieski et al., 2013; Epstein, Dantes, Magill, & Fiore, 2016). Evidence suggests that over 50% of all patient deaths following hospitalization may be due to sepsis (Engel et al., 2007; Liu et al., 2014). Sepsis costs the U.S. health system an estimated $14 billion annually (Healthcare Cost Utilization Project, 2008). For health care delivery systems, unrecognized worsening sepsis may result in higher treatment costs and longer stays with net loss due to bundled payments in the diagnosis-related group (DRG) payment model.

New or worsening sepsis is defined as signs of systemic inflammatory response plus acute organ dysfunction (Levy et al., 2003). Prior studies suggest that patients who receive early sepsis interventions have better survival outcomes (Levy et al., 2003; Dellinger et al., 2008; Liu, Whippy, & Morehouse, 2015). For example, the *2012 Surviving Sepsis* guidelines recommend fluid resuscitation of a minimum of 30ml/kg of crystalloid intravenous fluid completed within 3 hours and to consider additional fluid bolus therapy until hemodynamic stability is reached. In deteriorating sepsis patients,

the goal of timely fluid bolus therapy is to replenish volume because in the pathophysiological cellular inflammatory response of sepsis, volume from the blood vessels moves into the surrounding tissue (Dellinger et al., 2013; Glassford, Eastwood, & Bellomo, 2014). Multiple papers in the sepsis literature describe both typical problems of worsening sepsis (e.g., systemic inflammatory response syndrome and acute organ dysfunction such as respiratory distress or low blood pressure), as well as what should be done (e.g., fluid bolus, intravenous antibiotics, transfer to ICU) (Levy et al., 2003; Liu et al., 2013; Kramer, Cooke, Liu, Miller III, & Iwashyna, 2015; Fielding-Singh, Greene, Baker, Escobar, & Liu, 2016). It is known that the timing and volume of fluid bolus may be associated with better odds of surviving sepsis (Levy et al., 2003; Liu et al., 2013; Liu et al., 2014). Consequently, the Centers for Medicare and Medicaid Services core measure SEP-1 mandates the reporting of sepsis bundle compliance, also referred to as Early Goal Directed Therapy (Pepper et al., 2018). The measurement of sepsis bundle compliance includes two fluid bolus processes (FBPs): Time from presentation of sepsis to administration and total volume administered (Rhodes et al., 2017). These measures do not capture how hospitals facilitate the timely administration of fluid bolus therapy in patients with worsening sepsis, and how decompensation risk is identified. Both of these factors are crucially important for the effective delivery of sepsis care and will be discussed in the following paragraph.

Though not federally mandated, hospitals typically invest in clinical resources, such as rapid response teams (RRTs), to deliver the sepsis bundle on inpatient wards in a timely manner. Studies have documented that patients who experience worsening sepsis outside the intensive care unit (ICU) have worse outcomes than patients directly

admitted to intensive care. In acute care, there are two sepsis response options for patients facing clinical deterioration from worsening sepsis:

1. Rapid response team (RRT) activation: Typically staffed at minimum with an experienced critical care nurse, RRTs evaluate patients at-risk for worsening sepsis using a variety of strategies (e.g., proactive rounding or evaluating patients flagged by screening tools).
2. Code sepsis: A critical care team that includes RRT, physician and pharmacy responds to a positive sepsis screen indicating acute organ dysfunction. The team then delivers a bundle of sepsis interventions at optimal speed.

The evidence regarding the value of RRTs in averting clinical deterioration and unforeseen death is weak. Traditional RRT approaches are problematic because interventions may often not begin until a patient is already demonstrating an acutely worsening condition. The ongoing problem is not only failure to rescue sepsis patients who deteriorate on general hospital wards, but also failure to recognize early and scattered clues of worsening sepsis, so that clinical deterioration may be prevented entirely.

To that end, there is potential for early warning systems (EWSs) to support RRTs by identifying risk and by triggering sepsis evaluations and interventions earlier than traditionally possible. (Escobar et al., 2012b; Bates, Saria, Ohno-Machado, Shah, & Escobar, 2014; Linnen, 2016). EWSs are known under several other terms, including “track and trigger systems” (Smith, Prytherch, Meredith, Schmidt, & Featherstone, 2013) and “sniffers” (Herasevich, Pieper, Pulido, & Gajic, 2011). Historically, EWSs were paper-based point score instruments designed for fast manual calculation. Clinical

variables were typically limited to vital sign data (heart rate, respiratory rate, systolic blood pressure, SpO2, temperature), presence of oxygen support and level of consciousness. With point-score EWSs, most of the literature regarding their ability to improve survival has been inconclusive given their limited precision and because they may produce a large number of false positive alerts (Gao et al., 2007; McNeill & Bryden, 2013; Smith et al., 2014).

Alternatively, advanced EWSs, which we define as those using multivariate regression models or machine learning algorithms, use more complex data and more sophisticated analytic methods to quantify deterioration risk. Using modern data technology, these EWSs can compute risk scores in near real-time, and they can detect more subtle signs and trends of worsening sepsis and clinical deterioration. Research has demonstrated better discrimination and calibration performance in detecting patient deterioration risk up to 12 hours before a potential acute event (Escobar et al., 2012a; Escobar & Dellinger, 2016; Granich et al., 2016). However, despite knowing that advanced EWSs can facilitate a timelier identification of risk, it is not clear whether they also facilitate a better RRT response and whether such a response would improve sepsis survival. To date, no sepsis studies have investigated whether (or which) fluid bolus processes may improve sepsis survival relative to an advanced EWS alert. Such a gap is concerning given that health care delivery systems may expend large amounts of resources on these innovative analytic solutions and require value confirmation for adoption at scale.

Study Purpose

The purpose of the dissertation was to evaluate the predictive and clinical utility of advanced EWS and to evaluate whether there were differences in alert-driven sepsis interventions between sepsis survivors and decedents. The remainder of this chapter describes the research aims, hypotheses, theoretical framework, and significance of the dissertation study. Beyond this general introduction chapter, the organization of the dissertation is as follows. Chapter 2 presents results from a systematic review of the advanced EWS literature. Chapter 3 describes results from clinical chart reviews of alerted sepsis patients after admission to a general hospital ward in stable condition. Chapter 4 reports results of a retrospective matched pair cohort study that aimed to statistically quantify the impact of alert-driven FBPs performed by RRTs on survival outcomes. Chapter 5 offers a summary and synthesis of all results and discusses implications for future research.

Research Aims

This dissertation study had the following specific aims:

Aim 1) Conduct Chart Reviews of Electronic Health Records to Categorize Fluid Bolus Processes (FBPs) and other RRT Sepsis Interventions Relative to the Alert

The purpose of expert chart reviews was to 1) examine the RRT response after (and possibly before) the EWS alert; 2) locate discrete FBP data and other sepsis intervention data in the electronic health record; and 3) categorize fluid bolus processes in terms of elapsed time from EWS alert to bolus and total 24-hour bolus volume.

Aim 2) Characterize and Compare Sepsis Survivors and Decedents Admitted in Stable Condition and Evaluate the Impact of Fluid Bolus Processes on 30-day Survival in a Cohort of Patients in Hospitals with an advanced EWS

Among sepsis patients with a positive EWS alert, I conducted a retrospective matched pair cohort study using multivariate pair matching of identified decedents and matched survivors; I performed descriptive statistics and quantified the between-group differences of FBPs in survivors and decedents.

Hypotheses

By delivering sepsis interventions faster, the hypothesis was that clinical patient trajectories and sepsis survival would improve. The specific alternative hypotheses ($p < 0.05$) were:

H_1: Elapsed time from EWS alert to administration of IV fluid bolus will be shorter in survivors compared to decedents, after pair-matching adjusted for patient age, sex, EWS score, patient comorbidity [COmorbidity Point Score version 2 (COPS2)], and severity of illness [LAboratory-based acute Physiology Score (LAPS2)].

H_2: Total 24-hour fluid bolus volume following an alert will be larger in survivors compared to decedents, after pair-matching adjustment for patient age, sex, EWS score, length of stay, patient comorbidity (COPS2), and severity of illness (LAPS2).

Theoretical Framework

Underlying the study's hypotheses is a theoretical hybrid model (Figure 1: Petri Dish Model, Linnen, 2017), which combines Population Ecology Theory (Hannan & Freeman, 1977) and Human Factors Theory (Reason, 1995). Following is a brief description of each of these theories and a discussion of their merits, gaps and application to the phenomenon of alert-driven RRT sepsis interventions.

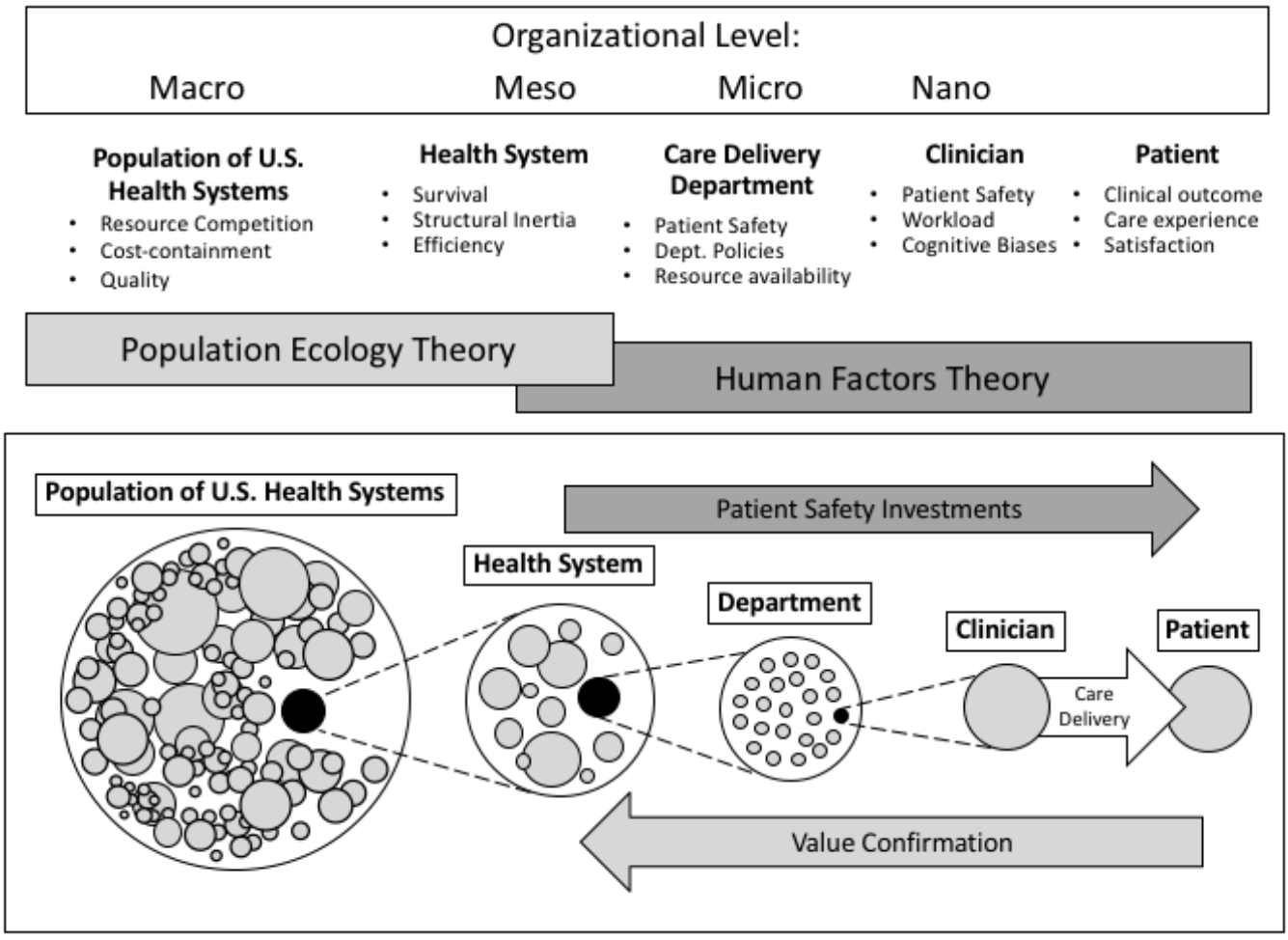
Population Ecology Theory (PET) aims to explain organizational growth rates and organizational behavior (strategic decisions and business behavior) in the context of survival in a competitive market. Due to a variety of constraining forces, organizations are described as exhibiting *structural inertia*, which is a conservative and innovation-averse stance. The theory can be aptly used to describe the underlying motivations of healthcare delivery systems to withhold investments in patient safety innovations such as an advanced EWS. Still, because of patient safety mandates, healthcare delivery systems are also known to expend resources despite their unclear benefit (e.g., RRT). PET does not attempt to reconcile this contradiction in its framework.

Human Factors Theory (HFT) has seen wide adoption in the healthcare domain of patient safety. HFT is a theory of accident causation and posits that medical errors (here: not recognizing worsening sepsis) are the result of four failure modes (fallible board decision and policy, line management problems, psychological precursors to unsafe acts, and unsafe acts). While HFT can explain why clinicians may not recognize worsening sepsis preemptively, it cannot explain how external and strategic forces might contribute to such a state.

The Petri Dish Model and Alert Driven Rescue

Both Population Ecology Theory and Human Factors Theory can be used to describe the context of RRT interventions in admitted sepsis patients with a positive EWS alert. However, to bridge the explanatory gaps of each theory, we developed the Petri Dish Model (Figure1, Linnen, 2017), which combines key aspects of these two theories and expands two interrelated organizational needs of survival in the market (cost) and optimal patient safety (quality).

Figure 1: The Petri Dish Model: Combining Population Ecology Theory and Human Factors Theory to Describe Value Return of Patient Safety Investments in Health Systems (Linnen, 2018)



The Petri Dish Model explains that care delivery systems (or “health systems”) are contained in a larger competitive macro environment of the U.S. healthcare system. At the meso level, the care delivery system provides the funding and allocates resources to facilitate the deployment of the EWS and RRTs. Acute care delivery departments staff the RRTs (micro level). EWS alerts and FBPs occur at the nano level of the individual acute care clinician and the patient. While it may appear that a value case could be made for tools improving early detection of worsening sepsis, it is unknown whether combining an advanced EWS with RRT has a meaningful impact on sepsis survival. To improve its survival stance in the market, the health system would require a value confirmation of such an investment. An econometric value confirmation would be achieved when the intervention benefits (in $) outweigh implementation and maintenance costs. Qualitatively, “value” may also be achieved if care delivery system leaders believe that the patient safety intervention has a probable, but immeasurable, benefit, for example by facilitating a selection advantage with insurance brokers.

The decentralized nature of hundreds of sepsis interventions occurring at multiple sites with varying personnel may result in variation in the delivery of fluid bolus processes. Hence, this dissertation aimed to aggregate fluid bolus processes, to describe how they were performed across KPNC hospitals with an advanced EWS, and to evaluate their impact on sepsis survival. For health systems, such information is important because it allows the meso-level business leaders to benchmark sepsis performance in the context of organizational targets and regulatory expectations (e.g., mortality). This new data-driven awareness closes the feedback loop (value confirmation) as these targets can now be measured in a more standardized and

reliable fashion, ultimately informing investment success of the EWS. In the Petri Dish Model, this knowledge may give the health system a selection advantage in the context of competitor performance and the wider market place.

Finally, the performance feedback loop starts over by allowing health systems to improve the intervention based on observed vs. expected outcome measurements, process measurements, and feedback from clinical staff. The Petri Dish Model predicts that the interrelated mechanisms of patient safety investment and value confirmation would motivate the use of an advanced EWS across the health system, and, long-term, across most hospitals in the U.S. This outcome could be measured by number of EWS deployment sites and annual deployment rates, though that is beyond the scope of this dissertation.

Significance of the Dissertation

The rescue of sepsis patients who are at risk of clinical decompensation on medical wards is hampered by a number of factors including the late detection of such risk using traditional screening methods resulting in delayed RRT interventions. Currently, at least in part because of a delayed RRT response, the evidence base for the impact of more timely administration of fluid bolus therapy on sepsis survival is weak. To date, no study has used an advanced EWS alert as a marker of Time Zero. This research is significant because it fills a critical gap in the literature. By being first to incorporate an advanced EWS alert and by categorizing fluid bolus processes and comparing survival outcomes, this dissertation study improved the empiric evidence base regarding the effective delivery of fluid bolus therapy and has the potential to inform clinical practice standards.

References

Angus, D. C., Linde-Zwirble, W. T., Lidicker, J., Clermont, G., Carcillo, J., & Pinsky, M. R. (2001). Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. *Crit Care Med, 29*(7), 1303-1310.

Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G. (2014). Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Aff (Millwood), 33*(7), 1123-1131. doi:10.1377/hlthaff.2014.0041

Dellinger, R. P., Levy, M. M., Rhodes, A., Annane, D., Gerlach, H., Opal, S. M., . . . Moreno, R. (2013). Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2012. *Crit Care Med, 41*(2), 580-637. doi:10.1097/CCM.0b013e31827e83af

Dombrovskiy, V. Y., Martin, A. A., Sunderram, J., & Paz, H. L. (2007). Rapid increase in hospitalization and mortality rates for severe sepsis in the United States: a trend analysis from 1993 to 2003. *Crit Care Med, 35*(5), 1244-1250. doi:10.1097/01.CCM.0000261890.41311.E9

Engel, C., Brunkhorst, F. M., Bone, H.-G., Brunkhorst, R., Gerlach, H., Grond, S., . . . John, S. (2007). Epidemiology of sepsis in Germany: results from a national prospective multicenter study. *Intensive care medicine, 33*(4), 606-618.

Epstein, L., Dantes, R., Magill, S., & Fiore, A. (2016). *Varying Estimates of Sepsis Mortality Using Death Certificates and Administrative Codes — United States, 1999–2014.* . Retrieved from

Escobar, G., & Dellinger, R. (2016). Early detection, prevention, and mitigation of critical illness outside intensive care settings. *J Hosp Med, 11 Suppl 1*, S5-s10. doi:10.1002/jhm.2653

Escobar, G., LaGuardia, J., Turk, B., Ragins, A., Kipnis, P., & Draper, D. (2012a). 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, 7*(5), 388-395. doi:10.1002/jhm.1929

Escobar, G. J., LaGuardia, J. C., Turk, B. J., Ragins, A., Kipnis, P., & Draper, D. (2012b). 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, 7*(5), 388-395. doi:10.1002/jhm.1929

Fielding-Singh, V., Greene, J., Baker, J. M., Escobar, G., & Liu, V. (2016). The Timing Of Early Antibiotics And Hospital Mortality In Sepsis *A105. CRITICAL CARE: VARIABILITY IN CARE, OPPORTUNITIES FOR ADOPTION OR DEADOPTION* (pp. A2741-A2741): Am Thoracic Soc.

Gaieski, D. F., Edwards, J. M., Kallan, M. J., & Carr, B. G. (2013). Benchmarking the incidence and mortality of severe sepsis in the United States. *Crit Care Med, 41*(5), 1167-1174.

Gaieski, D. F., Mikkelsen, M. E., Band, R. A., Pines, J. M., Massone, R., Furia, F. F., . . . Goyal, M. (2010). Impact of time to antibiotics on survival in patients with severe sepsis or septic shock in whom early goal-directed therapy was initiated in the emergency department. *Crit Care Med, 38*(4), 1045-1053.

Gao, H., McDonnell, A., Harrison, D. A., Moore, T., Adam, S., Daly, K., . . . Harvey, S. (2007). Systematic review and evaluation of physiological track and trigger warning systems for identifying at-risk patients on the ward. *Intensive Care Med, 33*(4), 667-679. doi:10.1007/s00134-007-0532-3

Glassford, N. J., Eastwood, G. M., & Bellomo, R. (2014). Physiological changes after fluid bolus therapy in sepsis: a systematic review of contemporary data. *Crit Care, 18*(6), 696. doi:10.1186/s13054-014-0696-5

Granich, R., Sutton, Z., Kim, Y. S., Anderson, M., Wood, H., Scharf, J. E., . . . Escobar, G. (2016). Early detection of critical illness outside the intensive care unit: Clarifying treatment plans and honoring goals of care using a supportive care team. *J Hosp Med, 11 Suppl 1*, S40-s47. doi:10.1002/jhm.2660

Hannan, M. T., & Freeman, J. (1977). The population ecology of organizations. *American journal of sociology, 82*(5), 929-964.

Healthcare Cost Utilization Project. (2008). HCUP Facts and Figures *HCUP Facts and Figures, 2006: Statistics on Hospital-Based Care in the United States*. Rockville (MD): Agency for Healthcare Research and Quality (US).

Herasevich, V., Pieper, M. S., Pulido, J., & Gajic, O. (2011). Enrollment into a time sensitive clinical study in the critical care setting: results from computerized septic shock sniffer implementation. *J Am Med Inform Assoc, 18*(5), 639-644. doi:10.1136/amiajnl-2011-000228

Institute of Medicine. (2015). *Dying in America: Improving Quality and Honoring Individual Preferences Near the End of Life*: National Academies Press.

Kramer, R. D., Cooke, C. R., Liu, V., Miller III, R. R., & Iwashyna, T. J. (2015). Variation in the Contents of Sepsis Bundles and Quality Measures. A Systematic Review. *Ann Am Thorac Soc, 12*(11), 1676-1684.

Levy, M. M., Fink, M. P., Marshall, J. C., Abraham, E., Angus, D., Cook, D., . . . Ramsay, G. (2003). 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Intensive Care Med, 29*(4), 530-538. doi:10.1007/s00134-003-1662-x

Linnen, D. (2016). The promise of big data: Improving patient safety and nursing practice. *Nursing, 46*(5), 28-34; quiz 34-25. doi:10.1097/01.nurse.0000482256.71143.09

Liu, V., Escobar, G., Greene, J., Soule, J., Whippy, A., Angus, D. C., & Iwashyna, T. J. (2014). Hospital deaths in patients with sepsis from 2 independent cohorts. *Jama, 312*(1), 90-92. doi:10.1001/jama.2014.5804

Liu, V., Morehouse, J. W., Soule, J., Whippy, A., & Escobar, G. J. (2013). Fluid volume, lactate values, and mortality in sepsis patients with intermediate lactate values. *Ann Am Thorac Soc, 10*(5), 466-473. doi:10.1513/AnnalsATS.201304-099OC

Martin, G. S., Mannino, D. M., Eaton, S., & Moss, M. (2003). The epidemiology of sepsis in the United States from 1979 through 2000. *N Engl J Med, 348*(16), 1546-1554. doi:10.1056/NEJMoa022139

McNeill, G., & Bryden, D. (2013). Do either early warning systems or emergency response teams improve hospital patient survival? A systematic review. *Resuscitation, 84*(12), 1652-1667. doi:10.1016/j.resuscitation.2013.08.006

O'Neill, R., Morales, J., & Jule, M. (2012). Early goal-directed therapy (EGDT) for severe sepsis/septic shock: which components of treatment are more difficult to implement in a community-based emergency department? *J Emerg Med, 42*(5), 503-510. doi:10.1016/j.jemermed.2011.03.024

Park, S. K., Shin, S. R., Hur, M., Kim, W. H., Oh, E. A., & Lee, S. H. (2017). The effect of early goal-directed therapy for treatment of severe sepsis or septic shock: A systemic review and meta-analysis. *J Crit Care, 38*, 115-122. doi:10.1016/j.jcrc.2016.10.019

Pastores, S. M., Dakwar, J., & Halpern, N. A. (2012). Costs of critical care medicine. *Crit Care Clin, 28*(1), 1-10, v. doi:10.1016/j.ccc.2011.10.003

Pepper, D. J., Jaswal, D., Sun, J., Welsh, J., Natanson, C., & Eichacker, P. Q. (2018). Evidence Underpinning the Centers for Medicare & Medicaid Services' Severe Sepsis and Septic Shock Management Bundle (SEP-1): A Systematic Review. *Ann Intern Med, 168*(8), 558-568. doi:10.7326/m17-2947

Prasad, P. A., Shea, E. R., Shiboski, S., Sullivan, M. C., Gonzales, R., & Shimabukuro, D. (2017). Relationship Between a Sepsis Intervention Bundle and In-Hospital Mortality Among Hospitalized Patients: A Retrospective Analysis of Real-World Data. *Anesthesia & Analgesia, 125*(2), 507-513.

Reason, J. (1995). Understanding adverse events: human factors. *Qual Health Care, 4*(2), 80-89.

Rhodes, A., Evans, L. E., Alhazzani, W., Levy, M. M., Antonelli, M., Ferrer, R., . . . Nunnally, M. E. (2017). Surviving sepsis campaign: international guidelines for

management of sepsis and septic shock: 2016. *Intensive care medicine, 43*(3), 304-377.

Rivers, E., Nguyen, B., Havstad, S., Ressler, J., Muzzin, A., Knoblich, B., . . . Tomlanovich, M. (2001a). Early goal-directed therapy in the treatment of severe sepsis and septic shock. *New England Journal of Medicine, 345*(19), 1368-1377.

Rivers, E., Nguyen, B., Havstad, S., Ressler, J., Muzzin, A., Knoblich, B., . . . Early Goal-Directed Therapy Collaborative, G. (2001b). Early goal-directed therapy in the treatment of severe sepsis and septic shock. *N Engl J Med, 345*(19), 1368-1377. doi:10.1056/NEJMoa010307

Smith, G. B., Prytherch, D. R., Meredith, P., Schmidt, P. E., & Featherstone, P. I. (2013). The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death. *Resuscitation, 84*(4), 465-470. doi:10.1016/j.resuscitation.2012.12.016

Smith, M. E., Chiovaro, J. C., O'Neil, M., Kansagara, D., Quinones, A. R., Freeman, M., . . . Slatore, C. G. (2014). Early warning system scores for clinical deterioration in hospitalized patients: a systematic review. *Ann Am Thorac Soc, 11*(9), 1454-1465. doi:10.1513/AnnalsATS.201403-102OC

Chapter 2

The Ability of Advanced Early Warning Systems vs. Point Score Tools to Detect Clinical Deterioration Risk: A Systematic Review

Daniel T. Linnen, PhD(c), MS, RN-BC[1,2]

Gabriel J. Escobar, MD[2]

Xiao Hu, PhD[3]

Elizabeth Scruth, PhD, MPH, CNS, RN[4]

Vincent Liu, MD, MS[2]

Caroline Stephens, PhD, RN, GNP[5]

[1] University of California San Francisco, School of Nursing, Department of Social & Behavioral Sciences

[2] Kaiser Permanente Northern California, The Permanente Medical Group, Inc., Division of Research

[3] University of California San Francisco, School of Nursing, Department of Physiological Nursing

[4] Kaiser Permanente Northern California, Department of Quality, Kaiser Foundation Hospitals, Inc.

[5] University of California San Francisco, Department of Community Health Systems

Abstract

Background: The clinical deterioration of patients admitted in stable condition to general medical-surgical hospital wards is a vexing patient safety issue. Few early warning systems (EWSs) can identify patients before clinical decline has ensued.

Objectives: We aimed to evaluate the literature regarding prognostic test accuracy and clinical evaluation workloads generated by advanced EWSs using multivariate regression or machine learning techniques vs. points-score tools to detect clinical deterioration risk in hospitalized adult patients on general wards.

Methods: We searched PubMed, CINHAL, and Google Scholar databases using terms that described clinical deterioration and use of advanced EWS model. The outcome was clinical deterioration of adult patients on general hospital wards (the composite of transfer to ICU and/or mortality). We included studies published in peer-reviewed journals from 2012 to 2017. Of 295 articles, we excluded 290 studies (different setting, population, or method). We selected 5 studies reporting model performance of EWSs using multivariate regression or machine learning. We used 2015 PRIMSA systematic review protocol guidelines and 2015 TRIPOD criteria for predictive model evaluation and the Cochrane Collaboration guidelines to assess the studies' methodological rigor and bias risk. We reported measures of model performance across studies, calculated pre-test probability, adjusted positive predictive value (PPV), and conducted simulations of workup-to-detection ratios. We then calculated means and graphed results of model performance between point-score tools and advanced EWSs and synthesized study results.

Results: Of the 5 studies, 4 were of good quality and employed rigorous measurement and modeling techniques. All advanced EWSs appeared to perform better at identifying patients at risk for clinical deterioration than point-score EWSs (mean AUC 0.80 vs 0.72). In adjusted analysis, advanced EWSs generated 5.4 alerts per hospital per day, compared to 8.0 alerts per hospital per day in point-score EWSs. The delta of 2.6 alerts per hospital per day equals a nearly 50% relative increase in RRT workload using point-score EWSs.

Conclusion and Implications: This systematic review is the first to evaluate the performance of advanced EWSs. The evidence is limited to a few studies. Compared to traditional point-score tools, advanced EWSs consistently demonstrated superior prognostic performance both in terms of accuracy and RRT workup demands. A standardized approach to reporting EWS model performance is needed, including pre-test probability, observed and adjusted PPV, alerts generated to find 1 true positive case, and alerts generated per 100 patients per day.

Introduction

Ensuring the delivery of safe and cost-effective care is the core mission of the U.S. acute care delivery system (National Academy of Medicine formerly the Institute of Medicine, 1999). Nearly 90% of all unplanned patient transfers to critical care may be the result of a new or worsening condition (Bapoje, Gaudiani, Narayanan, & Albert, 2011). For example, estimates suggest that nearly 200,000 patients suffer a cardiac arrest (Merchant et al., 2011) and 1,000,000 patients die of severe sepsis in U.S. hospitals annually. The combined national costs for the treatment of sepsis, respiratory failure and arrest are estimated to be $30.7 billion (8.1% of national costs) (Torio, 2015). As many as 44% of adverse events, and the associated costs of treatment in the intensive care unit (ICU), may be avoidable (Levinson & General, 2010). Clinical deterioration requires life supporting interventions by a rapid response team (RRT) or code blue team (Bellew, Cabrera, Lohse, & Bellolio, 2016) and transfer to the ICU. These interventions, however, may often be reactive and begin once a patient is already in acute distress, which is often too late. Evidence suggests that many hospitalized patients presenting with rapid decline showed warning signs 24-48 hours before the event (McGaughey et al., 2007). These signs may be subtle, making it difficult for nurses and physicians to recognize and synthesize their meaning leading up to a deterioration event (Silber, Williams, Krakauer, & Schwartz, 1992; Escobar et al., 2016; Hu, Bai, & Salas-Boni, 2016b).

Acute deterioration of patients admitted to general wards in stable condition may result in transfer to the intensive care unit (ICU), longer length of stay, and increased mortality risk (Escobar et al., 2012; Dahn et al., 2016). Life support, ICU stay and longer

hospitalization time also expose patients to additional iatrogenic risks, such as nosocomial infections (Richards, Edwards, Culver, & Gaynes, 2015), medical errors, hospital acquired pressure ulcers (Coyer & Tayyib, 2017), falls with injury (Bouldin et al., 2013), and ICU psychosis (Wolters et al., 2014). Across the over 5,500 hospitals in the U.S. (American Hospital Association, 2018), late recognition of impending deterioration may further result in financial burden due to longer hospital stays (Barwise et al., 2016), the high costs of critical care (Pastores, Dakwar, & Halpern, 2012), and litigation (Pascall, Trehane, Georgiou, & Cook, 2015).

Hospitals began staffing rapid response teams (RRTs) as an answer to these challenges. RRTs perform patient rounding, fulfill requests by staff nurses to evaluate patients, and respond to EWS alarms and acute deterioration events (Escobar & Dellinger, 2016). Still, two systematic reviews regarding the effectiveness of RRTs on patient outcomes reported inconclusive results (McGaughey et al., 2007; Winters et al., 2013). This may be at least in part because RRTs often respond to, rather than preempt, clinical deterioration.

As early as 1997, hospitals have used early warning systems (EWSs) to identify at-risk patients and to proactively inform clinicians (Morgan, Williams, & Wright, 1997). These EWSs, which can be automated in the electronic medical record, screen for deterioration risk in entire hospital populations (Dummett et al., 2016). There are two main types of EWS: 1) point-score EWSs and 2) EWSs that use computational methods of multivariate regression or machine learning (we will refer to the latter as “advanced” EWSs). Point-score EWSs identify patient deterioration risk using simple addition of a few clinical parameter scores, including vital signs and level of consciousness.

However, the lack of risk adjustment with point-score EWSs may produce a high false positive alarm rate (Romero-Brufau et al., 2014). Moreover, as these point-score EWSs become better at finding all true positive cases, they also increase the overall number of cases needed to evaluate. This is problematic because: 1) RRTs may be a costly resource; 2) RRTs do not have the capacity to evaluate a large number of false positive patients (which may take up to 30 minutes in total) on a routine basis; and 3) a high proportion of false positive alarms may cause alarm fatigue (Guardia-Labar, Scruth, Edworthy, Foss-Durant, & Burgoon, 2014). A systematic review (Smith, Chiovaro, & O'Neil, 2014a) found that too little is known regarding the effect of pairing point-score EWSs on clinical deterioration outcomes and RRT staffing costs. In contrast, advanced EWSs use computational methods to predict risk (Bates, Saria, Ohno-Machado, Shah, & Escobar, 2014) by adjusting for many clinical covariates and thereby reducing the degree of unexplained variance. While they are thought to be more precise and to generate fewer false positive alarms (Escobar et al., 2012; Churpek, Yuen, Park, Gibbons, & Edelson, 2014; Churpek et al., 2016; Kipnis et al., 2016), no review to date has synthesized and compared their performance against point-score EWSs systematically. To provide safe, timely and cost-effective care, health systems need to have a better understanding of how to precisely identify patients, who are at risk for clinical deterioration, while minimizing RRT evaluation workloads.

Purpose

To that end, the purpose of this systematic review was to evaluate the literature regarding prognostic test accuracy and RRT workloads generated by advanced EWSs

using multivariate regression or machine learning techniques vs. points-score systems to detect clinical deterioration risk in hospitalized adult patients on general wards.

Methods

Search Strategy

We searched the published peer-reviewed literature in PubMed, CINAHL and Google Scholar, as well as the gray literature (Agency for Healthcare Research and Quality and Institute for Healthcare Improvement websites) published in English and German language between January 1, 2012 and November 1, 2017. We selected this time frame because, compared to point-score EWSs, advanced EWSs using electronic medical record data are comparably new approaches. An expert PhD researcher independently confirmed the search results in a blinded search. Table 1 describes the search terms and search details. A search for “early warning score OR early warning system AND deterioration OR predict transfer ICU” returned 227 peer-reviewed articles. We then removed psychology references from the detailed search, resulting in 117 articles. A separate search in CINAHL using the same filters and query returned 175 academic journal articles, of which no additional articles met inclusion criteria. A Google Scholar search using the key terms “early warning score early warning system deterioration predict transfer ICU” returned 4,530 articles, and we deemed this query too broad. Narrowing the search to “early warning score early warning system deterioration predict transfer ICU regression machine learning” returned 295 results. Of these, we identified no additional journal articles.

Inclusion and Exclusion Criteria

Figure 1 shows the search strategy and study selection, following PRISMA protocol guidelines for systematic reviews. We included peer-reviewed journal articles that reported models predicting transfer to ICU and mortality, because these are the two most common proxies for clinical deterioration on wards (patients either die on the ward or get transferred to ICU). We included articles if they reported model performance of an advanced EWS for hospitalized adult patients on general medical-surgical wards (see Table 1), using the area under the receiver operator curve (AUC) (Zweig & Campbell, 1993) or the equivalent c-statistic (Romero-Brufau, Huddleston, Escobar, & Liebow, 2015). These are measures of model discrimination, and they compare a model's false positive rate (1-specificity) against its true positive rate (sensitivity). Our search did not require additional parameters of model performance (sensitivity, specificity, positive predictive value [PPV], model calibration, coefficient of determination [R^2] or work-up to detection ratio), but we captured or calculated them for analysis. Please see the appendix for detailed results.

Of the 295 articles, we excluded 281 in the initial abstract screen, and 9 additional articles during full-text review. We excluded studies if they only reported on a paper-based EWS or point-score EWS or if they used physiological monitor data because monitor use is ubiquitous in critical care and step-down units, but not general wards. We excluded pediatric and obstetric patient populations, as well as patients in intensive care units, emergency rooms, or specialty oncology units, and outcomes unrelated or only partially related to clinical deterioration. Of 4 reports published in conference proceedings, we excluded 3 machine learning papers because they

analyzed physiological patient monitor data and/or occurred in a critical care setting (Clifton, Clifton, Watkinson, & Tarassenko, 2011; Mao et al., 2012; Yoon, Alaa, Hu, & Van Der Schaar, 2016). The fourth conference report in the proceedings of the 2012 International Conference on Knowledge Discovery and Data Mining (Mao et al., 2011) did not meet the inclusion criterion of peer-reviewed journal publication. In total, 5 peer-reviewed research reports met eligibility criteria in this systematic review. A review of the reference lists of these 5 studies did not yield additional reports for inclusion in this review.

Data Abstraction

Following the TRIPOD guidelines for the transparent reporting of predictive models (Collins, Reitsma, Altman, & Moons, 2015), and the PRISMA and Cochrane Collaboration guidelines for systematic reviews (Moher, Liberati, Tetzlaff, Altman, & Group, 2009; Higgins, 2011), our data abstraction had three aims: 1) summarize the studies' basic characteristics; 2) compare EWS model performance using a systematic and transparent approach; and 3) assess bias and the level of scientific evidence.

Of the 5 number of studies reviewed, we extracted the following data:

1. **Study Characteristics (Table 2):** Year published, country, location and setting, study purpose, theoretical framework, research design, study time frame, sample size and attrition, and sample characteristics. Since heterogeneity of the studies is an important criterion when assessing the strength of the evidence, we captured how the studies differed in their design and sample selection (Polit & Beck, 2012).
2. **Predictive Model Characteristics and Performance (Appendix Table 2 and 3):** Surveillance method; predictors; reference standard; model performance [AUC/c-

statistic, sensitivity, specificity, positive predictive value (PPV), pre-test probability, model calibration]; workup-to-detection ratio, relevant findings, strengths, and limitations. We chose these measures to compare the predictor variables in each of the early warning system computer models and to facilitate a systematic comparison of EWS models.

3. **Level of Scientific Evidence and Risk of Bias Assessment (Appendix Table 4):** We compiled a scoring table assessing each study's level of scientific evidence. Factors included research design, measurement bias, detection bias, missing data bias, threats to external validity and total bias score (derived by summing the equally weighted sub-scores of each factor). We adapted these criteria from the Cochrane Collaboration's tool for assessing risk of bias in systematic reviews (Higgins et al., 2011) to minimize subjective interpretation of results. We used these scores to identify both the overall strength of the evidence and to identify common bias across studies.

Measures of Model Discrimination

The Area Under the Receiver Operator Curve (AUC) (Zweig & Campbell, 1993), also referred to as the c-statistic, plots a model's false positive rate (x-axis) and true positive rate (y-axis), with an ideal scenario of very high y-values and very low x-values (Hanley & McNeil, 1982). Models can be calibrated to produce fewer alarms at the tradeoff of identifying fewer true positive cases (and vice versa). Therefore, AUC alone is not a meaningful measure of model performance (Romero-Brufau et al., 2015) because it does not measure a model's RRT workup-to-detection ratio (WDR), or patient evaluation workload to find one true positive case. It is known that a population's

outcome prevalence (or pre-test probability) influences a model's positive predictive value (PPV) (Altman & Bland, 1994) and reciprocally WDR, which equals $\frac{1}{\text{PPV}}$. This makes it challenging to compare models systematically across varying research populations. However, PPV/NPV can be standardized using a simulated pre-test probability across studies. To standardize PPV (percent of true positive alerts among all alerts), we simulated a pre-test probability of 2% (Romero-Brufau et al., 2015) across studies. We did not capture or adjust Negative Predictive Value (NPV; detecting true negative tests among all negative tests) (Bewick, Cheek, & Ball, 2004) because the principal utility of EWSs is to find true cases, not true negatives.

Additionally, we captured sensitivity (the model's ability to detect a true positive case among all cases), and specificity (the model's ability to detect a true non-case among all non-cases (Bewick et al., 2004)). Because sensitivity and specificity describe characteristics of the test, rather than the outcome, AUROC/c-statistic are not influenced by pre-test outcome probability. Such measures of model discrimination are important because they answer how useful a model is at "finding cases." For example, if a model is not very sensitive (not good at recognizing true positive cases among all cases), it will have low practical utility for busy clinicians. Equally, if a model has low specificity (not good at detecting true negatives among all non-cases) it results in excessive alarming, wasteful RRT deployment, and alarm fatigue (Kho et al., 2007). Finally, we abstracted the models' reference standard (or "gold standard") to accurately identify a true case; the coefficient of determination (R^2), which is the percent variance explained by the model (Nagelkerke, 1991); and measures of model calibration, which

determine a model's observed performance against its expected score (e.g., the Hosmer-Lemeshow Goodness of Fit test (Steyerberg et al., 2010)).

Results

Following is an examination of the 5 advanced EWS studies meeting inclusion criteria.

Study Characteristics

There were a number of similarities across the selected studies (see Table 2). All reported on research conducted in the United States; all compared their model's performance against at least one point-score EWS model (Escobar et al., 2012; Alvarez et al., 2013; Churpek et al., 2014; Churpek et al., 2016; Kipnis et al., 2016); all used retrospective cohort/nested case-control designs; none of the studies reported a theoretical framework, though all studies discussed the significance and background in regard to patient safety, limited predictive capabilities of point-score EWS, or alarm fatigue. Of the 5 studies, 1 (Alvarez et al., 2013) took place in a single hospital; 2 (Churpek et al., 2014; Churpek et al., 2016) pooled data from 5 hospitals; and 2 (Escobar et al., 2012; Kipnis et al., 2016) occurred in a large integrated health care delivery system using data from 14 and, subsequently, 21 hospitals. Only 1 study (Alvarez et al., 2013) did not report sample demographics. The 4 studies using data from more than one medical center also had the longest timeframes, ranging from 3 to 4 years, compared to 11 months in the single center study. Studies reported the units of analysis heterogeneously, including hospitalizations, hospital episodes, and admitted patients. The largest study (Kipnis et al., 2016) had nearly 650,000 hospital episodes, while the smallest study (Alvarez et al., 2013) reported slightly less than 7,500 patient

admissions. Of the 5 studies, 4 used multivariate regression (Escobar et al., 2012; Alvarez et al., 2013; Churpek et al., 2014; Kipnis et al., 2016), and 1 study used machine learning techniques for outcome prediction (Churpek et al., 2016).

Outcome Variables

The primary outcome of interest was clinical deterioration. At minimum, all selected studies measured this outcome via the composite of transfer to ICU and some measure of mortality. This means, if patients on general hospital wards had a transfer to ICU or they died, they had the outcome, else they did not. Churpek et al. (2014) and Churpek at al. (2016) also included cardiac arrest, and Alvarez et al. (2013) included respiratory compromise in their outcome composite (though, clinically, these are contained in ICU transfer or death). In the interest of a systematic analysis, we compared the most frequently reported model outcome: transfer to ICU and mortality. Researchers used different definitions of mortality and varied in their conceptual view of when a patient death should or should not be counted as an outcome event. Definitions included “death outside the ICU in a patient whose care directive was “full code” (Escobar et al., 2012; Kipnis et al., 2016); “death on the wards without attempted resuscitation” (Churpek et al., 2014; Churpek et al., 2016); or “an in-hospital death in patients without a DNR order at admission that occurred on the medical ward or in ICU within 24 hours after transfer”. All studies conceptualized death as a potentially appropriate outcome (e.g., given a patient’s disease burden or end-of-life trajectory) and only included patients with a “Full Code” designation at the time of death on the ward.

Predictor Variables

We observed a broad assortment of predictor variables across studies. All models included vital signs (heart rate, respiratory rate, blood pressure, oxygen saturation, and mental state), laboratory data, age, and sex. Additional variables included comorbidity; shock index $\left(\frac{\text{heart rate}}{\text{systolic blood pressure}}\right)$ (Berger et al., 2013); a severity of illness score; length of stay; event time of day; season; admission category; and length of stay (Escobar et al., 2012; Kipnis et al., 2016); prior ICU stay (Churpek et al., 2014; Churpek et al., 2016); and "stat" orders (Alvarez et al., 2013), i.e., physician orders to be instantiated immediately.

Model Performance

All studies reported their EWS model's AUC or c-statistic (see Table 3 and Figure 2) as a minimum inclusion criterion for this systematic review. Overall, AUC values ranged from 0.77 to 0.85 (weighted mean = 0.80) in advanced EWSs indicating good model discrimination. Point-score EWS AUC's ranged from 0.70 to 0.76 (weighted mean = 0.72) indicating fair model discrimination.

The models' sensitivity ranged from 0.49 to 0.54 (weighted mean = 0.50) for advanced models and 0.39 to 0.50 (weighted mean = 0.42), though it is important to note that these results were based on chosen alert volume cutoffs. For example, Kipnis et al. (2016) selected a detection threshold that would not generate more than 1 alert per 35-bed hospital unit per day, while Churpek et al. (2016) reported various sensitivity and specificity results at different detection cutoffs. As such, it is incorrect to assume that a given model produces only one sensitivity result; for systematic comparison, we therefore selected the "overall" result or results in the 50% sensitivity range. Specificity

ranged from 0.90 to 0.94 (weighted mean = 0.92) in advanced EWS models, compared to 0.83 to 0.93 (weighted mean = 0.90) in point-score EWS models. Positive Predictive Value (PPV) ranged from 0.16 to 0.42 (mean = 0.21) in advanced models and 0.15 to 0.28 (mean = 0.15) in point-score EWS models. After adjustment for a simulated pre-test outcome probability (or prevalence) of 2% across studies (see Table 3), PPV ranged from 0.10 to 0.15 (mean = 0.11) in advanced EWS models vs 0.06 to 0.10 (weighted mean = 0.08) in point-score EWS models.

None of the studies reported a reference standard for the definition of a true positive, and none of the studies reported the coefficient of determination (R^2). Of the 5 studies, all studies reported adequate fit in their final models. Only two studies (Escobar et al., 2012; Kipnis et al., 2016) reported the workup-to-detection (WDR) metric (alerts generated to identify one true positive case). One study evaluating machine learning methods (Churpek et al., 2014; Hu, Wong, Correa, Li, & Deng, 2016a) did not report the Workup-to-Detection Ratio, but an “early warning score efficiency curve,” which plots the percent of positive screens against sensitivity. To make results comparable across studies, we adjusted RRT workload estimates by mean hospital count, weighted mean study timeframe (in months), and set a fixed outcome prevalence of 2%, a fixed sensitivity of 0.50 and observed specificity of 0.92 (weighted mean) in advanced EWS models vs a fixed sensitivity of 0.50 and assumed specificity of 0.87 in point-score EWSs. Using these (conservative) assumptions, advanced EWSs generated 5.4 alerts per hospital per day, and point-score EWS generated 8.0 alerts per hospital per day. The delta of 2.6 alerts per hospital per day equals a nearly 50% relative increase in RRT workload using point-score EWSs (Figure 3).

Risk of Bias Assessment

We examined potential threats to validity and reliability and assessed other sources of bias. Evidence-based practice relies on the strength of research findings to inform clinical practice standards (Polit & Beck, 2012). The evidence hierarchy places systematic reviews of randomized and nonrandomized trials at the top, however we cannot perform RCTs to answer whether EWSs and RRTs improve survival, as it would be unethical. The 5 selected studies were all situated in Level 4 (single observational study). Consequently, this systematic review appears to offer the highest attainable level of evidence.

Appendix Table 4 shows the bias assessment adopted from the Cochrane Collaboration tool for assessing risk of bias (Higgins et al., 2011). Of the 5 studies, 4 received total scores between 1.0 - 2.0 (indicating relatively low bias risk), and 1 study had a score of 3.5 (indicating higher bias risk). Low bias studies (Escobar et al., 2012; Churpek et al., 2014; Churpek et al., 2016; Kipnis et al., 2016) used large samples across multiple hospitals, discussed the choice of predictor variables and outcomes more precisely, and reported their measurement approaches and analytic methods in more detail, including imputation of missing data and model calibration.

One study (Alvarez et al., 2013) used a small sample from a single medical center, thereby introducing threats to external validity. While all selected studies had good face validity, this study also introduced detection bias and threats to measurement validity by defining the concept of clinical deterioration more broadly. In other words, Alvarez et al. (2013) may have had apparently better model performance because of more liberal outcome definitions (i.e., casting a wider net). This study (Alvarez et al.,

2013) also tolerated 2 alert-driven RRT calls per department per day, which may not be a manageable caseload for RRT, and included patients with “do not resuscitate” or “comfort care” orders in the “unexpected death” numerator. This approach is problematic because it includes patients for whom death may have been expected or considered probable. Evaluating such patients on a routine basis adds to RRT and hospitalist workloads without improving rescue.

Discussion

This systematic review assessed the predictive ability of advanced EWS models vs. point-score tools to detect clinical deterioration risk in hospitalized adults on general wards. From 2007 to 2017, at least 5 systematic reviews have examined point-score EWSs in adult inpatient settings (Subbe, Williams, Fligelstone, & Gemmell, 2005; Johnstone, Rattray, & Myers, 2007; McNeill & Bryden, 2013; Smith et al., 2014a; Smith et al., 2014b). No systematic review, however, has synthesized the evidence of advanced EWS models using regression techniques and machine learning algorithms. Our analysis suggests that advanced EWS models perform substantially better than point-score EWSs, and they generate considerably fewer alerts for RRTs to evaluate to identify a positive case. In fact, if one would extrapolate the absolute reduction in RRT workload of 2.6 fewer RRT evaluations per hospital per day across the ~5,500 U.S. hospitals (American Hospital Association, 2018), advanced EWSs would eliminate over 5 million unnecessary evaluations by RRT and the attending hospitalist. This efficiency benefit would save nearly $350 million in wasted clinician time annually , assuming 30 minutes per workup and a mean hourly wage of $35.36 for a registered RRT nurse

(Bureau of Labor Statistics, 2017b) and $101.63 for a hospitalist (Bureau of Labor Statistics, 2017a).

Favorable Properties of Advanced EWS Models

All studies included in this review demonstrated superior model performance of the advanced EWS compared to a point-score EWS, and at least 4 of the 5 studies employed high rigor in design, measurement, and analytic method. The AUC absolute difference between advanced EWSs and point-score tools was about 10% overall, moving model performance from fair to good (see Table 3). Given the concern that point-score EWSs are crude tools and often derived with questionable rigor (Smith et al., 2014a), results from our review appear to agree that advanced EWS models predict clinical deterioration risk with better precision. This is an important finding for three reasons: 1) *Earlier Rescue:* Better risk prediction can facilitate a speedier activation of rescue; 2) *Less waste:* Given federal mandates to maintain quality while curbing spending, the elimination of wasteful processes in health care is one chief strategy of high value care (Berwick & Hackbarth, 2012); and 3) *Reduced workloads and less cognitive burden:* Sikka, Morath, and Leape (2015) posited that Berwick and colleagues' Triple Aim of high value care should be expanded to account for the wellbeing of clinicians (the Quadruple Aim). Advanced EWSs appear to support clinicians by reducing unnecessary evaluation workloads and time pressures, which are associated with professional burnout (Maslach, Schaufeli, & Leiter, 2001). Further, advanced EWSs support clinicians by reducing alarm fatigue (alert desensitization) (Guardia-Labar et al., 2014; Ruskin & Hueske-Kraus, 2015). Finally, more precise alerting reduces

interruptions and the cognitive burden placed on clinicians, which are known antecedents to medical errors (Reason, 1995 2000).

Limitations of advanced EWS models

Even the most advanced EWS models appear to identify only about half of all positive cases (assuming manageable RRT workloads). Consequently, advanced EWS models can at best augment and support, but not replace, RRT rounding, physician workup, and vigilant frontline staff. To facilitate the best possible rescue response, RRT rounding and advanced EWSs are complementary solutions; sometimes advanced EWSs will identify patient risk before RRTs, while at other times frontline staff and RRT are the first to detect patient risk.

Another limitation of advanced EWSs lies in their measurement of an expected death. All advanced EWSs excluded deaths without resuscitation (DNR) to adjust for those patients for whom aggressive life supporting interventions would not have been a "good" outcome. For example, in the context of existing DNR status, an end-of life trajectory, or severe terminal illness, clinical deterioration may be expected. Nevertheless, it is unknown what might define a "good death after alert" in the context of patient preferences regarding intensity of treatment and palliative care (Granich et al., 2016) and knowing that not all patients who *can* be rescued may *want* to be rescued (Escobar & Dellinger, 2016). Instead, the National Academies of Medicine's (formerly the Institute of Medicine) report *Dying in America* (Institute of Medicine, 2015) stressed the need for palliative care teams to support patient-centered end-of-life issues in all care settings. In the future, advanced EWSs could not only facilitate faster RRT rescue

processes but also enable palliative care team referrals so that they may address patient-centered end-of-life decisions and patient comfort.

Comparison of Results with Prior Evidence

While this systematic review focused on advanced EWS models, the results agreed with several prior systematic reviews investigating point-score EWSs (Subbe et al., 2005; Johnstone et al., 2007; McNeill & Bryden, 2013; Smith et al., 2014a; Smith et al., 2014b). Both point-score EWSs and advanced EWSs succeed in their task of automating the identification of at–risk patients, but point-score EWSs generate more workload to identify such patients. Previously reported AUC results for the death outcome were better for point-score EWSs (AUC range 0.88 to 0.93) than results included in this systematic review (AUC range 0.77 to 0.85). However, these point-score models tolerated higher false positive rates or did not adjust for manageable RRT/hospitalist evaluation workloads. This means that predictive models cannot be judged purely on AUC (in fact, it would be ill-advised), but instead by their clinical utility. Precision is not meaningful if it comes at the expense of unmanageable evaluation workloads.

Limitations

Findings from this systematic review are subject to several limitations. Not all mortality definitions accounted for the reality that a patient death may be an appropriate outcome, assuming it was in concordance with a patient's treatment wishes in the context of severe illness or an end-of-life trajectory (Kim et al., 2016). However, past studies of point-scores EWSs did also not account for this nuance, and our review suggests that predictive analytics EWSs perform better than point-score EWSs,

meaning they identify patient risk prospectively and generate fewer alerts for evaluation while demonstrating greater precision.

Because of the relative dearth of current evidence, it was not possible to investigate specific hospital ward types, such as surgical wards or step-down units, individually. However, the feasibility and practicality of ward-specific risk models is questionable, given that there may be substantial variability across such wards across institutions and because hospital systems must invest considerable resources to develop, deploy and maintain even one EWSs.

We excluded studies using streaming data from physiological monitors, despite notable work in this field (Mao et al., 2011; Mao et al., 2012; Bai et al., 2015; Bai et al., 2016; Hu et al., 2016b). The use of streaming monitor data restricts model use to ICU and other monitored wards. The aim of this review, however, was to determine the capabilities of advanced EWSs as an evaluation tool for RRT on general hospital wards so that at-risk patients can be identified hours before they would acutely deteriorate.

Conclusion

Results from this systematic review point to three main areas of need for the field of predictive EWS and RRT research: 1) a standardized set of RRT rescue outcome measures; 2) a standardized set of RRT workup and alarm frequency measures; and 3) cost estimates of RRT workloads with and without advanced EWS deployment. In the future, EWS research may achieve a stronger evidence base by using standardized outcome measures (Amarasingham et al., 2016). Given the present divergence of outcome definitions, it appears that EWS research would benefit from a common "clinical deterioration" outcome standard, including transfer to ICU, inpatient/30-day/90-

day mortality, and death with DNR, comfort care or hospice. Finally, the field is lacking a standardized RRT work-up measure, such as workup-to-detection ratio and RRT evaluations per hospital per day. Clinically, our review suggests that health systems must scrutinize the workload implications of EWSs carefully, beyond focusing on AUC. Given the pressing issues of waste in health care, alert fatigue, clinician burnout and cognitive load, predictive models should be chosen for both their precision and resource efficiency.

By using predictive analytics, health systems may be better able to achieve the dual goals of high-value care (Berwick & Hackbarth, 2012) and patient safety (Linnen, 2016; Liu et al., 2016; Parikh, Kakad, & Bates, 2016) in support of the Quadruple Aim (Sikka et al., 2015). The advent of more sophisticated computational methods for outcome prediction appears to enable the use of very large datasets in real-time to develop more precise and efficient identification of patients at risk for clinical deterioration. Still, gaps in knowledge exist regarding the measurement of RRT processes triggered by EWSs, clinical outcomes, RRT workloads, and costs vs. benefits. Future research should employ a pair-matching design to study the balanced treatment effects of RRT interventions between decedents and survivors and investigate RRT process times relative to the time of an advanced EWS alert in the electronic medical records. In addition to reducing unexpected patient mortality and more effectively directing hospital resources, advanced EWS systems may also serve to prompt needed end-of-life discussions (Picker et al., 2017).

References

Altman, D. G., & Bland, J. M. (1994). Statistics Notes: Diagnostic tests 2: predictive values. *BMJ, 309*(6947), 102.

Alvarez, C. A., Clark, C. A., Zhang, S., Halm, E. A., Shannon, J. J., Girod, C. E., . . . Amarasingham, R. (2013). Predicting out of intensive care unit cardiopulmonary arrest or death using electronic medical record data. *BMC Med Inform Decis Mak, 13*, 28. doi:10.1186/1472-6947-13-28

Amarasingham, R., Audet, A. M., Bates, D. W., Glenn Cohen, I., Entwistle, M., Escobar, G., . . . Xie, B. (2016). Consensus Statement on Electronic Health Predictive Analytics: A Guiding Framework to Address Challenges. *EGEMS (Wash DC), 4*(1), 1163. doi:10.13063/2327-9214.1163

American Hospital Association. (2018). Fast Facts on U.S. Hospitals. Retrieved from https://www.aha.org/statistics/fast-facts-us-hospitals

Bai, Y., Do, D., Ding, Q., Arroyo Palacios, J., Shahriari, Y., Pelter, M., . . . Hu, X. (2016). Is the Sequence of SuperAlarm Triggers more Predictive than Sequence of the Currently Utilized Patient Monitor Alarms? *IEEE Trans Biomed Eng*. doi:10.1109/tbme.2016.2586443

Bai, Y., Do, D., Harris, P. R., Schindler, D., Boyle, N. G., Drew, B. J., & Hu, X. (2015). Integrating monitor alarms with laboratory test results to enhance patient deterioration prediction. *J Biomed Inform, 53*, 81-92.

Bapoje, S. R., Gaudiani, J. L., Narayanan, V., & Albert, R. K. (2011). Unplanned transfers to a medical intensive care unit: causes and relationship to preventable errors in care. *J Hosp Med, 6*(2), 68-72. doi:10.1002/jhm.812

Barwise, A., Thongprayoon, C., Gajic, O., Jensen, J., Herasevich, V., & Pickering, B. W. (2016). Delayed Rapid Response Team Activation Is Associated With Increased Hospital Mortality, Morbidity, and Length of Stay in a Tertiary Care Institution. *Crit Care Med, 44*(1), 54-63. doi:10.1097/ccm.0000000000001346

Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G. (2014). Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Aff (Millwood), 33*(7), 1123-1131. doi:10.1377/hlthaff.2014.0041

Bellew, S. D., Cabrera, D., Lohse, C. M., & Bellolio, M. F. (2016). Predicting early rapid response team activation in patients admitted from the emergency department: The PeRRT Score. *Acad Emerg Med*. doi:10.1111/acem.13077

Berger, T., Green, J., Horeczko, T., Hagar, Y., Garg, N., Suarez, A., . . . Shapiro, N. (2013). Shock index and early recognition of sepsis in the emergency department: pilot study. *Western Journal of Emergency Medicine, 14*(2), 168.

Berwick, D. M., & Hackbarth, A. D. (2012). Eliminating waste in us health care. *Jama, 307*(14), 1513-1516. doi:10.1001/jama.2012.362

Bewick, V., Cheek, L., & Ball, J. (2004). Statistics review 13: receiver operating characteristic curves. *Crit Care, 8*(6), 508-512. doi:10.1186/cc3000

Bouldin, E. L., Andresen, E. M., Dunton, N. E., Simon, M., Waters, T. M., Liu, M., . . . Shorr, R. I. (2013). Falls among adult patients hospitalized in the United States: prevalence and trends. *J Patient Saf, 9*(1), 13-17. doi:10.1097/PTS.0b013e3182699b64

Bureau of Labor Statistics. (2017a). *Occupational Employment and Wages, 29-1069 29-1069: Physicians and Surgeons*. Division of Occupational Employment Statistics Retrieved from https://www.bls.gov/oes/current/oes291069.htm.

Bureau of Labor Statistics. (2017b). *Occupational Employment and Wages, 29-1141 Registered Nurses*. Division of Employment Statistics website Retrieved from https://www.bls.gov/oes/current/oes291141.htm.

Churpek, M. M., Yuen, T. C., Park, S. Y., Gibbons, R., & Edelson, D. P. (2014). Using electronic health record data to develop and validate a prediction model for adverse outcomes in the wards*. *Crit Care Med, 42*(4), 841-848. doi:10.1097/ccm.0000000000000038

Churpek, M. M., Yuen, T. C., Winslow, C., Meltzer, D. O., Kattan, M. W., & Edelson, D. P. (2016). Multicenter Comparison of Machine Learning Methods and Conventional Regression for Predicting Clinical Deterioration on the Wards. *Crit Care Med, 44*(2), 368-374. doi:10.1097/ccm.0000000000001571

Clifton, L., Clifton, D. A., Watkinson, P. J., & Tarassenko, L. (2011). *Identification of patient deterioration in vital-sign data using one-class support vector machines.* Paper presented at the Computer Science and Information Systems (FedCSIS), 2011 Federated Conference on.

Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMC medicine, 13*(1), 1.

Coyer, F., & Tayyib, N. (2017). Risk factors for pressure injury development in critically ill patients in the intensive care unit: a systematic review protocol. *Syst Rev, 6*(1), 58. doi:10.1186/s13643-017-0451-5

Dahn, C. M., Manasco, A. T., Breaud, A. H., Kim, S., Rumas, N., Moin, O., . . . Feldman, J. A. (2016). A critical analysis of unplanned ICU transfer within 48 hours from ED admission as a quality measure. *Am J Emerg Med, 34*(8), 1505-1510. doi:10.1016/j.ajem.2016.05.009

Dummett, B. A., Adams, C., Scruth, E., Liu, V., Guo, M., & Escobar, G. J. (2016). Incorporating an early detection system into routine clinical practice in two community hospitals. *J Hosp Med, 11*(S1).

Escobar, G., & Dellinger, R. (2016). Early detection, prevention, and mitigation of critical illness outside intensive care settings. *J Hosp Med, 11 Suppl 1*, S5-s10. doi:10.1002/jhm.2653

Escobar, G., LaGuardia, J., Turk, B., Ragins, A., Kipnis, P., & Draper, D. (2012). 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, 7*(5), 388-395. doi:10.1002/jhm.1929

Escobar, G., Turk, B., Ragins, A., Ha, J., Hoberman, B., LeVine, S., . . . Kipnis, P. (2016). Piloting electronic medical record-based early detection of inpatient deterioration in community hospitals. *J Hosp Med, 11 Suppl 1*, S18-s24. doi:10.1002/jhm.2652

Finlay, G. D., Rothman, M. J., & Smith, R. A. (2014). Measuring the modified early warning score and the Rothman index: advantages of utilizing the electronic

medical record in an early warning system. *J Hosp Med, 9*(2), 116-119. doi:10.1002/jhm.2132

Granich, R., Sutton, Z., Kim, Y. S., Anderson, M., Wood, H., Scharf, J. E., . . . Escobar, G. (2016). Early detection of critical illness outside the intensive care unit: Clarifying treatment plans and honoring goals of care using a supportive care team. *J Hosp Med, 11 Suppl 1*, S40-s47. doi:10.1002/jhm.2660

Guardia-Labar, L. M., Scruth, E. A., Edworthy, J., Foss-Durant, A. M., & Burgoon, D. H. (2014). Alarm fatigue: the human-system interface. *Clin Nurse Spec, 28*(3), 135-137. doi:10.1097/nur.0000000000000039

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143*(1), 29-36.

Higgins, J. P. (2011). Green S. Cochrane handbook for systematic reviews of interventions version 5.1. 0. *The cochrane collaboration, 5*(0).

Higgins, J. P., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., . . . Sterne, J. A. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ, 343*. doi:10.1136/bmj.d5928

Hu, S. B., Wong, D. J., Correa, A., Li, N., & Deng, J. C. (2016a). Prediction of Clinical Deterioration in Hospitalized Adult Patients with Hematologic Malignancies Using a Neural Network Model. *PLoS One, 11*(8), e0161401. doi:10.1371/journal.pone.0161401

Hu, X., Bai, Y., & Salas-Boni, R. (2016b). Recognizing predictive patterns in the sequence of superalarm triggers for predicting patient deterioration: Google Patents.

Institute of Medicine. (2015). *Dying in America: Improving Quality and Honoring Individual Preferences Near the End of Life*: National Academies Press.

Johnstone, C. C., Rattray, J., & Myers, L. (2007). Physiological risk factors, early warning scoring systems and organizational changes. *Nurs Crit Care, 12*(5), 219-224. doi:10.1111/j.1478-5153.2007.00238.x

Kho, A., Rotz, D., Alrahi, K., Cárdenas, W., Ramsey, K., Liebovitz, D., . . . Watts, C. (2007). Utility of commonly captured data from an EHR to identify hospitalized patients at risk for clinical deterioration. *AMIA Annual Symposium Proceedings, 2007*, 404-408.

Kim, Y. S., Escobar, G., Halpern, S. D., Greene, J., Kipnis, P., & Liu, V. (2016). The Natural History of Changes in Preferences for Life-Sustaining Treatments and Implications for Inpatient Mortality in Younger and Older Hospitalized Adults. *J Am Geriatr Soc, 64*(5), 981-989. doi:10.1111/jgs.14048

Kipnis, P., Turk, B. J., Wulf, D. A., LaGuardia, J. C., Liu, V., Churpek, M. M., . . . Escobar, G. (2016). Development and validation of an electronic medical record-based alert score for detection of inpatient deterioration outside the ICU. *J Biomed Inform, 64*, 10-19. doi:10.1016/j.jbi.2016.09.013

Levinson, D. R., & General, I. (2010). Adverse events in hospitals: national incidence among Medicare beneficiaries. *Department of Health and Human Services Office of the Inspector General*.

Linnen, D. (2016). The promise of big data: Improving patient safety and nursing practice. *Nursing, 46*(5), 28-34; quiz 34-25. doi:10.1097/01.nurse.0000482256.71143.09

Liu, V., Morehouse, J. W., Baker, J. M., Greene, J. D., Kipnis, P., & Escobar, G. (2016). Data that drive: Closing the loop in the learning hospital system. *J Hosp Med, 11 Suppl 1*, S11-s17. doi:10.1002/jhm.2651

Mao, Y., Chen, W., Chen, Y., Lu, C., Kollef, M., & Bailey, T. (2012). *An integrated data mining approach to real-time clinical monitoring and deterioration warning.* Paper presented at the Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining.

Mao, Y., Chen, Y., Hackmann, G., Chen, M., Lu, C., Kollef, M., & Bailey, T. C. (2011). *Medical data mining for early deterioration warning in general hospital wards.* Paper presented at the Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on.

Maslach, C., Schaufeli, W. B., & Leiter, M. P. (2001). Job Burnout. *Annual Review of Psychology, 52*(1), 397-422. doi:10.1146/annurev.psych.52.1.397

McGaughey, J., Alderdice, F., Fowler, R., Kapila, A., Mayhew, A., & Moutray, M. (2007). Outreach and Early Warning Systems (EWS) for the prevention of intensive care admission and death of critically ill adult patients on general hospital wards. *Cochrane Database Syst Rev*(3), Cd005529. doi:10.1002/14651858.CD005529.pub2

McNeill, G., & Bryden, D. (2013). Do either early warning systems or emergency response teams improve hospital patient survival? A systematic review. *Resuscitation, 84*(12), 1652-1667. doi:10.1016/j.resuscitation.2013.08.006

Merchant, R. M., Yang, L., Becker, L. B., Berg, R. A., Nadkarni, V., Nichol, G., . . . Groeneveld, P. W. (2011). Incidence of treated cardiac arrest in hospitalized

patients in the United States. *Crit Care Med, 39*(11), 2401-2406. doi:10.1097/CCM.0b013e3182257459

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS medicine, 6*(7), e1000097.

Morgan, R., Williams, F., & Wright, M. (1997). An early warning score for the early detection of patients with impending illness. *Clin Intensive Care, 8*, 100.

Nagelkerke, N. (1991). A note on a general definition of the coefficient of determination. *Biometrika, 78*(3), 691-692.

National Academy of Medicine formerly the Institute of Medicine. (1999). *To err is human: building a safer health system*. Retrieved from Washington, DC:

Parikh, R. B., Kakad, M., & Bates, D. W. (2016). Integrating Predictive Analytics Into High-Value Care: The Dawn of Precision Delivery. *Jama, 315*(7), 651-652. doi:10.1001/jama.2015.19417

Pascall, E., Trehane, S. J., Georgiou, A., & Cook, T. M. (2015). Litigation associated with intensive care unit treatment in England: an analysis of NHSLA data 1995-2012. *Br J Anaesth, 115*(4), 601-607. doi:10.1093/bja/aev285

Pastores, S. M., Dakwar, J., & Halpern, N. A. (2012). Costs of critical care medicine. *Crit Care Clin, 28*(1), 1-10, v. doi:10.1016/j.ccc.2011.10.003

Picker, D., Dans, M., Heard, K., Bailey, T., Chen, Y., Lu, C., & Kollef, M. H. (2017). A Randomized Trial of Palliative Care Discussions Linked to an Automated Early Warning System Alert. *Crit Care Med, 45*(2), 234-240. doi:10.1097/ccm.0000000000002068

Polit, D. F., & Beck, C. T. (2012). *Nursing research: Generating and assessing evidence for nursing practice*: Lippincott Williams & Wilkins.

Reason, J. (1995). Understanding adverse events: human factors. *Qual Health Care, 4*(2), 80-89.

Reason, J. (2000). Human error: models and management. *BMJ : British Medical Journal, 320*(7237), 768-770.

Richards, M. J., Edwards, J. R., Culver, D. H., & Gaynes, R. P. (2015). Nosocomial Infections in Combined Medical-Surgical Intensive Care Units in the United States. *Infection Control & Hospital Epidemiology, 21*(8), 510-515. doi:10.1086/501795

Romero-Brufau, S., Huddleston, J. M., Escobar, G. J., & Liebow, M. (2015). Why the C-statistic is not informative to evaluate early warning scores and what metrics to use. *Critical Care, 19*(1), 285. doi:10.1186/s13054-015-0999-1

Romero-Brufau, S., Huddleston, J. M., Naessens, J. M., Johnson, M. G., Hickman, J., Morlan, B. W., . . . Santrach, P. J. (2014). Widely used track and trigger scores: are they ready for automation in practice? *Resuscitation, 85*(4), 549-552. doi:10.1016/j.resuscitation.2013.12.017

Ruskin, K. J., & Hueske-Kraus, D. (2015). Alarm fatigue: impacts on patient safety. *Curr Opin Anaesthesiol, 28*(6), 685-690. doi:10.1097/aco.0000000000000260

Sikka, R., Morath, J. M., & Leape, L. (2015). The Quadruple Aim: care, health, cost and meaning in work. *BMJ Quality & Safety*. doi:10.1136/bmjqs-2015-004160

Silber, J. H., Williams, S. V., Krakauer, H., & Schwartz, J. S. (1992). Hospital and patient characteristics associated with death after surgery. A study of adverse occurrence and failure to rescue. *Med Care, 30*(7), 615-629.

Smith, M., Chiovaro, J., & O'Neil, M. (2014a). *Early Warning System Scores: A Systematic Review* Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK259031/

Smith, M. E., Chiovaro, J. C., O'Neil, M., Kansagara, D., Quinones, A. R., Freeman, M., . . . Slatore, C. G. (2014b). Early warning system scores for clinical deterioration in hospitalized patients: a systematic review. *Ann Am Thorac Soc, 11*(9), 1454-1465. doi:10.1513/AnnalsATS.201403-102OC

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., . . . Kattan, M. W. (2010). Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology, 21*(1), 128-138. doi:10.1097/EDE.0b013e3181c30fb2

Subbe, C. P., Williams, E., Fligelstone, L., & Gemmell, L. (2005). Does earlier detection of critically ill patients on surgical wards lead to better outcomes? *Ann R Coll Surg Engl, 87*(4), 226-232. doi:10.1308/003588405x50921

Torio, C. (2015). Andrews RM (AHRQ). National inpatient hospital costs: the most expensive conditions by payer, 2011. HCUP Statistical Brief# 160. August 2013. Agency for Healthcare Research and Quality, Rockville, MD. *Agency for Healthcare Research and Quality*.

Winters, B. D., Weaver, S. J., Pfoh, E. R., Yang, T., Pham, J. C., & Dy, S. M. (2013). Rapid-response systems as a patient safety strategy: a systematic review. *Ann*

Intern Med, 158(5 Pt 2), 417-425. doi:10.7326/0003-4819-158-5-201303051-00009

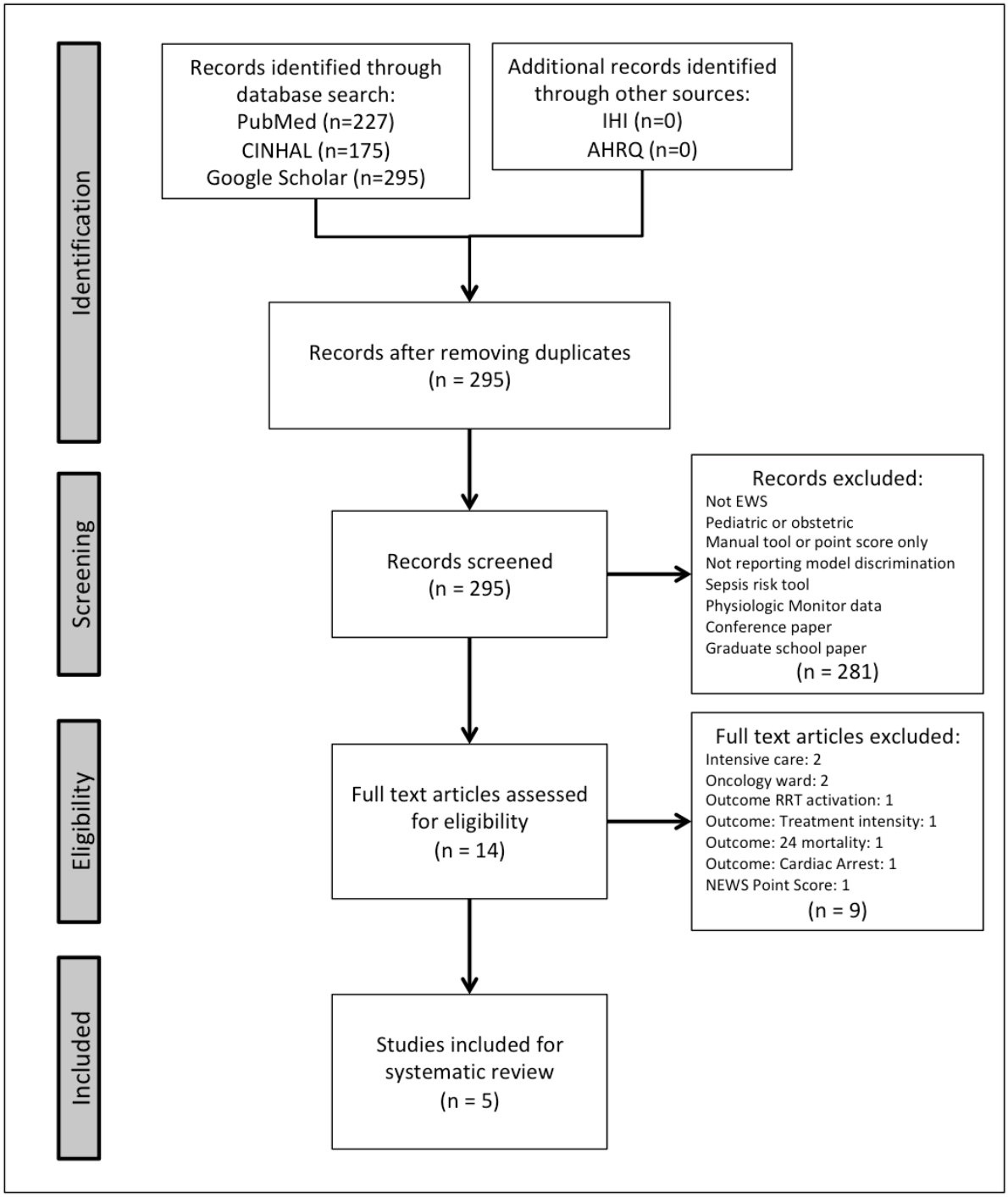
Wolters, A. E., van Dijk, D., Pasma, W., Cremer, O. L., Looije, M. F., de Lange, D. W., . . . Slooter, A. J. (2014). Long-term outcome of delirium during intensive care unit stay in survivors of critical illness: a prospective cohort study. *Critical Care, 18*(3), R125. doi:10.1186/cc13929

Yoon, J., Alaa, A., Hu, S., & Van Der Schaar, M. (2016). *ForecastICU: a prognostic decision support system for timely prediction of intensive care unit admission.* Paper presented at the Proceedings of The 33rd International Conference on Machine Learning.

Zweig, M. H., & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem, 39*(4), 561-577.

Tables and Figures

Figure 1. PRISMA Flow Diagram of Study Selection



Note. Adopted from Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA. Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement. Syst Rev. 2015;4(1):1. doi: 10.1186/2046-4053-4-1

Table 1. Screening Inclusion and Exclusion Criteria

Selection Criterion	Included	Excluded
Research Population	Hospitalized adults (≥18 years)	Adults under observation status Children and adolescents
Setting	General Medical-Surgical wards Step-Down wards	Intensive care unit Emergency room Labor & delivery Operating room Oncology ward Primary care
Timeframe	January 2012- November 2017	Prior to 2012
Method	Quantitative	Mixed method Qualitative Case reports Commentaries
Model	EMR[a]-based Multivariate regression Machine learning	Paper-based Point-score EWS[b] only
Predictors	Vitals signs Laboratory values Severity of illness scores Comorbidity scores Code Status and others	Monitor data (wave forms)
Outcome	Composite of ICU[c] transfer and mortality	RRT[d] activation Sepsis Cardiac arrest Mortality
Model Performance	AUC[e] (required) Sensitivity Specificity Positive Predictive Value RRT[d] workload (workup to detection ratio)	Risk ratios Odds ratios Chi Square ANOVA or other comparison of groups

Note.
[a] Eletronic Medical Record
[b] Early Warning System
[c] Intensive Care Unit
[d] Rapid Response Team
[e] Area Under the [Receiver Operator] Curve

Table 2. Characteristics of 5 Studies Using an Advanced Early Warning System for the Detection of Deterioration Risk

Study	Country	Location and Setting	Study Purpose/Primary Outcome	Theoretical Framework	Research Design/Analysis	Data Analysis Time Frame	Sample Size (missing data, %)	Sample Characteristics (%)
Escobar et al., 2012	United States	Northern California health system with 14 hospitals with EMR's deployed	Evaluation of Early Detection if Impending Physiologic Deterioration (EDIP) model using EMR data and comparing results against MEWS Transfer to ICU or death on wards when patient was full code	Not discussed	Retrospective case-control study	11/1/2006-12/31/2009 (3 years, 2 months)	145,335 hospitalizations in 102,422 patients, 4,036 events, 39,782 non-events (in 12h shift units)	Event/comparison: Age (mean): 67.2/65.4 Male (%): 49.7/44.5 COPS1 (mean): 116.3/100.8 LAPS1: 27.9/20.5 Predicted mortality risk (%): 1.9/5.2 Race not reported
Kipnis et al., 2016	United States	Northern California health system with 21 hospitals	Evaluation of Advanced Alarm Monitor (AAM) an automated electronic early warning system using EMR data, and comparing results against eCART (Churpek et al., 2014) and NEWS (Kovacs et al. 2016) Transfer to ICU or death on wards when patient was full code	Not discussed	Retrospective cohort study, predictive risk for death, unanticipated ICU transfer followed/not followed by a surgical intervention	1/1/2010-12/31/2013 (4 years)	649,418 episodes in 374,838 patients. 48.7 million hourly observations and 19,153 events (missing data were imputed, not discussed in detail)	Event/nonevent groups: Age (mean): 64.6/67.4 Male (%): 45.4/52.5 Comorbidity (COPS2, mean): 39.1/57.1 Physiologic instability score (LAPS2, mean): 59.1/88.5 30-day mortality (%): 4.4/24.5 Length of stay (LOS, mean): 3.6/11.8 Race not: reported

Study	Country	Location and Setting	Study Purpose/Primary Outcome	Theoretical Framework	Research Design/Analysis	Data Analysis Time Frame	Sample Size (missing data, %)	Sample Characteristics (%)
Churpek et al., 2014	United States	5 medical centers (one tertiary academic, four from a university health system)	Development and validation of a deterioration risk score using electronic medical record data; comparison of model performance against MEWS (point-score tool)	Not discussed	Retrospective cohort study	November 1, 2008 to January 2013 (4 years, 3 months)	269,999 admissions, 16,452 adverse outcomes (424 cardiac arrests, 13,188 ICU transfers, 2,840 deaths) Management/adjustment of missing: Carried previous value forward or imputed median value if no previous value was available	Derivation/validation cohort: Age (mean): 60/60 Female (%): 60/60 Race: Black (%): 17/20 White (%): 47/59 Other (%): 36/20 Mortality (%): 12/9
Churpek et al. 2016	United States	5 medical centers (one tertiary academic, two teaching, two community)	Determine different machine learning algorithms, regression models and compare results against MEWS Cardiac Arrest ICU transfer Death on ward without attempted resuscitation	Not discussed	Retrospective cohort study	November 1, 2008 to January 2013 (4 years, 3 months)	269,999 admissions, 16,452 adverse outcomes (424 cardiac arrests, 13,188 ICU transfers, 2,840 deaths on ward) Management/adjustment of missing data: Carried previous value forward or imputed median value if no previous value was available	Not discussed in-text; same dataset at Churpek (2014) above
Alvarez, et al., 2013	United States	One large academic medical center in Dallas, TX	Compare logistic regression model which adds additional predictors to MEWS vs. MEWS alone Composite: Cardiopulmonary arrest Acute respiratory compromise Unexpected death	Not discussed	Retrospective cohort study	May 2009 –March 2010 (11 months)	7,466 patients 585 cases, 6,881 controls Missing data not discussed	Not discussed for controls and cases

Table 3. Early Warning System Model Performance in 4 Studies Using Multivariate Regression or Machine Learning vs Point Scores Published Between 2012 and 2017

	Advanced Early Warning Systems				
	Kipnis et al. (2016)	**Churpek et al. (2014)**	**Churpek et al. (2016)**	**Alvarez et al. (2013)**	**Total or *Weighted mean***
Hospitalizations (n)	649,418	299,999	299,999	7,466	1,256,882
Cases (n)	19,153	16,452	16,452	585	52,642
Outcome probability[a]	0.03	0.045	0.061	0.078	0.04
Hospitals (n)	21	5	5	1	32
Time (months)	48	51	51	11	*49*
Cases per hospital per day	0.6	2.2	2.2	1.8	*1.4*
AUC[b] (95% CI)	0.82 (0.81-0.83)	0.77 (0.76-0.77)	0.8 (0.80-0.80)	0.85 (0.82-0.87)	*0.8*
Sensitivity	0.49	0.54	0.5	0.52	*0.5*
Specificity	0.92	0.9	0.93	0.94	*0.92*
PPV[c]	0.16	0.2	0.32	0.42	0.21
Workup:Detection Ratio (WDR)[d]	6.3	4.9	3.2	2.4	4.7
RRT[e] workload per hospital per day	4	10.6	6.8	4.2	6.6
Adjusted[f] PPV	0.11	0.1	0.13	0.15	0.11
Adjusted[f] WDR	9	10.1	7.9	6.7	9
Adjusted[f] RRT workload per hospital per day	3.9	7.9	6.2	3	*5.4*

	Point-Score Early Warning Systems					
	Kipnis et al. (2016)	**Churpek et al. (2014)**	**Churpek et al. (2016)**	**Alvarez et al. (2013)**	**Total or *Weighted mean***	**Simulated point score EWS estimate[g]**
Hospitalizations	649,418	299,999	299,999	7,466	1,256,882	1,256,882
Cases	19,153	16,452	16,452	585	52,642	52,642
Outcome probability[a]	0.03	0.045	0.061	0.078	0.04	0.04
Hospitals	21	5	5	1	32	32
Time (months)	48	51	51	11	49	*49*
Cases per hospital per day	0.6	2.2	2.2	1.8	1.4	*1.4*
AUC[b] (95% CI)	0.71 (0.70-0.72)	0.76 (0.75-0.78)	0.7 (0.70-0.70)	0.7 (0.70-0.70)	0.72	*0.72*
Sensitivity	0.395	0.39	0.5	0.42	0.42	**0.5**
Specificity	0.93	0.9	0.83	0.91	0.9	**0.87**
PPV[c]	0.15	0.16	0.16	0.28	0.15	0.14
WDR[d]	6.7	6.4	6.2	3.5	6.5	7.4
RRT[e] workload per hospital per day	4.3	13.9	13.4	6.3	9.1	10.4
Adjusted[f] PPV	0.1	0.07	0.06	0.09	0.08	0.07
Adjusted[f] WDR	9.7	13.6	17.7	11.5	12.8	14.7
Adjusted[f] RRT workload per hospital per day	4.2	10.6	13.9	5.2	6.9	*8*

Note.

[a] The outcome prevalence among the sample = $\frac{\text{n outcomes}}{\text{n hospitalizations}}$

[b] Area Under the [Receiver Operator] Curve

[c] Positive Predictive Value (PPV/precision): true positive cases among all positive EWS alerts; PPV changes as pre-test probability changes.

$$\text{PPV} = \frac{\text{sensitivity} * \text{prevalence}}{\text{sensitivity} * \text{prevalence} + (1-\text{specificity}) * (1-\text{prevalence})}$$

From: Altman, D. G., & Bland, J. M. (1994). Diagnostic tests 2: Predictive values. *BMJ : British Medical Journal*, *309*(6947), 102.

[d] Workup-to-Detection Ratio (WDR): the RRT evaluation workload generated to find one true positive case $(\text{WDR} = \frac{1}{\text{PPV}})$; WDR changes as pre-test probability changes

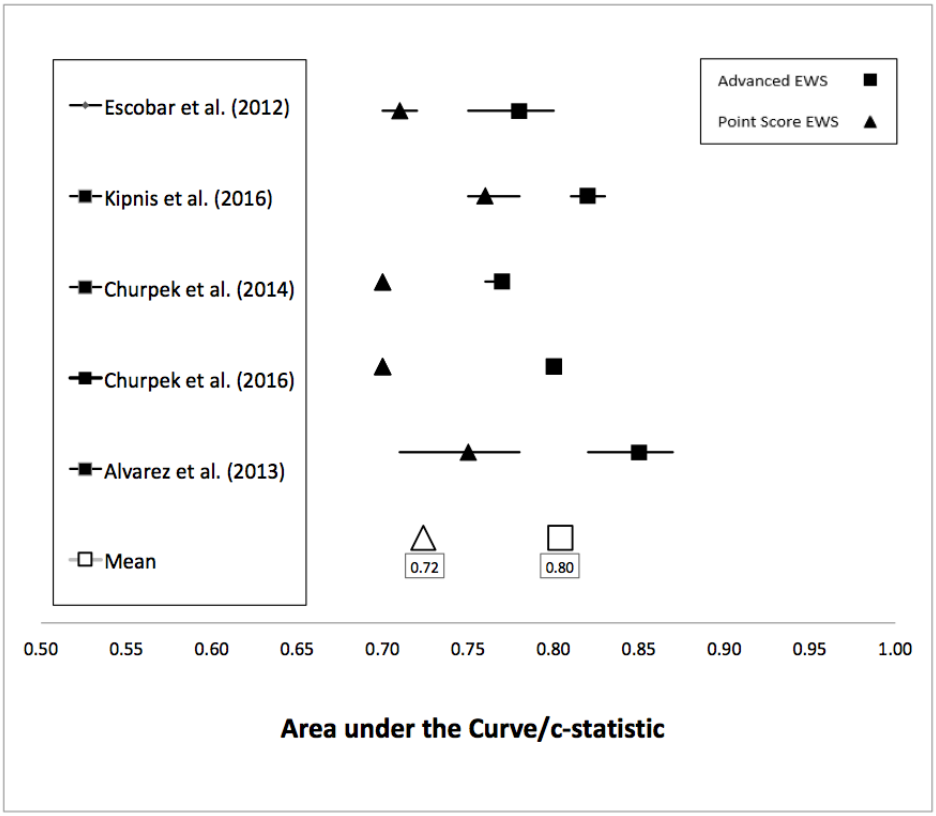
To make results comparable, we adjusted Positive Predictive Value (Precision) for an outcome prevalence (pretest-probability) of 2%

[e] Rapid Response Team

[f] To make results comparable, we adjusted WDR for an outcome prevalence (pretest-probability) of 2% across studies

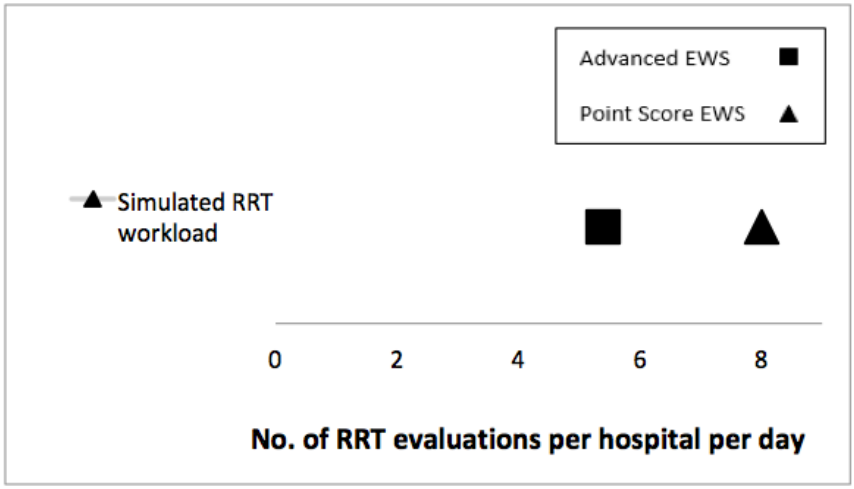
[g] We adjusted final point score EWS model workload results for a fixed outcome probability of 2%, fixed sensitivity of 50% (the mean of advanced model sensitivities), and specificity of 87%, The simulated/assumed specificity is a conservative estimate based on reported AUC graphs or reported model cutoffs.

Figure 2. Summary Results of 5 Studies Reporting Area Under the Curve (AUC) of Advanced Early Warning System Models vs Point-score Models from 2012 to 2017



Note. AUC describes the models' ability to predict an outcome accurately, with 0.50 indicating no ability to predict an outcome. Mean summary values are weighted according to the studies' sample size and time frame.

Figure 3. Adjusted Results of 5 Studies of Rapid Response Team (RRT) and Hospitalist Workload (Patient Evaluations per Hospital per Day) Generated by Advanced Early Warning System (EWS) Models vs Point-Score Tools from 2012 to 2017



Note. Results show the number of RRT evaluations generated per hospital per day to find one deterioration case adjusted for a fixed outcome prevalence of 2%, a fixed sensitivity of 50%, weighted mean specificity of 0.92 for advanced EWSs and estimated specificity of 87% for point score EWSs. To minimize bias, the estimate adjusts for the weighted mean of sample size and study timeframes.

Appendix

Table 1. Literature Search Keywords to Identify Early Warning System Studies Using Regression or Machine Learning for Inclusion in a Systematic Review (2012-2017)

PubMed and CINAHL
"early warning score OR early warning system AND deterioration OR predict transfer ICU" Search details: (early[All Fields] AND warning[All Fields] AND score[All Fields]) OR (early[All Fields] AND warning[All Fields] AND system[All Fields]) AND deterioration[All Fields] OR (predict[All Fields] AND ("transfer (psychology)"[MeSH Terms] OR ("transfer"[All Fields] AND "(psychology)"[All Fields]) OR "transfer (psychology)"[All Fields] OR "transfer"[All Fields]) AND ("intensive care units"[MeSH Terms] OR ("intensive"[All Fields] AND "care"[All Fields] AND "units"[All Fields]) OR "intensive care units"[All Fields] OR "icu"[All Fields])) AND ("2012/01/01"[PDAT] : "2017/11/01"[PDAT]) (227 results) After removal of "psychology" references: Search details: (early[All Fields] AND warning[All Fields] AND score[All Fields]) OR (early[All Fields] AND warning[All Fields] AND system[All Fields]) AND deterioration[All Fields] OR (predict[All Fields] AND "transfer"[All Fields]) AND ("intensive care units"[MeSH Terms] OR ("intensive"[All Fields] AND "care"[All Fields] AND "units"[All Fields]) OR "intensive care units"[All Fields] OR "icu"[All Fields]) AND ("2012/01/01"[PDAT] : "2017/11/01"[PDAT]) (117 results)
Google Scholar
"early warning system clinical deterioration"
"early warning system clinical deterioration machine learning"

Table 2. Measures of Model Performance

Measure Name	Description	Formula
Pre-test probability	Prevalence: % of those with the outcome among the sample	$\frac{\text{cases}}{\text{entire sample}}$
Pseudo-R^2 [a]	% of variation explained by the model	
Sensitivity	% true positive cases among all positive cases	$\frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$
Specificity	% true negative cases among all negative cases	$\frac{\text{true negatives}}{\text{true negatives} + \text{false positives}}$
PPV	% true positive cases among all positive tests	$\frac{\text{true positives}}{\text{true positives} + \text{false positives}}$ or $\frac{\text{sensitivity} * \text{prevalence}}{\text{sensitivity} * \text{prevalence} + (1 - \text{specificity}) * (1 - \text{prevalence})}$
AUC/c-stat	True positive (TP) rate plotted against false positive (FP) rate	$\frac{\text{Number of concordant pairs}}{\text{Total number of pairs}} + 0.5 * \frac{\text{Number of tied pairs}}{\text{Total number of pairs}}$
Workup-to-Detection	Workload measure: Number needed to evaluate to find one positive case	$\frac{\text{true positives} + \text{false positives}}{\text{true positives}}$ or $\frac{1}{\text{WDR}}$
RRT evaluations per hospital per day	Workload measure: The total number of patients RRTs need to evaluate	$\text{WDR} * \frac{\text{cases}}{\frac{\text{hospitals}}{\text{day}}}$

Note.
[a]Logistic regression does not use R^2 but Likelihood ratio R^2, Cox and Snell R^2, Nagelkerke R^2 or others.

Table 3. Predictive Model Characteristics and Model Performance of 5 Advanced Early Warning Systems Using Multivariate Regression or Machine Learning to Identify Clinical Deterioration Risk

Study	Early Warning prediction method/predictor variables	Reference standard used	Sensitivity (%)	Specificity (%)	Positive Predictive Value (PPV, %)	Negative Predictive Value (NPV, %)	AUROC/c-statistic	Calibration metric	Workup to Detection Ratio	Relevant Findings	Strengths	Limitations
Escobar et al., 2012	Laboratory tests, vitals signs, shock index, age, sex, LAPS1, COPS1, admission diagnosis, admission type, code status, length of stay	Non-events were comparison	Not reported	Not reported	Not reported	Not reported	.78 for EDIP (Ranging from .68 to .84 across diagnostic strata) .70 for MEWS (ranging from .54 to .79 across diagnostic strata)	Not discussed	Workup volume for MEWS threshold of >=6: EDIP: 14.5 false alarms for each ICU transfer MEWS: 34.4 false alarms for each transfer	EDIP outperformed manual MEWS scoring system in all models	Very large dataset, very complex risk adjustment, very precise variables and methods	Large integrated health system with fully integrated EMR, computational infrastructure limited, unable to determine if ward patient should have been a ICU admit
Kipnis et al., 2016	Laboratory test values, vital signs, comorbidity composite (COPS2), acute physiological instability index (LAPS2), Length of stay, age, sex, code status, time of day, season, admit category, hospital	Non-events were comparison	.38-.56 (across medical centers)	.88-.95 (across medical centers)	.11-.23 (across medical centers)	.97-.99 (across medical centers)	.82 for AAM .79 for eCART .76 for NEWS	Hosmer-Lemeshow p-value for calibration	Developed workup-to-detection ratio and also defined an operational alarm cutoff so that model was calibrated against maximum of one alarm per 35-bed unit per day	AAM performed better than eCART and NEWS, likely because model was more complex	Very large dataset, very complex risk adjustment, very precise variables and methods	Large integrated health system, computational infrastructure limited method (now working on more machine learning models)

Study	Early Warning prediction method/predictor variables	Reference standard used	Sensitivity (%)	Specificity (%)	Positive Predictive Value (PPV, %)	Negative Predictive Value (NPV, %)	AUROC/c-statistic	Calibration metric	Workup to Detection Ratio	Relevant Findings	Strengths	Limitations
Churpek et al., 2014	Patient demographics, vital signs, mental status, laboratory test values,	Non-events were comparison	.16-.89 depending on model risk score cutoff .54 at model score of ≥17	.52-.99 depending on model risk score cutoff .90 at model score of ≥17	Not reported	Not reported	.077 for eCART (combined outcomes) .70 for MEWS	Calculated predicted event probability. Did not discuss a calibration metric	Did not discuss workup or similar workload metric for selected risk score cutoff	eCART performed substantially better than MEWS, likely because model was more complex	Large dataset, complex set of covariates, very detailed analytic approach	Did not include comorbidity or severity of illness score, did not discuss workload generated by score
Churpek et al. 2016	Age, Length of stay, number of prior ICU stays, vital signs, laboratory values	Non-events were comparison	Not reported	Not reported	Not reported	Not reported	Random forest: .80 Gradient boosted machine: .79 Bagged trees: .79 Support vector machine: .79 Neural network: .78 Logistic regression: .77 K-nearest neighbor: .75 Logistic regression (linear): .74 Decision tree: .73 MEWS: .70	Hosmer-Lemeshow p-value for calibration and O/E plotting	At 75% sensitivity level Random Forrest model would screen 13% fewer than logistic linear model or more than 500,000 fewer screens out of a pool of 4.6 million observations	Machine learning algorithms were superior to traditional regression models and both RF and GBM had very good discrimination and calibration	Introduced novel “data science” machine learning methods that show superior performance to traditional supervised predictive analytics approaches (regression). Large sample size.	Black box output (clinicians cannot understand why a patient scores high). Composite outcome does not seem to account for expected deaths.

Study	Early Warning prediction method/predictor variables	Reference standard used	Sensitivity (%)	Specificity (%)	Positive Predictive Value (PPV, %)	Negative Predictive Value (NPV, %)	AUROC/c-statistic	Calibration metric	Workup to Detection Ratio	Relevant Findings	Strengths	Limitations
Alvarez, et al., 2013	Laboratory data, vital signs, level of consciousness, STAT orders, STAT medications, MEWS, "high risk floor"	Non-events were comparison	Regression Model .52 MEWS; .42	Regression Model .94 MEWS .91	Regression Model .10 MEWS .06	Regression Model .99 MEWS .99	Regression Model .85 MEWS .75	Hosmer-Lemeshow p-value for calibration	Median number of alarms per day: 9 Median number of RRT calls per day: 2	The automated EMR model performed better than MEWS alone and reduced number of false positive alarms. The model was twice as sensitive as manual RRT activation (.52 vs. .26) and trigger 5.7 hours sooner than RRT.	Provided important clinical comparison of RRT activation (human vigilance) and basic MEWS. Demonstrated that EWS accuracy can be improved by regression techniques.	Single center study with small cohort. Included all wards deaths as "unexpected"

Table 4. Level of Scientific Evidence and Risk of Bias Assessment

Study	Level of Scientific Evidence based on Research Design (1: High - 7: Low)	Measurement Bias Systematic differences in applying measurement	Detection Bias Systematic differences in outcome measurement	Missing data Bias Systematic differences in data sets	Threats to External Validity	Total Score Presence of Bias)
Escobar et al., 2012	4 (not used in score)	0: Used sophisticated adjustment techniques to account for confounding and validated model on hold-out dataset	0.5: Outcome was clearly defined and clinical variations in patient presentation were included. Though nearly impossible to design programmatically, a small fraction of the conceptualized events may not have been appropriate ward admissions (error in judgment). Patient may have entered as full code but became an "appropriate" death event (palliative care)	0: Discussed, models with imputed data and dropped observations were compared	0.5. Large health system study in integrated care delivery network in NCAL. Plan members may be receiving better care at baseline. NCAL demographic and income/SES not generalizable to all settings (limited to similar metropolitan regions with similar make-up)	1.0
Kipnis et al., 2016	4 (not used in score)	0: Used sophisticated adjustment techniques to account for confounding and validated model on hold-out dataset	0.5: Outcome was clearly defined and clinical variations in patient presentation were included. Though nearly impossible to design programmatically, a small fraction of patients may have had a first RRT call or code blue event without the outcome but subsequent deterioration. There were no data used for RRT activation or code blue.	0: Discussed, missing data were imputed	0.5. Large health system study in integrated care delivery network in NCAL. Plan members may be receiving better care at baseline. NCAL demographic and income/SES not generalizable to all settings (limited to similar metropolitan regions with similar make-up)	1.0
Churpek et al., 2014	4 (not used in score)	0.5: Used sophisticated adjustment techniques to account for confounding and validated model on hold-out dataset Data were from two different health systems, potential for different documentation standards was not discussed	1: Outcome was clearly defined and clinical variations in patient presentation were included. Though nearly impossible to design programmatically, a small fraction of the conceptualized events may have been misclassified Confirmation bias: Conflicts of interest: One researcher disclosed honoraria from a clinical alarm vendor	0: Discussed, missing data were imputed	0.5. Small-medium health system study (5 hospitals). Demographics reported in Table 1	2.0

Study	Level of Scientific Evidence based on Research Design (1: High - 7: Low)	Measurement Bias Systematic differences in applying measurement	Detection Bias Systematic differences in outcome measurement	Missing data Bias Systematic differences in data sets	Threats to External Validity	Total Score Presence of Bias)
Churpek et al. 2016	4 (not used in score)	0.5: Used sophisticated adjustment techniques to account for confounding and validated model on hold-out dataset Data were from two different health systems, potential for different documentation standards was not discussed	1: Outcome was clearly defined and clinical variations in patient presentation were included. Though nearly impossible to design programmatically, a small fraction of the conceptualized events may have been misclassified Confirmation bias: Conflicts of interest: Two researchers have a patent pending for a risk algorithm that may become commercially available. One researcher disclosed honoraria from a clinical alarm vendor	0: Discussed, missing data were imputed	0.5. Small-medium health system study (5 hospitals). Demographics not reported in text but same sample as 2014 paper	2.0
Alvarez, et al., 2013	4 (not used in score)	1: Used sophisticated adjustment techniques to account for confounding and validated model on hold-out dataset Unexpected death definition included those made DNR or comfort care (an inaccurate measurement) Neuro status was based on natural language processing search of nursing notes (validity and reliability not discussed)	0.5: Outcome was clearly defined and clinical variations in patient presentation were included. Though nearly impossible to design programmatically, a small fraction of the conceptualized events may have been misclassified	1: Missing data not discussed, though likely a concern	1: Single center study Did not report demographics of cases and controls	3.5

Note. Adopted from Higgins et al. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomized trials

Chapter 3

Alert-Driven Sepsis Interventions by Rapid Response Teams among Adult Patients on General Hospital Wards: Results of Clinical Chart Reviews

Daniel T. Linnen, PhD(c), MS, RN-BC[1,2]

Gabriel J. Escobar, MD[2]

John Greene, MA[2]

Vincent Liu, MD, MS[2]

Xiao Hu, PhD[3]

Caroline Stephens, PhD, RN, GNP[4]

[1] University of California San Francisco, School of Nursing, Department of Social & Behavioral Sciences

[2] Kaiser Permanente Northern California, The Permanente Medical Group, Inc., Division of Research

[3] University of California San Francisco, School of Nursing, Department of Physiological Nursing

[4] University of California San Francisco, School of Nursing, Department of Community Health Systems

Abstract

Background: Advanced early warning systems (EWS) can detect worsening sepsis in hospitalized patients sooner than traditionally possible. The timing and scope of clinical sepsis interventions relative an advanced EWS alert, however, are not well understood.

Objectives: To understand how sepsis decedents differed from survivors in terms of alert-driven receipt of fluid bolus therapy, new antibiotics, DNR status, and ICU transfer.

Methods: We conducted electronic chart reviews of 68 sepsis patients with an EWS alert admitted to Kaiser Permanente Northern California hospitals. In round 1, we reviewed the charts of 21 sepsis patients admitted to 3 hospitals between 8/1/2016 and 2/28/2017. In round 2, we reviewed charts of 47 patients admitted to 5 hospitals between 1/1/2017 and 7/31/2017. We abstracted demographic/clinical characteristics, and process measures of sepsis intervention, and performed summary statistics.

Results: Sepsis decedents were older and sicker at admission and alert time. Decedents received 32% less total fluid bolus volume and received the first bolus 4.6 hours later than survivors. Most first alerts occurred near admission; most sepsis interventions occurred before the first alert. Of 14 decedents, 12 (86%) had a DNR order before death.

Discussion: Survivors received more total fluid bolus volume, and sooner. Though limited by sample size, our results are consistent with prior studies regarding the protective effect of timely fluid bolus therapy on sepsis survival. Fluid bolus therapy and new intravenous antibiotics frequently occurred before the alert suggesting a potential treatment overlap between sepsis care initiated in the ED and the first alert within a few hours after admission. Our findings may also reflect that some decedents were on a

palliative/end-of-life trajectory. Demographics, clinical characteristics, sepsis interventions, and DNR orders of sepsis decedents differ from survivors, highlighting the need to adjust for these differences and other time process measures. Future research should exclude patients with a DNR order within the first 4 hours of admission and investigate the impact of RRT sepsis interventions on patient survival using adequately powered sample sizes and rigorous design and analytic methods.

Introduction

In US hospitals, sepsis claims more lives than any other medical condition. An estimated 900,000 to 3 million patients are diagnosed with sepsis in the United States annually (Gaieski, Edwards, Kallan, & Carr, 2013), and an estimated 180,000 to 750,000 patients die of sepsis each year (Gaieski et al., 2013; Epstein, Dantes, Magill, & Fiore, 2016). Evidence suggests that more than half of all inpatient deaths are the result of sepsis (Engel et al., 2007; Liu et al., 2014). Sepsis costs the U.S. health system an estimated $14 billion annually (Healthcare Cost Utilization Project, 2008). Sepsis patients may be exposed to prolonged hospitalizations, which puts some at risk for hospital acquired conditions, including infections (Richards, Edwards, Culver, & Gaynes, 2015), pressure injuries (Coyer & Tayyib, 2017), falls with injury (Bouldin et al., 2013), among others.

To prevent clinical deterioration in sepsis patients, two fundamental elements are early recognition and rapid clinical evaluation and treatment (Rozen & Butt, 2016). Various methods of detecting and responding to worsening sepsis exist. Traditional sepsis screening tools, which are either paper-based or automated in the electronic medical record, have used a limited set of clinical variables, including systematic inflammatory response syndrome criteria and evidence of organ dysfunction (e.g., hypotension, altered mental status) (Levy et al., 2003). Meeting such criteria may result in a “code sepsis” or activation of a rapid response team (RRT) to evaluate and treat the patient expediently. However, relying on the presentation of organ dysfunction for screening is problematic because it delays the recognition of worsening sepsis until a patient’s condition has already deteriorated. Consequently, RRTs and the

recommended sepsis interventions may have limited potential to stall a patient's clinical decline and death (Chan, Jain, Nallmothu, Berg, & Sasson, 2010; Solomon, Corwin, Barclay, Quddusi, & Dannenberg, 2016).

In contrast, advanced early warning systems (EWSs) using multivariate regression or machine learning can detect deterioration risk preemptively (Escobar et al., 2012b; Churpek et al., 2014; Linnen, 2018), hours before a patient would present with overt signs of worsening sepsis. A major benefit of such EWSs is that these algorithms permit RRTs to intervene proactively. We have shown in previous work that advanced EWSs are more precise and reduce the RRT workload by generating fewer total alerts per day (Linnen, 2018).

Current *Surviving Sepsis* guidelines (Dellinger et al., 2013) recommend an initial fluid resuscitation goal of 30ml/kg of intravenous crystalloids, such as Normal Saline or Lactated Ringer's solutions, following the time when a patient's clinical presentation suggests worsening sepsis. Early aggressive fluid bolus therapy may be associated with sepsis survival, although the evidence base is weak (Levy et al., 2003; Engel et al., 2007; Dellinger et al., 2008; Liu et al., 2014). Other common sepsis interventions include prescription and administration of a new intravenous antibiotic, transfer to the intensive care unit (ICU), and intensified surveillance (e.g., vital sign measurements and laboratory tests). However, the evidence base for the impact of advanced EWS alerts combined with proactive RRT interventions on sepsis survival, too, is weak. In addition, not every patient can be or wants to be rescued (Escobar & Dellinger, 2016) in the context of terminal disease or end of life (Granich et al., 2016). Little is known about how such sepsis patients might be appropriately and reliably excluded from mortality

measurements and how to retain a clean denominator of “rescuable” patients to measure sepsis care quality.

In this exploratory descriptive study, we summarize the results of two rounds of clinical chart reviews of the electronic medical records of hospitalized sepsis patients. Our study attempted to answer five questions to inform the design and methods of a quantitative study using a larger sample, which quantified the impact of EWS-based RRT interventions on sepsis survival: How do the demographics and clinical characteristics of sepsis decedents differ from those of survivors? How do sepsis interventions around the alert time differ in these groups? How frequently do RRT interventions occur before vs. after the alert? Could certain DNR order patterns be operationalized as a reliable exclusion criterion for “expected death” (i.e., where death was not a bad outcome)? Where are sepsis interventions and other process markers located in the electronic medical record for future data extraction? Finding answers to these questions is of twofold importance: First, gaining an understanding of the underlying data and the measurement concepts they represent, informs whether measurement approaches are valid and accurate. Second, the evidence base regarding whether (and which) specific alert-driven sepsis interventions may improve patient survival is very scarce. By examining measurement validity, chart reviews ensure that research findings in future studies do in fact produce the best attainable knowledge.

Methods

Setting and sampling

We conducted two rounds of clinical chart reviews of adult patients with a positive EWS alert who were admitted with sepsis to a general hospital ward. We deemed these patients clinically stable by virtue of their ward admission. We elected to divide chart reviews into 2 rounds. We anticipated that, based on findings from the first round, data abstraction and the way we defined concepts would evolve. Nevertheless, the data abstraction process was standardized in both rounds of reviews, meaning that all information was captured uniformly for all patients in the samples. The setting was Kaiser Permanente Northern California (KPNC), a large integrated health system with a feature-rich electronic medical record deployed in all hospitals as previously described (Escobar, LaGuardia, Turk, Ragins, Kipnis, & Draper, 2012). Of the 21 hospitals in Northern California, we randomly selected patients from hospitals that had implemented the advanced early warning system, *Advance Alert Monitor* (Escobar et al., 2012a; Escobar & Dellinger, 2016; Granich et al., 2016). In round 1 of chart reviews, we randomly selected 10 survivors and 10 decedents per hospital from a pool of patients (n=60) who were admitted to one of 3 KPNC hospitals between August 1, 2016 and February 28, 2017. This sample included all admission diagnoses, of which we excluded non-sepsis admissions (n=30) before chart reviews. During chart reviews, we flagged an additional 9 patients with a false positive EWS alert and excluded them before analysis. We defined false positive alerts as those where clinical notes indicated no worsening of the patient's condition and no sepsis interventions followed the alert. The final round 1 sample included 21 patients. In round 2, we randomly selected 5

survivors and 5 decedents with a sepsis admission diagnosis from 5 KPNC hospitals each (i.e., 10 patients per hospital) between January 1, 2017 and July 31, 2017. More hospitals had implemented the advanced EWS by that time. Of the 5 hospitals included in this time frame, 2 had fewer than 5 sepsis deaths; the final round 2 sample included 47 patients (22 sepsis decedents and 25 survivors).

Inclusion criteria for both samples were age ≥ 18 years, a positive EWS alert, Kaiser Permanente membership (it is not always possible to capture 30-day mortality in non-members), a hospital stay of at least 24 hours, and admission to a general hospital ward. We excluded patients with overnight surgery stays and patients admitted to labor and delivery or the ICU. We flagged patient records if, upon the alert, clinicians described the patient's condition as stable and no sepsis interventions followed. After data abstraction, we deemed these cases false positives and excluded them from analysis. The Institutional Review Boards of Kaiser Permanente Northern California and the University of California, San Francisco approved this study.

Composite measures

The study employed composite measures of *severity of illness at admission* [LAboratory-based acute Physiology Score (LAPS2)] (Escobar et al., 2012b) and *1-year comorbidity burden* calculated one month prior to admission [COmorbidity Point Score version 2 (COPS2)] (Escobar, Gardner, Greene, Draper, & Kipnis, 2013). LAPS2 measures a patient's physiological state including laboratory tests, vital signs, and the shock index (heart rate ÷ systolic blood pressure) (Berger et al., 2013). Van Walraven, Escobar, Greene, and Forster (2010) validated LAPS2 in an external patient population. The COPS2 score is a measure of a patient's comorbidity burden, defined as the

number of concurrent chronic health conditions documented in the previous 12 months. Stevens and Howell (2015) validated COPS2 externally. Previous health services studies have employed both LAPS2 and COPS2 scores (Escobar et al., 2012a; Escobar & Dellinger, 2016) and they appear to be valid and reliable measures of their respective constructs.

Data abstraction

In both rounds of chart reviews, the principal outcome was 30-day survival vs. mortality following an admission for sepsis. We defined sepsis admissions as those with ICD-9 codes of sepsis (995.91), severe sepsis (995.92), septic shock (785.52), septicemia (038) or bacteremia (790.7). While, the measurement of sepsis cases using administrative data is variable and a source of debate among clinicians and scholars (Sprung & Trahtemberg, 2017), we deemed this definition adequate to identify the majority of inpatient sepsis admissions. In addition, given the absence of a gold standard definition (Angus et al., 2016), using the above codes allowed us to extract patient samples for which sepsis was defined relatively unambiguously. After patient selection, we extracted the following data electronically: Hospital ID, 30-day survival, medical record number, patient ID, timestamps for admission and EWS alert time, time from admission to alert, severity of illness score (LAPS2) at ward admission and at alert time, the difference between these two LAPS2 scores (LAPS2 delta), comorbidity score (COPS2) at admission, and admission diagnosis. A clinical content expert [DL] then conducted the clinical chart reviews and entered data using a standardized electronic data abstraction tool.

In round 1, we manually captured patient age, sex, Do Not Resuscitate (DNR) order time, narrative clinical notes by RRT and the attending physician, sepsis intervention types, and intervention time relative to the alert. We captured RRT sepsis interventions in binary form as given vs. not given starting at alert time and up to 24 hours after the alert. In round 2, we evaluated time measures of sepsis interventions relative to alert time in continuous form. To account for all sepsis interventions that occurred in reasonable proximity to the alert, we included fluid bolus therapy and new intravenous antibiotic administrations within 12 hours before and 24 hours following the alert. We defined a new antibiotic as one that was not administered within 24 hours before the alert. In addition to the above variables, we abstracted time from alert to fluid bolus, total fluid bolus volume administered, new intravenous antibiotic administration, time from alert to new intravenous antibiotic administration, and time from alert to ICU transfer.

Results

Following are the results of round 1 and round 2 chart reviews. Table 1 summarizes demographics and clinical characteristics of sepsis decedents and survivors.

Round 1: Patient demographics and clinical characteristics

After we applied exclusions, 21 patients remained in the round 1 sample (14 sepsis decedents and 7 survivors). Men were somewhat overrepresented among sepsis decedents and underrepresented among survivors (64.3% vs. 42.9%, respectively). Sepsis decedents were slightly older than survivors, with a mean age (in years) of 78.4 vs. 75.0 and a median age of 82.5 vs. 81.0. Compared to survivors, decedents had

higher severity of illness scores (LAPS2) at admission and alert time (mean 125.9 vs. 96.0 at admission, and 153.1 vs. 124.6 at alert). The median score difference (LAPS2 delta) between these two measurements was higher in decedents (median LAPS2 delta of 22.5 vs. 12.0), though the mean LAPS2 was slightly higher in survivors (mean LAPS2 delta 27.2 vs. 28.6). Patient comorbidity burden scores (COPS2) were higher in decedents than survivors (mean 82.9 vs. 51.4).

Round 1: Sepsis interventions and DNR orders

Table 2 summarizes RRT sepsis interventions and DNR order patterns across decedents and survivors. Among decedents, somewhat more time (in days) elapsed between admission and the first advanced EWS alert (mean 1.6 vs. 1.0), though the median time was similar between decedents and survivors (median 0.6 vs. 0.5). In round 1, we counted sepsis interventions exclusively if they followed the alert and did not include interventions that occurred before the alert. Using this methodology, 7 decedents (50%) received a fluid bolus vs. none in the survivors. New antibiotic administration after the alert occurred in 1 decedent (7.1%) vs. in 2 survivors (28.6%). Transfer to the intensive care unit occurred in none of the decedents vs. in 2 survivors (28.6%). New intravenous antibiotic administrations documented in the electronic medication administration record were vancomycin, piperacillin-tazobactam, and ampicillin-sulbactam.

Of the 14 sepsis decedents, 12 (85.7%) had a DNR order before death. Of those with a DNR order, 6 (50%) had an order in place before the alert fired, and 3 (25%) had a DNR order within 3 days after the alert (Figure 1). None of the survivors had a DNR order placed during their hospitalization. In sepsis decedents, the time from alert to

DNR order (in days) was 1.7 days (median 1.1), and the range was wide (about -2 days to 7 days relative to the alert). The mean time from DNR to death was 10.2 days (median 4.4), and the range was wide (less than 1 day to 30 days).

Round 2: Patient demographics and clinical characteristics

In round 2, we reviewed the clinical charts of 43 patients (21 sepsis decedents and 22 sepsis survivors). Men were underrepresented among both decedents and survivors (42.9% vs. 36.4). Sepsis decedents were older, and the age differential was more pronounced than in round 1 (mean age 75.2 vs. 69.8). Similar to round 1, severity of illness scores (LAPS2) at admission and alert time, the LAPS2 delta, and comorbidity burden scores (COPS2) were higher in decedents compared to survivors (mean LAPS2 of 115.6 vs. 88.1 at admission and 155.7 vs. 117.0 at alert; mean LAPS2 delta of 40.1 vs. 28.9; mean COPS2 of 73.5 vs. 63.4).

Round 2: Sepsis interventions

The time from admission to first alert (in days) was similar to that observed in round 1 (mean 1.6 vs. 0.6; median 0.6 vs. 0.4). Of the 21 sepsis decedents, 14 (66.6%) received at least one fluid bolus compared to 17 survivors (77.3%). On average, the time from alert to the first fluid bolus was negative (i.e., it occurred before the alert). Fluid bolus administration began earlier in survivors than decedents, and frequently occurred before the alert (mean -5.3 hours in decedents vs. -9.9 hours in survivors). The mean time from alert to the second bolus (in days) was short in both groups, and it appeared that in at least half of survivors the 2nd fluid bolus preceded the alert. (0.4 vs. 0.7). In total, sepsis decedents received a total of 23 liters of fluid bolus therapy vs. 37.5 liters in survivors.

Nearly all patients received a new intravenous antibiotic within the alert time frame, and most received a new antibiotic before the alert. 17 sepsis decedents (81.0%) received a new antibiotic compared to 22 survivors (100.0%). The mean time from alert to new intravenous antibiotic (in hours) was -2.8 in decedents vs. -7.6 in survivors. Transfer from the ward to the intensive care unit occurred infrequently. Of the 21 sepsis decedents, we observed 6 transfers (28.6%) compared to 3 transfers in survivors (13.6%). The mean time from alert to transfer (in days) was 3.2 in decedents vs. 1.0 in survivors.

Discussion

Results of this clinical chart review study suggest that, on average, sepsis decedents are somewhat more frequently male, older, and present with higher severity of illness both at admission and alert time. LAPS2 scores suggest that decedents in this study were sicker at baseline and declined more rapidly; their presentation at admission time was about as severe or more severe than the presentation of survivors at alert time. Given that the median time from admission to alert was about 12 hours, it is conceivable that the higher severity of illness in sepsis decedents is at least in part the result of a longer duration of untreated sepsis before emergency department entry. The higher LAPS2 delta scores in decedents also suggest that their clinical decline may occur somewhat more rapidly. Sepsis decedents had substantially higher comorbidity burden scores (COPS2); a 46% relative median increase in decedents compared to survivors in round 1 and a 29% relative median increase in round 2. These findings are plausible and consistent with prior research demonstrating an association of comorbidity with sepsis mortality (Whiles, Deis, Miller, & Simpson, 2016).

Sepsis interventions

Perhaps the most pronounced finding was that decedents received substantially less fluid bolus volume and that administration time occurred later, compared to survivors. In total, sepsis decedents received 32% less fluid bolus volume than survivors, though it is important to note that these results are unadjusted for potential differences between decedents and survivors. On average, decedents received the first fluid bolus 4.6 hours later than survivors, and the median time difference was 1.4 hours. Although at least some of the decedents may have been on a palliative end-of-life trajectory, this finding appears to agree with prior evidence regarding the protective effect of timely fluid bolus therapy on sepsis survival.

Results from round 2 suggest that both fluid bolus therapy and new intravenous antibiotic administration frequently occur before the alert. This finding appears to contradict the main purpose of studying the impact of an advanced EWS, which is to predict deterioration risk prospectively, to alert clinicians, and to motivate additional sepsis interventions. Our results suggest that many first alerts on the ward occurred immediately following admission from the emergency department (see Figure 2). One explanation may be that newly admitted patients receive an alert after laboratory data become available for the first time. However, the initial bundle of sepsis treatment may have just occurred in the ED, a few hours before admission. It is known that the initial presentation of sepsis (and the first bundle of sepsis treatment) occurs in the emergency department, at times hours after ED entry because of an unclear clinical presentation (Villar et al., 2014). Thus, there may be a potential treatment overlap between sepsis treatment in the emergency department and the first alert within a few

hours after admission to the ward. To that end, in order to evaluate the impact of an EWS alert and the additional RRT sepsis interventions it may motivate, we recommend an alert exclusion time period of 12 hours following emergency department entry (see Figure 2). With regard to length of stay, our results suggest that measuring any occurrence of "worsening sepsis" relative to an advanced alert on any hospital day may introduce considerable measurement bias. For example, patients admitted with sepsis in stable condition with an alert on day 2 may present with a different clinical picture (worsening sepsis is an acute problem) than patients with an alert on day 7 (sepsis is a reoccurring problem).

DNR order patterns

We observed that nearly all decedents in round 1 received a DNR order before their death, and half of all decedents in round 1 had a DNR order in place before the alert. There was a wide range of alert-to-DNR and DNR-to death process measure times. While this range somewhat limits interpretability, it generally appeared that patients had a DNR order in close proximity to the alert time and then again 2 to 3 days after the initial alert (Figure 1). From a measurement perspective, it is desirable to exclude patients for whom death is expected because their treatment goals may be palliative in nature (Escobar & Dellinger, 2016) and because it is known that patients may be unclear about their treatment wishes before hospital admission (Stephens et al., 2015; Stephens et al., 2018). In in the majority of decedents, clinicians appeared to address patient-centered decisions regarding code status and goals of care before death. Yet, knowing that nearly all decedents received a DNR order before death, an unrestricted exclusion of such patients would not result in a valid measure of "expected

death." However, it appears reasonable to exclude patients who received a DNR order within the first 4 hours following admission.

Our study had several limitations. First, both chart review samples were small, and we did not intend or power them to detect statistically significant differences between decedents and survivors. The goal of this report was not to conduct statistical difference testing but to answer preliminary questions regarding sepsis interventions following the EWS alert and to inform the design of a quantitative study using a larger sample. Second, our results must be viewed with some caution because we did not employ methods to ensure alike cohorts. Decedents and survivors may have differed in terms of their severity of illness and comorbidity, among other factors. However, we employed random sampling to minimize selection bias and all patients were stable enough for admission to a general hospital ward. Third, we did not exclude patients for whom death may have been expected. This approach again limits the interpretation of our results because the decedent group may contain patients who did not wish to receive aggressive RRT interventions. However, one of our goals was to evaluate DNR order patterns relative to alert and death. In addition, measuring the quality of mortality and survival in acute care health services research is a vexing issue (Granich et al., 2016) and not unique to this study. Fourth, our sampling strategy relied on common ICD-9 sepsis codes and may have not captured all true sepsis cases. The identification of sepsis patients using administrative data is known to be variable across studies (Jolley et al., 2015). Our sampling procedures yielded results for alerted patients with known or suspected sepsis, and results are limited to similar patient populations. Lastly, our results have limited applicability outside of large integrated health systems with an

advanced EWS using multivariate regression or machine learning plus rapid response teams. Advanced EWSs are known to generate fewer false positive alerts at comparable risk thresholds than basic point-score warning systems [e.g., Modified Early Warning Score, (MEWS)] (Linnen, 2018). Consequently, the RRT workload to find and treat a true positive case in health systems using a point-score system would likely be considerably higher than in our setting.

Conclusion

Our study addressed the novel combination of advanced EWSs and RRTs in the field of patient safety. Rather than responding to overt signs of clinical deterioration, an advanced EWS can preemptively determine worsening sepsis, which is among the chief precursors of inpatient and 30-day mortality in the U.S. We have demonstrated that clinical chart reviews are an essential step before conducting quantitative outcomes research. Chart reviews inform fundamental conceptual measurement questions, which ultimately reduces the potential for bias. To measure the impact of the EWS/RRT combination on sepsis survival of ward patients, we discovered important exclusion criteria for the first eligible alert (treatment overlap with emergency department care; see Figure 2), new worsening vs. reoccurring sepsis, as well as DNR status (patients for whom death was expected). For future data extraction, our chart reviews defined discrete locations of sepsis intervention data in the electronic medical record.

This study also identified that the demographics, clinical characteristics, sepsis interventions, and DNR orders of sepsis decedents differ from those of survivors, highlighting the need to adjust for these differences. Future research should investigate the impact of RRT sepsis interventions on patient survival using adequately powered

sample sizes and a rigorous design and analytic methods. For example, a retrospective matched-pair cohort study can approximate conditions of a randomized controlled trial by comparing the between-group differences of sepsis interventions in like decedents and survivors. Such a study would fill an important gap in the sepsis literature and offer the highest attainable level of evidence. Currently, the evidence base for the benefit of pairing advanced EWSs with RRT interventions in terms of patient survival is weak, and a randomized controlled trial to study the impact of sepsis interventions in a no-treatment group would be unethical. Future research should also evaluate timeline differences of hospitalization process times among sepsis decedents and survivors (e.g., emergency department entry, ward admission, alert time, intervention times, discharge and 30-day survival vs. death; see Figure 2) and evaluate whether the alert motivated closer surveillance (e.g., more vital signs measurements or laboratory orders), in addition to treatment.

References

Angus, D. C., Seymour, C. W., Coopersmith, C. M., Deutschman, C., Klompas, M., Levy, M. M., . . . Watson, R. S. (2016). A framework for the development and interpretation of different sepsis definitions and clinical criteria. *Crit Care Med, 44*(3), e113.

Berger, T., Green, J., Horeczko, T., Hagar, Y., Garg, N., Suarez, A., . . . Shapiro, N. (2013). Shock index and early recognition of sepsis in the emergency department: pilot study. *Western Journal of Emergency Medicine, 14*(2), 168.

Bouldin, E. D., Andresen, E. M., Dunton, N. E., Simon, M., Waters, T. M., Liu, M., . . . Shorr, R. I. (2013). Falls among Adult Patients Hospitalized in the United States: Prevalence and Trends. *Journal of patient safety, 9*(1), 13-17. doi:10.1097/PTS.0b013e3182699b64

Chan, P. S., Jain, R., Nallmothu, B. K., Berg, R. A., & Sasson, C. (2010). Rapid Response Teams: A Systematic Review and Meta-analysis. *Arch Intern Med, 170*(1), 18-26. doi:10.1001/archinternmed.2009.424

Churpek, M. M., Yuen, T. C., Winslow, C., Robicsek, A. A., Meltzer, D. O., Gibbons, R. D., & Edelson, D. P. (2014). Multicenter development and validation of a risk stratification tool for ward patients. *American journal of respiratory and critical care medicine, 190*(6), 649-655.

Coyer, F., & Tayyib, N. (2017). Risk factors for pressure injury development in critically ill patients in the intensive care unit: a systematic review protocol. *Syst Rev, 6*(1), 58. doi:10.1186/s13643-017-0451-5

Engel, C., Brunkhorst, F. M., Bone, H.-G., Brunkhorst, R., Gerlach, H., Grond, S., . . . John, S. (2007). Epidemiology of sepsis in Germany: results from a national prospective multicenter study. *Intensive care medicine, 33*(4), 606-618.

Epstein, L., Dantes, R., Magill, S., & Fiore, A. (2016). *Varying Estimates of Sepsis Mortality Using Death Certificates and Administrative Codes — United States, 1999–2014.* . Retrieved from

Escobar, G., & Dellinger, R. (2016). Early detection, prevention, and mitigation of critical illness outside intensive care settings. *J Hosp Med, 11 Suppl 1*, S5-s10. doi:10.1002/jhm.2653

Escobar, G., LaGuardia, J., Turk, B., Ragins, A., Kipnis, P., & Draper, D. (2012a). 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, 7*(5), 388-395. doi:10.1002/jhm.1929

Escobar, G. J., Gardner, M. N., Greene, J. D., Draper, D., & Kipnis, P. (2013). Risk-adjusting hospital mortality using a comprehensive electronic record in an integrated health care delivery system. *Med Care, 51*(5), 446-453. doi:10.1097/MLR.0b013e3182881c8e

Escobar, G. J., LaGuardia, J. C., Turk, B. J., Ragins, A., Kipnis, P., & Draper, D. (2012b). 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, 7*(5), 388-395. doi:10.1002/jhm.1929

Gaieski, D. F., Edwards, J. M., Kallan, M. J., & Carr, B. G. (2013). Benchmarking the incidence and mortality of severe sepsis in the United States. *Crit Care Med, 41*(5), 1167-1174.

Granich, R., Sutton, Z., Kim, Y. S., Anderson, M., Wood, H., Scharf, J. E., . . . Escobar, G. (2016). Early detection of critical illness outside the intensive care unit: Clarifying treatment plans and honoring goals of care using a supportive care team. *J Hosp Med, 11 Suppl 1*, S40-s47. doi:10.1002/jhm.2660

Healthcare Cost Utilization Project. (2008). HCUP Facts and Figures *HCUP Facts and Figures, 2006: Statistics on Hospital-Based Care in the United States*. Rockville (MD): Agency for Healthcare Research and Quality (US).

Jolley, R. J., Sawka, K. J., Yergens, D. W., Quan, H., Jetté, N., & Doig, C. J. (2015). Validity of administrative data in recording sepsis: a systematic review. *Critical Care, 19*(1), 139.

Levy, M. M., Fink, M. P., Marshall, J. C., Abraham, E., Angus, D., Cook, D., . . . Ramsay, G. (2003). 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Intensive Care Med, 29*(4), 530-538. doi:10.1007/s00134-003-1662-x

Linnen, D. (2018). *Doctoral Dissertation Chapter 2: Comparing the Ability of Advanced vs. Point-Score Early Warning Systems to Detect Deterioration Risk in Patients on General Hospital Wards:*

A Systematic Review. University of California, San Francisco.

Liu, V., Escobar, G., Greene, J., Soule, J., Whippy, A., Angus, D. C., & Iwashyna, T. J. (2014). Hospital deaths in patients with sepsis from 2 independent cohorts. *Jama, 312*(1), 90-92. doi:10.1001/jama.2014.5804

Richards, M. J., Edwards, J. R., Culver, D. H., & Gaynes, R. P. (2015). Nosocomial Infections in Combined Medical-Surgical Intensive Care Units in the United States. *Infection Control & Hospital Epidemiology, 21*(8), 510-515. doi:10.1086/501795

Rozen, T., & Butt, W. (2016). Rapid response teams: how are they best used? *Critical Care, 20*(1), 253. doi:10.1186/s13054-016-1425-z

Solomon, R. S., Corwin, G. S., Barclay, D. C., Quddusi, S. F., & Dannenberg, M. D. (2016). Effectiveness of rapid response teams on rates of in-hospital cardiopulmonary arrest and mortality: A systematic review and meta-analysis. *J Hosp Med, 11*(6), 438-445. doi:10.1002/jhm.2554

Sprung, C. L., & Trahtemberg, U. (2017). What definition should we use for sepsis and septic shock? *Critical care medicine, 45*(9), 1564-1567.

Stephens, C. E., Halifax, E., Bui, N., Lee, S. J., Harrington, C., Shim, J., & Ritchie, C. (2015). Provider Perspectives on the Influence of Family on Nursing Home Resident Transfers to the Emergency Department: Crises at the End of Life. *Curr Gerontol Geriatr Res, 2015*, 893062. doi:10.1155/2015/893062

Stephens, C. E., Hunt, L. J., Bui, N., Halifax, E., Ritchie, C. S., & Lee, S. J. (2018). Palliative care eligibility, symptom burden, and quality-of-life ratings in nursing home residents. *JAMA Internal Medicine, 178*(1), 141-142. doi:10.1001/jamainternmed.2017.6299

Van Walraven, C., Escobar, G. J., Greene, J. D., & Forster, A. J. (2010). The Kaiser Permanente inpatient risk adjustment methodology was valid in an external patient population. *Journal of clinical epidemiology, 63*(7), 798-803.

Villar, J., Clement, J. P., Stotts, J., Linnen, D., Rubin, D. J., Thompson, D., . . . Fee, C. (2014). Many emergency department patients with severe sepsis and septic shock do not meet diagnostic criteria within 3 hours of arrival. *Ann Emerg Med, 64*(1), 48-54. doi:10.1016/j.annemergmed.2014.02.023

Whiles, B., Deis, A., Miller, P., & Simpson, S. (2016). Comorbid Conditions Predict Outcomes in Patients With Severe Sepsis. *CHEST, 149*(4), A170. doi:10.1016/j.chest.2016.02.176

Tables and Figures

Table 1. Sample Demographics and Clinical Characteristics of Sepsis Decedents and Survivors in Two Rounds of Clinical Chart Reviews

Clinical Chart Reviews Round 1			
	Entire sample (n=21)	**Decedents (n=14)**	**Survivors (n=7)**
Age, median, mean $\pm SD$	82.5 77.1 $\pm$ 14.3	82.5 78.4 $\pm$ 9.9	81.0 75.0 $\pm$ 15.7
Male, No. (%)	12 (57.1)	9 (64.3)	3 (42.9)
Do Not Resuscitate, No. (%)	12 (40.0)	12 (85.7)	0 (0)
Severity of illness score (LAPS2)[a] at ward admission, median, mean $\pm SD$	109.0 120.4$\pm$ 35.8	130.5 125.9 $\pm$ 26.5	106.0 96.0 $\pm$ 21.7
Severity of illness score (LAPS2)[a] at alert, median, mean $\pm SD$	119.5 145.1 $\pm$ 25.1	159.5 153.1 $\pm$ 25.2	115.0 124.6 $\pm$ 26.0
LAPS2[a] delta, median, mean $\pm SD$	14.5 24.7 $\pm$ 34.3	22.5 27.3 $\pm$ 21.4	12.0 28.6 $\pm$ 28.6
Comorbidity Score at admission (COPS2)[b], median, mean $\pm SD$	34.0 74.6 $\pm$ 32.0	70.0 82.9 $\pm$ 61.9	48.0 51.4 $\pm$ 35.6
Clinical Chart Reviews Round 2			
	Entire sample (n=43)	**Decedents (n=21)**	**Survivors (n=22)**
Age, median, mean $\pm SD$	78.0 72.4 $\pm$ 16.0	82.0 75.2 $\pm$ 18.3	71.0 69.8 $\pm$ 13.1
Male, No. (%)	17 (39.5)	9 (42.9)	8 (36.4)
Severity of illness score (LAPS2)[a] at ward admission, median, mean $\pm SD$	100.0 101.5 $\pm$ 45.7	123.0 115.6 $\pm$ 47.8	87.0 88.1 $\pm$ 40.7
Severity of illness score (LAPS2)[a] at alert, median, mean $\pm SD$	126.0 135.9 $\pm$ 41.2	149.0 155.7 $\pm$ 43.1	114.0 117.0 $\pm$ 28.8
LAPS2[a] delta, median, mean $\pm SD$	20.0 34.4 $\pm$ 39.6	20.0 40.1 $\pm$ 44.0	21.0 28.9 $\pm$ 36.4
Comorbidity Score at admission (COPS2)[b], median, mean $\pm SD$	63.0 68.3 $\pm$ 53.3	74.0 73.5 $\pm$ 50.8	57.5 63.4 $\pm$ 56.1

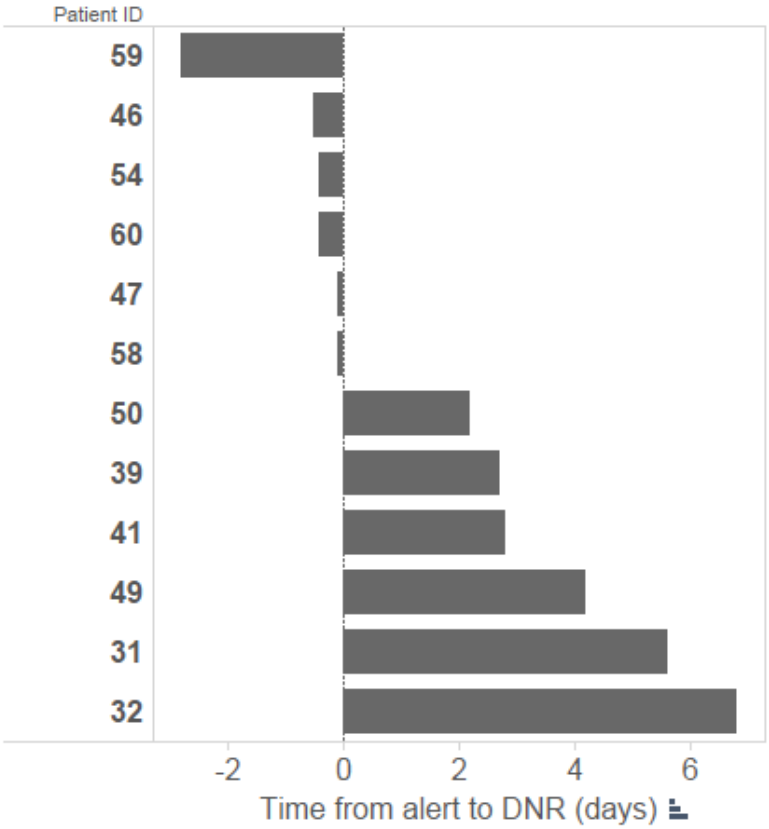
Note. Excludes false positive alerts (i.e., the clinical notes did not indicate a concern for worsening sepsis).
[a] LAboratory-based Acute Physiology Score (LAPS2) is a severity of illness instrument using laboratory values and vital signs
[b] COmorbidity Point Score version 2 (COPS2 measures a patient's 1-year comorbidity burden calculated one month prior to admission

Table 2. Summary of Sepsis Intervention Process Times Relative to an Advanced Early Warning System Alert between Sepsis Decedents and Survivors

Sepsis Interventions (Round 1)			
	Entire sample (n=21)	**Decedents (n=14)**	**Survivors (n=7)**
Time from admission to alert (days), median, mean $\pm SD$	0.5 1.4 ± 1.9	1.6 ± 2.0	1.0 ± 1.6
Time from alert to DNR (days), median, mean $\pm SD$	-	n=12 1.7 ± 2.9	-
Time from DNR to death (days), median, mean $\pm SD$	-	n=12 10.2 ± 12.6	-
Received fluid bolus after alert, No. (%)	7 (33.3)	7 (50.0)	0 (0)
Received new antibiotic after alert, No. (%)	3 (14.3)	1 (7.1)	2 (28.6)
Transfer to ICU, No. (%)	2 (9.5)	0 (0)	2 (28.6)
Sepsis Interventions (Round 2)			
	Entire Sample (n=43)	**Decedents (n=21)**	**Survivors (n=22)**
Time from admission to alert (days), median, mean $\pm SD$	0.5 1.1 ± 1.9	0.6 1.6 ± 2.6	0.4 0.6 ± 0.7
Fluid Bolus administered, No. (%)	31 (72.1)	14 (66.6)	17 (77.3)
Time from alert to fluid bolus 1 (hours), median, mean $\pm SD$	-5.4 -7.8 ± 7.1	-5.1 -5.3 ± 3.2	-6.5 -9.9 ± 8.9
Fluid bolus 1 volume (liters), median, mean $\pm SD$	1.3 1.4 ± 0.8	1.5 1.3 ± 0.8	1.0 1.4 ± 0.8
Time from alert to fluid bolus 2 (hours), median, mean $\pm SD$	0.2 0.6 ± 5.5	-0.2 0.4 ± 4.0	0.4 0.7 ± 6.6
Fluid bolus 2 volume (liters), median, mean $\pm SD$	1.0 0.9 ± 0.5	0.5 0.7 ± 0.3	1.0 1.0 ± 0.6
New antibiotic administered, No. (%)	39 (90.7)	17 (81.0)	22 (100.0)
Time from alert to new IV antibiotic (hours), median, mean $\pm SD$	-4.3 -5.4 ± 14.5	-1.1 -2.8 ± 9.2	-4.4 -7.6 ± 17.7
Transfer to ICU, No. (%)	9 (20.9)	6 (28.6)	3 (13.6)
Time from alert to transfer to ICU (days), median, mean $\pm SD$	1.1 2.5 ± 2.6	2.6 3.2 ± 2.9	1.0 1.0 ± 0.1

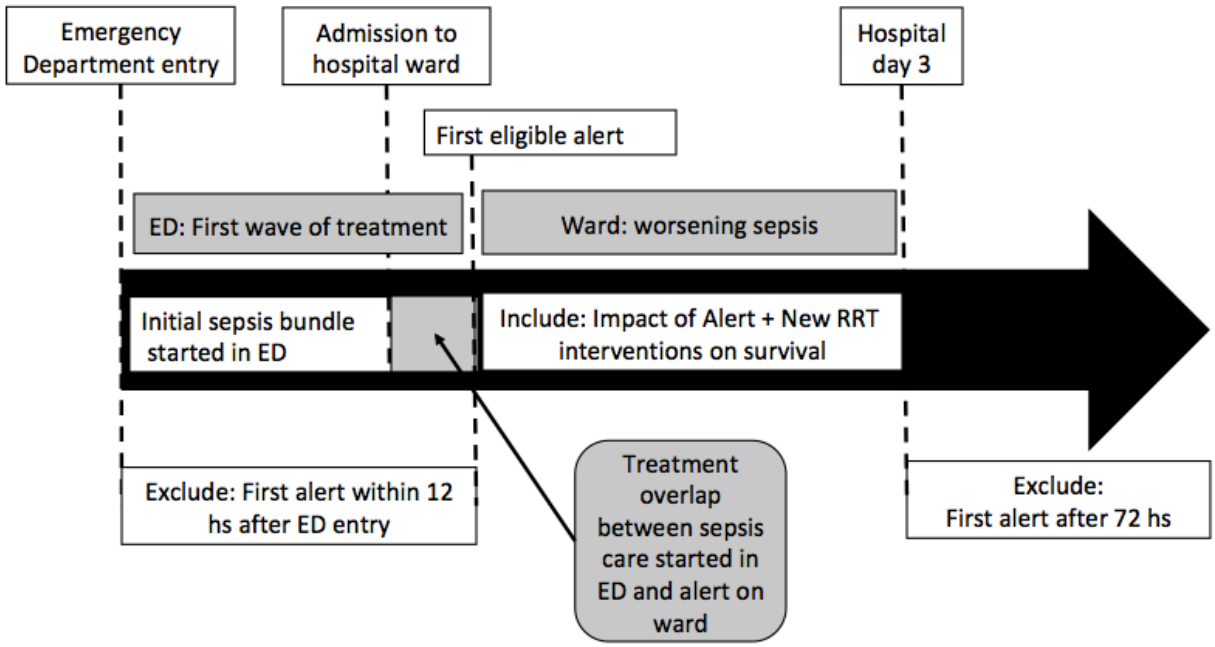
Note. Excludes false positive alerts (i.e., the clinical notes did not indicate a concern for worsening sepsis).

Figure 1. Time of Do Not Resuscitate Order Relative to an Advanced Early Warning System Alert in Sepsis Decedents Admitted to a General Hospital Ward (n=12)



Note. Includes true positive patients only; of 14 decedents, 12 had a Do Not Resuscitate order before death.

Figure 2. Timeline of Hospitalization Time Markers of Sepsis Patients with an Advanced EWS[a] Alert and Treatment Overlap Between Sepsis Care Started in ED[b] and Upon Ward Admission



Note. Not to scale; for illustration purposes only.
[a] Early Warning System
[b] Emergency department

Chapter 4

Do in-hospital Sepsis Interventions Following an Advanced Early Warning System Alert Differ Substantially Between Decedents and Survivors?

Daniel T. Linnen, PhD(c), MS, RN-BC[1,3]

Caroline Stephens, PhD, RN, GNP[2]

John Greene, MA[3]

Vincent Liu, MD, MS[3]

Xiao Hu, PhD[4]

Gabriel J. Escobar, MD[3]

[1] University of California San Francisco, School of Nursing, Department of Social & Behavioral Sciences

[2] University of California San Francisco, School of Nursing, Department of Community Health Systems

[3] Kaiser Permanente Northern California, The Permanente Medical Group, Inc., Division of Research

[4] University of California San Francisco, School of Nursing, Department of Physiological Nursing

Abstract

Background: Advanced early warning systems can predict deterioration risk, but few studies have evaluated the effects of sepsis interventions relative to an advanced early warning system alert.

Objectives: To identify alert-driven interventions that impact survival, we quantified sepsis interventions between comparable groups of sepsis decedents and survivors and measured their differences.

Methods: We employed pair-matching to derive two comparable groups of sepsis decedents and survivors among adults at 3 hospitals in Northern California between 2016 and 2017 (n=1,767). We tested for differences in demographics and clinical characteristics, hospital transfer process times, sepsis interventions, vitals sign documentation, laboratory orders, and time of day (shift) when the alert occurred.

Results: Matching procedures derived like pairs for 42 decedents and 92 survivors. Substantially more survivors than decedents received a new intravenous antibiotic within 6 hours before the alert (10.9% vs. 0.0%, $p = 0.026$). Decedents received more than twice as much fluid bolus therapy within 24 hours following the alert (2.1 liters vs. 1.0 liters, $p = 0.052$), had a greater number of vital sign documentations (mean 86.3 vs. 56.9, $p = 0.011$) and laboratory orders within 24 hours following the alert (15.4 vs. 11.0, $p = 0.004$), more transfers to ICU (42.9% vs. 29.6%, $p = 0.005$), and more DNR or comfort care orders after the alert (69.1% vs. 14.2 , $p < 0.001$).

Discussion: Although limited by sample size, out results suggest that earlier additional antibiotic coverage may lead to better survival. In alerted sepsis patients without receipt of new antibiotics, the alert could motivate clinicians to consider additional antibiotic

therapy. Intensification of fluid bolus therapy and other sepsis interventions in decedents may occur because they present with a more apparent downward trajectory at alert time. More than half of decedents (57%) succumbed to sepsis on the ward. Mortality outcomes may include patients for whom death was expected or unavoidable. **Conclusion:** Future research should validate findings in a larger cohort, control for outcome quality (expected vs. unexpected mortality) and clinical deterioration, and employ more computationally intensive matching procedures.

Introduction

As many as half of all hospital deaths occur in patients with sepsis (Engel et al., 2007; Liu et al., 2014). The preemptive detection of clinical deterioration in hospitalized patients on general hospital wards has historically been challenging (Silber, Williams, Krakauer, & Schwartz, 1992; Silber et al., 2007). However, advanced early warning systems (EWSs) using multivariate regression or machine learning have recently emerged as promising tools (Churpek et al., 2016; Escobar et al., 2016; Kipnis et al., 2016). Previous methods of identifying sepsis patients at risk for deterioration included basic point-score EWSs [e.g., Modified Early Warning Score (MEWS), National Early Warning Score (NEWS)] and sepsis "sniffer" tools—algorithms that use systemic inflammatory response syndrome criteria plus evidence of organ dysfunction, such as hypotension or hyperlactatemia, or the Sequential Organ Failure Assessment (SOFA) score. These tools have limited prognostic accuracy and predictive precision (Linnen et al., 2018) because they use only a limited set of input variables and because a positive screen depends on evidence of deterioration. Consequently, point-score EWSs and sepsis sniffers may only facilitate a clinical response, rather than prevention of a worsening condition (Chan, Jain, Nallmothu, Berg, & Sasson, 2010; Bellew, Cabrera, Lohse, & Bellolio, 2016; Escobar & Dellinger, 2016). Recent evidence suggests that more advanced EWSs using multivariate regression or machine learning can predict clinical deterioration risk with better precision, while creating less workload burden, and, consequently, facilitate a more efficient use of rapid response teams (RRTs) (Linnen et al., 2018).

Since 2008, an international panel of experts has developed evidence-based treatment guidelines for *Surviving Sepsis* (Dellinger et al., 2008; Dellinger et al., 2013; Rhodes et al., 2017). These guidelines include the timely delivery of initial fluid resuscitation with 30ml/kg of intravenous crystalloids, followed by additional fluid bolus therapy (FBT) to maintain hemodynamic stability, and the timely administration of intravenous antibiotics. The evidence base for the impact of fluid bolus therapy on sepsis survival remains weak (Glassford, Eastwood, & Bellomo, 2014; Liu et al., 2014; Park et al., 2017), and evidence suggest that in patients with severe sepsis and septic shock, aggressive FBT of more than 5 liters of intravenous crystalloid solution may in fact be associated with increased mortality (Marik, Linde-Zwirble, Bittner, Sahatjian, & Hansell, 2017). Randomized controlled trials that could offer more definitive evidence would be unethical because they would require the withholding of FBT. However, observational designs using pair-matching can approximate conditions of a controlled trial (Staffa & Zurakowski, 2018).

Many sepsis patients present with symptoms in the emergency department (ED) (Villar et al., 2014), and the initial sepsis bundle of care is measured relative to the *Time Zero* of ED entry time. However, on hospital wards —the setting of our study—such a clean time marker does not exist, and preemptively identifying and treating worsening sepsis would require the use of an advanced EWS alert. This approach holds promise, because it moves identification of worsening sepsis on the ward (Time Zero) several hours ahead—before a patient noticeably destabilizes. Currently, no studies have evaluated whether FBT and other sepsis interventions improve sepsis survival following an advanced EWS alert. However, Kaiser Permanente Northern California (KPNC), a

large integrated healthcare delivery system, is deploying an advanced EWS in 21 hospitals (Dummett et al., 2016). In this study, we report early results from the first 3 hospitals.

The purpose of this observational matched pair cohort study was to evaluate the treatment effects of FBT and other sepsis interventions following an EWS alert, that is, whether alert-driven interventions differed between sepsis survivors and decedents admitted from the ED in stable condition. By matching decedents and survivors on demographics and clinical characteristics, our design mimicked that of a randomized controlled trial and tested the treatment effects of alert-driven RRT sepsis interventions between two comparable groups.

Methods

Setting and Sample

Our research cohort consisted of patients admitted to a general ward in any of 3 Kaiser Permanente community hospitals in Northern California (KPNC) between 08/01/16 and 08/31/17 (13 months). KPNC is deploying an advanced EWS, called *Advance Alert Monitor* (AAM) (Escobar et al., 2016; Kipnis et al., 2016), to all of its 21 hospitals. AAM uses a sophisticated multivariate regression algorithm to predict deterioration risk in patients on general wards up to 12hs before an acute event. After alert triage by a registered nurse in a remote regional surveillance unit (eHospital), the alert is routed to a hospital's rapid response team. In this study, we report early results from the 3 hospitals in which KPNC deployed AAM. The Institutional Review Boards of Kaiser Permanente Northern California and the University of California, San Francisco approved the study.

Inclusion and Exclusion Criteria

The initial research population was 1,768 alerted adult patients admitted in stable condition to a general ward (any medical diagnosis). Inclusion criteria were age ≥*18* years, a positive AAM alert in a medical-surgical or step-down unit that occurred within 12 hours after ED entry and 72 hours following admission. Because patients may transfer between KPNC hospitals, we combined related hospital encounters into one hospital episode as previously described (Escobar et al., 2011). We excluded patients with an alert in the first 12 hours following ED entry because the initial sepsis bundle (e.g., fluid bolus, intravenous antibiotics) in the ED may create a treatment overlap in the first hours of ward admission (i.e., in these patients it is not always clear whether additional FBT or antibiotics would be indicated). To mitigate this issue, we included alerts after the initial sepsis bundle could be considered complete.

We excluded the following patients (see Figure 1): those without KP plan membership (n=177), because it is not always possible to reliably collect 30-day mortality in non-members; direct admissions to the ward or ICU, because these hospital episodes started in a non-KPNC hospital (n=284); patients with a length of stay less than 24 hours, discharge under observation status, and overnight ambulatory surgery stays (n=157); patients without ICD-10 sepsis admission codes (see appendix) (n=519); and patients with a “do not resuscitate” (DNR) or “comfort care” order within 4 hours following admission (n=174), because death may have been imminent or was expected. After applying exclusions, a total of 517 patients remained in the sample, including 56 sepsis decedents and 416 survivors.

Matching procedures

To minimize measurement bias, we created comparable groups of decedents and survivors using multivariate matching procedures in SAS (Kosanke & Bergstralh, 2004). Given a specified set of matching criteria, the GREEDY algorithm (*gmatch*) identifies x matching controls for each case. A matching control (here: survivor) is identified by virtue of the sum of total differences of the covariates to a given case (here: decedent) (Bamman, 2016). GREEDY retains the first match it encounters. Matching criteria included age, sex, care directive at alert, AAM score (Escobar & Dellinger, 2016; Kipnis et al., 2016), a patient's 12-month comorbidity burden (COmorbidity Point Score 2 [COPS2]) (Escobar, Gardner, Greene, Draper, & Kipnis, 2013), severity of illness score (LAboratory-Based Acute Physiology Score 2 [LAPS2]) at admission (Escobar et al., 2012), and length of hospital stay. In total, we matched 42 decedents with 92 survivors (2 to 3 matching survivors for each decedent; see the appendix for detailed matching results). We then assessed match quality via standardized difference scores using Stata14's *pbalchk* program (Lunt, 2013). Standardized difference scores evaluate match quality (balance) and identify whether the covariates used for matching are stable between outcome groups (Garrido et al., 2014).

Of the 56 sepsis decedents and 461 survivors in our cohort, matching procedures derived like pairs for 42 decedents and 92 survivors. Matching occurred in 4 steps using successively wider inclusion limits (see appendix). In step 1, we identified 19 survivors matching on age within 5 years, sex, care directive at alert, AAM score within 10 points, LAPS2 score at admission within 10 points, COPS2 score within 10 points, and length of stay at alert with 24 hours. In step 2, we dropped length of stay at

alert, which identified an additional 11 survivors. In step 3, we widened the inclusion limits for LAPS2 (within 15 points) and COPS2 (within 15 points), resulting in 21 additional survivor matches. Finally, we widened age (with 10 years), LAPS2 (within 20 points) and COPS2 (within 20 points), resulting in 41 additional survivor matches. We identified no matches for 14 sepsis decedents (25%) and 369 survivors (80%).

Research Variables

The outcome variable was 30-day survival or death following admission. We also captured hospital mortality (death at discharge) for comparison. We extracted patient demographics and clinical data from the electronic medical record: age; sex; race; LAPS2 scores at admission and alert and the change in these two measurements; COPS2; hospital length of stay; length of stay at alert; admission type; care directive at alert, presence of hypotension or hyperlactatemia within 6 hours before the alert. We extracted relevant hospitalization process times: time from ED entry to admission; time from admission to alert; and time from alert to discharge.

We also collected data on key sepsis intervention processes for FBT and antibiotics administered. These included elapsed time from alert to fluid bolus; total fluid bolus volume administered within 6 hours before and 24 hours after the alert; number of fluid bolus administrations; and number of administrations preceding the alert. Because we were concerned that in certain sepsis patients FBT may contraindicated, we flagged sepsis patients with potential fluid overload (an elevated brain natriuretic peptide level (BNP) >400 pg/mL within 12 hours before the alert and/or intravenous furosemide administration within 12 hours before and 2h after the alert.

We examined number of *new* intravenous antibiotic administrations (defined as an antibiotic not administered within 24 hours before the alert); and number of antibiotic administrations preceding the alert.

Finally, we examined number of ICU transfers; number of Do Not Resuscitate (DNR) or comfort care orders after the alert; sepsis surveillance measures (number of vital signs, and laboratory orders within 24 hours after the alert), and characteristics of the work shift at alert time (night/day/evening shift; weekend shift).

Statistical Analysis

We described the data and tested statistical differences between sepsis decedents and survivors using Stata14 (StataCorp. 2015. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP). After confirming normality of continuous variables using histograms and p-norm/q-norm plots (Wilk & Gnanadesikan, 1968; Filliben, 1975) (results omitted), we performed Student's t-test for continuous data and Pearson's chi-squared test for categorical data. Matching adjusted the treatment effects for age, sex, advanced EWS score, severity of illness at admission (LAPS2), patient comorbidity burden (COPS2), code status, and, in a subsample of matches, length of stay at alert. We deemed the groups of sepsis decedents and survivors independent, because it was implausible that sepsis interventions in a given decedent could influence those in a given survivor.

Results

Comparison of matched vs. unmatched groups

On average, compared to the matched groups, unmatched decedents were younger (mean age 69.2 years), sicker at admission (mean LAPS2 130.5) and at alert time (mean LAPS2 152.9), had a substantially higher comorbidity burden (COPS2 110.4), higher AAM alert score (16.2), had fewer full code orders (78.6%), and had more hypotension within 6 hours before the alert (64.3%). Matching procedures dropped 369 survivors (80%), for which we could not identify a suitable pairing. Compared to the matched groups, these unmatched survivors were younger (mean age 66.4), less sick at admission (mean LAPS2 101.2) and at alert time (mean LAPS2 118.4), had had a lower comorbidity burden (mean COPS2 67.7), and fewer patients had hyperlactatemia within 6 hours before the alert (mean 14.3%).

Overall, standardized difference scores of the matching covariates showed good match quality, with difference scores ranging from 5% to 11% in 5 of the 7 covariates and 16-21% in the remaining 2 covariates (length of stay at alert, which we removed after the first round of matching, and AAM score; see Figure 2). The mean difference of AAM scores between matched decedents and survivors was marginal (13.7 vs. 12.5, $p = 0.269$), indicating reasonable match quality. The p-values of all between-group difference tests were also not statistically significant, ranging from 0.203 for length of stay at alert to 0.783 for age at admission (see Table 1).

Cohort characteristics of matched decedents and survivors

Sepsis decedents did not statistically differ from survivors across any of the demographic variables or clinical characteristics. In addition to the matching variables,

the matched decedents and survivors were alike in terms of race, LAPS2 delta (the difference between LAPS2 severity of illness score at alert vs. admission), admission type, and presence of hypotension or hyperlactatemia within 6 hours before the alert. All of the survivors and decedents had the correct 30-day mortality classification. Among the sepsis decedents (n=42), we observed in-hospital mortality in 25 cases (59.5%), and in 1 case (1.1%) among the survivors (n=92); this patient survived to 30 days after admission, but ultimately died in the hospital.

Hospitalization process measures

None of the hospitalization process times differed statistically between matched sepsis decedents and survivors (see Table 2). These included time from ED entry to ward admission (mean 8.8 hours in decedents vs. 7.6 hours in survivors, p = 0.233), time from hospital admission to alert (mean 17.0 hours vs. 13.6 hours, respectively, p = 205), time from alert to discharge (mean 6.6 days vs. 7.8 days, p = 0.402), and hospital length of stay (mean 7.3 days vs. 8.4, p = 0.456).

Sepsis interventions and surveillance

We did not detect statistically significant differences in time from alert to fluid bolus administration (174 minutes vs. 123 minutes, p = 0.714), percent of patients with at least one fluid bolus administration (42.9% vs. 32.6%, p = 0.251), percent of fluid bolus administrations preceding the alert (28.6% vs. 17.4%, p = 0.140), and total fluid bolus volume administered within 6 hours before the alert (0.7 liters vs. 0.7 liters, p = 0.896). Decedents received more than twice as much fluid bolus volume within 24 hours following the alert (2.1 liters vs. 1.0 liters, p = 0.052). We performed a sensitivity analysis by excluding patients with potential evidence of fluid overload (high BNP or

furosemide administration). This approach dropped 25 cases. In the remaining cohort, decedents again received more fluid bolus volume within 24hs following the alert and the difference reached statistical significance (2.25 liters vs. 1.0 liter, p = 0.034).

Comparing decedents to matched survivors, a greater percentage of survivors received a new intravenous antibiotic, although the difference was not statistically significant (14.3% vs. 20.7%, p = 0.380). However, substantially more survivors received a new intravenous antibiotic within 6 hours before the alert (0.0% vs. 10.9%, p = 0.026). More sepsis decedents received a DNR or comfort care order after the alert (69.1% vs. 14.1%, p < 0.001) and experienced a transfer to ICU (42.9% vs. 19.6%, p = 0.005). Sepsis decedents had more vital sign documentations within 24 hours after the alert (mean 86.3 vs. 56.9, p = 0.011), while the number of complete vital sign sets was not substantially different (5.7 vs. 5.2, p = 0.262). Decedents also had more laboratory orders within 24 hours following the alert (15.4 vs. 11.0, p = 0.004).

Characteristics of the work shift at alert time

Finally, we did not detect a difference in regard to the alert's shift type (p = 0.230 across three shift types – day, evening, night). We did, however, observe a greater proportion of survivors with alerts on day shift (21.4% of decedents vs. 43.5% of survivors had an alert on day shift) and a greater proportion of decedents with alerts on evening shifts (31.0% of decedents vs. 12.0% of survivors had an alert on evening shift). Alerts on weekends were somewhat more frequent among survivors, although this difference, too, was not statistically significant (23.8% vs. 32.6%, p = 0.302).

Discussion

Our study is among the first to evaluate whether sepsis interventions following an advanced EWS alert differ substantially between sepsis decedents and survivors. Our design mimicked that of a randomized controlled trial, in an effort to test the treatment effects of alert-driven sepsis RRT interventions between two comparable groups. We used a retrospective matched pair design, which is a useful and sound method of causal inference testing (Staffa & Zurakowski, 2018) when a comparison of treatment effects in an actual no-treatment group would be unethical. Our results suggest that we derived comparable groups of decedents and survivors. Because very few studies have investigated the impact of pairing an advanced EWS with RRT on sepsis survival, our analysis deliberately focused on sepsis patients who were admitted in stable condition (by virtue of admission to a ward, rather than ICU) with a positive deterioration alert within 3 days.

Consistent with a systematic review investigating the treatment effects of FBT on hemodynamic stability (Glassford et al., 2014) and the *Surviving Sepsis* guidelines (Rhodes et al., 2017), we hypothesized that survivors would receive earlier and more FBT (as well as antibiotics, transfer to ICU, and sepsis surveillance). However, our results suggest there is no difference in time from alert to fluid bolus between decedents and survivors, and that decedents received substantially more fluid bolus therapy (especially after excluding those for whom FBT may have been contraindicated), transfer to ICU, and DNR or comfort care orders following the alert. We also observed substantially more surveillance activity (vital sign measurements and laboratory tests) in sepsis decedents. These findings suggest a reverse causal path; rather than being

protective factors, the observed intensification of sepsis interventions is more pronounced in decedents because these patients may present with a more apparent downward trajectory. This explanation, then, raises the question: which mechanisms of sepsis deterioration may remain unmeasured after accounting for a large and nuanced portfolio of clinical characteristics? One plausible explanation is that the (unmeasured) pre-hospital sepsis exposure time may be longer in decedents, which then may worsen their overall clinical course. Even though sepsis decedents appeared clinically stable and comparable to sepsis survivors in the ED, the slope of their deterioration trajectory may have been steeper: The mean severity of illness scores (LAPS2) at alert time and LAPS2 delta score were more pronounced in decedents, and although not statistically significant at alpha = 0.05, these 2 markers appeared to be the least balanced between the two groups.

In this cohort of sepsis decedents with a mean age of 76.0 years, we observed higher ICU utilization (doubled in decedents compared to survivors) and more frequent DNR or comfort care orders (nearly 5-fold in decedents compared to survivors). More than half of decedents (57%) succumbed to sepsis on the ward. These findings suggest that 1) patients may have chosen limited interventions in the context of terminal disease (e.g., cancer), and 2) transfer to ICU may have been the result of unclear treatment wishes leading to heroic, but ultimately unsuccessful, attempts of life support. This study did not aim to evaluate the quality of the observed outcomes (i.e., whether death was expected or unavoidable), but the observed ICU transfer and DNR order patterns appear to offer circumstantial support for the notion that death was an expected or unavoidable outcome.

We found that substantially more survivors received a new intravenous antibiotic within 6 hours before the alert, with no difference in time from ED entry to admission and from admission to alert. Consistent with prior evidence (Rivers et al., 2001; Gaieski et al., 2010; Prasad et al., 2017), receiving antibiotic coverage early may have a favorable therapeutic effect. Our findings suggest that earlier additional antibiotic coverage may have a bigger impact on sepsis survival than FBT, which is an important discovery in the light of antibiotic stewardship (Pollack & Srinivasan, 2014). While the judicious prescription of antibiotics is an ever-present concern for clinicians, our results suggest that sepsis survival is improved with early additional coverage. This finding is consistent with the evidence regarding the treatment effect of early goal directed therapy. Given that fewer decedents received a new antibiotic, our results suggest that the alert time could be used to prompt the consideration for antibiotic coverage, if it did not occur in the hours before the alert.

Finally, though not statistically significant in this sample, we observed that decedents alerted more frequently during the evening shift and survivors alerted more frequently during day shift. The hospitals in this sample routinely staffed RRTs on both shift types and adhered to state-mandated nurse-to-patient staffing ratios (e.g., 1 registered nurse for up to 5 patients on general wards). It is possible, though speculative, that the presence of fewer hospitalists on evening shift may result in different antibiotic prescription patterns after the RRT has responded. Given that the shift difference was not statistically significant in this study, confirmation would require follow-up analysis in a larger sample.

We also wish to address the study's limitations. First, the sample size of our cohort was relatively small, and matching procedures further reduced the pool of comparable decedents and survivors. This limitation could not be overcome, because a total of 3 KPNC hospitals had implemented an advanced EWS during the research timeframe and because the study's matched pair design required rigorous exclusion criteria and matching procedures to minimize measurement bias. Despite limited power, our work found strong effects for a number of research variables. Second, our study was not powered to detect potential variation across hospitals or hospital-level effects because, again, a hierarchical analysis of treatment effects would require a larger number of hospitals. However, previous work has shown that KPNC hospitals are relatively alike across a wide range of characteristics (Linnen et al., 2018). That is, by belonging to an integrated health system, they may employ relatively similar clinical workflows, staffing practices, and clinical practice standards. Third, we acknowledge that by using GREEDY matching procedures, we could not identify a complete set of matches for all decedents and survivors. As a result, our findings have limited generalizability beyond sepsis patients who resemble the final matched groups. The GREEDY program identifies and keeps the first possible match, rather than the best possible one. This approach is similar to propensity score matching (Rosenbaum & Rubin, 1983), which also lacks the ability to control for imbalanced matching outputs. However, our matches appeared to be of suitable quality, as assessed by both standardized difference scores (Austin, 2009) and between-group difference testing. Fourth, while we excluded patients with an immediate DNR or comfort care order after admission, it is possible that some of the sepsis decedents with Full Code at admission

were on a fixed end-of-life trajectory. If death was foreseeable in such patients, their treatment focus may have been palliative, even as life-saving interventions continued (Kim et al., 2016). This measurement challenge is not unique to this study; it is known that elderly or terminally ill patients lack clarity regarding treatment wishes (Stephens et al., 2015), which then complicates a clear measurement of outcome quality. Finally, our analytic approach (between-group testing) may not have fully risk-adjusted for all potential effects between independent variables. However, we controlled for measurement bias by using externally validated risk-adjusted composite scores of severity of illness, comorbidity burden, and deterioration risk and by using multivariate matching. Our methods generated the closest achievable approximation of comparable groups.

Conclusion

While advanced EWSs aim to support clinical decision-making, it appears that death is often an unavoidable outcome after the identification of worsening sepsis, despite the alert-driven intensification of treatment. Our finding that substantially more survivors received a new intravenous antibiotic within 6 hours prior to the EWS alert is consistent with previous findings in its message—earlier additional antibiotics may lead to better survival and may be more important to surviving sepsis than fluid bolus therapy. Reliable methods of quality measurement of expected vs. preventable mortality are needed because these may influence the direction of treatment effects of sepsis interventions following an alert. Future research should aim to validate findings in a larger cohort (e.g., one that includes all admissions stratified by diagnostic groups) to overcome power limitations, derive and validate a regression model which adjusts and

controls for the slope of clinical deterioration between admission and alert time, and employ more computationally intensive matching procedures to overcome matching limitations. Such confirmatory analyses would provide more definitive answers. Given the strength of matched pair cohort studies, results have the potential to inform evidence-based guidelines for the treatment of worsening in-hospital sepsis, which remains the number 1 cause of death in U.S. hospitals.

References

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity‐score matched samples. *Statistics in medicine, 28*(25), 3083-3107.

Bamman, B. (2016). Deconstructing data science: Lecture 17: Distance Models . Retrieved from http://courses.ischool.berkeley.edu/i290-dds/s16/dds/slides/17_distance_models_classification.pdf

Bellew, S. D., Cabrera, D., Lohse, C. M., & Bellolio, M. F. (2016). Predicting early rapid response team activation in patients admitted from the emergency department: The PeRRT Score. *Acad Emerg Med*. doi:10.1111/acem.13077

Chan, P. S., Jain, R., Nallmothu, B. K., Berg, R. A., & Sasson, C. (2010). Rapid Response Teams: A Systematic Review and Meta-analysis. *Arch Intern Med, 170*(1), 18-26. doi:10.1001/archinternmed.2009.424

Churpek, M. M., Yuen, T. C., Winslow, C., Meltzer, D. O., Kattan, M. W., & Edelson, D. P. (2016). Multicenter Comparison of Machine Learning Methods and Conventional Regression for Predicting Clinical Deterioration on the Wards. *Crit Care Med, 44*(2), 368-374. doi:10.1097/ccm.0000000000001571

Dellinger, R. P., Levy, M. M., Carlet, J. M., Bion, J., Parker, M. M., Jaeschke, R., . . . Beale, R. (2008). Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock: 2008. *Intensive care medicine, 34*(1), 17-60.

Dellinger, R. P., Levy, M. M., Rhodes, A., Annane, D., Gerlach, H., Opal, S. M., . . . Moreno, R. (2013). Surviving sepsis campaign: international guidelines for

management of severe sepsis and septic shock: 2012. *Crit Care Med, 41*(2), 580-637. doi:10.1097/CCM.0b013e31827e83af

Dummett, B. A., Adams, C., Scruth, E., Liu, V., Guo, M., & Escobar, G. J. (2016). Incorporating an early detection system into routine clinical practice in two community hospitals. *J Hosp Med, 11*(S1).

Engel, C., Brunkhorst, F. M., Bone, H.-G., Brunkhorst, R., Gerlach, H., Grond, S., . . . John, S. (2007). Epidemiology of sepsis in Germany: results from a national prospective multicenter study. *Intensive care medicine, 33*(4), 606-618.

Escobar, G., & Dellinger, R. (2016). Early detection, prevention, and mitigation of critical illness outside intensive care settings. *J Hosp Med, 11 Suppl 1*, S5-s10. doi:10.1002/jhm.2653

Escobar, G., Greene, J., Gardner, M., Marelich, G., Quick, B., & Kipnis, P. (2011). Intra-hospital transfers to a higher level of care: contribution to total hospital and intensive care unit (ICU) mortality and length of stay (LOS). *J Hosp Med, 6*(2), 74-80. doi:10.1002/jhm.817

Escobar, G., LaGuardia, J., Turk, B., Ragins, A., Kipnis, P., & Draper, D. (2012). 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, 7*(5), 388-395. doi:10.1002/jhm.1929

Escobar, G., Turk, B., Ragins, A., Ha, J., Hoberman, B., LeVine, S., . . . Kipnis, P. (2016). Piloting electronic medical record-based early detection of inpatient deterioration in community hospitals. *J Hosp Med, 11 Suppl 1*, S18-s24. doi:10.1002/jhm.2652

Escobar, G. J., Gardner, M. N., Greene, J. D., Draper, D., & Kipnis, P. (2013). Risk-adjusting hospital mortality using a comprehensive electronic record in an integrated health care delivery system. *Med Care, 51*(5), 446-453.

Filliben, J. J. (1975). The probability plot correlation coefficient test for normality. *Technometrics, 17*(1), 111-117.

Gaieski, D. F., Mikkelsen, M. E., Band, R. A., Pines, J. M., Massone, R., Furia, F. F., . . . Goyal, M. (2010). Impact of time to antibiotics on survival in patients with severe sepsis or septic shock in whom early goal-directed therapy was initiated in the emergency department. *Crit Care Med, 38*(4), 1045-1053.

Garrido, M. M., Kelley, A. S., Paris, J., Roza, K., Meier, D. E., Morrison, R. S., & Aldridge, M. D. (2014). Methods for constructing and assessing propensity scores. *Health services research, 49*(5), 1701-1720.

Glassford, N. J., Eastwood, G. M., & Bellomo, R. (2014). Physiological changes after fluid bolus therapy in sepsis: a systematic review of contemporary data. *Crit Care, 18*(6), 696. doi:10.1186/s13054-014-0696-5

Kim, Y. S., Escobar, G., Halpern, S. D., Greene, J., Kipnis, P., & Liu, V. (2016). The Natural History of Changes in Preferences for Life-Sustaining Treatments and Implications for Inpatient Mortality in Younger and Older Hospitalized Adults. *J Am Geriatr Soc, 64*(5), 981-989. doi:10.1111/jgs.14048

Kipnis, P., Turk, B. J., Wulf, D. A., LaGuardia, J. C., Liu, V., Churpek, M. M., . . . Escobar, G. (2016). Development and validation of an electronic medical record-based alert score for detection of inpatient deterioration outside the ICU. *J Biomed Inform, 64*, 10-19. doi:10.1016/j.jbi.2016.09.013

Kosanke, J., & Bergstralh, E. (2004). gmatch: Match 1 or more controls to cases using the GREEDY algorithm. *Google Scholar*.

Linnen, D., Escobar, G. J., Hu, X., Scruth, E. A., Liu, V., & Stephens, C. (2018). *Advanced Early Warning Systems Detect Deterioration Risk in Patients on General Hospital Wards with Greater Accuracy Than Point-Score Tools and Reduce the RRT Evaluation Workload: A Systematic Review*. Doctoral Dissertation. University of California, San Francisco. San Francisco, CA.

Liu, V., Escobar, G., Greene, J., Soule, J., Whippy, A., Angus, D. C., & Iwashyna, T. J. (2014). Hospital deaths in patients with sepsis from 2 independent cohorts. *Jama, 312*(1), 90-92. doi:10.1001/jama.2014.5804

Lunt, M. (2013). PBALCHK: Checking covariate balance. *accessed on May, 23*, 2013.

Marik, P. E., Linde-Zwirble, W. T., Bittner, E. A., Sahatjian, J., & Hansell, D. (2017). Fluid administration in severe sepsis and septic shock, patterns and outcomes: an analysis of a large national database. *Intensive care medicine, 43*(5), 625-632.

Park, S. K., Shin, S. R., Hur, M., Kim, W. H., Oh, E. A., & Lee, S. H. (2017). The effect of early goal-directed therapy for treatment of severe sepsis or septic shock: A systemic review and meta-analysis. *J Crit Care, 38*, 115-122. doi:10.1016/j.jcrc.2016.10.019

Pollack, L. A., & Srinivasan, A. (2014). Core elements of hospital antibiotic stewardship programs from the Centers for Disease Control and Prevention. *Clinical Infectious Diseases, 59*(suppl_3), S97-S100.

Prasad, P. A., Shea, E. R., Shiboski, S., Sullivan, M. C., Gonzales, R., & Shimabukuro, D. (2017). Relationship Between a Sepsis Intervention Bundle and In-Hospital Mortality Among Hospitalized Patients: A Retrospective Analysis of Real-World Data. *Anesthesia & Analgesia, 125*(2), 507-513.

Rhodes, A., Evans, L. E., Alhazzani, W., Levy, M. M., Antonelli, M., Ferrer, R., . . . Nunnally, M. E. (2017). Surviving sepsis campaign: international guidelines for management of sepsis and septic shock: 2016. *Intensive care medicine, 43*(3), 304-377.

Rivers, E., Nguyen, B., Havstad, S., Ressler, J., Muzzin, A., Knoblich, B., . . . Tomlanovich, M. (2001). Early goal-directed therapy in the treatment of severe sepsis and septic shock. *New England Journal of Medicine, 345*(19), 1368-1377.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41-55.

Silber, J. H., Romano, P. S., Rosen, A. K., Wang, Y., Even-Shoshan, O., & Volpp, K. G. (2007). Failure-to-rescue: comparing definitions to measure quality of care. *Med Care, 45*(10), 918-925. doi:10.1097/MLR.0b013e31812e01cc

Silber, J. H., Williams, S. V., Krakauer, H., & Schwartz, J. S. (1992). Hospital and patient characteristics associated with death after surgery. A study of adverse occurrence and failure to rescue. *Med Care, 30*(7), 615-629.

Staffa, S. J., & Zurakowski, D. (2018). Five Steps to Successfully Implement and Evaluate Propensity Score Matching in Clinical Research Studies. *Anesthesia and analgesia*.

Stephens, C. E., Halifax, E., Bui, N., Lee, S. J., Harrington, C., Shim, J., & Ritchie, C. (2015). Provider Perspectives on the Influence of Family on Nursing Home Resident Transfers to the Emergency Department: Crises at the End of Life. *Curr Gerontol Geriatr Res, 2015*, 893062. doi:10.1155/2015/893062

Villar, J., Clement, J. P., Stotts, J., Linnen, D., Rubin, D. J., Thompson, D., . . . Fee, C. (2014). Many emergency department patients with severe sepsis and septic shock do not meet diagnostic criteria within 3 hours of arrival. *Annals of Emergency Medicine, 64*(1), 48-54.

Wilk, M. B., & Gnanadesikan, R. (1968). Probability plotting methods for the analysis for the analysis of data. *Biometrika, 55*(1), 1-17.

Tables and Figures

Table1. Demographic and Clinical Characteristics of a Cohort of Sepsis Patients Identified by an Advanced Early Warning System

	Entire Cohort (n=517)	Matched Decedents (n=42)	Matched Survivors[a] (n=92)	P value	Unmatched Survivors (n=369)	Unmatched Decedents (n=14)
Age at admission (years), median, mean ±SD	70.0, 68.4 ± 14.1	76.0, 74.1 ± 12.7	74.5, 73.5 ± 9.7	0.783	68.0, 66.4 ± 14.3	81.0, 69.2 ± 22.8
Male, No. (%)	265 (51.3)	22 (52.4)	45 (48.9)	0.139	190 (51.5)	8 (57.1)
Race				0.539		
White, No. (%)	308 (59.6)	24 (57.1)	50 (54.4)		228 (61.8)	6 (42.9)
African American, No. (%)	27 (5.2)	1 (2.4)	5 (5.4)		18 (4.9)	3 (21.4)
Asian, No. (%)	75 (14.5)	6 (14.3)	13 (14.1)		54 (14.6)	2 (14.3)
Hispanic, No. (%)	71 (13.8)	9 (21.4)	13 (14.1)		46 (12.5)	3 (21.4)
Other, No. (%)	36 (7.0)	2 (4.8)	11 (12.0)		23 (6.2)	0 (0.0)
Admission Type				0.707		
ED Surgical, No. (%)	78 (15.1)	6 (14.3)	11 (12.0)		61 (16.5)	0 (0.0)
ED Medical, No. (%)	439 (84.9)	36 (85.7)	81 (88.0)		308 (83.5)	14 (100.0)
Full Code at alert, No. (%)	505 (97.7)	41 (97.6)	91 (98.9)	0.567	362 (98.1)	11 (78.6)
AAM score[b]	11.6, 15.1 ± 10.7	12.1, 13.7 ± 5.7	10.4, 12.5 ± 5.6	0.269	12.0, 15.6 ± 11.1	16.2, 24.5 ± 23.3
Severity of illness (LAPS2 score)[c]						
At ward admission, median, mean ±SD	105.0, 105.9 ± 29.9	121.0, 118.3 ± 26.0	115.5, 115.6 ± 22.6	0.541	100.0, 101.2 ± 29.8	131.5, 130.5 ± 46.9
At alert, median, mean ±SD	120, 122.4 ± 29.3	137, 136.4 ± 29.9	127.5, 127.4 ± 23.0	0.059	117, 118.4 ± 29.1	147.5, 152.9 ± 35.3
LAPS2 delta, median, mean ±SD	12, 16.5 ± 21.6	15, 18.1 ± 21.4	10.5, 11.8 ± 19.3	0.093	12, 17.3 ± 21.7	13, 22.4 ± 29.8
Comorbidity burden (COPS2 score)[d], median, mean ±SD	58.0, 71.4 ± 55.9	71.5, 79.8 ± 45.7	72.0, 76.2 ± 44.2	0.665	51.0, 67.7 ± 58.6	108.0, 110.4 ± 66.5
Length of stay (days), median, mean ±SD	5.0, 7.3 ± 7.6	6.75, 7.3 ± 4.4	4.85, 8.4 ± 9.5	0.456	4.8, 7.1 ± 7.5	5.6, 6.0 ± 4.3
Length of stay at alert (hours), median, mean ±SD	10.1, 16.1 ± 20.5	11.4, 17.0 ± 16.3	9.2, 13.6 ± 13.5	0.205	10.1, 16.5 ± 22.1	9.3, 21.4 ± 24.6
SBP[e] <100 within 6hs before alert, No. (%)	150 (29.0)	12 (28.6)	28 (30.4)	0.827	101 (27.4)	9 (64.3)
Serum lactate ≥2.0 within 6hs before alert, No. (%)	60 (11.6)	9 (21.4)	17 (18.5)	0.698	32 (8.7)	2 (14.3)
Hospital mortality, No. (%)		25 (59.5)	1[f] (1.1)	n/a	3 (0.8)	11 (78.6)
30-day mortality, No. (%)		42 (100.0)	0 (0.0)	n/a	0 (0.0)	14 (100.0)

Note.
[a] Matching parameters were age, sex, care directive at alert, AAM score, COPS2 score, and LAPS2 score

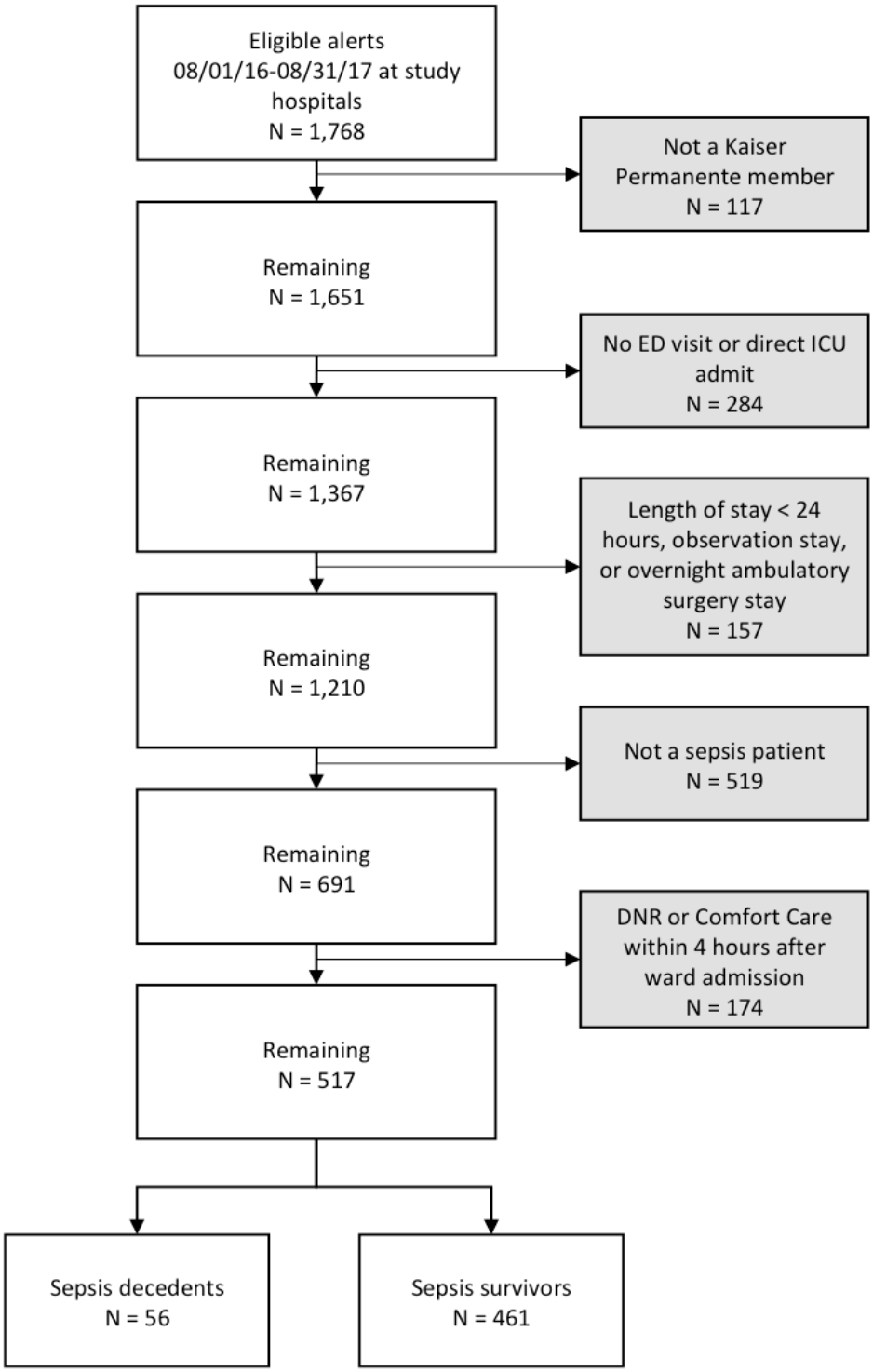
[b] Advanced Alert Monitor score predicting clinical deterioration risk (a higher score indicates higher risk)
[c] LAboratory-based Acute Physiology Score (LAPS2) is a severity of illness instrument using laboratory values and vital signs
[d] COmorbidity Point Score version 2 (COPS2 measures a patient's 1-year comorbidity burden calculated one month prior to admission
[e] Systolic Blood Pressure (mmHg)
[f] One matched survivor died in the hospital after 30 days

Table 2. Comparison of RRT Sepsis Interventions, Hospitalization and Alert Process Times, and Alert Shift Characteristics between Sepsis Decedents and Matched Survivors

	Matched[a] Decedents (n=42)	Matched[a] Survivors (n=92)	P value
Received fluid bolus, No. (%)	18 (42.9)	30 (32.6)	0.251
Time from alert to fluid bolus (minutes), median, mean±SD	-60 174 ± 541	-21 125 ± 392	0.714
Fluid bolus preceded[d] the alert, No. (%)	12 (28.6)	16 (17.4)	0.140
Total fluid bolus volume 6 hours before alert (liters), median, mean±SD	0.5 0.7 ± 0.3	0.5 0.7 ± 0.3	0.896
Total fluid bolus volume 24 hours after alert (liters), median, mean±SD	1.5 2.1 ± 2.5	0.75 1.0 ± 0.7	0.052
Received new intravenous antibiotic[e], No. (%)	6 (14.3)	19 (20.7)	0.380
Antibiotic preceded[d] the alert, No. (%)	0 (0)	10 (10.9)	0.026
Transfer to ICU, No. (%)	18 (42.9)	18 (19.6)	0.005
Number of vital signs following alert, median, mean±SD	51.5 86.3 ± 81.1	41.5 56.9 ± 49.2	0.011
Number of vital sign sets[c] following alert, median, mean±SD	6 5.7 ± 2.4	5 5.2 ± 2.0	0.262
Number of lab orders following alert, median, mean±SD	12 15.4 ± 11.3	9 11.0 ± 5.7	0.004
DNR[f] or Comfort Care after alert, No. (%)	29 (69.1)	13 (14.1)	< 0.001
Time from ED[b] entry to ward admission (hours), median, mean±SD	6.6 8.8 ± 6.1	6.2 7.6 ± 4.6	0.233
Time from ward admission to alert (hours), median, mean±SD	11.4 17.0 ± 16.3	9.1 13.6 ± 13.5	0.205
Time from alert to discharge (days), median, mean±SD	6.2 6.6 ± 4.4	4.3 7.8 ± 9.4	0.402
Length of hospital stay (days), median, mean±SD	6.8 7.3 ± 4.4	3.3 8.4 ± 9.5	0.456
Alert shift[g]			0.230
Night, No. (%)	20 (47.6)	41 (44.6)	
Day, No. (%)	9 (21.4)	40 (43.5)	
Evening, No. (%)	13 (31.0)	11 (12.0)	
Alert on weekend, No. (%)	10 (23.8)	30 (32.6)	0.302

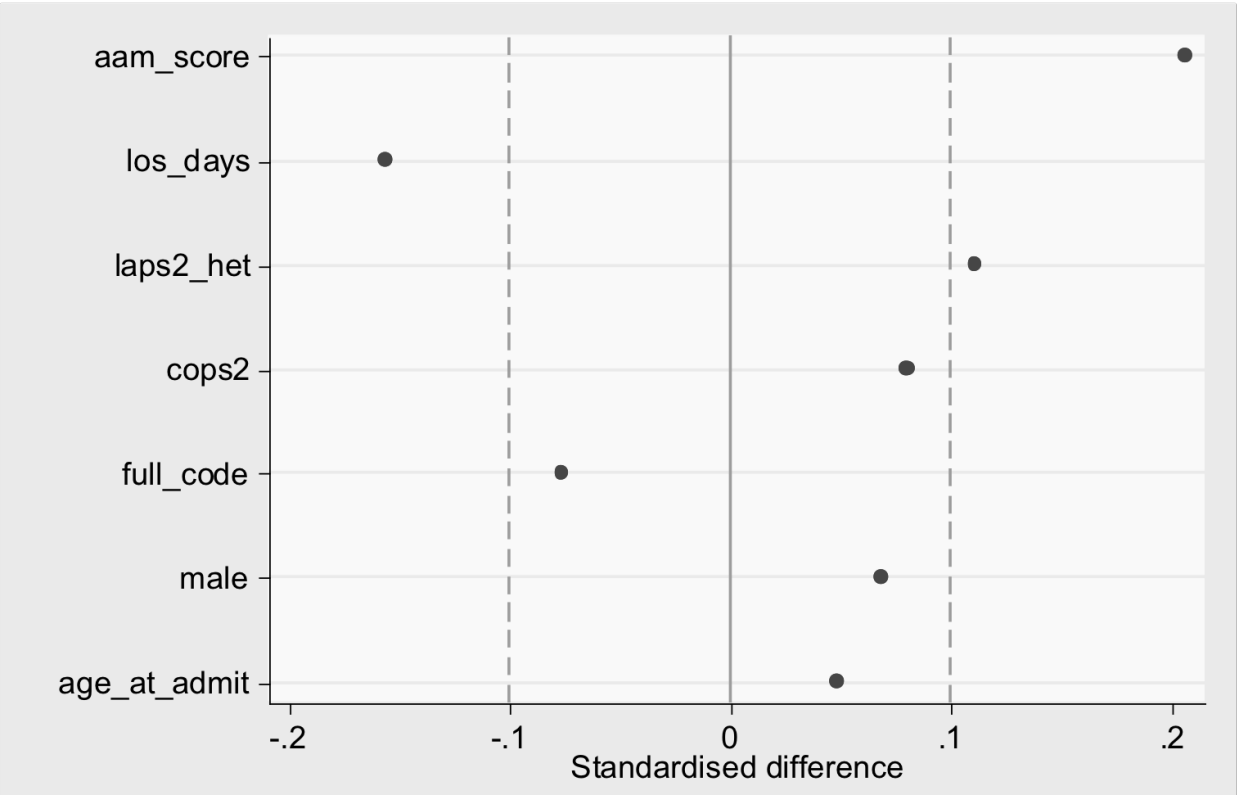
Note.
[a] Matching parameters were age, sex, care directive at alert, AAM (EWS) score, COPS2 score, and LAPS2 score
[b] Emergency Department
[c] Temperature, heart rate, blood pressure, respiratory rate, blood oxygen saturation level (SpO2);
[d] Preceding the alert: Intervention occurred within 6hs before the alert
[e] New antibiotic = not administered in the past 24 hours before the alert
[f] Do Not Resuscitate
[g] Shift times: night 00:00-07:59, day 08:00-15:59, evening 16:00-23:59;

Figure 1. Flow Diagram of Cohort Selection and Patient Exclusion Criteria in a Research Population of Patients on general wards (n=1,768) in 3 Community Hospitals



Note. Alerted patients were those with an advanced EWS risk score above a predetermined threshold. This threshold was selected to generate no more than one alert per 35-bed unit per day; DNR = Do Not Resuscitate

Figure 2. Graph of Standardized Difference Scores of Matching Variables between Sepsis Decedents and Survivors after Matching Procedures



Note. Stata14's *pbalchk* program computes standardized difference scores to demonstrate the balance of covariates between two matched groups. Scores below 0.1 are desirable. There were no statistically significant differences in any of the matching covariates between decedents and survivors. Matching on AAM (EWS) score occurred within 10 points between decedents and survivors. We removed length of stay as a matching variable after the first round of matching did not yield an adequate number of pairs.

Appendix

Table 1. List of ICD-10 Sepsis Codes Used for Inclusion of Patients for Matching Procedures

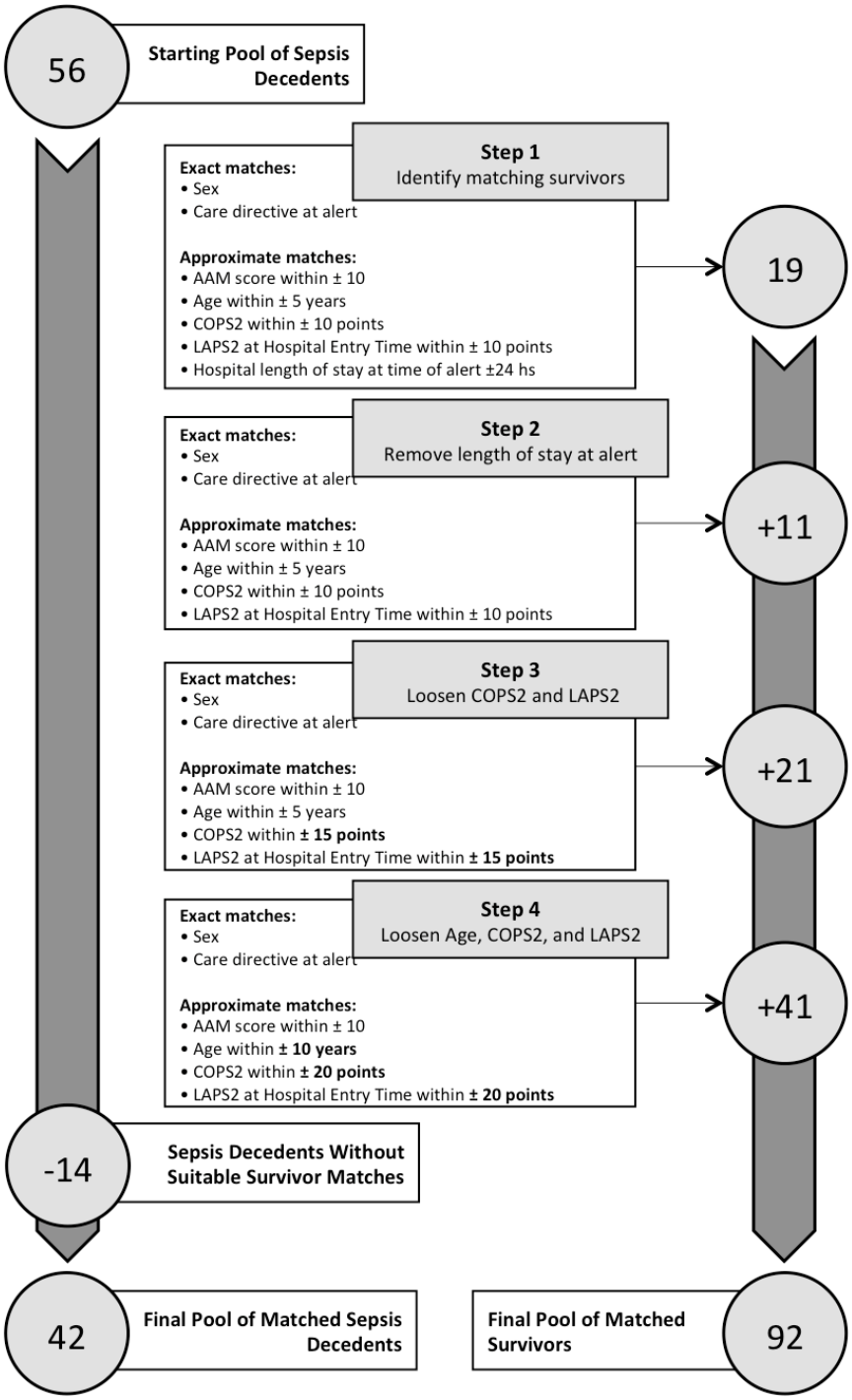
ICD-10 code	Description
A40.0	Sepsis due to streptococcus, group A
A40.1	Sepsis due to streptococcus, group B
A40.3	Sepsis due to Streptococcus pneumoniae
A40.8	Other streptococcal sepsis
A40.9	Streptococcal sepsis, unspecified
A41.0	Sepsis due to Staphylococcus aureus
A41.02	Sepsis due to Methicillin Resistant Staphylococcus aureus
A41.2	Sepsis due to unspecified staphylococcus
A41.3	Sepsis due to Hemophilus influenza
A41.4	Sepsis due to anaerobes
A41.50	Gram-negative sepsis, unspecified
A41.51	Sepsis due to Escherichia coli
A41.52	Sepsis due to Pseudomonas
A41.53	Sepsis due to Serratia
A41.59	Other Gram-negative sepsis
A41.81	Sepsis due to Enterococcus
A41.89	Other specified sepsis
A41.9	Sepsis, unspecified organism
R65.20	Severe sepsis without septic shock
R65.21	Severe sepsis with septic shock

Table 2. Standardized Difference Tests Between Matched Sepsis Decedents and Survivors Demonstrate Adequate Match Quality

	Total (n=134)	**Matched Decedents (n=42)**	**Matched Survivors (n=92)**	**Standardized difference score**	**P value**
Age at Admission [median, mean ± SD]	75.0, 73.7 ± 10.7	76.0, 74.1 ± 12.7	74.5, 73.5 ± 9.7	0.049	0.783
Male [n (%)]	67 (50.0)	22 (52.4)	45 (48.9)	0.069	0.710
Full Code at 1st Alert [n (%)]	132 (98.5)	41 (97.6)	91 (98.9)	-0.076	0.567
AAM score[a] [median, mean ± SD]	10.6, 12.9 ± 5.6	12.1, 13.7 ± 5.7	10.4, 12.5 ± 5.6	0.206	0.266
LAPS2 score[b] at HET [median, mean ± SD]	117.0, 116.4 ± 23.7	121.0, 118.3 ± 26.0	115.5, 115.6 ± 22.6	0.111	0.541
COPS2 score[c] [median, mean ± SD]	72.0, 77.3 ± 44.5	71.5, 79.8 ± 45.7	72.0, 76.2 ± 44.2	0.080	0.665
Length of Stay at Alert (Hours) [median, mean ± SD]	10.2, 14.7 ± 14.4	11.3, 17.0 ± 16.3	9.1, 13.6 ± 13.5	-0.155	0.203

Note.
[a] Advanced Alert Monitor score predicting clinical deterioration risk (a higher score indicates higher risk)
[b]LAboratory-based Acute Physiology Score (LAPS2) is a severity of illness instrument using laboratory values and vital signs
[c] COmorbidity Point Score version 2 (COPS2 measures a patient's 1-year comorbidity burden calculated

Figure 1. Flow Diagram of the Matching Process to Derive Comparable Groups of Sepsis Decedents and Survivors



Note. Advanced Alert Monitor score predicting clinical deterioration risk (a higher score indicates higher risk); LAboratory-based Acute Physiology Score (LAPS2) is a severity of illness instrument using laboratory values and vital signs; COmorbidity Point Score version 2 (COPS2 measures a patient's 1-year comorbidity burden)

Chapter 5

Conclusion

Overview

This final chapter offers a summary and synthesis of the dissertation's findings, a description and interpretation of results, and a discussion of implications for research and practice. We conceived this dissertation to answer an important clinical question, which previous rescue research has not addressed: Do sepsis interventions by rapid response teams (RRTs) - relative to an advanced early warning system (EWS) alert - improve survival of sepsis? The dissertation had two primary aims: 1) to categorize fluid bolus processes (FBPs) and other sepsis interventions relative to an advanced EWS alert via chart reviews of electronic medical records; and 2) to characterize and compare alike sepsis survivors and decedents admitted in stable condition and to evaluate the impact of fluid bolus processes and other sepsis interventions on 30-day survival in hospitals with an advanced EWS.

Fulfilling the first aim, we performed expert chart reviews to examine sepsis interventions performed by RRTs relative to EWS alert time, mapped the discrete locations of FBP data and other sepsis intervention data in the electronic medical record, and categorized FBPs in terms of elapsed time from EWS alert to bolus and total 24-hour bolus volume (see Chapter 3). In fulfillment of the second aim, we conducted a matched-pair cohort study using multivariate pair matching procedures of sepsis decedents and matched survivors, described the groups using summary statistics, and quantified the between-group differences of fluid bolus processes and other sepsis interventions in decedents and survivors (see Chapter 4).

We evaluated the following alternative hypotheses for statistical significance ($p < 0.05$):

H_1: Elapsed time from EWS alert to administration of IV fluid bolus will be shorter in survivors compared to decedents, after pair-matching adjusted for patient age, sex, EWS score, patient comorbidity [COmorbidity Point Score version 2 (COPS2)], and severity of illness [LAboratory-based acute Physiology Score (LAPS2)].

H_2: Total 24-hour fluid bolus volume following an alert will be larger in survivors compared to decedents, after pair-matching adjustment for patient age, sex, EWS score, length of stay, patient comorbidity (COPS2), and severity of illness (LAPS2).

Summary of results

The following section will offer a summary and synthesis of the results of the three dissertation papers.

Chapter 2: Systematic review of the advanced EWS literature

Prior to conducting studies to quantify the impact of alert-driven sepsis interventions on patient survival, it was imperative to confirm whether advanced EWSs using multivariate regression or machine learning algorithms appear to demonstrate superior precision compared to traditional point-score EWS tools. Little benefit would come from this dissertation unless we could first determine that advanced EWSs are a sound patient safety investment for health systems, given that basic EWS tools are already in use in many hospitals. We systematically compared model performance of advanced EWSs vs. point-score EWSs. Then, we evaluated the models' ability to

correctly identify cases and quantified the RRT workload to find (and treat) one positive case.

We found that studies using advanced EWSs were relatively scarce. Ours is the first systematic review that compared the diagnostic performance of EWSs using multivariate regression or machine learning against point-score warning systems. Only 5 studies met inclusion criteria, of which 1 demonstrated considerable threats to validity and lack of measurement rigor. The remaining 4 studies were of good quality; they used large sample sizes by pooling data from multiple hospitals, employed sophisticated modeling and model validation techniques, and used rigorous measurement methods. It is an illustration of the novelty of the field, that these 4 studies occurred at only 2 research centers: The University of Chicago and the Kaiser Permanente Northern California (KPNC) Division of Research.

Advanced EWSs have higher prognostic accuracy and appear to offer better case detection properties (precision) than point-score EWSs, while generating fewer alerts for RRTs to evaluate. Consequently, advanced EWSs may improve resource efficiency and appear to reduce clinical evaluation workloads by reducing false positive case identification. Our conservative estimate of waste reduction (fewer unnecessary patient evaluations by RRTs and hospitalists) is ~$65,500 per hospital per year, or nearly $350 million per year across the entirety of U.S. hospitals. This benefit offers indirect savings, because RRTs and hospitalists are staffed roles; however, their time may be better utilized by performing other essential functions, with less distraction and less risk for burnout. Despite these benefits, it is important to note that even advanced

EWSs are calibrated to detect about half of all actual cases. Therefore, they can at best augment and support, but not replace, clinical monitoring and vigilance.

Chapter 3: Chart reviews

In Chapter 3, we reviewed the medical records of 68 sepsis patients admitted to Kaiser Permanente Northern California (KPNC) hospitals to evaluate how RRTs performed sepsis interventions relative to the alert and to ensure measurement rigor in the subsequent study. During two rounds of clinical chart reviews, we captured patient demographics, clinical patient characteristics of severity of illness (LAPS2) and comorbidity burden (COPS2), and the timing of intravenous antibiotics, transfer to ICU, and Do Not Resuscitate orders relative to the alert. We then summarized results using descriptive statistics and graphed results.

Regarding the use of “DNR order before death” as an exclusion criterion for “expected death”, we found that a very large share of sepsis decedents (86%) had a DNR order in place before their death. As such, a DNR order before death did not appear to reflect the question whether a patient was on a fixed end-of-life trajectory. We also noted a second critical problem: In patients with an EWS alert immediately following admission, sepsis interventions tended to precede the alert. These patients often received the initial bundle of sepsis care in the emergency department (ED), meaning that additional sepsis interventions at alert time may not have been indicated again. Since the overall purpose of the dissertation was to quantify new interventions motivated by the alert, we elected to exclude patients with alerts within 12 hours after ED entry in the subsequent study.

Results from this study had important implications for the design and measurements of the subsequent study (Chapter 4). Our findings demonstrated the crucial importance of clinical chart reviews prior to embarking on data collection and analysis, and they contributed to the body of sepsis and RRT science by highlighting and proposing solutions to distinct challenges of measurement and patient selection.

Chapter 4: Matched pair cohort study

Satisfying Aim 2, the evaluation of the between-group differences of fluid bolus processes and other alert-driven sepsis interventions delivered a novel understanding of the impact of alert-driven FBPs on sepsis survival. By categorizing sepsis interventions and comparing sepsis survival outcomes, this dissertation study aimed to improve the empiric evidence base regarding the effective delivery of fluid bolus therapy. We ensured internal and external validity of measurements and results by using a rigorous sampling strategy, well circumscribed measurement definitions, sophisticated composite measurement tools with external validation, and by following a research plan that went through a total of three distinct development stages.

Using a cohort of alerted patients admitted to general wards in 3 KPNC hospitals, we performed multivariate matching procedures to derive alike groups of sepsis decedents and survivors. We then statistically compared demographics, clinical patient characteristics, hospital transfer times, sepsis interventions, intervention times, and characteristics of the event shift between these two groups. Decedents and survivors were comparable across all demographics and clinical characteristics. All patients were admitted in stable condition, their EWS risk scores adjusted for severity of illness, and the mean EWS scores did not differ substantially.

We found that, after matching adjusted for patient demographics and clinical characteristics, sepsis decedents received statistically significantly more fluid bolus volume, more vital sign measurements, more laboratory tests, more transfers to the intensive care unit (ICU), and do not resuscitate (DNR) orders after the alert. Survivors received a new intravenous antibiotic earlier, and often before the alert. We did not detect statistically significant differences in terms of hospital transfer times from ED to alert to ward to discharge, timeliness of fluid bolus therapy, and characteristic of the alert shift.

Given these early results, we failed to reject the dissertation's null hypotheses of H_1 (no difference in elapsed time from alert to fluid bolus administration between sepsis decedents and survivors) and H_2 (no difference in total fluid bolus therapy within 24 hours after alert between sepsis decedents and survivors). Because we adjusted for a wide range of clinical covariates during the matching procedures, including severity of illness, comorbidity, EWS alert score, among others, our findings are generalizable for similar patients, but not all patients on general wards.

Interpretation of results

This dissertation offers important new insights for the field of predictive health analytics and data science in patient-centric outcomes research and patient safety. Advanced EWSs appear to possess favorable properties, including better prognostic precision, which translates into RRT and physician efficiencies. While we excluded patients for whom sepsis care in the ED may have created treatment overlap with sepsis interventions at the alert time, we were unable to confirm timelier or more complete alert-driven fluid bolus processes in survivors. Instead, we found that

decedents received more fluids overall and often before the alert, though not reaching statistical significance. Despite matching, the severity of illness score (LAPS2) at admission and alert time were higher in decedents. The 10-point difference in LAPS2 scores between decedents and survivors (though not statistically significant at alpha of 0.05) suggests that decedents were considerably sicker at the time of alert. Given that this difference was less pronounced between the groups at the time of admission, the severity of illness slope appears to be stepper in decedents. This would then explain, at least in part, their receipt of more fluid bolus therapy. Since statistical significance is influenced by sample size, it would be desirable to evaluate total fluid bolus volume 24 hours after the alert and LAPS2 slopes in future research using a larger cohort.

Compared to 10 out of 92 survivors, none of the decedents received a new antibiotic before the alert ($p = 0.026$). This finding is congruent with prior evidence regarding the impact of timely antibiotic therapy on sepsis survival. We hypothesized that an advanced EWS would have the utility of motivating additional sepsis therapy following the alert; however, the protective effect of antibiotics appears to occur before the alert. Nevertheless, because we observed a protective effect of earlier antibiotic coverage and fewer new antibiotics in decedents overall, it would be reasonable to prompt clinicians to consider additional antibiotics at alert time.

It appeared that many patients expired despite appropriate care and treatment intensification following the alert. Therefore, findings from this dissertation suggest that there may be unmeasured contexts that could explain the observed differences. For example, we did not account for the potential of longer sepsis exposure time before ED entry in decedents (because such an exposure measure is not reliably documented), or

the relative increase in illness severity following admission, and the degree to which a patient's death may have been expected or could be considered unavoidable. These factors may confound or influence treatment intensification in decedents.

In Chapters 3, 4, and 5, we discussed the dissertation's limitations in detail. The greatest limitation for this dissertation study was the relatively small sample size; some treatment effects may prove to be statistically significant simply by having access to larger sample sizes in future research. For example, post-hoc power calculations showed that the treatment effect of total fluid bolus therapy ($p = 0.052$) would have reached statistical significance with only 4 additional decedents (needed n=46, had n=42). A second important limitation was the inability to measure mortality outcome quality. We excluded patients with a Do Not Resuscitate or Comfort Care order immediately following hospital admission and observed that nearly 70% of decedents had a DNR order in place after the alert. These results suggest that there may be additional unmeasured differences in decedents in terms of their potential predetermined end-of-life trajectory. In such patients, palliative care may be an equally important intervention.

Implications for research and practice

The dissertation marks an important foundational milestone, because no research to date has evaluated the impact of the pairing of RRTs and an advanced EWS alert on sepsis survival outcomes. We successfully addressed all aims and research questions. Nonetheless, the overall scientific objective of evaluating the impact of alert-driven RRT interventions has at least five future-facing implications. We will discuss these implications in the following section.

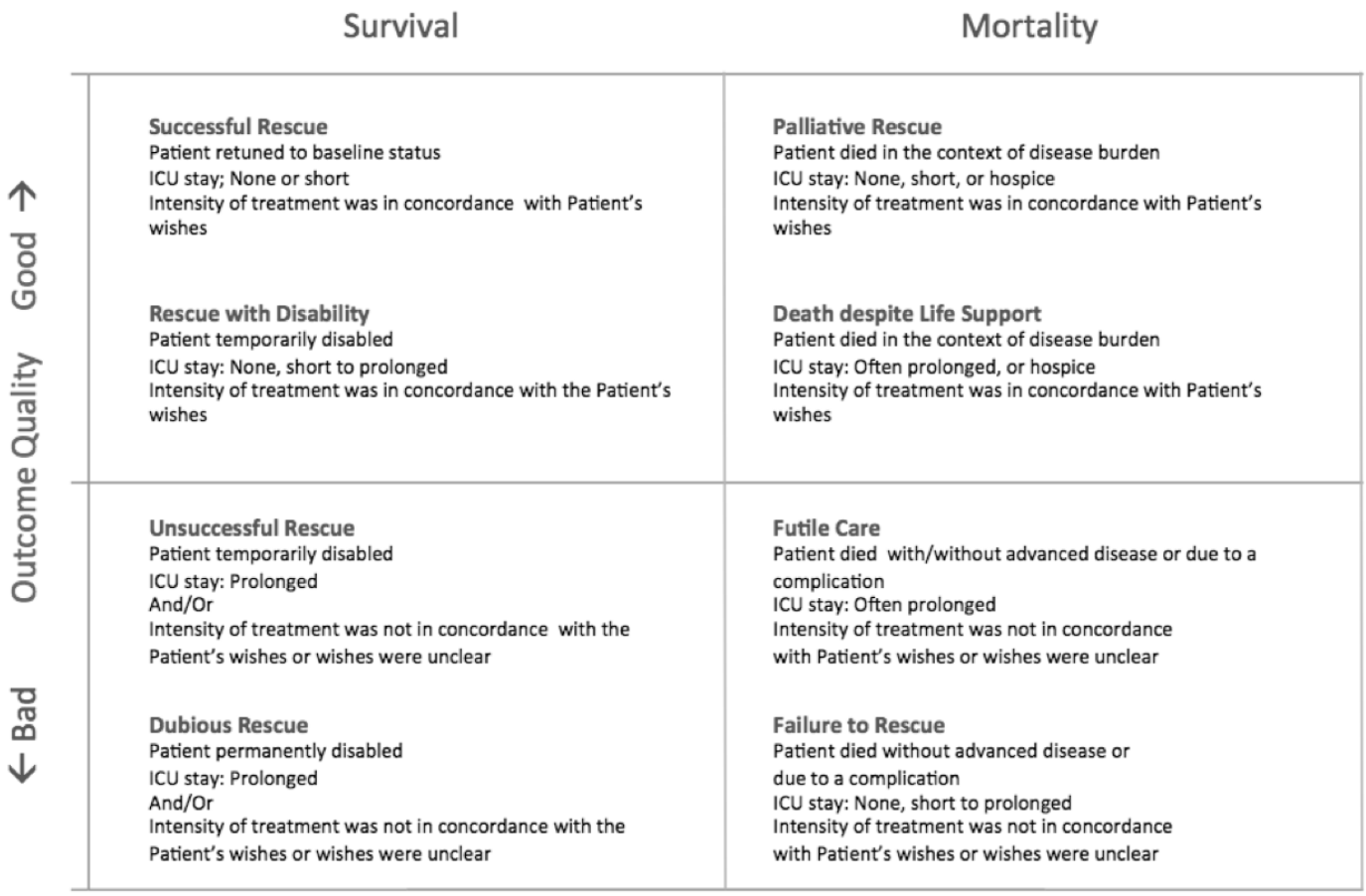
1. Outcome quality measurement

Studies examining the quality of patient survival and mortality are needed in the field of alert-driven rescue, because an apparent outcome of lived vs. died may not necessarily reflect true outcome quality. Not every patient who survives sepsis returns to their baseline function (Iwashyna, Ely, Smith, & Langa, 2010) and not every patient who dies always experiences a “bad death” (Costello, 2006). Treatment concordance with the patient’s wishes and unfavorable sequela resulting from life support (e.g., disability) moderate the quality of survival and mortality (see Figure 1).

For example, patient survival with long-term disability may be an unfavorable outcome if the patient’s treatment wishes were unclear or if the patient was terminally ill. Similarly, not every death is always a bad outcome, given treatment concordance with the patient’s wishes and presence of terminal illness. This lack of definitive measurement of outcome quality is problematic for two reasons: 1) the fundamental task of predictive early warning systems is to identify patients for rescue; and 2) value confirmation of alert driven rescue requires a reliable quantification of observed vs. expected outcomes.

Clearly, a patient-centric approach to rescue would include consideration for the patient’s ultimate treatment wishes, although patients may not always define these wishes clearly or completely (Stephens et al., 2018). While patients may not forgo all therapy in the light of a worsening condition, alert-driven outcome measurement would be more complete if it included palliative care interventions and a marker for terminal disease, life expectancy, frailty and quality of life.

Figure 1. Patient Scenarios of Outcome Quality of Survival or Mortality Following an Advanced Early Warning System Alert and Clinical Rescue Interventions



2. Larger sample size

This dissertation offered early results using a limited number of KPNC hospitals with a deployed advanced EWS. Once the EWS is fully implemented across the region's 21 hospitals, future KPNC studies would benefit from this larger pool of hospitals. Additionally, by pooling data between institutions, collaborative studies could be performed that would increase the sample size and the number of the hospital clusters for hierarchical analysis. Such larger datasets would facilitate greater statistical power and may confirm subtle treatment effects of sepsis interventions on patient survival.

3. Inclusion of additional diagnostic strata

Rather than sepsis alone, it would be desirable to evaluate the impact of RRT interventions on patient survival in the general population of all alerted patients. By developing diagnostic strata (e.g., surgical, cardiac, respiratory, sepsis), such a study could facilitate a greater understanding of the areas of highest benefit and value of RRTs and predictive EWS in terms of patient safety.

4. Hospital-level treatment effects and intra-system variation

For quality measurement purposes, health systems would benefit from future health services research that examines the degree of outcome variation between hospitals. Such knowledge would be crucial as health systems are under the dual pressures of financial stewardship and quality improvement to maximize reimbursements and gain a market advantage. By adjusting for patient-level and hospital-level characteristics, such a study could identify "hospital fingerprints" and could motivate targeted performance improvement in outlier hospitals.

5. Value confirmation of Alert Driven Rescue

The field of delivery science examines how research findings can be successfully implemented in a healthcare delivery system (Grant & Schmittdiel, 2015). Health systems engaging in this translational implementation of research are also called Learning Health Systems (Liu et al., 2016). This dynamic is consistent with the Petri Dish Model, which describes the dual mandates of organizational survival and safeguarding patient safety: Delivery science represents the information feedback loops that send value confirmation of locally deployed interventions upstream to the regional level. This function, ultimately, supports an organization's efforts of achieving the quadruple aim of optimal patient experience, population health, cost-reduction, and workforce well-being (Sikka, Morath, & Leape, 2015).

In this light, health economics studies investigating the costs, benefits, and return on investment (ROI) of advanced EWSs are needed. Findings from this dissertation could lead to additional econometric analyses. For example, to date, the financial value of predictive health analytics is not well-established, although notable use cases include the identification of readmission risk (tied to federal reimbursement penalties) and the identification of patients with patterns of high service utilization. Two main obstacles are large investments in, and maintenance of, electronic medical records and substantial up-front costs of computationally intensive analytic solutions. Health systems may have little motivation to invest in such analytic solutions unless they have a favorable ROI that goes beyond qualitative benefits or indirect costs saving. Future delivery science studies of the advanced EWS at KPNC could inform such an ROI, and thus more widespread adoption in other health systems.

Implications for Clinical Practice

To minimize treatment variation across clinicians, the ultimate goal of decision support tools is not only to alert, but also to advise. For RRTs, it would be desirable to derive specific algorithmic treatment recommendations following an alert (e.g., in patients with risk score of x, administer y ml of fluid within z minutes). Knowing that decedents did not receive any new antibiotics before the alert, the EWS could prompt an evaluation for new antibiotic coverage. This dissertation has produced early insights; what is needed now are confirmatory analyses that could lead to more prescriptive practice standards. Such work has the potential to advance the scientific evidence base of sepsis interventions and to standardize clinical practice and RRT workflows.

Conclusion

In closing, the use of advanced analytics using multivariate regression and machine learning to predict health outcomes has gained undeniable momentum. Many hospital-based “nurse-sensitive” outcomes (e.g., falls with injury, pressure injuries, catheter associated urinary tract infections, central line associated blood stream infection, patient satisfaction with nurse communication and pain management) are under the professional purview of nursing, in collaboration with the allied health professions of medicine, pharmacy and others. By lacking risk-adjustment and pattern detection, traditional methods of quality measurement and data analysis have limited ability to identify high-risk patients, or in health systems, to identify hospitals with outlier performance. Consequently, the sector needs clinical academicians who can traverse the domains of quantitative research, advanced computational methods, and health systems leadership.

References

Costello, J. (2006). Dying well: nurses' experiences of 'good and bad'deaths in hospital. *Journal of Advanced Nursing, 54*(5), 594-601.

Grant, R. W., & Schmittdiel, J. A. (2015). Building a career as a delivery science researcher in a changing health care landscape: Springer.

Iwashyna, T. J., Ely, E. W., Smith, D. M., & Langa, K. M. (2010). Long-term cognitive impairment and functional disability among survivors of severe sepsis. *JAMA, 304*(16), 1787-1794.

Liu, V., Morehouse, J. W., Baker, J. M., Greene, J. D., Kipnis, P., & Escobar, G. (2016). Data that drive: Closing the loop in the learning hospital system. *J Hosp Med, 11 Suppl 1*, S11-s17. doi:10.1002/jhm.2651

Sikka, R., Morath, J. M., & Leape, L. (2015). The Quadruple Aim: care, health, cost and meaning in work. *BMJ Quality & Safety*. doi:10.1136/bmjqs-2015-004160

Stephens, C. E., Hunt, L. J., Bui, N., Halifax, E., Ritchie, C. S., & Lee, S. J. (2018). Palliative care eligibility, symptom burden, and quality-of-life ratings in nursing home residents. *JAMA Internal Medicine, 178*(1), 141-142. doi:10.1001/jamainternmed.2017.6299

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Author Signature  Date: 6/5/2018