RESEARCH ARTICLE

A continuously updated predictive analytics model for the timely detection of critically ill patients with a high risk of mortality

Authors
Andrew A. Kramer, PhD FCCM1
1 Prescient Healthcare Consulting, Charlottesville, VA, USA.
E-mail address: akramer@prescient-healthcare.com
Disclosure: Prescient Healthcare Consulting has a patent submission on the technology described in this manuscript

Abstract

Intensivists routinely encounter seemingly stabilized patients who expire before discharge. Current early warning systems have not proven effective for identifying these patients in sufficient time for clinical intervention. A novel algorithm is described here that is timely and accurate for recognizing patients in an intensive care unit (ICU) that have a substantial mortality risk.
Population: 59,400 admissions at 13 adult ICUs from 1/1/2012-9/30/2018.
Outcome: Mortality before discharge from the ICU. Overall rate was 6.9%.
Methods: All heart rate, respiratory rate and oxygen saturation values were obtained, as well as the start and stop times for those patients receiving mechanical ventilation. Data from the first two hours post-admission were used to find the cut points that maximized variability rates across ranges of vital signs values. These ranges were subsequenty mapped to a letter. A letter was then assigned to the median of each vital sign over consecutive 30-minute periods. Four consecutive letters were concatenated to form a pattern, and these were candidates for triggers (i.e. risk alerts). Using a genetic algorithm that weighted the outcome of mortality, we acquired a set of patterns that increased risk. Those patterns were then validated as triggers for increased risk.
Results: Patients with zero or one triggers had a mortality rate of 0%; patients with two to four triggers had a mortality rate of between 2.8% and 5.5%; five or more triggers were seen in patients with a 20.2% to 25.6% mortality rate.
Conclusion: Distinctive patterns in vital signs and whether a patient received mechanical ventilation can identify patients that have a high risk of mortality. This methodology could be prospectively used in ICUs to identify high-risk patients in a timely enough manner to effect remedial treatment.

Keywords: Intensive care unit, pattern recognition, mortality, vital signs, mechanical ventilation

1. Introduction

Predictive models of outcomes in the intensive care unit have been around for decades. The outcome most commonly predicted is mortality, either before a unit or hospital discharge. Currently, APACHE IV[1], SAPS 3[2], ACUITY 2016[3], and ICNARC[4] are the models of greatest use throughout the world. All of these models use data collected throughout the first 24 hours after admission to produce a patient's probability of mortality. The probabilities across patients within an intensive care unit (ICU) are then compared with the number of outcomes to arrive at a Standardized Mortality Ratio (SMR). SMRs tracked over time within a unit or among ICUs during the same time interval are useful ways of measuring an ICU's risk-adjusted performance. APACHE and similar models are quite accurate at the group level. However, they are not precise in assessing an individual's likelihood of dying before unit discharge[5]. The reason is that these models do not include variables such as processes of care or change in physiology over time, and thus contain a high degree of noise. This noise becomes canceled out when groups of patients, such as all patients within an ICU, are considered.

There has been an increasing demand for predictions that are actionable at the bedside[5]. These systems are called "predictive analytic solutions" or "clinical decision support systems". This type of system would enable clinicians to provide timely care for patients who might seemingly be stable but are actually at high risk for an adverse event. Early attempts to create such a metric involved looking at extremes in physiology at discrete and often infrequent time points. Systems developed for patients on general care units include MEWS[6], NEWS[7], and NEWS2[8]. However, their accuracy has been questioned[9].

More robust predictive analytic solutions are available for non-critical venues within acute care hospitals. The Visensia Index[10] was developed using data from patients admitted to step-down units. It incorporates data from heart rate, respiratory rate, temperature, blood pressure, and SaO_2 over time and looks for multidimensional outliers. A clinical study showed that it generated alarms an average of 5.8 hours before a medical emergency team activation[11]. Pera Health's Rothman Index produces a constantly updated acuity score but relies heavily on nursing assessments and the Braden score. These are collected sporadically and are subjective. The Rothman Index has not been thoroughly tested in ICUs, and its use to predict readmissions to an ICU was based on a retrospective study at a single ICU[12]. Thus there is no present system capable of providing a bedside predictions for patients in the ICU.

To be beneficial, a predictive analytics system for the ICU needs to meet several criteria[13]. First, it needs to incorporate data that were acquired electronically. Using manual data entry is not timely enough for a fluid environment such as an ICU. Second, the system must be able to generate near real-time alerts for patients at an elevated risk of an adverse outcome. These alerts must be capable of identifying seemingly normal patients who are actually at risk and not just signal patients who are decidedly ill. Finally, the alerts need activation hours before a pre-defined adverse event takes place.

Previously a prototype predictive analytics solution for the ICU was described for use in making discharge decisions from the ICU[3]. Called SIGNIPHY™, it used vital signs data starting at two hours before discharge to identify patients whose physiologic patterns suggested either a high risk for dying on the floor post-ICU discharge or being discharged to a hospice. This paper greatly expands on SIGNIPHY to provide a continuously updated risk analysis of a patient expiring before ICU discharge: rSIGNIPHY. We describe a study where rSIGNIPHY was successfully applied to a large ICU database, and discuss future directions that such a predictive analytics solution for the ICU could take.

2. Methods

2.1 Data used in this study

We obtained the following vital signs data from a large commercial database ("Phoenix", Medical Decision Network, Charlottesville, VA, USA): heart rate, respiratory rate, and mean arterial pressure (map). We also extracted data on the start date-time and stop date-time for patients placed on mechanical ventilation, as well as ICU admission date-time and ICU discharge date-time. Finally, each patient's discharge status (alive, dead) was collected. The dataset was split 2:1 by admission date-time to allocate patients to a development data set or a validation data set, respectively. The validation data set took the parameters created in the development data set and verified them on a new set of data.

2.2 Assigning symbols to ranges of vital signs that maximize mortality variation

The first part of rSIGNIPHY involved splitting each vital sign into ranges where mortality was optimized. We extracted vital signs from the first two hours after ICU admission and then calculated the median value across that time. A genetic algorithm[14] was employed to find the bins (i.e., ranges of the vital sign's distribution) that maximized variation in mortality. Genetic algorithms are multi-parameter optimization algorithms that can handle boundaries and constraints. Formula 1 shows the fitness function optimized by the genetic algorithm:

Formula 1.

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (\|m_i - m_j\|) \Big/ n$$

Where n = the number of bins, m_i is the mortality rate in bin "i', and m_j is the mortality rate in bin j. Formula 1 had constraints on the number of bins (3, 4, or 5) and the minimum bin had to have at least 250 patients. The reason for the latter was to guard against the trivial situation where one bin contained all mortality. Each bin was assigned a symbol, which in itself had no quantitative value, but was used to designate a specific range of values for a vital sign. Table 1 gives an example of the bins and the corresponding symbols assigned to them.

Table 1. Hypothetical assignment of letters to bins of median respiratory rates

Respiratory Rate Range	Symbol Assigned
3-18	X
19-31	N
32-36	A
37-90	Q

2.3 Temporal data mining of vital signs and ventilator status

We attached symbols to each patient's 30-minute median heart rate, respiratory rate, and map, respectively, across their ICU stay. At the end of each two-hour interval, the four symbols from the four preceding 30-minute periods were concatenated to form a "word". This process was repeated until the patient either was discharged from the unit, died before discharge, or was still in the unit seven days after discharge; whichever came first. Consequently patients had varying numbers of words, capped off at a maximum of 84 (168 hours in a week). Each word was randomly assigned a weight = {0, 1, or 2}.

Every two-hour interval also included whether or not a patient received mechanical ventilation. If the patient had never been on a mechanical ventilator, then he was given an MV score = 0; currently on a mechanical ventilator yielded an MV score = 2; being previously on a ventilator resulted in an mv score = 1. The reason for the latter circumstance receiving a point is that having been on a ventilator, even if currently weaned off of it, still presents risks, such as ventilator-associated pneumonia.

A patient's score for a particular two-hour interval consisted of the points assigned to heart rate, respiratory rate, map, and MV score. Summing the weights resulted in a score ranging from 0 to 8. Starting with time period two, a patient's score became the maximum of the current and immediately preceding time period to introduce some lag in physiologic effects. A smoothing factor was added to account for the decreasing effect of more distant time periods (see Formula 2).

Formula 2

Final $Score_t$ = ω*($Score_t$) + (1-ω)*(Final $Score_{t-1}$),

Where $0.1 \leq \omega \leq 0.9$. A value of $\omega = 0.5$ means that equal weight is placed on the score at this time period and the final score from the previous time periods. ω is not pre-determined but obtained as a result of the optimization process described below. The smoothing in Formula 2 also converts the Final Score from being an integer to a continuously distributed metric, ranging from 0.00 to 8.00.

We selected the maximum Final Score over all time periods: Score(max). Then a genetic algorithm was used to find a weight for each word {0, 1, or 2} that optimized a patient's Final Score (max) with mortality. That entailed optimizing hundreds of weights simultaneously. The fitness function for the genetic algorithm had a further component, in which patients with a low score who died (false negative) were penalized more heavily than patients with a high score who survived (false positive). The result of the genetic algorithm is that every heart rate word, respiratory rate word, and mean arterial pressure word has a final weight of 0, 1, or 2, and ω maximized at a value between 0.1-0.9.

Patients were grouped into 5% percentiles according to their Final Score(max). The mortality rate and median of the Final Score(max) for each five-percentile group were obtained and then graphed. Inflection points where the mortality rate accelerated were chosen as alarms: the first inflection point indicated a "moderate risk", and the second inflection point was associated with a "high risk".

Each patient in the validation data set had their Final Score(max) calculated. A confusion matrix was created that looked at the number of patients alive and dead, respectively, at discharge according to their risk profile: none, moderate, and high, as defined from the development data set. Sensitivity and specificity were calculated based on having either a moderate or high risk.

3. Results

The Phoenix database consisted of 59,400 patients admitted to thirteen adult ICUs in the U.S. between 1/1/2012 and 9/30/2018. Patients admitted before 1/1/2016 (40,047; 67.4%)

made up the development data set, and the 19,353 (32.6%) patients admitted from 1/1/2017 onwards were allocated to the validation data set. The overall ICU mortality rate was 6.9%.

Table 2 shows the bins for each vital sign yielding the maximum mortality validation, along with the symbol mapped to each.

Table 2. Bins from calculating the maximum mortality variation and their corresponding symbol

Heart Rate		Respiratory Rate		Mean Arterial Pressure	
5-85	R	4-18	F	30.0 – 53.8	S
86-122	A	19-32	Z	53.9 – 61.9	D
123-140	E	33-37	Q	62.0 – 73.6	B
> 140	M	> 37	P	73.7 – 81.5	N
				≥ 81.6	W

There were four bins for heart rate, four bins for respiratory rate, and five bins for map. From the map bins' values, it is clear that hypotension rather than hypertension carries significant risk.

Each four consecutive time periods yielded words. This process yielded 132 words for heart rate, 125 words for respiratory rate, and 293 words for map. (Some words did not appear in this data set.) Thus the genetic algorithm had 551 weights to optimize, one more than the sum of the words due to the need to optimize the smoothing parameter.

The genetic algorithm found 270 words that had a weight > 0. These words and the patient being/had been on a mechanical ventilator contributed to a patient's score at each time period. Table 3 gives the maximum score for each patient vs. mortality for the development and validation data sets, respectively. Scores are rounded to the nearest integer

.

Table 3. Mortality Rate by Maximum Score

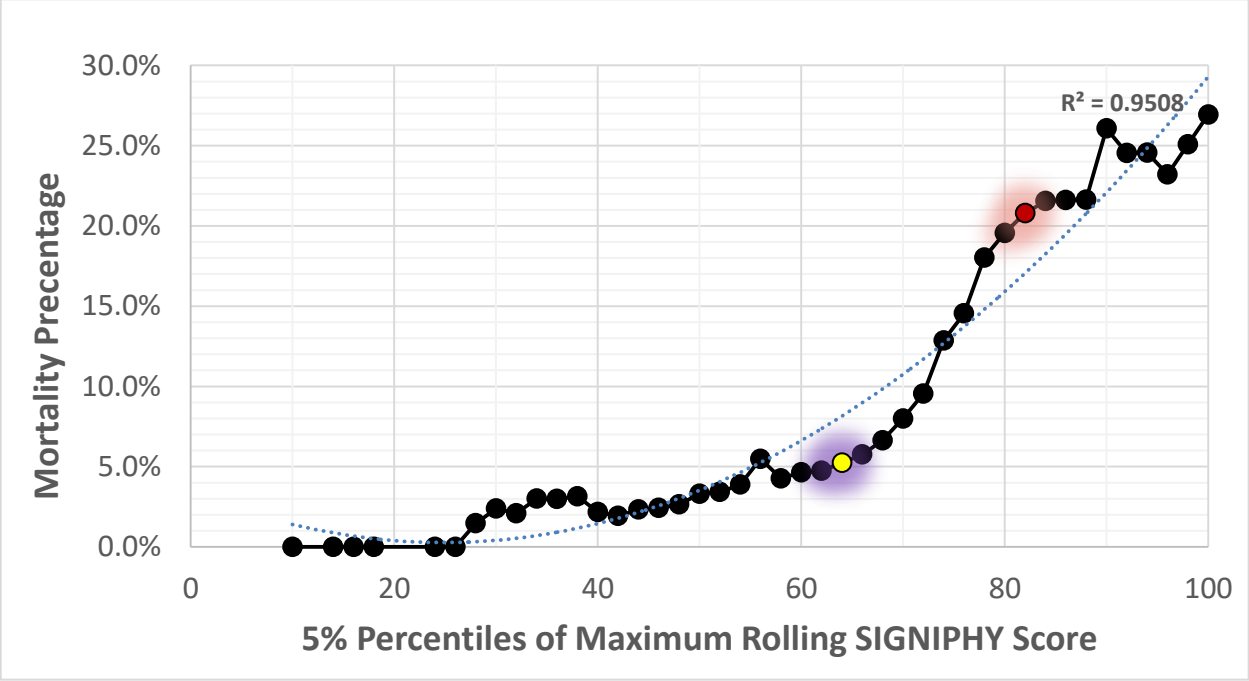
Development Data Set		Validation Data Set	
Maximum Score	Mortality rate	Maximum Score	Mortality Rate
0 – 1	0.0%	0 – 1	0.0%
2	1.8%	2	2.8%

3	2.3%	3	3.3%
4	5.1%	4	5.5%
5	17.8%	5	20.2%
≥ 6	21.9%	≥ 6	25.6%

There was a strict monotonic relationship between patients' maximum score and the corresponding mortality rate. This relationship was almost identical to that found using the validation data set, affirming the success of rSIGNIPHY to identify patients at an increased risk of mortality.

The Final Score(maximum) for each patient was rescaled to lie between 0 – 100. Figure 1 shows increasing 5% percentiles of the Final Score(maximum) for patients versus the corresponding mortality rate in the validation data set.

Figure 1. Mortality by Maximum Rolling SIGNIPHY Score: Validation Data Set

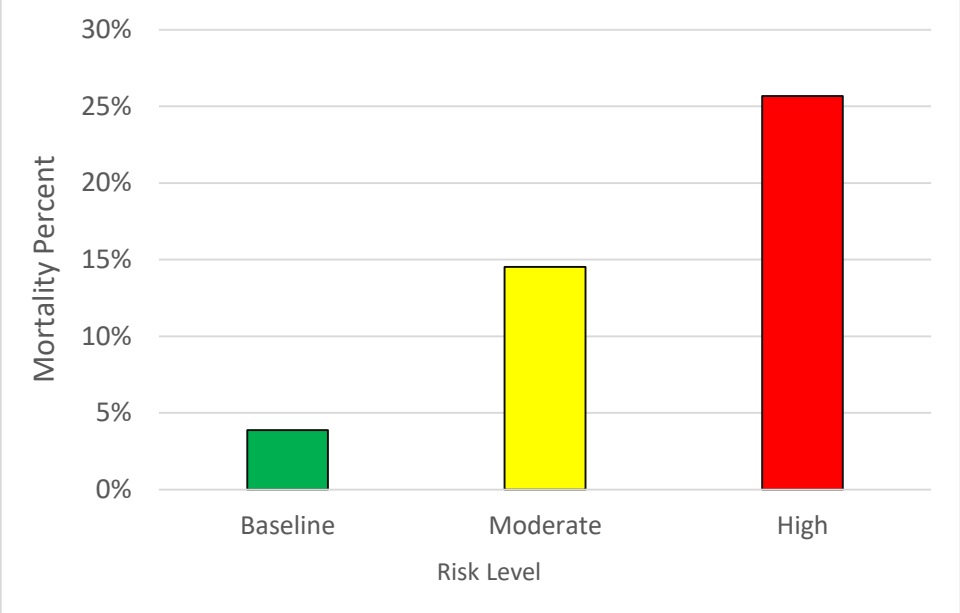


There is a clear relationship between a patient's score, with $r^2 = 0.9508$. From the graph it can be discerned that two inflection points exist where a large increase in the mortality rate occurs: 64 and 82, respectively. These inflection points became thresholds for sending alarms. If a patient's score rises to 64 then a "moderate risk" signal would be sent to the

patient's clinicians. However, if the score rises to 82, then a "high risk" alert would be sent to the patient's clinicians.

Figure 2 shows the results from using the alerts to signal an increased risk of mortality in the validation data set. Overall mortality in the validation data set was 7.6%. For patients with baseline risk, the mortality rate was 3.9%. When there was a moderate risk alert, mortality was increased 3.8-fold to 14.5%. At the highest risk level, patients' mortality was increased 6.6-fold to 25.7%.

Figure 2. Baseline Risk, Moderate Risk, and High Risk versus Mortality Before ICU Discharge



A high-risk alert level was achieved at least seven hours in advance of mortality in 90% of the patients. Thus there would have been ample time to administer remedial treatment. The average APACHE day one prediction of mortality showed significant overlap among the three risk groups (not shown here), indicating that many of the patients with moderate or high risk were not obviously in a downward trend.

4. Discussion

This paper describes a novel method for generating alerts using physiologic patterns and clinical data from patients admitted to adult ICUs: rSIGNIPHY. When applied to a large ICU database, rSIGNIPHY showed the ability to clearly distinguish patents at a 6.6-fold increased risk of mortality before unit discharge. Further, rSIGNIPHY can be used to produce an acuity score that is highly associated with ICU mortality. This

relationship was almost identical in a hold-out validation set of patients.

There are several reasons why rSIGNIPHY might be an important predictive analytic systems tool in the ICU. First, the inclusion of mechanical ventilation was important. Almost 1/3rd of adult patients admitted to ICUs in the U.S. are placed on mechanical ventilation within 24 hours post-admission[1]. Receiving active ventilation is an important indicator of an adverse outcome[15] as well as influencing a patient's respiratory rate. Even having been on a ventilator has attendant risks such as ventilator-associated pneumonia, which is why patients continued to receive a weight (albeit a lower one) after being weaned off a ventilator. Second, the system was calibrated using data from patients in adult ICUs exclusively. This patient population is more physiologically deranged than patients in other levels of care within a hospital, and analytic tools need to take this situation into account. Third, the use of medians to aggregate data rather than means decreases the bias from outliers in the data. Fourth, deleterious patterns in vital signs are important for rSIGNIPHY. Other mortality risk systems simplistically look for vital signs exceeding an outlier or utilize APACHE's method of selecting the highest or lowest value within 24 hours. Finally, rSIGNIPHY had a "memory", in that scores from previous periods had an impact on current values.

rSIGNIPHY belongs to a general class of metrics previously called early warning systems. But that nomenclature is confining for dynamic solutions such as rSIGNIPHY. Thus we elected to call rSIGNIPHY a predictive analytic solution. Such a system has to contain a continually updated acuity score, which would be beneficial in inter-shift handoffs and telemedicine. Acuity scores for each patient could be displayed simultaneously. This type of user interface allows intensivists to see all of their patients at a glance, aiding in identifying patients at the most urgent need of remedial care. Ideally, the alerts based on the acuity score would be color-coded and sent to clinicians' mobile devices. Earlier we stated that predictive analytics solutions need to have the following characteristics: obtaining all data electronically; providing scores and alerts in near real-time; identification of non-obvious patients, i.e. not clearly in sharp decline; and providing alerts hours before a patient's demise. rSIGNIPHY met all of these requirements.

While there are no existing predictive analytic solutions for critical care, a fairly sophisticated system has been established for intermediate care unit: the Visensia Index[10]. It shares most of the same date elements as rSIGNIPHY, namely streaming vital signs. Where rSIGNIPHY looks for patterns across time that are indicative of subsequent patient detriration,

the Visensia Index tries to identify instances where multidimensional outliers exist. However it has not been calibrated for critically ill patients, nor does it contain information on patients receiving mechanical ventilation.

There were limitations to this study. First, the data consisted of vital signs that had been confirmed retrospectively by a nurse. Data captured "live" will be messier. However, rSIGNIPHY's use of patients' median values to assign a "letter", which were based on ranges of values protects against the effects of erroneous and outlier data. But in order to successfully validate the methodology, a multi-center prospective clinical study is needed. Second, as described here, an rSIGNIPHY score is updated every two hours. With more high-frequency data a smaller time window might be sought. Presently a clinical study using rSIGNIPHY in live environments is being initiated. This study will capture data at one-minute intervals, and scores will be updated every five minutes. Third, while sensitivity and specificity were good, both can be improved. While using high-frequency data would aid in this, clinical interventions beyond mechanical ventilation might be informative. Fourth, although mortality is an important endpoint, other adverse events should be examined. Finally, the methodology for generating the weights for "words" in rSIGNIPHY is complex. However, once these weights are obtained then using these values in new data sets is straightforward.

5. Conclusions

A novel method was described for analyzing patterns in vital signs, augmented by a patient receiving mechanical ventilation. The method, called rSIGNIPHY, was applied to an adult ICU database. The results show that rSIGNIPHY can generate mortality risk alerts to intensivists that have high sensitivity and specificity. Future applications of rSIGNIPHY include other adverse outcomes in the ICU.

6. References

1. Zimmerman JE, Kramer AA, et al. 2006. "Acute Physiology and Chronic Health Evaluation (APACHE) IV: Hospital mortality assessment for today's critically ill patients", *Crit Care Med* ; 34(5):1297-1310.
2. Moreno RP, Metnitz PGH, Almeida E, et al. 2005. "SAPS 3: from evaluation of the patient to evaluation of the intensive care unit. Part 2: Development of a prognostic model for hospital mortality at ICU admission". *Intensive Care Med* ; 31:1345–1355.
3. Kramer AA. 2017 "A novel method using vital signs information for assistance in making a discharge a discharge decision from the intensive care unit: identification of those patients at highest risk of mortality on the floor or discharge to a hospice". *Med Res Archives* ; v. 5, n. 12, Dec. 2017.
4. Harrison DA, Parry GJ, Carpenter JR, et al. 2007 "A new risk prediction model for critical care: the Intensive Care National Audit & Research Centre (ICNARC) model" *Crit Care Med* ; 35:1091–1098.
5. Kramer AA, Sebat F, Lissauer M. 2019 "A review of early warning systems for prompt detection of patients at risk for clinical decline". *J Trauma Acute Care Surg* ; 87[1]: S67-S73.
6. Subbe CP, Kruger M, Rutherford P, Gemmel L. 2001 "Validation of a modified Early Warning Score in medical admissions", *Q J Med*. 2001; 94(10):521-526.
7. Royal College of Physicians. "National Early Warning Score (NEWS): Standardising the assessment of acute-illness severity in the NHS: Report of a working party", https://www.rcplondon.ac.uk/projects/outputs/national-early-warning-score. Published July 2012. Accessed October 15, 2018
8. Royal College of Physicians. "National Early Warning Score (NEWS) 2: Standardising the assessment of acute-illness severity in the NHS: Updated report of a working party", https://www.rcplondon.ac.uk/projects/outputs/national-early-warning-score-news-2 Published December 2017. Accessed October 15, 2018.
9. Hodgson LE, Congleton J, Venn R, Forni LG, Roderick PJ. 2018 "NEWS 2 – too little evidence to implement?", *Clin Med* ; 18[5]:371-373.
10. Hravnak M, Edwards L, Clontz A, et al. 2008 "Defining the Incidence of Cardiorespiratory Instability in Patients in Step-down Units Ysing an Electronic Integrated Monitoring System", *Arch Int Med*; 168[12]:1300-1308.
11. Hravnak M, DeVita MA, Clontz A, et al. 2011 "Cardiorespiratory instability before and after implementing an integrated monitoring system", *Crit Care Med* ; 39[1]:65-72.
12. Alarhayem AQ, Muir MT, Jenkins DJ, et a. 2019 "Application of electronic medical record-derived analytics in critical care: Rothman Index predicts mortality and readmissions in surgical intensive care patients", *J Trauma Acute Care Surg* ; 86[4]:635-641.
13. Kramer AA. 2018 "When Using Biomarkers in Alerts, Timing Is Everything", *Crit Care Med* ; 46[12]:2050-2051.
14. Mitchell M. 1997 "An Introduction to Genetic Algorithms", MIT Press, Cambridge, MA.

15. Zilberberg MD, Kramer AA, Higgins TL, Shorr AF. 2009 “Prolonged Acute Mechanical Ventilation”, *Chest* ; 135:1157-1162.

