A novel method using vital signs information for assistance in making 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

Andrew A. Kramer, PhD

Authors’ Affiliation:
1 Prescient Healthcare Consulting, Charlottesville, VA, USA.

Corresponding Address: E-mail address: akramer@prescient-healthcare.com

Disclosure: Prescient Healthcare Consulting has a patent submission on the technology described in this manuscript.

ABSTRACT

Intensivists face a difficult decision deciding when a patient should be discharged from the intensive care unit (ICU) to a unit of lesser acuity. Accurate algorithms don’t exist to aid in that decision. Therefore, we developed and validated a stratification tool using physiologic data obtained from a patient’s electronic medical record for assignment to risk categories of mortality or hospital discharge to hospice.

Population: 33,039 admissions at 41 adult ICUs using an electronic medical record from 1/1/2012-6/30/2017.

Outcome: Mortality on a unit of lesser acuity or discharge to a hospice after leaving the ICU ("mortality/hospice").

Methods: Vital signs from three hours to one hour before actual discharge were obtained. We used the 12 most proximate values for heart rate, respiratory rate, and mean arterial pressure, respectively. A letter was assigned to the median of every three measurements based on the underlying distribution of the vital sign. Four consecutive letters were concatenated to form a pattern, which were candidates for triggers (i.e. risk alerts). A patient could have three triggers, one for whether or not each vital sign contained a word that increased risk, along with a fourth trigger if a patient received mechanical ventilation. Using a genetic algorithm that weighted the outcome of mortality/hospice, we acquired a set of patterns that maximally increased risk. Those patterns were then validated as triggers for increased risk.

Results: The overall mortality/hospice rate was 4.7%. Fifteen patterns were identified that increased risk. Patients without triggers had a mortality/hospice rate of 3.2%, while patients with one, two, and three to four triggers had rates of 7.5%, 13.3%, and 27.5%, respectively

Conclusion: It’s possible to use vital signs proximate to when a discharge decision is made to identify patients with an increased risk of either mortality or discharge to hospice after leaving the ICU.

KEYWORDS: Intensive care unit, discharge decision, pattern recognition, mortality, hospice
1. INTRODUCTION

Clinicians have to make many decisions while a patient is in the intensive care unit (ICU). Arguably the most difficult decision is whether a patient is ready for discharge to a unit of lesser acuity\(^1\). If the patient’s discharge is unduly delayed, then a prolonged ICU stay with the attendant problems of cost\(^2\) and risk of infection\(^3\) is possible. Conversely, a patient discharged prematurely might be at high risk for mortality on the floor or readmission to the ICU.\(^4,5\)

Many factors go into the decision to discharge a patient from the ICU. Availability of a bed on the wards, existence of a step-down unit in the hospital, and current patient census in the ICU are examples of factors that do not involve a patient’s acuity of illness, and are thus, beyond the clinician’s control. Therefore, the clinician may add his/her own perception of “readiness” to ultimately guide the decision whether or not to discharge at a particular point.\(^6\)

Predictive models have been effectively used for retrospective benchmarking of ICUs for outcomes such as mortality,\(^7,8,9,10\) ICU length of stay,\(^11\) and duration of mechanical ventilation.\(^12\) Their success is most likely due to the strong influence that physiologic derangement and admission diagnosis have on a patient’s outcome. Predictive models such as APACHE IV\(^8\) and MPM\(_0\)-III\(^9\) have been explicitly stated by their developers to be used for comparing ICUs, not for driving care decisions at the patient level. One reason for this caution is the models’ inaccuracy at the patient level. Further, retrospective benchmarking models are heavily influenced by day 1 physiology and administrative data. As such, predictive models that impact discharge decisions cannot rely on these previously established models.

Unfortunately, an informative algorithm doesn’t exist to aid in making a discharge decision. Predicting readmission to the ICU using existing severity of illness scores has been unsuccessful,\(^13,14,15,16\) even though risk factors associated with that outcome have been well documented.\(^17,18,19\) Mortality in the hospital after discharge from the unit has received less study than readmission, and predictive tools have shown mixed results.\(^20,21,22\) The problem is that with a low incidence rate, mortality will have a low specificity, meaning a large number of false alarms in any decision-making algorithm.

The mortality rate by itself may be a bit misleading. It is well-recognized that a number of patients are discharged from the ICU for end-of-life care, many ending up in a hospice facility.\(^23\) An outcome that includes either mortality or subsequent discharge from the hospital to a hospice better reflects a target group of individuals for whom ICU discharge decisions might have an impact.

Intensivists could benefit greatly when making a decision about a patient’s ICU discharge by having a tool available that identifies patients at high risk for either subsequent mortality or being placed in hospice. In order to construct such a tool three things are required: 1) accuracy at the patient level; 2) timely identification of high risk patients; and 3) reliance on manually entered data. This means creating a new way of utilizing the copious amount of clinical data available from critically ill patients. Additionally, such a tool would be of more value if it assigned patients to risk groups rather than giving each patient a distinct prediction probability. Therefore, the objective of this paper is to describe the development and validation of a stratification tool that uses physiologic data obtained from a patient’s electronic medical record (EMR) to assign patients to risk categories for mortality or hospital discharge to hospice.
2. MATERIALS & METHODS

2.1. Data set in this study

Data used for this study were obtained from Medical Decision Network’s Phoenix ICU database (www.mdnllc.net). Phoenix collects data using HL7 feeds from a variety of sources within the ICU, assembles and cleans the data, and produces a data mart that can be used for ad hoc reporting and quality improvement initiatives. We used information on patients admitted to 41 adult ICUs at 20 hospitals in the Phoenix database during the time period 1/1/2012 – 6/30/2017. To be included in the study, an ICU had to be collecting vital signs information electronically throughout a patient’s stay in the unit. ICUs that manually recorded vital signs values were not considered, as the vital signs data are not timely.

The data encompassed demographics, vital signs, laboratory measurements, arterial blood gasses, Glasgow Coma Score (GCS), comorbidities, diagnosis, and clinical events such as a patient receiving mechanical ventilation. In this study we used all heart rate, respiratory rate, and mean arterial pressure measurements, whether the patient received mechanical ventilation (MV) after admission to the ICU, whether the patient died after discharge from the ICU to a unit of lesser acuity, and the discharge disposition from the hospital.

Data were split into development and validation data sets dependent upon date of ICU admission: patients admitted to the ICU prior to July 1, 2015 were used in the development data set, and those admitted from July 1, 2015 through June 30, 2017 made up the validation data set. The development data set was used to create and initially test the analytic tool, while the validation data set served to properly affirm the accuracy of this tool. Patients who were admitted post-operatively from coronary artery bypass graft surgery were excluded, as their outcomes were radically different from other patients. We also excluded patients who died during their stay in the ICU, were discharged from the ICU directly to another hospital or directly home. This was done as these patients would not be capable of having an event in the hospital post-ICU discharge.

2.2. Temporal data mining of vital signs

In order to assign patients to one of four stratification groups (low risk, average risk, high risk, very high risk) we developed a novel methodology for looking at patterns in vital signs. First, we gathered heart rate, respiratory rate, and mean arterial pressure measurements from the time period of three hours before discharge until one hour prior to discharge. This was done since discharge decisions are normally made at least one hour before actual discharge, and we wanted to replicate this in our methodology. Beginning with one hour before discharge, we took the preceding 12 measurements for each vital sign. The value furthest back was considered the first measurement, the next value the second measurement, etc… up to the 12th value; the 12th value is that closest to one hour before discharge. We then obtained the median value of each vital sign for each set of three consecutive measurements; this gave us four median values for each vital sign. The median was chosen over the mean because occasionally vital signs can have erroneous values recorded.

Next we obtained the 10th percentile, 25th percentile, 75th percentile, and 90th percentile for the first three data points’ median. These values constituted cut-points for demarcating the five bins of values. The five bins (i.e. ranges of values) were each allocated a letter: A, B, M, Y, or Z (Figure 1). Each person’s median values were assigned a letter. The letters from the four time periods (i.e. 12 data points) constituted a word. Thus, each patient had a word for heart rate, respiratory rate, and mean arterial pressure, respectively.
Figure 1. Assignment of letters and word formation

<table>
<thead>
<tr>
<th>LETTERS</th>
<th>CONSECUTIVE SET OF 4 LETTERS FORMS A WORD</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤ 10 pct</td>
<td>MYAB, AABA, MMMM, MZYA, etc...</td>
</tr>
<tr>
<td>11-25 pct</td>
<td></td>
</tr>
<tr>
<td>26-75 pct</td>
<td>ABBY, ZZZZ, MYBB, ZZZZ, etc...</td>
</tr>
<tr>
<td>76-90 pct</td>
<td></td>
</tr>
<tr>
<td>&gt; 90 pct</td>
<td>YZBA, AAMA, MYMY, BBBy, etc...</td>
</tr>
</tbody>
</table>

There was one more modification that needed to be made. A word has four letters, each having one of five values (A, B, M, Y, or Z). Thus, there are $5^4 = 625$ possible combinations. As letters do not have the same probability, some combinations (i.e. words) will appear infrequently. Thus, each word with a frequency < 100 was combined with other such words to form a generic word of “XXXX”, in other words a miscellaneous designation.

The aim was to collectively identify over all three vital signs which set of words corresponded to an increased risk of mortality on the floor or discharge from the hospital to a hospice: hereafter referred to as “mortality/hospice”. In order to accomplish that, the following process was carried out. Each word was initially given a random value of either zero or one. A value of one denotes a word that would prompt a “trigger” and zero meant no trigger would be sent. We also took into account whether or not the patient received mechanical ventilation as an additional trigger. So a patient could have {0, 1, 2, 3, or 4} triggers. The number of triggers for each patient was paired with their outcome of mortality/hospice. Each specific pair of number of triggers along with mortality/hospice was given a “score” such that false negatives were more heavily punished than false positives (Table 1). This score was summed over all patients.
Table 1. Scores given for combinations of number of triggers and outcome (mortality/hospice)

<table>
<thead>
<tr>
<th># of triggers</th>
<th>Outcome of mortality/hospice</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>NO</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>YES</td>
<td>-2</td>
</tr>
<tr>
<td>1</td>
<td>NO</td>
<td>-1</td>
</tr>
<tr>
<td>1</td>
<td>YES</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>NO</td>
<td>-1</td>
</tr>
<tr>
<td>2</td>
<td>YES</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>NO</td>
<td>-1</td>
</tr>
<tr>
<td>3</td>
<td>YES</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>NO</td>
<td>-1</td>
</tr>
<tr>
<td>4</td>
<td>YES</td>
<td>8</td>
</tr>
</tbody>
</table>

Using a genetic algorithm, the initial value of zero or one given to each word was either flipped or remained the same in order to maximize the sum of scores over all patients. (Genetic algorithms are a heavily utilized tool for optimizing a set of outcomes given a complex mix of predictors.) They do not make any assumption about the data’s underlying distribution, making them superior to linear models for many types of biomedical analysis.) Once the sum of scores was maximized, the words with a value equal to one were considered triggers. I.e. they increased the risk of mortality/hospice.

2.3. Outcome assessment and validation of the results

All of the above was carried out solely for patients in the development data set. The set of words ending up designated as triggers were then applied to data on patients in the validation data set. Every patient had from 0-4 triggers and an outcome (mortality/hospice = 1, otherwise = 0). Since very few patients had four triggers, we combined the categories of three and four triggers, respectively into a single category. A contingency table was constructed with four rows corresponding to 0, 1, 2, or 3-4 triggers, respectively and two columns for a patient’s outcome. Pearson’s Chi-square test was used to assess the statistical significance of the results, with a Mantel-Haenszel trend test applied to determine if there was a trend in the mortality rate corresponding to an increasing number of triggers.

3. RESULTS

A total of 59,432 admissions were available for analysis. A total of 33,039 admissions were at ICUs that collected vital signs electronically. From this cohort there were 19,046 admissions for the development data set and 13,993 admissions for the validation data set.

Table 2 presents some characteristics about this population. The mortality rate in the development data set (3.5%) was slightly higher than in the validation data set (2.9%). This held true for the percentage of patients discharged to a hospice: 3.1% in the development set, 1.8% in the validation data set. Thus, the combined outcome of mortality/hospice occurred in 6.6% of the development data set and 4.7% of the validation data set (p<0.001). However, patients in the validation data set were on average more severely ill. They had a higher mean acute physiology score, a greater percentage receiving mechanical ventilation, higher readmission rate, and a longer total hospital length of stay.
Table 2. Characteristics of patients included in the study’s analysis

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Development Data Set (n=22,679)</th>
<th>Validation Data Set (n=9,406)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mortality &gt; ICU discharge</td>
<td>3.5%</td>
<td>2.9%</td>
<td>0.002</td>
</tr>
<tr>
<td>Hospital discharge to hospice</td>
<td>3.2%</td>
<td>1.8%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Mortality or hospital discharge to hospice</td>
<td>6.6%</td>
<td>4.7%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Patient received mechanical ventilation</td>
<td>13.3%</td>
<td>15.9%</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Female gender</td>
<td>47.0%</td>
<td>44.4%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Patient admitted post-operatively</td>
<td>20.5%</td>
<td>19.3%</td>
<td>0.007</td>
</tr>
<tr>
<td>Readmission to the ICU</td>
<td>3.3%</td>
<td>6.1%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Age (mean ± std err)</td>
<td>64.69 ± 0.13</td>
<td>64.06 ± 0.14</td>
<td>0.001</td>
</tr>
<tr>
<td>Acute Physiology Score* (mean ± std err)</td>
<td>41.97 ± 0.16</td>
<td>44.77 ± 0.20</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>ICU length of stay (mean ± std err)</td>
<td>2.88 ± 0.03</td>
<td>2.95 ± 0.03</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>Hospital length of stay (mean ± std err)</td>
<td>8.62 ± 0.07</td>
<td>10.13 ± 0.11</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

The frequency distribution of heart rate, respiratory rate, and mean arterial pressure in the development data set are shown in Figures 2a, 2b, and 2c. Heart rate is slightly skewed to the right, while respiratory rate is highly skewed to the right and leptokurtic. Mean arterial pressure is symmetrically distributed and resembles a Gaussian distribution. The cut points for each vital sign corresponding to the 10th percentile, 25th percentile, 75th percentile, and 90th percentile, respectively are given in Table 3. These were used to assign a letter to each median value.

Table 3. Cut points for heart rate, respiratory rate, and mean arterial pressure, respectively. Based on the 10th, 25th, 75th, and 90th percentiles.

<table>
<thead>
<tr>
<th>Percentile</th>
<th>Heart Rate</th>
<th>Respiratory Rate</th>
<th>Mean Arterial Pressure</th>
</tr>
</thead>
<tbody>
<tr>
<td>10th</td>
<td>61</td>
<td>13</td>
<td>67</td>
</tr>
<tr>
<td>25th</td>
<td>69</td>
<td>16</td>
<td>75</td>
</tr>
<tr>
<td>75th</td>
<td>92</td>
<td>22</td>
<td>95</td>
</tr>
<tr>
<td>90th</td>
<td>104</td>
<td>27</td>
<td>105</td>
</tr>
</tbody>
</table>
Figures 2a, b, & c. Frequency distribution of median values for the first three: a) heart rate measurements; b) respiratory rate measurements; and c) mean arterial pressure measurements.
Figure 3. Rates of mortality or hospice after discharge from the intensive care unit to a unit of lesser acuity, stratifying by the number of triggers issued

Across the three vital signs, there were 127 words with a frequency > 100. Including the miscellaneous groups, there were 130 values to be optimized. To minimize false negatives we scored each patient who received no triggers but subsequently had an outcome a value of -2.0 (see Table 1). Therefore, theoretically assigning all patients to not having the outcome of mortality/hospice results in an objective function score = -2,530 (1,265 patients actually died or discharged to hospice * -2.0 points for each). This became the value under the null hypothesis. The genetic algorithm we used resulted in a maximum objective function = -1,775, demonstrably improving upon generating a prediction of no outcome for all patients. Various “seed” values were tried and all arrived at the same result, suggesting that the solution might be global rather than local.

A total of 15 words (11.5%) were classified as triggers: six heart rate words, five respiratory rate words, and four mean arterial pressure words. Figure 3 shows the incidence of mortality/hospice by the number of triggers in the development and validation data sets, respectively. In the development data set, patients having no triggers had an outcome in just 4.7% of admissions, below the overall rate of 6.6%. Patients with one, two, and three to four triggers had mortality/hospice rates of 10.8%, 19.6%, and 26.7%, respectively. The difference across the percentage of admissions with an outcome was highly significant (p< 0.001), as was the test for a trend in the mortality/hospice rates (p< 0.001). There was a similar pattern in the validation data set, albeit at lower rates as the overall mortality/hospice rate was 4.7%, less than the 6.6% in the development data set (p< 0.001). Patients with no triggers had outcomes in 3.2% of admissions. Patients with one, two, and three to four triggers had mortality/hospice rates of 7.5%, 13.3%, and 27.5%, respectively. Again, the difference in
mortality rates was highly significant (p<0.001), as was the test for a trend in the rates (p<0.001). If the mortality/hospice rate for zero triggers, 3.2% is considered the baseline category, the relative increase in risk was as follows: one trigger resulted in a 2.4-fold increased risk; two triggers resulted in a 4.2-fold increased risk; and three to four triggers resulted in an 8.7-fold increased risk. The sensitivity of having one or more triggers was 52.1%, and the specificity was 72.4%.

4. DISCUSSION

The results of this study show that patients for whom a discharge decision is pending can be successfully stratified to risk categories. The four categories here had mortality/hospice rates of 3.2%, 7.5%, 13.3%, and 27.5% in the validation data set. The group with three or four triggers had an 8.7-fold increased risk over admissions with no alerts. These results were similar to what was obtained from the development data set, even though those patients were less severely ill yet had a higher rate of mortality/hospice.

Variability in vital signs has been well studied, as has analyzing serial data in the ICU. However, successfully using vital signs to predict subsequent patient deterioration post-ICU discharge has proven difficult. There are reasons for vital signs analysis has been intractable. Many studies consider a single vital sign, usually heart rate. Those studies that do include multiple vital signs may lack a sufficient number of patients, or assume a Gaussian distribution for vital signs. Most studies use a variety of linear model methodology in generating their predictions. Finally, avoiding the use of summary measures such as the mean or median to represent all vital signs measurements as a single value is also important. The methods proposed in this study did not have these limitations, and introduced a novel technique for risk stratification.

How could this analytical tool be effectively used in the ICU? One possibility is to have the underlying electronic medical record system generate an alert when a patient has two or more triggers. This alert might be interpreted as a caution sign on discharging the patient within the next couple of hours. Given that the average ICU bed census in the U.S. is approximately 80%, keeping a high risk patient in the ICU a bit longer might not block the admission of another patient. Although this might result in a non-trivial increased cost for this patient, it would more than be offset by a reduction in the time spent post-discharge on the general floor.

The methodology described here is completely different from existing quantitative discharge readiness measures. It does not involve any manual data detection and can easily be imbedded within a hospital’s existing electronic medical records system. Although generating a signal is predicated upon summing the number of triggers, the underlying methodology for generating triggers is completely non-additive. Trends in the median value of three vital signs (as well as whether or not a patient received mechanical ventilation) were derived using a genetic algorithm, which is a technology that searches over a large set of weights to arrive at the set that optimally explains the outcomes across patients. Another distinguishing feature of this methodology is the inclusion of differential penalties for outcomes. While subjective in some respects, this allowed for false negatives (patients not having any triggers yet dying after ICU discharge) to be penalized more heavily than false positives (patient having > 1 trigger but alive at hospital discharge).

Our study did not have access to information on limitations of medical care such as “do not resuscitate” orders, which are important...
It’s possible that having a high number of triggers was merely a proxy for a very ill patient being transferred to the floor for end-of-life care. However, patients directly discharged from the ICU to post-acute care facilities, including hospices, were not included in the analyses. Also, patients in the highest risk group actually remained in the hospital slightly longer than patients in the lowest risk group (data not shown), refuting the notion that these patients were pushed to the floor for an expected death within days.

The data source used here was multi-institutional and had complete electronic data capture. Still, there were “only” 33,039 admissions available to us in this analysis. Therefore, out of approximately 1,875 possible word combinations across three vital signs, there were a sufficient number of admissions for generating only 127 distinct words; the rest were put into a miscellaneous category for each vital sign. A larger data set would increase the number of words used as predictors (i.e. lower the words in the miscellaneous group). This should result in a higher sensitivity for the triggers. If the database included more variables such as SaO2 and temperature, as well additional treatments other than mechanical ventilation, the results might be improved.

There are several other limitations to the research described here. First, the initial work that went into determining the cut-points for each level, assignment of letters and subsequent word formation, and determining which words were indicators of increased risk was quite involved. However, once the words designated as triggers are selected, the amount of additional work is relatively small. Second, we used a database retrospectively. For a realistic appraisal of the method proposed here its use in near-real time in a number of hospitals is warranted. Third, we preset the number of letters for each word to five. This needn’t be the case, especially for a clinical data point such as SaO2, which is non-symmetric in relation to risk. (High values are not problematic, only low values.) We have an ongoing research project that is discovering objective ways of selecting the number of letters and their corresponding cut-points. Finally, the choice of 12 data points does not represent an equivalent time period across patients. Our current research is also looking at fixing the number of time periods (e.g. 5 minutes, 10 minutes, etc…) in which to assign a letter.

5. CONCLUSION
We have demonstrated that evaluating a series of vital signs proximate to when a discharge decision might be made can accurately identify patients with an increased risk of either mortality or discharge to hospice after leaving the unit. The process described here resulted in patients with 3-4 triggers having an 8.7-fold increased risk compared to patients with no triggers. Data all came from electronic data feeds, so that the methodology can be adopted into a hospital’s existing electronic medical records system.

References


death and readmission after intensive care discharge”, Brit J Anaesth, 100[5]:656-662.


