Mortality rates among adult critical care patients with unusual or extreme values of vital signs and other physiological parameters: a retrospective study

Charles Harding¹, Marybeth Pompei², Dmitriy Burmistrov³, Francesco Pompei²

¹Harding Research, LLC, Seattle, WA, USA
²Exegen Corp., Watertown, MA, USA
³Independent Researcher, Waltham, MA, USA

Background: We evaluated relationships of vital signs and laboratory-tested physiological parameters with in-hospital mortality, focusing on values that are unusual or extreme even in critical care settings.

Methods: We retrospectively studied Philips Healthcare–MIT eICU data (207 U.S. hospitals, 2014–2015), including 166,959 adult-patient critical care admissions. Analyzing most-deranged (worst) value measured in the first admission day, we investigated vital signs (body temperature, heart rate, mean arterial pressure, and respiratory rate) as well as albumin, bilirubin, blood pH via arterial blood gas (ABG), blood urea nitrogen, creatinine, FiO₂, glucose, hematocrit, PaO₂, PaCO₂, ABG, sodium, 24-hour urine output, and white blood cell count (WBC).

Results: In-hospital mortality was ≥50% at extremes of low blood pH, low and high body temperature, low albumin, low glucose, and low heart rate. Near extremes of blood pH, temperature, glucose, heart rate, PaO₂, and WBC, relatively. Small changes in measured values correlated with several-fold mortality rate increases. However, high mortality rates and abrupt mortality increases were often hidden by the common practice of thresholding or binning physiological parameters. The best predictors of in-hospital mortality were blood pH, temperature, and FiO₂ (scaled Brier scores: 0.084, 0.063, and 0.049, respectively).

Conclusions: In-hospital mortality is high and sharply increasing at extremes of blood pH, body temperature, and other parameters. Common-practice thresholding obscures these associations. In practice, vital signs are sometimes treated more casually than laboratory-tested parameters. Yet, vitals are easier to obtain and we found they are often the best mortality predictors, supporting perspectives that vitals are undervalued.

Key Words: acidosis; body temperature; fever; hypothermia; physiological parameters; vital signs

INTRODUCTION

In scoring systems, patient risk of mortality is commonly evaluated by combining information on vital signs, other physiological parameters, medical histories, and other attributes [1-5]. However, it is also of interest to understand how values of individual measurements are...
associated with mortality rates, particularly when the value of a physiological parameter is unusual or extreme. For example, if a measurement shows very unusually low blood pH, body temperature, or albumin level, how is that quantitatively associated with mortality rates? Here, we sought to address this question for adult critical care patients and the outcome of in-hospital death, focusing on physiological values that are unusual or extreme even relative to the standards of the critical care setting.

Many studies evaluate mortality associations for common and moderately unusual values of physiological parameters, but few studies evaluate extreme values [6-8]. We used the large-scale Philips Healthcare–MIT eICU database of electronic health records (EHR) to provide the especially large numbers of patient records that are necessary to study extreme values.

Several studies report that vital signs are often measured and recorded with less care than other physiological parameters, including laboratory-tested physiological parameters [9-11]. Consequently, it has been argued that vital signs are undervalued in nursing practice [9,10,12]. In light of this research, we also sought to compare results between vital signs and other physiological parameters.

The objective of this study was to evaluate associations between major physiological parameters and in-hospital mortality among adult critical care patients, focusing especially on unusual physiological values and the comparison of vital signs with other physiological parameters.

MATERIALS AND METHODS

Setting
We retrospectively studied adult critical care stays from the Philips Healthcare–MIT eICU database (version 2.0), which includes critical care stays at 208 U.S. hospitals during 2014–2015. The dataset consists of EHR routinely collected in the Philips eICU program, a multihospital telehealth system providing remote support to bedside clinical teams in the critical care setting. Year 2014–2015 Philips eICU EHR were prepared for research purposes by the Philips eICU Research Institute and the MIT Lab for Computational Physiology [13-17].

Ethical Approval
As a retrospective, secondary study of deidentified data made available publicly through the eICU program, this research required no ethics committee approval or additional patient consent.

Participants
The eICU database includes 175,091 stays that were (1) by patients aged ≥18 years and (2) admissions, readmissions, or transfers, rather than stepdown stays or other stay types. We excluded stays lacking discharge survival status (n=1,597, 0.9%) and/or collected physiological parameters for Acute Physiology Score III (APS-III) (n=6,535, 3.7%) [1,17]. At one of the 208 eICU hospitals, no stays met the study selection criteria, so this hospital was excluded. Altogether, application of study selection criteria left 166,959 stays by 132,513 patients at 207 hospitals for analysis.

Data Collection
All objective physiological parameters from APS-III were analyzed, including the four traditional vital signs (blood pressure, body temperature, heart rate, and respiratory rate) and 12 other physiological parameters [1]. For each physiological parameter, we analyzed the most deranged value measured during the first 24 hours of admission. Following APS-III, most deranged was defined as furthest from reference-range midpoints of albumin, 13.5 g/dl (de facto lowest considered worst); bilirubin, 0 mg/dl (highest considered worst); blood pH via arterial blood gas (ABG), 7.4; blood urea nitrogen, 0 mg/dL (highest considered worst); body temperature, 38 °C (100.4 °F); creatinine, 1.0 mg/dl; FiO2, ABG, 21%; glucose, 130 mg/dl; heart rate, 75 beats/min; hematocrit, 45.5%; mean arterial pressure (average of systolic and diastolic pressure), 90 mm Hg; PaO2 ABG, 80 mm Hg; PaCO2 ABG, 40 mm Hg; respiratory rate, 19
breaths/min; sodium, 145 mEq/L; 24-hour urine output, 10,00 mL (stays <24 hour excluded); and white blood cell count (WBC), 11.5 1000/μl.

Data Availability
The PhysioNet program (MIT Laboratory for Computational Physiology, MIT) makes all datasets analyzed in the current study publicly available subject to research use agreements and training requirements. The data sharing policy for this data source can be found at https://eicu-crd.mit.edu/ (date accessed: 2023-01-28). Code used in this study is available at https://doi.org/10.6084/m9.figshare.25596504.

Data Analysis
Splines were used to allow for nonlinear relationships [18,19]. Penalized cubic splines were fit in univariate logistic regressions using the R mgcv package, modeling the response variable as quasibinomial to account for overdispersion [20]. Confidence bands are 95% (pointwise) and account for uncertainty in spline degrees of freedom. In addition to analyses that used splines, a binned (i.e., categorized) analysis was performed to demonstrate the consequences of binning for risk prediction.

Binning is standard practice in risk prediction systems that use physiological parameters. In the binned analysis, the values of physiological parameters were binned by quintiles and the mortality rate was evaluated in each bin.

For any analysis of big data, there is a question of what data range to plot since rare typos and mis-entries lead to a wide set of values beyond those that are relevant or clinically verifiable. Our aim was to plot the central range of the data where observations were sufficiently dense to support statistically informative estimates, and we selected this range by applying the criterion that widths of confidence bands from the main (continuous) analysis were required to be ≤25 percentage points.

Predictive performance was evaluated from spline estimates using scaled Brier scores, which are Brier scores rescaled to account for differences in outcome prevalence. These resemble Pearson’s $R^2$ [21]. Higher values indicate better performance.

RESULTS

Overall Associations with In–hospital Morality
Figure 1 shows the associations between 16 major physiological parameters and in-hospital mortality among adult patients during 166,959 critical care visits at 207 U.S. hospitals. When analyzing actual physiological values (continuous analysis), the relationships were commonly nonlinear and often bathtub shaped—mortality rates increased at both very low and very high values of physiological parameters. For several parameters, unusual values were associated in-hospital mortality rates that exceeded 50%, including at extremes of blood pH, body temperature, glucose, heart rate, and albumin level.

In addition to analyzing values of physiological parameters as measured, we analyzed values binned (categorized) by quintiles (Figure 1). This was done to illustrate consequences of binning, which is often applied to physiological parameters. After binning, the bathtub-shaped relationships with mortality were no longer evident for several physiological parameters, including heart rate and hematocrit. Moreover, after binning, in-hospital mortality rates did not reach 50% for any physiological parameter. Binning reduced the maximum observed mortality rate by more than half for several physiological parameters, including blood pH, body temperature, glucose, heart rate, hematocrit, and WBC.

Abruptness of Changes in In–hospital Morality
For several physiological parameters, relatively small changes in measured values were associated with abrupt, several-fold increases in in-hospital mortality rates (Figure 1). These abrupt increases were most common at extreme physiological values. For example, reductions in glucose level from 50 to 10 mg/dl were associated with a 2.9-fold increase in mortality (21% [95% CI, 20%–22%] to 60% [95% CI, 56%–66%]), reductions in heart rate from 40 to 25 beats/min were associated with a 4.0-fold increase in mortality rate (12% [95% CI, 11%–13%] to 47% [95% CI, 44%–51%]), and reductions in WBC from 4,000 to 1,000/μl were associated with a 3.7-fold increase in mortality rate (9% [95% CI, 8%–9%] to 33% [95%, 31%–6%]). When binning was applied to the physiological parameters, all of these sharp changes in mortality rates were hidden.

Predictiveness for In–hospital Mortality
Of the physiological parameters, blood pH, body temperature, and FiO₂ were most predictive of in-hospital death (scaled Brier scores: 0.084, 0.063, and 0.049, respectively) (Table 1). However, predictive performance fell after binning (median performance loss, 36%) (Table 1). Binning caused the largest reduction of predictive performance for body temperature (performance loss, 68%) and the smallest reduction for urine output (performance loss, 4%).

Comparison of Vital Signs with Other Physiological Parameters

On average, predictiveness for in-hospital mortality was better for traditional vital signs (body temperature, blood pressure, heart rate, and respiratory rate) than for other physiological parameters (mean of scaled Brier scores: 0.041 vs. 0.033). However, binning caused larger average reductions in predictive performance for vital signs than for other physiological parameters, with the result that binned vital signs had less overall predictive ability for in-hospital mortality than did other physiological parameters (mean of scaled Brier scores after binning: 0.021 vs. 0.023).

As expected, measurement availability in the first 24 hours of critical care admission was invariably better for vital signs than for other physiological parameters (Table 1). Each vital sign was measured and recorded in EHR for at least 94% of patients. (It is probable that vital signs were taken for more patients, but the measurements may not have reached EHR.) Of the physiological parameters that are not vital signs, albumin level, urine output, bilirubin level, blood pH, FiO₂, PaCO₂, and PaO₂ were each available for less than 40% of patients. Considering the three physiological parameters that were most predictive of in-hospital mortality, availability was 94% for body temperature but only 23% for blood pH and FiO₂.

DISCUSSION

In this study, we observed that in-hospital mortality rates were
### Table 1. Physiological parameters measured in the first 24 hours of critical care admission: availability of measurements and predictive performance for in-hospital mortality

<table>
<thead>
<tr>
<th>Physiological parameter</th>
<th>Availability (%)</th>
<th>Predictive performance</th>
<th>Performance loss by binning (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Continuous value</td>
<td>Binned value</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Albumin (g/dl)</td>
<td>39.6</td>
<td>0.048</td>
<td>0.038</td>
</tr>
<tr>
<td>Bilirubin (mg/dl)</td>
<td>35.9</td>
<td>0.029</td>
<td>0.022</td>
</tr>
<tr>
<td>Blood pH</td>
<td>23.2</td>
<td>0.084</td>
<td>0.044</td>
</tr>
<tr>
<td>Blood urea nitrogen (mg/dl)</td>
<td>79.1</td>
<td>0.041</td>
<td>0.039</td>
</tr>
<tr>
<td>Body temperature (°C)</td>
<td>94.0</td>
<td>0.063</td>
<td>0.020</td>
</tr>
<tr>
<td>Creatinine (mg/dl)</td>
<td>79.4</td>
<td>0.043</td>
<td>0.037</td>
</tr>
<tr>
<td>FiO₂ (%)</td>
<td>23.2</td>
<td>0.049</td>
<td>0.044</td>
</tr>
<tr>
<td>Glucose (mg/dl)</td>
<td>87.7</td>
<td>0.019</td>
<td>0.007</td>
</tr>
<tr>
<td>Heart rate (beats/min)</td>
<td>99.7</td>
<td>0.039</td>
<td>0.021</td>
</tr>
<tr>
<td>Hematocrit (%)</td>
<td>77.7</td>
<td>0.011</td>
<td>0.006</td>
</tr>
<tr>
<td>Mean arterial pressure (mm Hg)</td>
<td>99.5</td>
<td>0.048</td>
<td>0.031</td>
</tr>
<tr>
<td>PaCO₂ (mm Hg)</td>
<td>23.2</td>
<td>0.017</td>
<td>0.012</td>
</tr>
<tr>
<td>PaO₂ (mm Hg)</td>
<td>23.2</td>
<td>0.012</td>
<td>0.007</td>
</tr>
<tr>
<td>Respiratory rate (breaths/min)</td>
<td>99.2</td>
<td>0.014</td>
<td>0.011</td>
</tr>
<tr>
<td>Sodium (mEq/L)</td>
<td>79.8</td>
<td>0.018</td>
<td>0.007</td>
</tr>
<tr>
<td>Urine output (mL/day)</td>
<td>36.9</td>
<td>0.019</td>
<td>0.018</td>
</tr>
<tr>
<td>White blood cell count (1,000/µl)</td>
<td>75.6</td>
<td>0.039</td>
<td>0.020</td>
</tr>
</tbody>
</table>

a) The proportion of stays for which the physiological parameter was recorded in the first 24 hours of critical care admission; b) Predictive performance for in-hospital mortality is evaluated in terms of scaled Brier scores. Higher values indicate better predictive performance; c) In the analysis of continuous values, the actual value of the physiological parameter is used for risk prediction; d) In the analysis of binned values, the quintile-binned value of the physiological parameter is used for risk prediction instead of the actual value, mirroring common binning approaches in risk scoring systems. In both cases, performance is evaluated for the most deranged value of the physiological parameter measured in the first 24 hours of critical care admission; e) Performance loss is calculated as the percentage reduction in the scaled Brier score after binning.

especially elevated for adult critical care patients with very low blood pH, very low or high body temperature, very low glucose level, very low heart rate, or very low albumin level in the first 24 hours of admission to critical care. In-hospital mortality reached at least 50% in the extreme ranges of these physiological parameters. We also observed that, in some ranges of physiological parameters, relatively small changes in the physiological parameter value were associated with abrupt increases in in-hospital mortality rates. For example, tripling or quadrupling of in-hospital mortality rates was associated with reductions in glucose level from 50 to 10 mg/dl, heart rate from 40 to 25 beats/min, and WBC from 4,000 to 1,000/µl. Further, vital signs were on average better predictors of in-hospital mortality than were other physiological parameters, and vital signs had consistently better availability.

The novelty of these findings is their extension of previous research to unusual and extreme values of physiological parameters. Most studies are not large enough to analyze such unusual values. Others miss associations with high mortality because they apply binning (i.e., categorization) procedures that mix extreme and more typical physiological parameters values together, diluting effects of the extreme values. Studies of patient risk scoring regularly bin, categorize, dichotomize, threshold, cutoff, or linearize the physiological parameters [1-5,22]. For example, Acute Physiology and Chronic Health Evaluation (APACHE) IV, Simplified Acute Physiology Score (SAPS) 3, and Mortality Probability Model 0-III (MPM0-III) all bin physiological values [1-4]. Our result demonstrate that binning obscures associations with high mortality (Figure 1), conceals bathtub-shaped (U-shaped) associations with mortality (Figure 1), hides abrupt mortality-rate increases that accompany small changes in some physiological parameter values (Figure 1), and greatly reduces the predictiveness for in-hospital mortality (Table 1). Further, binning produced larger average reductions in predictive performance for vital signs than for other physiological parameters. These findings support the view that it is best to avoid binning in mortality risk prediction and early warning score systems [19], especially for vital signs.
Despite the limitations of binning, sophisticated risk scoring systems can still provide well-calibrated estimates of mortality risk because patients with an extreme value of one physiological parameter usually also have other mortality risk factors that are accounted for in the scoring systems. Nonetheless, we think there is value in seeing how high mortality rates climb at extremes of individual physiological parameters—for example, how in-hospital mortality rates more than tripled from 39 °C to 41.5 °C body temperature (15% vs. 50% mortality)—because these physiological parameter changes are observed in clinical practice before evaluation in risk scoring systems.

Binned values are also common in research studies, textbook introductions, and everyday practice. Perhaps because binning is ubiquitous, nurses, physicians, and researchers sometimes misinterpret results reported for binned physiological values as if they applied to all values of the underlying parameter that were binned together. For example, based on binned research, it is sometimes stated that body temperatures above thresholds like >38 °C and 40 °C are associated with no increase in mortality rates or with only moderate increases [12,23]. Yet, by performing unbinned analyses, we found that about half of critical care patients died at the highest ranges of body temperature (41.5 °C) (Figure 1). This is four times greater than the mortality rate suggested by binned analysis of the same data (Figure 1), demonstrating how binning can produce major errors in risk assessment.

Several articles conclude that vital signs are undervalued in clinical practice, especially in comparison with laboratory-tested physiological parameters [9,10,12]. Our results support the value of vital signs. For example, vital signs had better average predictiveness for in-hospital mortality than did other physiological parameters, even though measurements of vital signs cost less and are simpler to perform. However, binning reduced the average predictive ability of vital signs below other physiological parameters, which suggests that the ubiquity of binning may be one reason that vital signs are undervalued.

Prior studies also found that vital signs were taken and recorded with less care and accuracy than other physiological parameters, especially laboratory-tested parameters [9,11]. This suggests the actual predictiveness of vital signs might increase if they were measured more carefully. Better nursing practices and protocols may be able to improve vital sign measurement quality [10,11,24]. Of the four vital signs, only respiratory rate was not strongly predictive of in-hospital mortality in our study. Earlier research found that well-measured respiratory rate strongly indicates patient risk, but that frequent measurement errors reduce its predictiveness in everyday practice [25-27], which may explain our study findings.

In our study, the most abrupt increases in mortality occurred at very low glucose levels, heart rates, and WBCs. These sudden changes emphasize the need for accurate and precise measurements in these parameter ranges. Separate studies of individual physiological parameters would be needed to characterize specific causes of mortality patterns observed here.

This study has several limitations. One limitation is that this study only examined associations between physiological parameters and in-hospital mortality. In-hospital mortality is also influenced by many other factors, such as the underlying disease or injury that led to admission, preexisting conditions, and treatments administered. If these other factors are not considered and the patient’s mortality risk is evaluated solely on the basis of physiological parameter values, then substantial bias or error may occur. Another limitation is that other mis-entries of values and/or units, which appear in any large database of hand-entered data, owing to human fallibility. Typos and mis-entries may be especially common among records with extreme values. However, we see no means by which these errors could produce artificially high mortality rates—the errors consist of mis-entering extreme values for patients whose physiological status is actually more normal or typical, and who therefore do not have rates of death that are as heightened. Therefore, the elevated mortality rates that we report must be due to patients who actually have extreme physiological parameter values. Indeed, the true mortality rate at physiological parameter extremes is only expected to be diluted by the presence of mis-entries, and the actual mortality elevations may therefore be even higher than we have reported. Erroneous values also present a limitation to our analyses of predictive performance because they may misrepresent the patient’s condition and thereby degrade predictive performance.

Other limitations include test misclassification, which can occur in large databases for tests with similar names, and the possibility of system and database rules that reject the inclusion of unusual values for some measurements. Given the large number of hospitals contributing to the data (n=207), we are not able to exclude effects of rules like these, which can be hospital-specific. In addition, our analysis was limited to the most deranged value of each physiological parameter in the first 24 hours of admission, but this single value may provide an incomplete picture of the patient’s condition. In future research, it may be worth extending analyses to include tem-
poral trends of physiological parameters. Another limitation is that the definition of the most deranged measurement treats the midpoint of APS scoring categories as the central value [1]. While we followed this definition for consistency with eICU practice, it sometimes resulted in non-typical values being selected as the midpoint (e.g., 38.0 °C instead of the typical body temperature of about 36.8 °C), which may somewhat underestimate parameters’ associations with mortality. Lastly, a limitation is that assessments of predictive performance can differ by scoring rule choice [21].

In conclusion, in-hospital mortality rates are greatly elevated for critically ill adults who have extreme values of some physiological parameters, for example reaching at least 50% for patients with very low blood pH, very low or high body temperature, very low glucose level, or very low heart rate measured in the first 24 hours of critical care admission. These relationships are obscured by thresholding and binning of physiological parameters. In clinical practice, vital signs are sometimes treated more casually than laboratory-tested parameters. Yet, our results show vital signs often have better predictiveness for mortality and universally have better availability, supporting opinions that vital signs are currently undervalued.

CONFLICT OF INTEREST

Charles Harding received consulting fees from Exergen, including related to this report. Marybeth Pompei is Senior Vice President of Exergen and has thermometry patents. Dmitriy Burmistrov reports no conflicts of interest. Francesco Pompei is CEO of Exergen and has thermometry patents. Exergen, Corp. is a manufacturer of thermometers. No other potential conflicts of interest relevant to this article were reported.

FUNDING

This work was supported by Exergen Corp. (Watertown, MA, USA), which played roles in study design; collection, analysis, and interpretation of data; writing of the manuscript; and the decision to submit for publication.

ACKNOWLEDGMENTS

None.

REFERENCES