Research and Applications

Discharge recommendation based on a novel technique of homeostatic analysis

Jacob S Calvert,1 Daniel A Price,1 Christopher W Barton,2 Uli K Chettipally,2,3 and Ritankar Das1

1 Dascena Inc., Hayward, CA, USA, 2 Department of Emergency Medicine, University of California San Francisco, San Francisco, CA, USA, and 3 Kaiser Permanente South San Francisco Medical Center, South San Francisco, CA, USA

Correspondence to Ritankar Das, 1135 Martin Luther King Dr, Hayward, CA 94541, USA; ritankar@dascena.com; Tel: 872-228-5332

Received 6 September 2015; Revised 30 December 2015; Accepted 21 January 2016

ABSTRACT

Objective: We propose a computational framework for integrating diverse patient measurements into an aggregate health score and applying it to patient stability prediction.

Materials and Methods: We mapped retrospective patient data from the Multiparameter Intelligent Monitoring in Intensive Care (MIMIC) II clinical database into a discrete multidimensional space, which was searched for measurement combinations and trends relevant to patient outcomes of interest. Patient trajectories through this space were then used to make outcome predictions. As a case study, we built AutoTriage, a patient stability prediction tool to be used for discharge recommendation.

Results: AutoTriage correctly identified 3 times as many stabilizing patients as existing tools and achieved an accuracy of 92.9% (95% CI: 91.6–93.9%), while maintaining 94.5% specificity. Analysis of AutoTriage parameters revealed that interdependencies between risk factors comprised the majority of each patient stability score.

Discussion: AutoTriage demonstrated an improvement in the sensitivity of existing stability prediction tools, while considering patient safety upon discharge. The relative contributions of risk factors indicated that time-series trends and measurement interdependencies are most important to stability prediction.

Conclusion: Our results motivate the application of multidimensional analysis to other clinical problems and highlight the importance of risk factor trends and interdependencies in outcome prediction.

Key words: patient discharge, medical informatics, length of stay, computer-assisted diagnosis, clinical decision support systems

BACKGROUND AND SIGNIFICANCE

The increased use of electronic health records in clinical practice has provided an opportunity to translate knowledge into practice at the point of care. Clinical decision support (CDS) systems provide patient-specific information, presented to the care provider at appropriate times during the clinical workflow, to help with decisions that enhance the health of the patient. Based on the clinical context, CDS systems take on a variety of forms, including alerts, reminders, and risk scores.1

Risk-scoring systems quantify a patient’s degree of homeostatic stability and augment the clinician’s ability to determine the appropriate care setting for treatment. However, the most common CDS scoring systems can fail to identify nearly 60% of unstable patients.2 As a result, many unstable patients are not transferred to appropriate care until after they have experienced a life-threatening homeostatic failure like anaphylaxis, heart failure, or kidney failure. Moreover, patients are sometimes misidentified as stable and are transferred to less acute care or are discharged and subsequently...
undergo homeostatic failure. Collectively, these mistakes are referred to as patient transfer and discharge (TD) errors.

Each year, TD errors cost the US health care system billions of dollars and are partially responsible for at least 46,000 deaths. Existing CDS scoring systems are weakly tuned to the underlying homeostatic physiologies of patients, which limits their ability to support TD decisions. In particular, these systems assume risk factors are independent of one another. Moreover, they are unable to account for chronic conditions, variations in physiology among patients, and trends in patient information. Here, we present a general methodology for homeostatic analysis that addresses the shortcomings of existing CDS scoring systems by considering interactions between risk factors as well as their changes over time. We demonstrate this technique with AutoTriage, a patient stability prediction tool. We compare AutoTriage results to the Medical Emergency Team (MET) calling criteria, the most prevalent tool for identifying at-risk patients, as well as the Modified Early Warning Score (MEWS 4), a state-of-the-art CDS scoring system in clinical use.

**METHODS**

**Preparation of datasets for training and testing**

We used a dataset of 4000 patient records from the MIMIC II database. This subset consisted of anonymized clinical documentation of adult patients admitted to an ICU at the Beth Israel Deaconess Medical Center (BIDMC) with a variety of chief complaints. The Institutional Review Boards of BIDMC and the Massachusetts Institute of Technology waived the requirement for individual patient consent, as the study did not impact clinical care and all data were de-identified.

Inclusion criteria for this study were:

1. Adult (ie., age ≥18 years) admitted to the cardiac, medical, surgical, or trauma Intensive Care Unit (ICU)
2. Documented length of stay and survival for at least 48 hours following admission

The first 48 hours of each patient’s stay were available for analysis, and we utilized both patient demographic information and dynamic physiological measurements with 1-hour resolution. In particular, we used age, blood oxygen saturation, Glasgow Coma Score, heart rate, respiration rate, systolic blood pressure, temperature, and white blood cell count. These measurements were chosen due to the practical considerations of implementing a discharge recommendation tool in a clinical setting and to facilitate comparison of our results with MET and MEWS, which also rely on vital signs. In particular, these measurements were chosen because they are frequently sampled and widely available, ensuring that our predictions are current and less susceptible to missing data.

**Gold standard**

The length of stay in the ICU and the number of days survived following ICU admission were known for each patient and were used to define the gold standard for patient stability. In particular, patients were defined to be stable if they (1) were discharged within the 24 hours following our last observation period, and (2) lived for at least 30 days after discharge (Table 1). A random selection of half the patients formed the training set; the remaining patients were reserved for testing. The partitioning was done with a built-in MATLAB (MathWorks, Natick, MA, R2014a) function, which randomized the patients being placed in each group based on their anonymized medical record number (AMRN). In MIMIC-II, AMRNs were assigned in a randomized fashion to meet the Health Insurance Portability and Accountability Act (HIPAA) privacy rule in de-identified clinical data.

**Identifying and weighting groups of risk factors**

To discretize the possible ranges of each biological measurement of interest, the measurement ranges were first partitioned into a finite number of bins. In this study, we partitioned the 6 vital signs according to a custom heuristic table that incorporates recent findings in the scientific literature and is similar in structure to the MEWS partitioning table. For measurements not covered by existing heuristic tables (eg, white blood cell count), partitions were set to make the number of patients in each bin roughly equal, which is sometimes referred to as statistical binning. Patient trajectories (time-parameterized evolution or coevolution of physiological parameters) were then constructed from the progression of the binned measurements and inserted into a finite discrete hyper-dimensional (FDHD) space (Figure 1). Further, to analyze trends in patient physiology over time, we computed and binned the trends in measurements hour by hour. These differences were treated in the same way as other patient observations. For example, just as a patient would have a heart rate (HR) observation every hour, they would also have a ΔHR (change in trend in HR) observation every hour (difference in HR between the current hour and the previous one).

We then used custom MATLAB scripts to search each subset of the FDHD space for groupings of risk factors that have significant correlation with patient stability. We searched all individual measurements and groupings of 2 measurements (denoted \( S_1 \) and \( S_2 \)) and the equivalent for trends (denoted \( D_1 \) and \( D_2 \)) for the combinations that were most significantly correlated with patient stability (see Figure 1 for examples of \( S_2 \) and \( D_2 \) analyses). Specifically, we searched through the time-parameterized evolution of 8 S1 measurements (the 8 physiological values used in this study) and 28 S2 measurements (all possible pairs of physiological values). \( D_1 \) and \( D_2 \) also had 8 and 28 measurements, respectively, because there is 1 trend measure (eg, ΔHR) for each physiological value (eg, HR). Each measurement of \( S_1 \), \( S_2 \), \( D_1 \), or \( D_2 \) type was then assigned a weight between 0 and 1 that indicated its relative impact on patient stability. The weights of the 16 individual measurements and the 56 pairs of measurements were fixed after this step.

**Assigning AutoTriage scores for patient stability**

After weighting individual measurements by their correlation with patient stability, S-types and D-types (categories) were scaled and combined according to Eq. 1.
Stability score

\[ \text{Stability score} = \sum_{i \in G} \alpha_i P_i \]  

MEWS, where normalized values of each vital sign are summed, motivated Eq. 1. The normalized sum-based structure is extended in Eq. 1 to include interdependencies and trends among measurements.

The stability score is the sum of trajectory groupings, weighted by coefficient \( \alpha_i \). \( P_i \) denotes the correlation between measurement grouping \( i \) and patient stability. The index, \( i \), takes on values from the set of possible groupings \( G = \{S_1, S_2, D_1, D_2\} \). Note that while Eq. 1 is a linear combination, the \( P_i \) values already contain information on higher-dimensional correlations between risk factors. The \( \alpha_i \) was chosen by an interior-point optimization algorithm to maximize the partial area under the training set receiver operator characteristic curve, corresponding to specificities above 90\%, which are within the clinical operating range for discharge specificity.\(^4,12\) This optimization criterion was selected to ensure that AutoTriage would achieve high sensitivity (to facilitate early discharge for as many stable patients as possible) with a low false positive rate (to ensure that patients who are not stable are not erroneously discharged). In order to better ensure that the optimized solution was a global maximum, the optimization was performed for 1000 random initializations of coefficients, with values ranging from \(-0.5 \) to \( 0.5 \). Each coefficient was constrained to lie within \(-1 \) to \( 1 \).

**RESULTS**

**Prediction performance comparison**

When applied to the test set, AutoTriage assigned stability scores ranging from 0.0020 to 0.1533, with an average score of 0.1282. The discharge prediction comparison is shown in Figure 2, yielding area under the ROC curve (AUC) values of 0.81, 0.68, and 0.68 for AutoTriage, MET, and MEWS, respectively. To illustrate AutoTriage performance within the range of acceptable erroneous discharge rates, a stability score cutoff of 0.1312 identified 57.7\% (95\% CI: 55.4–59.8\%) of all stable patients while maintaining an erroneous discharge rate of \( \approx 5.5\% \) (95\% CI: 4.5–6.6\%). AutoTriage identified roughly 3 times the number of stable patients recognized by MET (18.5\%) and MEWS (14.1\%) at the same discharge error rate. K-fold cross-
Table 2. Confusion matrix for patient stability analysis

<table>
<thead>
<tr>
<th></th>
<th>Y</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>53</td>
<td>104</td>
</tr>
<tr>
<td>N</td>
<td>39</td>
<td>1804</td>
</tr>
</tbody>
</table>

Validation (with K = 10) yielded a similar improvement in performance over MET and MEWS, with AutoTriage identifying 55.2% of stable patients while maintaining an erroneous discharge rate of 93.4%. This corresponded to an average AUC of 0.75, as well as average positive and negative predictive values of 29.7% and 96.7%, respectively. Notably, the AUC performance metric was robust to changes in the time binning interval; choices of 30-, 90-, or 120-minute intervals yielded AUCs similar to that of our default 60-minute bin.

Table 2 lists the correctly and incorrectly identified patients from the test set, using the stability score cutoff described above. Yindicates the number of patients predicted to be discharged by AutoTriage, while Ydenotes the set of patients satisfying the gold standard. Thus, the top-left table entry (Y,Y), for example, lists the number of true positives. At a score cutoff of 0.1312, AutoTriage accurately identifies 92.9% (95% CI: 91.6–93.9%) of test set patients.

Contributions of individual trajectory groupings to patient score

In order to determine the most important measurement types (categories) for stability prediction, we analyzed the distribution of static (S) and trend (D) type score contributions on a patient-by-patient basis. Averaging the fraction of each patient score from S-type and D-type over all patients revealed that, for AutoTriage, patient stability scores were mostly determined by higher-dimensional and time-series trend measurement types, as opposed to MET and MEWS, which are S1-type (1-dimensional and static) analyses (Figure 3). In particular, S1 accounted for 18.8% (95% CI: 18.6–19.0%), D1 for 19.1% (95% CI: 18.9–19.3%), D2 for 28.3% (95% CI: 28.2–28.5%), and S2 for 33.8% (95% CI: 33.6–34.0%) of each patient’s score.

DISCUSSION

The homeostatic analysis framework described here has expanded functionality as compared to existing CDS systems. Diverse patient information sources are made interoperable in FDHD space and used to explore risk factor interdependencies. Moreover, while existing risk scoring systems focus on predicting mortality, our framework can be used to predict any binary clinical outcome. This was demonstrated with the creation of AutoTriage, a patient stability prediction tool that can quantify imminent patient discharge 24 hours in advance.

Making sound discharge recommendations requires high specificity, to protect against the erroneous discharge of an unstable patient. Because typical 30-day unplanned rehospitalization rates tend to range from 10% to 20%, we aimed to keep the AutoTriage erroneous discharge rate below 10%.\(^4,^{12}\) However, while maintaining patient safety, improving sensitivity for discharge recommendations could greatly reduce costs from unnecessary testing and extended length of stay in the ICU and reduce the risk of contracting hospital-acquired infections. Further, early discharge can accelerate patient recovery in some cases, because rest at home is unhindered by the cacophony of medical alarm noise ever present in the ICU.\(^13\) The concern for patient safety and the benefit of early discharge were both incorporated during stability score training, where measurement types were weighted to improve sensitivity while maintaining high specificity. The resulting receiver operator characteristic curve demonstrated that AutoTriage recommendations could ensure patient safety while doubling the number of patients correctly recommended for early discharge. High specificity will also encourage the acknowledgment of AutoTriage recommendations in the clinical setting, as low specificity can cause an estimated 96% of computerized alerts to be disregarded.\(^14\)

AutoTriage, while built for patient stability prediction, sits atop a general CDS framework. Within this framework, the many possible parameter combinations are searched for those that show the greatest improvement in prediction metrics. This flexibility is critical for developing robust prediction tools, as was demonstrated by the differential weighting of discharge predictor types. Notably, the resulting parameter weightings for S2 and D2 analyses were much higher than the simpler 1-dimensional analyses. These results can help explain the difference in predictive capabilities between AutoTriage, which considers low- and high-dimensional trends, and MET/MEWS, which rely solely on 1-dimensional S-type predictors. In addition, the D-type trend measurements played a large role in stability scoring, which demonstrates the value of including time-series data in predictions.

While AutoTriage was specifically used to recommend discharge, the general CDS framework may also be well suited for applications where early intervention in serious conditions could mitigate morbidity and mortality. For example, it has been shown that mortality from traumatic brain injuries is doubled when patients experience more than 2 hypotensive episodes in the emergency department.\(^15\) A time-series analysis of vital sign data could predict a trajectory trending toward a second episode of hypotension and warn the

---

Figure 3. Average percent contribution of each measurement type on patient score. S1 are individual measurements and D1 are the trends of individual measurements. S2 are the groupings of 2 measurements and D2 are the equivalent groupings of 2 trends. 95% confidence interval error bars are shown.
The methodology described here integrates information from multiple systems to better inform the clinician of the patient’s trajectory by elucidating patterns among complex correlations among large sets of medical data. While AutoTriage was used to reduce ICU length of stay, we believe that the same CDS framework could identify dangerous patient trajectories in a myriad of other medical syndromes and disease cases and empower the clinician to intervene more effectively.

FUNDING
This material is based upon work supported by the National Science Foundation under Grant No. 1549867. The funder had no role in the conduct of the study; collection, management, analysis, and interpretation of data; preparation, review, and approval of the manuscript; and decision to submit the manuscript for publication.

COMPETING INTERESTS
The authors have no competing interests to declare.

CONTRIBUTORS
J.S.C. designed the study, contributed to the data analysis, interpreted results, and drafted and revised the paper. D.A.P. contributed to the data acquisition and data analysis, and made revisions to the paper. C.W.B. contributed to the study design and data interpretation, and made significant revisions to the paper. U.K.C. contributed to the study design and results interpretation, and made substantive revisions to the paper. R.D. helped to conceive the study, and contributed to the study design, data analysis, and paper revisions.

ACKNOWLEDGEMENTS
We would like to thank Mitchell D. Feldman, MD, MPhil (Department of Medicine, University of California, San Francisco, CA, USA) and Jana Hoffman, PhD (Dascena Inc., Hayward, CA, USA) for their suggestions and help with editing.

REFERENCES


