# Development and Evaluation of Predictive Alerts for Hemodynamic Instability in ICU Patients

**Larry J. Eshelman, PhD[1], K. P. Lee, PhD[1], Joseph J. Frassica, MD[2], Wei Zong, PhD[2], Larry Nielsen, MS[2], Mohammed Saeed, MD, PhD[2]**

**[1]Philips Research North America, Briarcliff Manor, NY and [2]Philips Healthcare, Andover, MA**

## Abstract

*This paper describes an algorithm for identifying ICU patients that are likely to become hemodynamically unstable. The algorithm consists of a set of rules that trigger alerts. Unlike most existing ICU alert mechanisms, it uses data from multiple sources and is often able to identify unstable patients earlier and with more accuracy than alerts based on a single threshold. The rules were generated using a machine learning technique and were tested on retrospective data in the MIMIC II ICU database, yielding a specificity of approximately 0.9 and a sensitivity of 0.6.*

## Introduction

The modern intensive care unit (ICU) generates large amounts of patient data from an increasing number of monitoring devices and laboratory tests. These data must be interpreted in a timely fashion by ICU clinicians and nurses to provide proper patient care. There is a growing body of evidence that supports the notion that early intervention in the development of certain disease processes has a positive impact on patient outcome. Recently, such diverse disease states as septic shock, stroke, and acute myocardial infarction have come to the fore as disease entities whose outcome may be significantly modified by early goal-directed therapies [1, 2]. Many of the conditions which result in critical illness result in hemodynamic instability. To aid in the process of early detection and initiation of early goal-directed therapies, it would be desirable to have automated algorithms to aid in the identification of patients who are likely to become physiologically unstable. Most alerting mechanisms currently available to clinicians monitor a single parameter and generate an alert if the value crosses a threshold. We postulate that by correlating data from multiple sources, better determination of the stability of a patient can be achieved [3].

This paper describes an algorithm for identifying ICU patients that are likely to become hemodynamically unstable and provides a preliminary evaluation using retrospective ICU data. A machine learning approach was used with data from the MIMIC II database [4] collected from ICU patients at the Beth Israel Deaconess Medical Center, Boston MA. In the first section we describe the method for selecting the cases used in the machine learning approach, i.e., the intervention-based criteria used for differentiating stable from unstable. Next we describe the rules that constitute the algorithm and how they were derived. Then we describe two methods used for evaluating the algorithm and the results of our evaluation. In the conclusion section we suggest areas for further refinement of the algorithm.

## Materials and Methods

### Stability/Instability Criteria

Patient records from the MIMIC II database were utilized to create and test the predictive algorithms developed in this study. The MIMIC II project is an ongoing effort to capture de-identified ICU patient records consisting of physiologic monitoring and clinical information system data [4]. The physiologic data contained in the MIMIC II database was obtained from the Philips CareVue clinical information system in the participating ICUs. The Philips' CareVue clinical information system includes a relational database that contains nurse-charted vital signs, laboratory data, continuous medication administration, and fluid balance [5]. In the present study, 12,695 adult patient records collected between 2001 and 2005 were utilized to develop and evaluate predictive alerts to indicate impending physiologic instability [5].

We divided patient records into three major categories: (1) those that are stable throughout their stay, (2) those that become unstable, and (3) those that don't fit either category – the non-stable. The criteria are based on therapies given, reflecting physicians' judgment. Patients qualify as *unstable* if they received certain medications or procedures that are potentially indicative of the treatment of acute hemodynamic instability. The medications (any dosage) are Dobutamine, Dopamine, Epinephrine, Levophed or Neosynephrine. The procedures are (1) Intra-Aortic Balloon Pump, (2) >= 1100cc of Packed Red Blood Cells in 2 hours or (3) >2000cc of total IV-in in 1 hour.

If a patient meets any one of these criteria, then the patient is categorized as **unstable**. An unstable patient's *critical time* (ct) is the first time that the patient meets one of these criteria. **Stable** patients are those that throughout their stay do not meet any of the instability criteria and also do not meet any of the non-stability criteria. The non-stability criteria are interventions that are less severe than those that qualify patients as unstable, but severe enough to disqualify a patient as stable. These include any dosage of Lidocaine, Labetolol, nitroglycerine, nitroprusside, amiodarone, milrinone, and Esmolol; or the patient receiving >= 750cc of Packed Red Blood Cells in 24 hours or > 1500cc of total IV-in in 1 hour. Patients who do not meet any of the instability criteria but meet at least one of the non-stability criteria are categorized as **non-stable**. The non-stable patients were eliminated from the rule generation analysis due to our inability to decisively declare their condition with the available data but were included in the test set. Table 1 shows the counts of patients in the various categories.

**Table 1. Patient Count**

| Category | Count |
| --- | --- |
| Total | 12695 |
| Stable | 4063 |
| Unstable | 4867 |
| Non-stable | 3280 |
| Exp/DNR Stable | 485 |

The table includes a category that we have not mentioned so far: patients that qualify as stable according to our criteria, but who expired or are designated as Do-Not-Resuscitate (DNR) or Comfort-Measures-Only (CMO) patients. Many of these patients are unstable, but do not meet our intervention-based criteria either because they died before an intervention could be taken or a decision was made not to resuscitate. This patient group was removed from the rule generation analysis as well.

**Algorithm Development**

The algorithm consists of a set of 15 rules. Each rule has a list of conditions that must be satisfied in order for it to "fire". The conditions are feature values and thresholds. If all the conditions of any rule are satisfied, then an alert is issued. The features used in the rules are: BUN (blood urea nitrogen), WBC (white blood cell count), PTT (partial thromboplastin time), hematocrit, HR, systolic BP (arterial if available, otherwise noninvasive), and OxI (oxygenation index = Fraction of Inspired Oxygen*Mean Airway Presssure/$PaO_2$). The rules use the most recently measured value of any feature.

*The Rules*

The current version of the algorithm consists of the set of 15 rules shown in Table 2:

**Table 2. Rules**

| Rule# | BUN | HCT | PTT | WBC | SBP | OxI | HR |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | >18 | | | | <92 | | >111 |
| 2 | >30 | | | >11.5 | <81 | | |
| 3 | >56 | | >28.5 | | | >6.43 | |
| 4 | >26 | <24.3 | | | <89 | | |
| 5 | | <29.5 | | >13.1 | <92 | | |
| 6 | | <26.4 | >32.8 | | <86 | | |
| 7 | | <27.6 | >37.3 | | | >6.0 | |
| 8 | | | >39.4 | | <89 | | |
| 9 | | | >50.6 | | <92 | | |
| 10 | | | >45.8 | | | >5.3 | |
| 11 | | | >32.8 | | | >9.5 | |
| 12 | | | >32.5 | >10.5 | <104 | >6.0 | |
| 13 | | | | | <83 | >5.6 | |
| 14 | | | | | <88 | >9.0 | |
| 15 | | | | | <60 | | |

BUN= Blood Urea Nitrogen, HCT = Hematocrit, PTT = Partial Thromboplastin Time, WBC = White Blood Cell Count, SBP = Systolic Blood Pressure, OxI = Oxygenation Index, HR = Heart Rate

## Algorithm Development: Feature Selection and Rule Discovery

Algorithm development is a two stage process: selection of the features and discovery of the rules. We examined over forty features before settling on the seven used in the rules. Three criteria were used for choosing features. First, and most importantly, it needed to have some discriminating power, significantly differentiating stable from unstable patients. In particular, using either the minimum or maximum value, there was a signal-to-noise ratio (difference of means divided by sum of standard deviations) of at least 0.20. Second, it preferably was measured for most patients, i.e., at least two-thirds of them. Third, the more often it was measured for a patient the better. Most of the above features have strong discriminating power. The weakest is heart rate. It was kept, however, because it is measured for most patients and is measured often. On the other hand, some of the values used for calculating OxI are not available for many patients (it can be calculated for only about 25% of the stable patients), but OxI was selected because of its strong discriminating power.

During the rule discovery phase, we tried a number of machine-learning classification algorithms including such powerful techniques as support vector machines (SVM) and neural nets (NN). These techniques produced similar results. Since our preference was for techniques that produce classifier algorithms that are clinically transparent, we ultimately settled for one based on a rule learning technique called RIPPER (Repeated Incremental Pruning to Produce Error Reduction)[6].

The classifiers produced were validated using standard ten-fold cross-validation. Furthermore, we repeated the process several times using varying patient data sets. These data sets, consisting of several hundred stable and unstable patients, differed according to the average length of period from which we selected the maximum or minimum value (e.g., 12 hours versus 2.5 days) and where the period ended for unstable patients (e.g., up to but not including the critical time, or two hours before critical time of the first intervention).

The rules listed above are based on the results from several classifier runs using these different data sets. Typically, each run produced a classifier consisting of five or six rules. We merged these rules, eliminating those with low specificity or that were redundant. Merging the rules increased sensitivity at the expense of specificity. However, since the algorithm tended to produce rule sets with high specificity, we were still able to achieve our target specificity of at least 0.9. Throughout this process, in order to avoid over-fitting, we used about half of the available stable and unstable patients for training/tuning, reserving the other half for validation.

## Evaluation of the Algorithm

We evaluated the algorithm's performance in two different ways. First we wanted to know the sensitivity and specificity of the algorithm when comparing stable and unstable patients under ideal conditions, i.e., for similar lengths of time under conditions where all the data are available. Second we wanted to know how often there would be alerts in an ICU setting, and how these rates of alerts compared among different categories of patients. Toward this end, we computed the rates of alerts per patient-day using data from all the patients (stable, unstable and non-stable) in our data base of 12,695 patients.

## Results for Stable and Unstable Patients under Ideal Conditions

We first compare stable and unstable patients under ideal conditions, calculating the sensitivity and specificity. We selected those patients for which all the features used in the rules are available and for which the length of stay in the ICU is at least 12 hours in the case of stable patients and 12 hours prior to going unstable in the case of the unstable patients. In both cases the maximum period examined was 24 hours. Table 3 summarizes the patient population and the result. The 15-rule algorithm has a sensitivity of 0.6067, a specificity of 0.9285, and a positive predictive value of 0.7970. Keep in mind that our predictive alerts do not replace existing failsafe mechanisms (e.g. alarms that sound when blood pressure plummets) that have high sensitivity.

**Table 3. Results under Ideal Conditions**

| Category | # of Patients | Average Time Analyzed | Num/Fraction Positive |
|---|---|---|---|
| **Stable -** require all data | 769 | 0.94 days | 55   0.0715 (fp) |
| **Unstable** - require all data – 24 hrs prior critical time | 356 | 0.84 days | 216   0.6067 (tp) |

The results for the unstable patients are calculated using data up to, but not including, the critical time of the first intervention. On average the last point measured is about 25 minutes before critical time. However, because the recorded time typically is a rounded value, these times are not very precise. In order to get a better idea of the prediction power of the algorithm, we have calculated the sensitivity up to varying points of time before the critical time. The results are summarized in Figure 1 for 0.25 hours, 0.5 hours and hourly up to 12 hours before critical time. (These are minimum times – the average times are 15-25 minutes longer.)
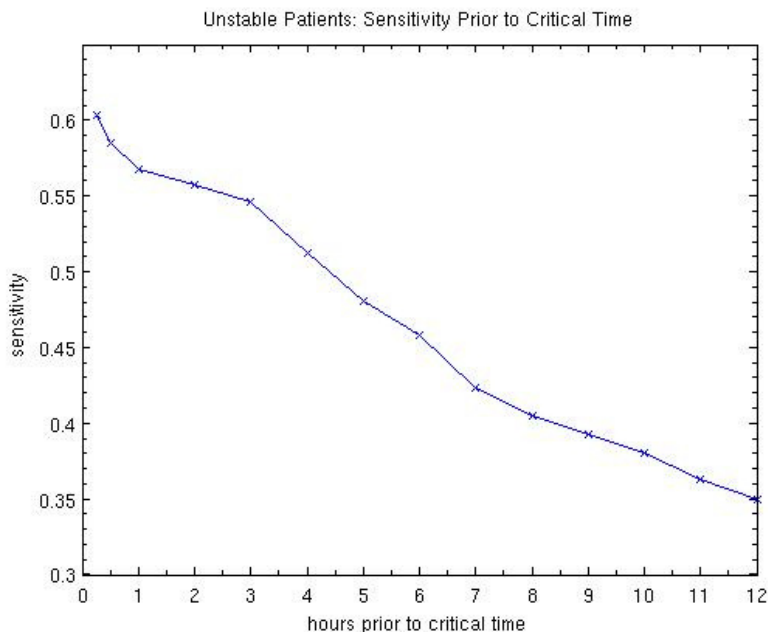


Unstable Patients: Sensitivity Prior to Critical Time

**Figure 1. Predictive Power of 15-rule Algorithm**

*Alert Rates*

Of course, the conditions in an ICU are not as ideal as the above analysis assumes. Often some of the features used in rules have not been measured and the length of the period for which data is available can vary considerably. A more realistic indicator of performance is to measure the alert rates for various categories of patients. However, this raises the question of what counts as a separate alert. When determining sensitivity and specificity, one or more rule firings during the period under consideration counted as a positive (whether true or false). In order to determine the alert rate, we needed to differentiate between two different alerts for the same patient. Intuitively, if the conditions that triggered rules to fire continue to hold over time, then two adjacent rule firings are part of the same alert. Therefore we count any two rule firings (either the same or different rules) within two hours of each other as part of the same alert.

**Table 4. Alert Rates**

| Category | # of patients | Ave. length period | Alerts/pat-day |
|---|---|---|---|
| All | 12679 | 5.73 days | 0.2757 |
| Stable | 4048 | 2.51 days | 0.1078 |
| "Stable": expired/DNR | 485 | 3.50 days | 0.2656 |
| Non-stable | 3279 | 5.18 days | 0.1721 |
| Unstable: full period | 4867 | 9.00 days | 0.3815 |
| Unstable: 24hrs prior to ct | 3431 | 0.33 days | 0.8214 |
| Unstable: beginning to 24hrs prior to ct | 712 | 4.93 days | 0.2010 |
| Unstable: after ct | 4867 | 8.06 days | 0.3894 |

Table 4 shows the alert rate per patient day for different categories of patients. The unstable patients are further divided into three sub-categories: (1) the 24-hour period prior to going

unstable, (2) the period prior to this 24-hour period, and (3) the period after the critical time. Ideally, the alert rate should be highest during the 24-hour period prior to critical time. It should also be noted that for the period after the critical time, alerts are counted only for the time when there is no intervention that qualifies the patient as unstable (and the length of period is calculated only for the time when there is no intervention).

As can be seen from the table the alert rate for unstable patients during the crucial 24 hour period prior to critical time is four times higher than for the previous period and about eight times higher than for stable patients. The relative rates for the other categories seem reasonable. For example, the rate for non-stable patients falls in between those for stable and unstable patients. Similarly, the rate for "stable" patients who died or were designated DNR or CMO is about 2.5 times higher than truly stable patients.

## Discussion

Our evaluation of the 15-rule hemodynamic instability-alert algorithm indicates that the alert rate in an ICU setting would be fairly low for patients that are not clearly unstable – about one alert per 10 patient days on average -- and so would not be a nuisance. Furthermore, the results indicate that the rate for unstable patients would be considerably higher than for stable patients – about 8 times higher.

It should be stressed that the algorithm presented here could potentially be improved. With regard to feature selection there are several things we would like to try in the future that might improve the resulting algorithm. First, we would like to try substituting mean blood pressure for systolic blood pressure. Second, we would like to construct an index that combines PTT with PT (prothrombin time) or INR (international normalized ratio) and use this instead of PTT alone. This would help identify cases where PTT is high because of the administration of heparin. We have examined the cases where high PTT values are associated with heparin, and they are nearly evenly distributed over stable and unstable patients and so do not seem to have biased our results. However, using an index that combines PTT and PT might be more accurate.

It should also be noted that the rules used in the algorithm have not been fully optimized. One possible method of improving the performance would be to convert the sharp thresholds to "fuzzy" ones.

This algorithm needs to be validated within other ICUs and tuned if necessary. Apart from tuning, there are many other ways of enhancing the performance of the algorithm. The current algorithm uses "absolute" values rather than relative values or deltas or rates of changes. Adding rules with conditions that monitor relative changes might improve the performance of the algorithm. Also the algorithm currently does not take into account the effect of interventions (other than using them as criteria for instability). The fact that a patient has just been given a powerful sedative, for example, might be incorporated into the conditions of some of the rules. Another way of improving the algorithm would be to take into account demographic data – e.g., adjust heart rate and blood pressure thresholds according to the patient's age. These enhancements would make the rules more sensitive to a patient's unique conditions, and thus improve performance.

### References

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

2. Kumar A, Roberts D, Wood KE, Light B, Parrillo JE, Sharma S, et al. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. Critical Care Medicine 2006;34(6):1589-1596.

3. Chambrin M-C. Alarms in the intensive care unit: how can the number of false alarms be reduced? Critical Care 2001;5(4):184-188.

4. Saeed M, Lieu C, Raber G, Mark R. MIMIC II: A massive temporal ICU patient database to support research in intelligent patient monitoring. Computers in Cardiology 2002;29:641-644.

5. CDM Product Documentation: Information Support Mart User's Guide, Version B.0. CD-ROM: Philips Medical Systems, Andover, MA 2001.

6. Cohen WW. Fast efficient rule induction. In: Prieditis A, Russell S, editors. Twelfth International Conference on Machine Learning; 1995; Tahoe City, CA: Morgan Kaufmann; 1995. p. 115-123.