A Clinical Database-Driven Approach to Decision Support: Predicting Mortality Among Patients with Acute Kidney Injury

Leo Anthony G. Celi, MD, MS, MPH^{1,*}; Robin J. Tang, MBA²; Mauricio C. Villarroel³; Guido A. Davidzon, MD, MS⁴; William T. Lester, MD⁵; Henry C. Chueh, MD, MS⁵

¹*Division of Pulmonary, Critical Care and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, MA USA*

²*Columbia University, College of Physicians and Surgeons, New York, NY USA*

³*Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, UK*

⁴*Department of Radiology (Nuclear Medicine), Stanford University Medical Center, Stanford, CA USA*

⁵*Laboratory of Computer Science, Massachusetts General Hospital, Boston, MA USA*

Submitted January 2010. Accepted for publication November 2010.

ABSTRACT

In exploring an approach to decision support based on information extracted from a clinical database, we developed mortality prediction models of intensive care unit (ICU) patients who had acute kidney injury (AKI) and compared them against the Simplified Acute Physiology Score (SAPS). We used MIMIC, a public de-identified database of ICU patients admitted to Beth Israel Deaconess Medical Center, and identified 1400 patients with an ICD9 diagnosis of AKI and who had an ICU stay ≥ 3 days. Multivariate regression models were built using the SAPS variables from the first 72 hours of ICU admission. All the models developed on the training set performed better than SAPS (AUC = 0.64, Hosmer-Lemeshow $p < 0.001$) on an unseen test set; the best model had an AUC = 0.74 and Hosmer-Lemeshow $p = 0.53$. These findings suggest that local customized modeling might provide more accurate predictions. This could be the first step towards an envisioned individualized point-of-care probabilistic modeling using one's clinical database.

Keywords: modeling, data mining, collective experience, decision support, ICU, mortality prediction, acute kidney injury

*Corresponding Author: Leo Anthony Celi, Beth Israel Deaconess Medical Center, Division of Pulmonary Critical Care and Sleep Medicine, 330 Brookline Avenue, Boston, MA 02215. Phone: (617) 710-3114. Fax: (617) 258-7859. E-mail: lceli@bidmc.harvard.edu. Other authors: rjt2112@columbia.edu; davidzon@stanford.edu; mauricio.villarroel@eng.ox.ac.uk; wlester@partners.org; hchueh@partners.org.

1. INTRODUCTION

We introduce an approach to decision support using one's own clinical database as an alternative to built-in expert systems derived from published large, usually multi-center, interventional and observational studies. Clinical trials tend to enroll heterogeneous group of patients in order to maximize external validity of their findings. As a result, recommendations that arise from these studies tend to be general and applicable to an "average" patient. Similarly, predictive models developed using this approach perform poorly when applied to specific subsets of patients or patients from a different geographic location as the initial cohort [1]. For these reasons, clinicians complement guidelines and protocols with their knowledge base and anecdotal experience. However, there is considerable variation in clinicians' prior experience, and in addition, clinician recall is often biased, with recent patients and patients with adverse outcomes being recalled most readily [2].

We propose using the information stored in one's own clinical database that represents the collective experience of numerous clinicians. The idea is to build models on more homogeneous group of patients with the explicit understanding that generalizability is traded-off for accuracy. We envision individualized probabilistic modeling in order to predict a diagnosis, treatment outcome or prognosis for an index patient. The modeling is performed on a cohort of similar patients as regards the features that are important in predicting the outcome variable. To illustrate, consider a 77-year old Hispanic lady admitted for community-acquired pneumonia with a serum hemoglobin on admission of 7.2 g/dL. She has known poor left ventricular function with an ejection fraction of 45%. Her vital signs, physical findings, urine output and laboratory results are available. Current guidelines recommend blood transfusion for hemoglobin levels between 7.0 and 10.0 g/dL "in the presence of end-organ dysfunction". In reality, some clinicians will transfuse while others will not, depending on both individual and local practice. For those who elect to transfuse, the number of units of packed red blood cells will likewise vary. The idea is to extract a cohort similar to the index patient from the local database and predict a certain outcome, with the knowledge that the model will only be useful to the index patient.

Before such a data-driven decision support system is designed, it is important to first determine whether models built on one's own database and on a more homogeneous patient subset perform better than current standard models developed from large, multi-center studies. We evaluate this approach in mortality prediction.

While there are a number of intensive care unit (ICU) and pre-operative mortality predictive models that are accurate in predicting mortality in a heterogeneous patient group, they are poorly calibrated; i.e., their prognostic accuracy at the individual patient or patient subgroup level is poor [1]. For these reasons, these severity scoring systems are currently used primarily for case-mix determination and benchmarking purposes, and not for individual patient prognostication.

One area in which generalized severity scoring systems have consistently performed poorly is among patients admitted to the ICU who then develop acute kidney injury

(AKI). AKI develops in approximately 6% of critically ill patients; two-thirds of these patients require renal support therapy [3]. The largest worldwide multi-center prospective study found that the observed mortality among these patients was substantially greater than predicted by the Simple Acute Physiologic Score (SAPS) II (60.3% vs. 45.6%, $p < 0.001$). In another UK-wide study of ICU patients who develop AKI, the APACHE II score under-predicted the number of deaths [4]. In this study, the null hypothesis of perfect calibration was strongly rejected ($p < 0.001$) by both the Hosmer-Lemeshow and Cox's calibration regression. More recently, the risk, injury, failure, loss, end-stage kidney disease (RIFLE) criteria were evaluated in predicting clinical outcomes among patients who develop AKI [5–9]. Although the RIFLE classification was found to be an independent predictor of mortality, it was not assessed with regard to discrimination and calibration.

We aim to determine whether a customized mortality prediction model for patients admitted to the ICU who are admitted with or develop acute kidney injury (AKI) will perform better than traditional scoring systems such as the SAPS. A second goal is to evaluate whether capturing the evolution of physiologic variables over time will produce a better model compared to the customary physiologic snapshot during the first 24 hours in the ICU.

2. METHODS

The Laboratory of Computational Physiology at Massachusetts Institute of Technology (MIT) developed and maintains the Multi-parameter Intelligent Monitoring for Intensive Care (MIMIC II) database, a high resolution database of ICU patients admitted to the Beth Israel Deaconess Medical Center (BIDMC) since 2003, that has been de-identified by removal of all Protected Health Information.

BIDMC is a 621-bed teaching hospital of the Harvard Medical School with 28 medical, 25 surgical (including neurosurgical), 16 cardiothoracic surgical and 8 cardiology ICU beds. An Institutional Review Board (IRB) approval was obtained from both MIT and BIDMC for the development, maintenance and public use of a de-identified ICU database. The MIMIC II database currently consists of data from more than 25,000 patients that has been de-identified and formatted to facilitate data-mining. The 3 sources of data are waveform data collected from the bedside monitors, hospital information systems and other third party clinical information systems.

Using the MIMIC II database, we identified the patients who had an ICD-9 diagnosis of acute renal failure (584.9) and who survived at least 72 hours in the ICU. We verified whether the patients developed acute kidney injury at around the time of ICU admission by looking at the serum creatinine and urine output during the first 72 hours in the ICU. Patients whose serum creatinine determinations were less than 1.0 mg/dl and who had an average urine output of 0.5 ml/kg/hr during the first three days of their ICU stay were excluded from the cohort, as they are unlikely to have sustained acute kidney injury at around the time of ICU admission.

The SAPS of each patient was converted to predicted mortality using the following formula:

$$\text{Predicted Death Rate} = e^{(\text{Logit})} / (1 + e^{(\text{Logit})}),$$

where $\text{Logit} = -7.7631 + 0.0737 * \text{SAPS} + 0.9971 * \ln(\text{SAPS} + 1)$ [10]. The predicted death rate was compared against the true outcome, and the Area under the Receiver Operating Characteristic Curve (AUC) and Hosmer-Lemeshow goodness-of-fit test p values were calculated. This AUC and Hosmer-Lemeshow p values were used as the benchmark to compare the performance of the customized models.

The MIMIC II database contains all the vital sign determinations, laboratory values, ventilatory settings, hemodynamic measurements, medications, interventions and treatment response as captured in the clinical notes, and diagnoses. Rather than performing automated feature selection, the physiologic variables included in SAPS from the first 72 hours in the ICU were used to build logistic regression models to predict hospital mortality.

Instead of replacing the missing values with the mean for variables with Gaussian distribution or the median for all other variables, we applied a set of rules derived from clinical experience. The rules are as follows:

1. If the values for a certain variable are missing for all three days, they are replaced by the middle value of the normal range of that variable. If a variable is not measured in the ICU, e.g., serum bilirubin, there is a good likelihood that there is no concern that this may be abnormal.
2. A missing value on the second ICU day was replaced with the average of the first and third day values. A missing value on the first or third ICU day was replaced with the second day value.
3. If values on two of the first three days in the ICU are missing, they are both replaced with the value that is present.

A univariate logistic regression was performed on each of the variables to determine whether they are correlated with hospital mortality.

Models were built using the SAPS variables from Day 1, Day 2 and Day 3, and from all three days using the SAS software (SAS Version 9.2, SAS Institute Inc., Cary, NC, USA). A forward selection algorithm was performed on the SAPS variables from the first 72 hours of ICU admission, and a separate model was built with the significant predictors selected by the algorithm. Each of the model building was done using threefold cross-validation repeated three times. The mean AUC and Hosmer-Lemeshow p value were calculated for each model. The effect size estimate and the p values of the covariates of the best-performing of the three cross-validation models are reported. A diagram of the steps in constructing the mortality prediction models is illustrated in Figure 1.

We explored whether the physiologic variables were better represented as continuous or categorical, and whether a quadratic or cubic relationship exists between the physiologic variables and the log odds of dying in the multivariate analysis. Effect modification was also explored between age and the minimum systolic blood pressure,

Figure 1. Construction of the mortality prediction models.

and between age and the maximum heart rate. This is to test the hypothesis that the effect of the minimum blood pressure and the maximum heart rate on mortality varies at different age groups. A p value of 0.05 was used to determine the significance of the interaction term.

3. RESULTS

There were a total of 1400 patients with an ICD-9 diagnosis of acute renal failure who survived at least 72 hours in the ICU out of the 25,642 ICU patients in the database at the time of the study. Among them, 970 survived and 430 died in the hospital (30.7% mortality). Of the 1400 patients, 265 were admitted to the Coronary Care Unit, 121 from the Cardiac Surgery Recovery Unit, 800 from the Medical Intensive Care Unit, and 214 from the Surgical Intensive Care Unit. Among the patients who survived, the median ICU length-of-stay was 5.9 days (1st, 3rd quartiles = 4, 10). Among those who died, the median ICU length-of-stay was 7.5 days (1st, 3rd quartiles = 4.9, 11.9). The distributions of selected variables in the entire patient cohort among survivors and non-survivors are shown in Figure 2.

(a) Age distribution among survivors and non-survivors

Red:

Mean (survivors) = 68.0

Standard deviation (survivors) = 16.0

Blue:

Mean (non-survivors) = 71.9

Standard deviation (non-survivors) = 15.6

(b) Maximum serum creatinine on day 1 among survivors and non-survivors

Table 1. Univariate analyses of the SAPS variables from Day 1, Day 2 and Day 3

Day 1	Day 2	Day 3
Worst Systolic Blood Pressure ($p < 0.0001$)*	Worst Systolic Blood Pressure ($p < 0.0001$)*	Worst Systolic Blood Pressure ($p < 0.0001$)*
Worst Heart Rate ($p = 0.066$)	Worst Heart Rate ($p = 0.027$)*	Worst Heart Rate ($p = 0.013$)*
Worst Temperature ($p = 0.221$)	Worst Temperature ($p = 0.093$)	Worst Temperature ($p = 0.917$)
Urine Output ($p = 0.001$)	Urine Output ($p = 0.001$)	Urine Output ($p = 0.001$)
Worst Blood Urea Nitrogen ($p = 0.002$)*	Worst Blood Urea Nitrogen ($p < 0.0001$)*	Worst Blood Urea Nitrogen ($p < 0.0001$)*
Worst White Blood Count ($p = 0.127$)	Worst White Blood Count ($p < 0.0001$)*	Worst White Blood Count ($p < 0.0001$)*
Worst Serum Potassium ($p = 0.163$)	Worst Serum Potassium ($p = 0.187$)	Worst Serum Potassium ($p = 0.403$)
Worst Serum Sodium ($p = 0.546$)	Worst Serum Sodium ($p = 0.699$)	Worst Serum Sodium ($p = 0.283$)
Worst Serum Bicarbonate ($p = 0.134$)	Worst Serum Bicarbonate ($p = 0.011$)*	Worst Serum Bicarbonate ($p = 0.002$)*
Worst Serum Bilirubin ($p < 0.0001$)*	Worst Serum Bilirubin ($p < 0.0001$)*	Worst Serum Bilirubin ($p < 0.0001$)*
Age ($p = 0.012$)		

respectively. The null hypothesis that the SAPS model is perfectly calibrated for this patient subgroup is strongly rejected.

The variables were represented as either continuous or categorical in various combinations using the training set, and the resulting model was assessed in the test set (data not shown). The best models in terms of AUC and Hosmer-Lemeshow p value were obtained when the maximum temperature, the urine output and the minimum serum bicarbonate were represented as categorical variables and the rest as continuous variables. The interaction terms between age and the minimum systolic blood pressure as well as between age and the maximum heart rate had non-significant p values in all the models.

Tables 2A, 2B and 2C show the effect size estimates, standard errors and p values of the covariates of the best fitted models using the SAPS variables from Day 1, Day 2 and Day 3, respectively. Table 3 pertains to the best fitted logistic regression model using the SAPS variables from all three days, while Table 4 shows

Table 2. Best fitted model using SAPS variables. (2A) Day 1. (2B) Day 2. (2C) Day 3 (2A)

Variable	Effect Size	Standard Error	P Value
Worst Systolic Blood Pressure	−0.015	0.004	<0.0001*
Worst Heart Rate	0.007	0.003	0.0139*
Worst Temperature	−0.113	0.073	0.1246
Presence/Absence of Oliguria	0.562	0.126	<0.0001*
Worst Blood Urea Nitrogen	0.003	0.002	0.1854
Presence/Absence of Leukocytosis	0.536	0.151	0.0004*
Worst Serum Potassium	0.144	0.111	0.1965
Worst Serum Sodium	−0.012	0.012	0.3189
Presence/Absence of Acidosis	0.013	0.159	0.9315
Worst Serum Bilirubin	0.067	0.015	<0.0001*
Age	0.005	0.002	0.0038*

(2B)

Variable	Effect Size	Standard Error	P Value
Worst Systolic Blood Pressure	−0.015	0.004	0.0001*
Worst Heart Rate	0.007	0.003	0.0109*
Worst Temperature	−0.135	0.081	0.0977
Presence/Absence of Oliguria	0.491	0.140	0.0004*
Worst Blood Urea Nitrogen	0.007	0.002	0.0021*
Presence/Absence of Leukocytosis	0.768	0.168	<0.0001*
Worst Serum Potassium	0.082	0.117	0.4864
Worst Serum Sodium	−0.00009	0.013	0.9947
Presence/Absence of Acidosis	0.324	0.175	0.0644
Worst Serum Bilirubin	0.064	0.015	<0.0001*
Age	0.005	0.002	0.0063

(2C)

Variable	Effect Size	Standard Error	P Value
Worst Systolic Blood Pressure	−0.019	0.002	<0.0001*
Worst Heart Rate	0.007	0.004	0.0127*
Worst Temperature	0.076	0.003	0.3562
Presence/Absence of Oliguria	0.591	0.082	<0.0001*
Worst Blood Urea Nitrogen	0.013	0.151	<0.0001*
Presence/Absence of Leukocytosis	0.822	0.002	<0.0001*
Worst Serum Potassium	0.106	0.179	0.4017
Worst Serum Sodium	−0.006	0.126	0.6307
Presence/Absence of Acidosis	0.180	0.013	0.3660
Worst Serum Bilirubin	0.068	0.198	<0.0001*
Age	0.005	0.15	0.0062*

Table 3. Best fitted model using Day 1, Day 2 and Day 3 SAPS variables

	Variables	Effect Size	Standard Error	P Value
Day 1	Worst Systolic Blood Pressure	−0.009	0.005	0.0734
	Worst Heart Rate	0.002	0.004	0.5426
	Worst Temperature	−0.040	0.093	0.6713
	Presence/Absence of Oliguria	0.547	0.139	<0.001*
	Worst Blood Urea Nitrogen	−0.008	0.006	0.2138
	Presence/Absence of Leukocytosis	0.181	0.205	0.3678
	Worst Serum Potassium	0.164	0.142	0.2496
	Worst Serum Sodium	−0.026	0.022	0.2387
	Presence/Absence of Acidosis	−0.183	0.197	0.3515
	Worst Serum Bilirubin	−0.130	0.085	0.1258
Day 2	Worst Systolic Blood Pressure	−0.002	0.006	0.7876
	Worst Heart Rate	0.002	0.004	0.6368
	Worst Temperature	−0.300	0.118	0.0107*
	Presence/Absence of Oliguria	0.044	0.167	0.7906
	Worst Blood Urea Nitrogen	−0.020	0.012	0.0954
	Presence/Absence of Leukocytosis	0.272	0.268	0.3100
	Worst Serum Potassium	−0.163	0.161	0.3102
	Worst Serum Sodium	0.072	0.030	0.0176*
	Presence/Absence of Acidosis	0.327	0.229	0.1529
	Worst Serum Bilirubin	−0.062	0.134	0.6435
Day 3	Worst Systolic Blood Pressure	−0.017	0.005	0.0004*
	Worst Heart Rate	0.004	0.004	0.3317
	Worst Temperature	0.281	0.110	0.0104*
	Presence/Absence of Oliguria	0.370	0.170	0.0296*
	Worst Blood Urea Nitrogen	0.038	0.008	<0.0001*
	Presence/Absence of Leukocytosis	0.532	0.254	0.0363*
	Worst Serum Potassium	0.094	0.152	0.5381
	Worst Serum Sodium	−0.037	0.023	0.1056
	Presence/Absence of Acidosis	−0.002	0.235	0.9942
	Worst Serum Bilirubin	0.248	0.100	0.0128*
	Age	0.005	0.002	0.0147*

the best model using only the significant predictors when a forward selection algorithm (significant level of entry of $p = 0.05$) was employed on the Day 1, Day 2 and Day 3 SAPS variables.

Finally, Table 5 presents the average AUC's and Hosmer-Lemeshow p values for each of the best fitted models. The AUC and Hosmer-Lemeshow p value of SAPS on this patient cohort are provided in the first row for comparison.

Table 4. Best fitted model using the significant predictors from Day 1, Day 2 and Day 3 through forward selection

Variable	Effect Size	Standard Error	P Value
Day 3 Worst Blood Urea Nitrogen	0.026	0.004	<0.0001*
Day 3 Worst Systolic Blood Pressure	-0.021	0.004	<0.0001*
Day 3 Presence/Absence of Leukocytosis	0.818	0.180	<0.0001*
Day 1 Presence/Absence of Oliguria	0.554	0.131	<0.0001*
Day 3 Worst Serum Bilirubin	0.192	0.060	0.0015*
Day 1 Worst Blood Urea Nitrogen	-0.015	0.004	0.0001*
Day 3 Presence/Absence of Oliguria	0.443	0.153	0.0038*
Day 3 Worst Heart Rate	0.007	0.003	0.0162*
Day 1 Worst Serum Bilirubin	-0.137	0.006	0.0291*
Age	0.005	0.002	0.0110*

Table 5. Comparison of the logistic regression models

	Area under the ROC Curve (AUC)	Hosmer-Lemeshow P Value
SAPS	0.642	<0.0001
Day 1 SAPS Variables	0.682	0.4201
Day 2 SAPS Variables	0.695	0.8717
Day 3 SAPS Variables	0.726	0.8691
Day 1 + Day 2 + Day 3 (All Variables)	0.762	0.1649
Day 1 + Day 2 + Day 3 (Forward Selection)	0.743	0.5321

4. DISCUSSION

The goal of this paper is to determine whether local customized modeling can provide more accurate mortality prediction as compared to current standard severity scoring systems such as SAPS. The accuracy of mortality prediction models is assessed based on discrimination between survivors and non-survivors and on correspondence between observed and predicted mortality across the entire range of risk (calibration). We reported the AUC and the Hosmer-Lemeshow p value as measures of discrimination and calibration, respectively, for all the logistic regression models.

All the models developed on the training set using the MIMIC database performed better than SAPS on an unseen test set with regard to both discrimination and calibration. When SAPS was used to predict death among the MIMIC patients with AKI, the null hypothesis of a well-calibrated model was rejected. In contrast, all the local customized models had non-significant p values in predicting mortality in this

patient subset, suggesting better calibration, in addition to higher AUCs compared to the AUC of SAPS.

We observed an increasing AUC and Hosmer-Lemeshow p value when comparing Day 1, Day 2 and Day 3 models. This suggests that the evolution of physiologic variables over time is more predictive of the clinical outcome than the physiologic snapshot of the ICU patient on admission which is the basis of current severity scoring systems. This is concordant with what we observe clinically; how a patient responds to treatments, rather than his initial presentation, eventually determines whether he/she survives or not. The model with the highest AUC is the one that incorporated the SAPS variables from the first 72 hours in the ICU. The Hosmer-Lemeshow p value of this model is lower than the individual day models but remained non-significant, suggesting goodness-of-fit. The more parsimonious model using only the significant predictors from the first three days in the ICU as identified by a forward selection algorithm had a higher Hosmer-Lemeshow p value but slightly lower AUC. It is of note that majority of the significant covariates in the Day 1 + Day 2 + Day 3 model and in the forward selection model was from day 3 of the patients' ICU stay. The clinical status of an ICU patient on day 3 reflects how the patient has responded to the interventions which might be more predictive of the outcome than his/her presentation on admission.

Despite a significant improvement in the AUC and Hosmer-Lemeshow p value using the SAPS physiologic variables from the first 72 hours in the ICU, a much higher AUC would have been more convincing evidence to support our hypothesis. There are several reasons why a higher AUC was not seen with the logistic regression models that were presented. First, patients admitted to the ICU who develop AKI may still represent a heterogeneous group. Variables apart from the SAPS components may be required to build a more accurate predictive model. Feature selection may be applied to all the variables in the database rather than using the SAPS variables to build the models. Second, we did not capture quality of care and the physiologic reserve, two important determinants of clinical outcome apart from the severity of illness. Third, assumptions of logistic regression analysis may not be true; i.e., the effect of the SAPS variables on the log odds of dying may not be monotonic and additive. Other machine learning algorithms might provide more accurate mortality prediction models among this patient subset, and are explored in a separate paper. Finally, data noise as a result of inaccurate data capture or entry was evident during pre-processing, and likely contributed to suboptimal model performance.

We describe here the framework of a dynamic decision support tool based on empiric data, a departure from the typical rule-based algorithms that dominate current computer-assisted decision support systems (CDSSs). Such an approach is only made possible now by the increasing use of electronic medical records (EMRs). In addition to assisting clinicians in providing care to individual patients, EMRs have the potential to serve as a repository of highly granular clinical information. An EMR documents per patient every diagnostic test and intervention ordered, as well as outcomes. We envision using these data to perform probabilistic modeling on homogeneous patient subsets as the basis of a decision support system.

Increasingly, large patient databases are being mined to build and validate clinical models [11]. We coin the term “Collective Experience” for this approach, a name that reflects the growing emphasis on CDSSs as aggregators of necessary clinical knowledge rather than replacements for human judgment. This system can provide diagnostic assistance, treatment guidance and prognostic capabilities personalized at the patient level, by only including similar patients in the modeling cohort. There is a long road ahead before our vision of probabilistic modeling at the point-of-care to assist clinicians with contextual questions regarding individual patients becomes a reality. Non-trivial pre-processing and processing issues when mining large high-resolution clinical databases abound. More importantly, impact studies are necessary to evaluate whether this approach will influence clinician behavior and improve patient outcomes.

5. CONCLUSION

In summary, we introduce an approach to decision support that mines the local database to provide truly representative patient data. While evidence-based medicine has overshadowed empirical therapies, we argue that each patient interaction, particularly when recorded with granularity in an easily accessible and computationally convenient electronic form, has the potential to tailor recommendations for individual patients. Even though prospective randomized controlled trials have become gold standard in establishing the link between science and medical practice, important insights continue to arise from routine patient care. “Collective Experience” can fuel a dynamic decision support system where recorded health care transactions - clinical decisions linked with patient outcomes - are constantly uploaded to support point-of-care probabilistic modeling towards a new era of personalized medicine.

The present study attempts to help clinicians answer just such a question: “What is the prognosis of this patient in my ICU with acute kidney injury?” This approach is not intended to replace clinician judgment or autonomy – indeed, it cannot and should not – but to expand and expedite their access to the hard-to-find data they need.

ACKNOWLEDGEMENTS

This work is supported by National Library of Medicine (NLM) Training Grant # 2T15LM007092-16. The MIMIC II database was funded by the National Institute of Health (NIH), and its National Institute of Biomedical Imaging and Bioengineering (NIBIB) under Grant 2R01 EB001659.

REFERENCES

- [1] Strand K. and Flaaten H., Severity Scoring in the ICU: A Review, *Acta Anaesthesiologica Scandinavica*, 2008, 52(4), 467–478.
- [2] Groopman J., *How Doctors Think*. Houghton Mifflin, Boston, 2007.
- [3] Uchino S., Kellum J., Bellomo R., Doig G.S., Morimatsu H., Morgera S., Schetz M., Tan I., Bouman C., Macedo E., Gibney N., Tolwani A., Ronco C., Acute Renal Failure in Critically Ill Patients: A Multinational, Multicenter Study, *JAMA*, 2005, 294(7), 813–818.
- [4] Kolhe N., Stevens P., Crowe A., Lipkin G.W., Harrison D.A., Case Mix, Outcome and Activity for Patients with Severe Acute Kidney Injury During the First 24 hours After Admission to an Adult General Critical Care Unit: Application of Predictive Models from a Secondary Analysis of the ICNARC Case Mix Programme Database, *Critical Care*, 2008, 12 Suppl 1, S2.

- [5] Park W.Y., Hwang E.A., Jang M.H., Park S.B., Kim H.C., The Risk Factors and Outcome of Acute Kidney Injury in the Intensive Care Units, *Korean J Intern Med*, 2010, 25(2),181–187.
- [6] Li W.X., Chen H.D., Wang X.W., Zhao S., Chen X.K., Zheng Y., Song Y., Predictive Value of RIFLE Classification on Prognosis of Critically Ill Patients with Acute Kidney Injury Treated with Continuous Renal Replacement Therapy, *Chin Med J*, 2009, 122(9),1020–1025.
- [7] Abosaif N.Y., Tolba Y.A., Heap M., Russell J., El Nahas A.M., The Outcome of Acute Renal Failure in the Intensive Care Unit according to RIFLE: Model Application, Sensitivity, and Predictability, *Am J Kidney Dis*, 2005, 46(6),1038–1048.
- [8] Ostermann M., Chang R., Riyadh ICU Program Users Group, Correlation between the AKI Classification and Outcome, *Crit Care*, 2008, 12(6),R144.
- [9] Hoste E.A., Clermont G., Kersten A., Venkataraman R., Angus D.C., De Bacquer D., Kellum J., RIFLE Criteria for Acute Kidney Injury are Associated with Hospital Mortality in Critically Ill Patients: a Cohort Analysis, *Crit Care*, 2006, 10(3),R73.
- [10] Le Gall J., Lemeshow S., Saulnier F., A New Simplified Acute Physiology Score (SAPS II) based on a European/North American Multicenter Study, *JAMA*, 1993, 270(24),2957–2963.
- [11] Eddy D.M., Linking Electronic Medical Records to Large-Scale Simulation Models: Can We Put Rapid Learning on Turbo?, *Health Affairs*, 2007, 26(2),w125-w136.