Capstone Project

Predictive Analytics in the Pediatric ICU
Using Electronic Health Record Data:
Clinical rationale, data science techniques, and evaluation of different prediction models to predict acute kidney injury

Master of Biomedical Informatics

Department of Medical Informatics & Clinical Epidemiology

Oregon Health & Science University

Portland, Oregon

Term: Fall 2015

Student: L. Nelson Sanchez-Pinto, MD

Capstone Advisor: Michael Mooney, PhD

School of Medicine

Oregon Health & Science University

CERTIFICATE OF APPROVAL

This is to certify that the Master's Capstone Project of

L. Nelson Sanchez-Pinto, MD

"*Predictive Analytics in the Pediatric ICU Using Electronic Health Record Data: Clinical rationale, data science techniques, and evaluation of different prediction models to predict acute kidney injury*"

Has been approved

Michael Mooney, PhD

Introduction

Clinicians, like every other human, are not good prognosticators. With the ever-increasing complexity of healthcare, the inherent limitations of the human brain, and a myriad of cognitive biases exacerbated by interruptions, fatigue and multitasking, it's no wonder that we often fail in our clinical predictions (1). The advances in medical technology and the advent of Big Data have increased the complexity of clinical care, but have also opened the door to data science at a scale not possible a few years ago (2). Predictive analytics is perhaps the most promising area of data science in healthcare, and coupled with our need for better predictions it's ripe to potentially make a big impact in the way we take care of our patients. Being able to predict a disease course, a complication or the response to a therapy could help us shift from a reactive to a proactive approach in healthcare. Coupled with advance clinical decision support (CDS) systems, data-driven prediction models can interface with the clinician's workflow to improve decision-making. Indeed, these models are not intended to replace the clinician's qualitative reasoning but supplement it in order to achieve the best possible decisions in patient care (3).

All technological advances have risks, however, and the possibility of unreliable research fueled by poorly applied data science techniques is real (4). Therefore, it's paramount that clinicians understand the general principles of data science and predictive modeling to make better use of the technology, recognize its promises, and acknowledge its limitations.

The intensive care unit (ICU) perhaps represents one of the areas of medicine where the need for accurate predictions and the complexity of care are most manifest. Diagnostic prediction modeling, in particular, can be very helpful in this setting where early recognition and early intervention are usually associated with improved outcomes (5). In the pediatric ICU, a diagnosis that oftentimes eludes clinicians and could benefit from an accurate diagnostic prediction model is acute kidney injury (AKI). AKI is independently associated with increased morbidity and mortality in critically ill children (6,7), but unfortunately the main biomarker of AKI, serum creatinine, can take several days to peak after the injury has occurred. This reliance on serum creatinine can delay the diagnosis of AKI and therefore the implementation of potentially beneficial interventions and preventive strategies (8).

In this paper, we will review the key components of the development and validation of diagnostic prediction models using the example of AKI in the pediatric ICU. We will apply clinical data quality and data cleaning techniques to deal with missing

values, age-dependant variables, unbalanced datasets, and non-linear associations. We will review and apply feature selection techniques and various statistical learning methods, such as multivariate adaptive regression splines, random forest, and support vector machine. We will also discuss model evaluation techniques that are pertinent to the prediction models themselves, but also to the medical stetting and clinical realities in which the models will be used.

Key Words

Predictive modeling, data science, machine learning, pediatric critical care, intensive care unit, acute kidney injury, electronic health records.

Purpose of the Study

Our purpose is to:

1) Identify candidate variables to predict acute kidney injury in the pediatric ICU in a large retrospective cohort using EHR data.
2) Derive and validate three prediction models (multivariate adaptive regression splines, random forest, and support vector machine) and two ensemble models.
3) Evaluate the diagnostic performance and clinical utility of the models.

Previous Work Done in this Area

1. Acute Kidney Injury (AKI)

a. Risk stratification models for AKI

In their 2005 study, Chawla et al. (9) designed a risk stratification model of adult patients at risk to develop AKI in the ICU based on the presence of cancer, decreased albumin, and increased A-a gradient. This model was later tested by Malhotra et al. (10) in 63 patients and was proven to be effective in predicting AKI development and mortality (40.2% of patients in the high-risk group developed AKI vs. 6.4% in the rest of the cohort). In the pediatric literature there are several descriptive, epidemiologic studies of AKI (6,7,11-13) but none of them have attempted to create a similar stratification of at-risk patients. Of note, many of these studies have shown that, as opposed to adult patients, children usually develop AKI early in their hospital course, in a majority of cases within the first 48 hours (6,7,11,12). This further emphasizes the importance of early recognition of those at risk for AKI.

In the wake of the growing interest in AKI biomarkers, Goldstein and Chawla proposed the concept of "Renal Angina" (14) suggesting that since AKI doesn't hurt in the way that myocardial angina does, trying to find the "renal equivalent of troponin I" without defining which patients should be tested would undermine the utility of such a biomarker. With this in mind, they reviewed the current literature in AKI and developed a pediatric and an adult "renal angina syndrome equivalent". The pediatric model, specifically for ICU patients, is based on acute co-morbidities (mechanical ventilation, inotrope use) and chronic diagnosis (stem cell transplant, heart failure) that place the patients in different hazard tranches. The estimated creatinine clearance and fluid overload of the patient then triggers the "renal angina" at different levels depending on the hazard tranche in which he falls. This risk stratification model was validated in several subsets of pediatric ICU patients and had good discrimination to detect severe AKI by day 3 of ICU (AUC 0.74 to 0.81) (14). In combination with novel AKI biomarkers, the model has had a good performance in pediatric patients with sepsis (AUC 0.84 to 0.88), but its performance in a general, larger pediatric ICU population with a lower pre-test probability remains to be tested (15,16). The main limitations of this stratification tool is that it was not specifically designed to be implemented in the form of an automated CDS system and that it requires additional testing in the form of novel AKI biomarkers in order to achieve good discrimination.

b. Risk factors for AKI

In the pediatric literature, several risk factors have been associated with AKI, notably: mechanical ventilation (7,11,17), hypoxia (17,18), hypotension (18), nephrotoxic drug exposure (19,20), malignancy (17,18,21), stem cell transplantation (20,22), sepsis (23), metabolic acidosis (17), thrombocytopenia (18), congenital heart disease (17) and heart surgery requiring cardiopulmonary bypass (11). As previously mentioned, it's worth noting that none of these studies have attempted to prospectively test the presence of these risk factors as predictors of AKI in children. In the adult AKI literature, notable additional risk factors identified include: high CRP (24), changes in systolic blood pressure (25), increased A-a gradient (9), active cancer (9), low albumin (9), low prealbumin (24), anemia (26) and anemia requiring blood transfusions (27).

2. Prediction models in medicine

Prediction models and prediction rules have been developed in medicine for a long time for things such as estimating the likelihood of acute myocardial infection or different types of cancer (28). The reality, though, is that few models are externally validated, fewer are clinically implemented, and their clinical impact is rarely measured (1). Recently, there has been a growing interest in the role of data mining and predictive analytics in healthcare with the increased adoption of electronic health records (4,29). In other industries, data mining has been successful were data is in abundance (30), and now –more than ever before– data abounds in medicine. Specifically in the ICU we have "critical data" (4) and a need to perform better diagnostic prediction to improve prevention and early interventions (5).

Data mining, machine learning, and predictive analytics is slowly making its way into the healthcare system with successful models to prevent fraud and increase revenue (31), predict morbidity in the emergency department (32), forecast heart failure progression (33), and to process the vast amounts of data that genetic and proteomic analyses are producing in cancer research (34). The next step is to build models with both clinical effectiveness and clinical utility (35) in order to improve patient outcomes.

Use Case: AKI Prediction Model in the Pediatric ICU Workflow

When developing a clinical prediction model, perhaps one of the most important aspects of the design process is to determine how the model will be used in the clinical setting. For our use case, these were the leading clinical characteristics of AKI in the pediatric ICU that determined the design of our model (6):

1) Most patients who are diagnosed with AKI in the pediatric ICU are admitted with some degree of AKI or develop AKI within 72 hours of being admitted to the pediatric ICU.
2) Those who develop new AKI, have progression of their AKI, and those who have no improvement of their AKI during the first week of ICU care have the worst outcomes amongst all ICU patients.
3) Serum creatinine is a late biomarker of AKI, taking up to 48h-72h to peak after the kidneys are injured (figure 1).

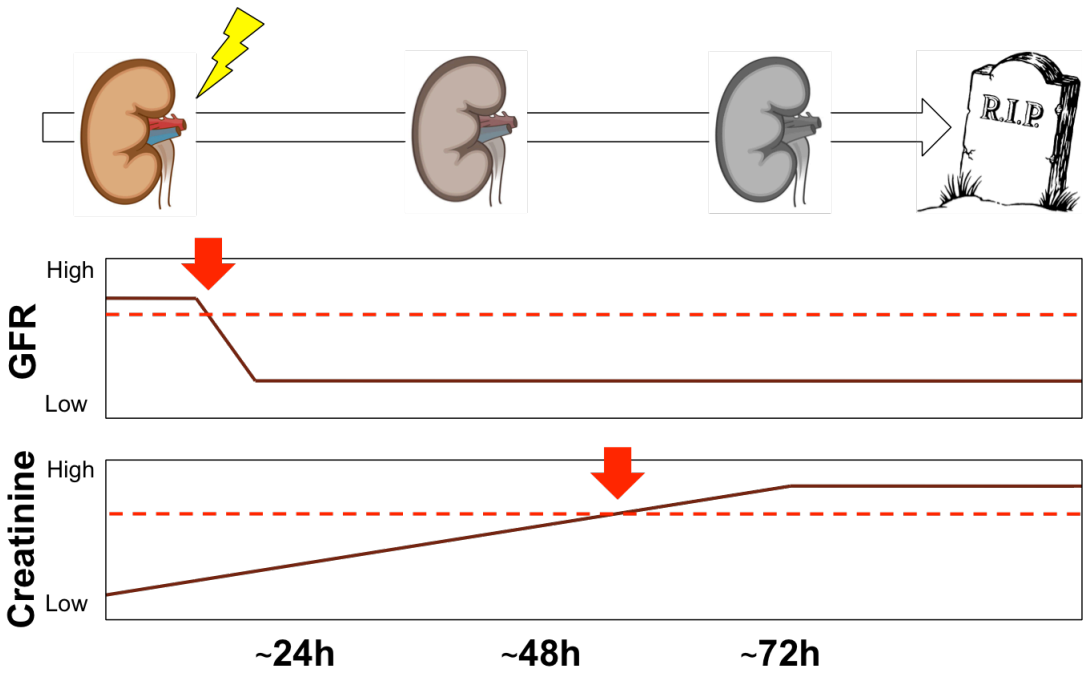

Figure 1. After an injury the kidney's glomerular filtration rate (GFR), the main physiologic measure of kidney function, drops rapidly. Serum creatinine is a surrogate of GFR, but it takes time to accumulate in the blood after an injury, essentially causing a delay in diagnosis. The red line denotes the threshold of normal values and the red arrow represent the point at which each of the two measure of kidney function becomes abnormal.

In addition to these AKI characteristics, an important factor influencing the design of our prediction model was the optimal time when it would be most useful. The most active clinical decision-making in the pediatric ICU by the medical team takes place in two different situations: at the time of admission of the patient to the unit and during medical team rounding in the morning of each of the days the patient remains in the ICU. The timing of a prediction model-enabled clinical decision support would ideally align with one of these two time points in order to magnify the impact of the prediction in the decision-making.

For the reasons enumerated above, we decided to use new or progressing AKI by 72 hours, or persistent AKI as the predicted outcome (see Methods) and the first 12 hours of ICU data as the data source for model training. This would allow the prediction output (and triggered CDS tool) to be available by the first morning of ICU stay the day after admission for most patients (figure 2).

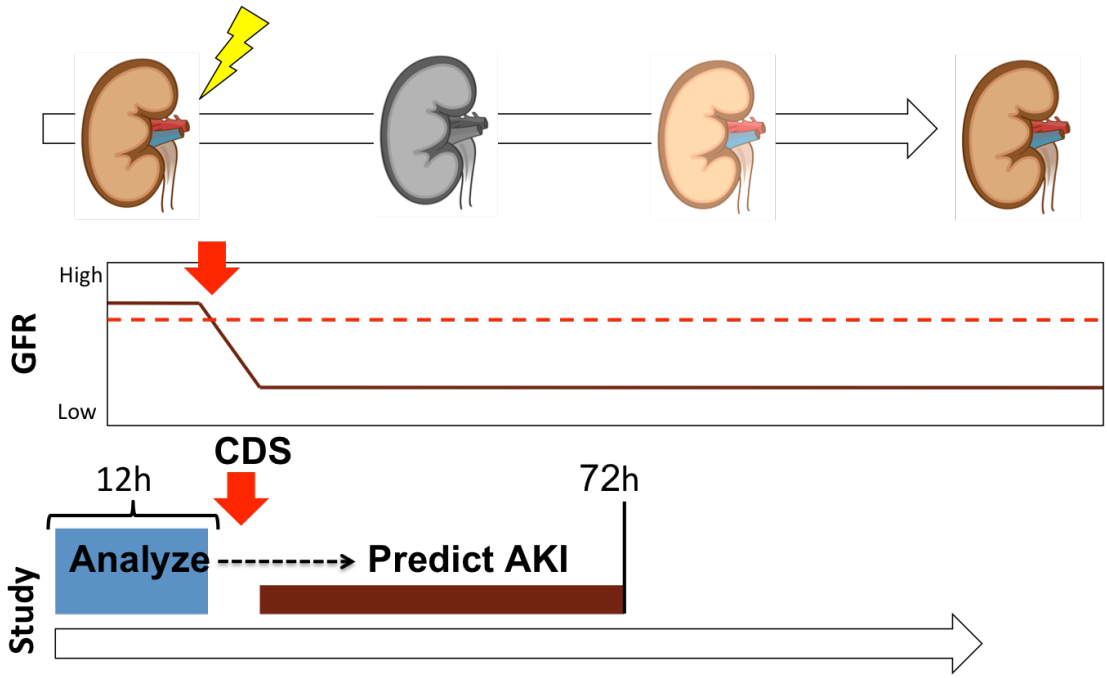

Figure 2. Using data from the first 12 hours of ICU to predict AKI by 72 hours would theoretically reduce the amount of time between kidney injury and the initiation of preventive or therapeutic intervention, which could improve the outcome of the kidney injury. *AKI*, acute kidney injury.

Methods

1. Patients and Data Sources

We performed a retrospective analysis of all patients admitted to a multidisciplinary, tertiary pediatric ICU between May 2003 and March 2012 who were in the ICU and alive for at least 24 hours. This 24-bed pediatric ICU serves a mixed population of medical, surgical, trauma, and solid-organ and stem cell transplantation patients, but not postoperative cardiac patients. Data were extracted from our EHR clinical databases (Cerner Kids, Kansas City, MO; Philips/CareVue, Waltham, MA) and a locally developed quality improvement and clinical database (Microsoft Access, Seattle, WA) maintained by the ICU physicians delivering care.

Patients were excluded if they were younger than 1 month or older than 21 years of age, had documented chronic kidney disease, were peri-operative for a kidney transplant, or had no serum creatinine levels measured. Each ICU admission was treated

independently. This study was approved with a waiver of informed consent by the Institutional Review Board of Children's Hospital Los Angeles.

2. Derivation and Validation Groups

The population was divided into two cohorts, one for training the prediction models and one for validation of the models. A random sample of 60% of the patients was used for model training (Derivation set), and the remaining 40% were used for the validation set (Validation set).

3. AKI Definition

AKI was defined by the serum creatinine (SCr) Kidney Disease Improving Global Outcome (KDIGO) staging criteria (36). SCr levels measured during the ICU stay were compared to a baseline creatinine, which was the most recent documented SCr within 6 months of the ICU admission. If a prior SCr was unavailable, the upper limit of normal for age and gender was used (6). Initiation of renal replacement therapy (RRT) was also considered as a criterion for AKI stage 3 as per the KDIGO guidelines. Additional details regarding this cohort of patients and methodology have been previously published (6).

4. Outcome

Our goal was to develop a model to predict a composite outcome of AKI progression or persistence. "AKI progression/persistence" was defined as new or progressing AKI (increase in the KDIGO AKI stage over the first 72 hours of ICU stay), or persistent AKI from admission until death, discharge or 1 week of ICU stay (whichever came first). The criteria for AKI progression/persistence were based on the results of a prior AKI outcomes study in our patient population (6). Examples of patients with AKI

progression/persistence would be: a patient with no AKI on admission who develops AKI (any stage) before 72 hours of ICU (figure 3, point A); a patient with AKI on admission who progresses to a higher stage AKI by 72 hours of admission (figure 3, point B); or a patient with AKI on admission who has persistent AKI for more than a week or death occurs (figure 3, point C).

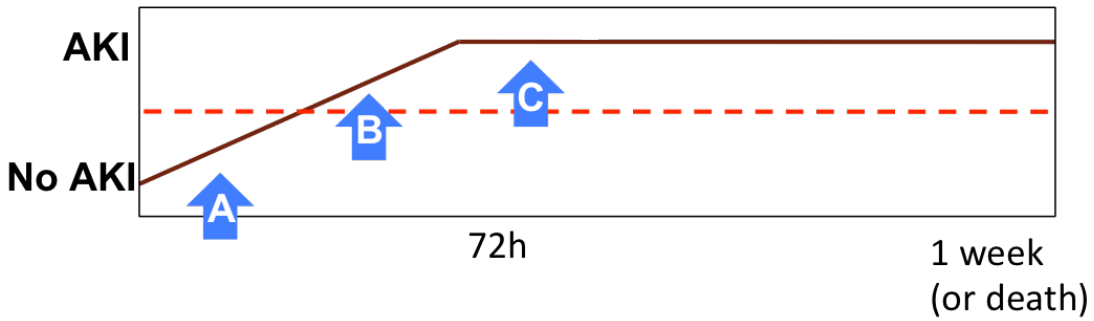

Figure 3. The AKI composite outcome of AKI progression/persistence conceptually represents the idea that patients may be in different phases of the illness at the time of ICU admission. Point A: patient with new AKI and progression; Point B: AKI on admission with progression; Point C: AKI on admission with persistence. *AKI*, acute kidney injury; *AKI progression/persistence*, new or progressing AKI by 72 hours of ICU or persistent AKI until discharge, death or 1 week of ICU;

5. Candidate Variables

The derivation dataset was used to select variables for inclusion in the final model. Candidate variables were considered for inclusion in the multivariate model if they were generally available from the EHR, were generalizable across ICUs, and had previously been identified as potential risk factors for AKI based on the pathophysiology of kidney injury and the medical literature (7,9,11,17-26). Since we sought to identify early progression of AKI, analysis of candidate variables was limited to data available within the first 12 hours of ICU admission. 20 candidate variables represented by a total of 22 features were studied (Table 1).

a. Missing values

Missing values were imputed using predictive mean matching with a K-nearest neighbor of 5 using the MI package in STATA version 14 (StatCorp, College Station,

TX). The Pediatric Index of Mortality-2 score, a measure of severity of illness, was included as a covariate for matching in the imputation process.

b. Age normalization

The age-dependant variables (systolic blood pressure, diastolic blood pressure, heart rate and serum creatinine level) were normalized using a t-statistic standardization using the mean and standard deviation of the subsample of patients who had no AKI, did not die and were in the ICU less than 72 hours ("healthy" patients), using the following equation:

$$\text{Normalized value} = \frac{\text{Non-normalized value} - \text{mean of healthy patients}}{\text{Standard deviation of healthy patients}}$$

Risk factor groups	Features	% Missing
Hemodynamic Instability	Lowest systolic blood pressure	0
	Systolic blood pressure range	0
	Lowest diastolic blood pressure	0
	Highest heart rate	0
	Heart rate range	0
	Highest vasoactive-inotrope score	0
Hypoxemia	Invasive mechanical ventilation (Yes/No)	0
	Lowest SpO_2	0
	Lowest SpO_2/FiO_2 ratio	59.8
Anemia	Lowest hemoglobin level	21.4
Inflammation	Highest white blood cell count	23
Thrombocytopenia	Lowest platelet count	21.8
Liver failure	Highest total bilirubin level	63.7
	Lowest albumin level	63.7
Acidosis	Lowest pH level	45.7
Renal/metabolic derangements	First serum creatinine	17.8
	Highest blood urea nitrogen level	17.8
	Highest glucose level	18.4
Demographic and admission characteristics	Age	0
	Gender	0
	Post-operative status (Yes/No)	0
	Pre-admission cardiac arrest (Yes/N)	0

Table 1. List of features representing the 20 candidate variables studied in the first 12 hours of ICU care, grouped by risk factor groups.

6. Feature selection

Correlation amongst variables was studied using Spearman's correlation coefficient. Features were then selected using a random forest approach, called VSURF (37) in the derivation cohort. VSURF is a three step feature selection procedure with the first step dedicated to eliminate irrelevant variables, the second step selecting all features related to the response and the third step refining the selection by eliminating redundancy in the set of features selected by the second step.

Because of the unbalanced nature of the dataset with a class distribution of 1:16 between cases and controls, the derivation cohort dataset was rebalanced by random under-sampling of the majority class to improve the performance of the random forest algorithm (38,39). To reduce bias in the feature selection process, five different datasets with randomly under-sampled majority classes were generated and VSURF was applied to each one of them. Features selected in at least 80% of the datasets were used for model training.

7. Prediction Model Training

The prediction models were trained using R version 3.2.2 (R Foundation for Statistical Computing, Vienna, Austria). We trained three different prediction models with the selected features: multivariate adaptive regression splines (MARS) (40), random forest (RF) (41), and support vector machine (SVM) (42). MARS was trained using the complete training dataset, while RF and SVM models were trained in a rebalanced training dataset with random under-sampled majority class to improve performance (38,39).

The RF model was trained with 1000 trees. The MARS model was trained with second-degree interactions. The SVM and RF models were trained for two-class prediction. The MARS model probability score output had a single cut-off point determined using the highest Youden's Index (Sensitivity + Specificity – 1) to generate a two-class prediction. In addition, the MARS model was tested for two-class prediction using other cut-off points determined by the 50^{th}, 75^{th}, and 90^{th} percentiles of the probability score.

In addition to the three models, we derived two ensemble models: one which considered patients high-risk for AKI progression/persistence (test +) if any of the three models were positive ("Any model +") and one were patients were considered test + if all three models were positive ("All models +").

8. Prediction Model Performance

a. Diagnostic Performance

The three prediction models, the two ensemble models and the MARS model with three alternative cut points were tested as diagnostic tests with a single output, test + or test –. Sensitivity, specificity, positive and negative predictive value, and positive and negative likelihood ratios were calculated.

b. Net Benefit

Net benefit (NB) is a measure of clinical utility that incorporates the relative benefit of correctly identifying patients that will have the disease (true positives) with the cost of the false positive results (35). NB is calculated with the following equation:

$$NB = P(TP) - P(FP)w$$

where $P(TP)$ is the proportion of true positives, $P(FP)$ is the proportion of false positives and w is the weight of cost of identifying a false positive case for each true positive case. We used three theoretical scenarios in which an additional screening tool, such as an AKI urine biomarker, is applied to the 'test positive' patients. Even though NB can be calculated using only clinical value metrics, these are difficult to assign at a population level, and hence we use a cost-benefit approach to determine the value of the different models in the three scenarios. We assign a plausible cost of $500 to the new test and the average economic benefit of identifying a true positive in three possible outcomes: (1) $500 (for a modest saving in care costs for mild improvement in the outcome, e.g. less frequent laboratory checks); (2) $5,000 (for a larger benefit, e.g. decreased length of ICU stay); and (3) $50,000 (for an even larger benefit, e.g. avoidance of a major complication). In these three scenarios the weights would be: w_1= 500/500 = 1; w_2= 500/5000 = 0.1; w_3= 500/50,000) = 0.01.

Results

1. Epidemiology and Demographics

7,029 patients were included in the analysis, of which 6.4% (440 patients) had AKI progression/persistence. Patients with AKI progression/persistence had significantly higher mortality than those without AKI (28.9% vs. 2.6%, p <0.001) and those with AKI who did not meet AKI progression/persistence criteria (28.9% vs. 18.6%, p = 0.001).

Table 2 summarizes the demographic characteristics of the derivation and validation groups. Both groups were very similar.

	Derivation	Validation
Total N	4,200	2,829
AKI progression/persistence (%)	261 (6.2)	179 (6.3)
No AKI on admission	161 (3.8)	102 (3.6)
AKI with progression	80 (1.9)	71 (2.5)
AKI with persistence	20 (0.5)	6 (0.2)
AKI total (%)	531 (12.6)	366 (12.9)
Age (Years, IQR)	7.3 (1.7, 13.6)	7 (1.6, 13.5)
Male (%)	2,271 (54.1)	1,504 (53.2)
Mechanical Ventilation (%)	1,846 (44)	1,311 (46.3)
Length of Stay (days, IQR)	3.1 (1.8, 6.7)	3.1 (1.9, 7)
PIM-2 %ROM (IQR)	0.8 (0.1, 3.1)	0.8 (0.2, 3.5)
28-day Mortality (%)	200 (4.8)	160 (5.7)

Table 2. **Demographic and clinical characteristics of the patient groups.** Averages expressed in medians with inter-quartile ranges (IQR). *AKI*, acute kidney injury; *AKI progression/persistence*, new or progressing AKI by 72 hours of ICU or persistent AKI until discharge, death or 1 week of ICU; *PIM-2 %ROM*, Pediatric Index of Mortality-2 Risk of Mortality

2. Feature selection

The random forest-based feature selection procedure selected 14 different features across five different rebalanced training datasets with under-sampled majority class. Of these features, only 6 features were consistently selected across all datasets, none of the others were selected at least 80% of the time. Figure 4 shows the 6 features (highest blood urea nitrogen (BUN) level, first serum creatinine level, lowest platelet count, lowest pH level, post-operative status, highest vasoactive-inotrope score), plus the variable 'age', that was included for interpretability and model stability purposes. The selected features had low correlation (all $r < 0.5$) and missing values (all < 50%).

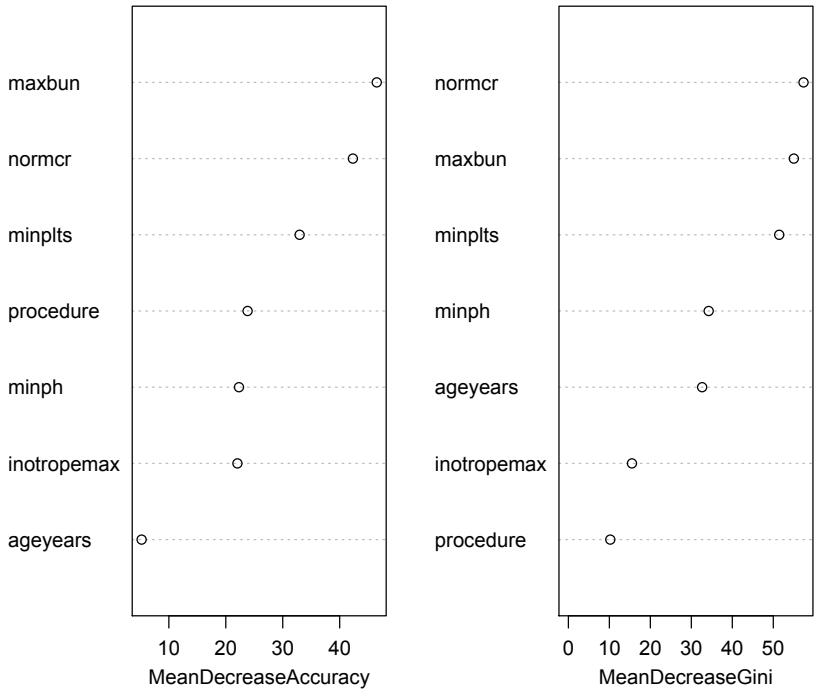

Figure 4. **Features selected for the final model.** Features selected and their relative importance based on mean decreased accuracy and mean decreased Gini in a random forest model. *maxbun*, highest blood urea nitrogen (BUN) level; *normcr*, first serum creatinine level; *minplts*, lowest platelet count; *minph*, lowest pH level; *procedure*, post-operative status; *inotropemax*, highest vasoactive-inotrope score; *ageyears*, age.

3. Prediction Model Performance

a. Diagnostic Performance

The diagnostic performance of the three prediction models is shown in table 3. RF had the highest sensitivity and lowest negative likelihood ratio. SVM had the highest accuracy, specificity and positive likelihood ratio.

	Random Forest	MARS*	SVM
Test + (%)	729 (25.8)	661 (23.3)	577 (20.4)
True Positives	140	135	128
No AKI on admission	68 (67)	66 (65)	60 (59)
AKI with progression	66 (93)	63 (89)	62 (87)
AKI with persistence	6 (100)	6 (100)	6 (100)
Accuracy % (95% CI)	78 (77, 79)	80 (79, 81)	82 (81, 83)
Sensitivity % (95% CI)	78 (72, 84)	75 (69, 81)	72 (65, 78)
Specificity % (95% CI)	78 (77, 78)	80 (80, 81)	83 (83, 84)

PPV % (95% CI)	19 (18, 21)	20 (19, 22)	22 (20, 24)
NPV % (95% CI)	98 (98, 99)	98 (97, 99)	98 (97, 98)
LR+ (95% CI)	3.5 (3.2, 3.7)	3.8 (3.4, 4.2)	4.2 (3.7, 4.7)
LR- (95% CI)	0.28 (0.21, 0.37)	0.31 (0.23, 0.4)	0.34 (0.27, 0.43)

Table 3. Performance of the three different prediction models in the validation set. *The cut point in the MARS model presented in this table was determined by the highest Youden's index. *MARS*, multivariate adaptive regression splines; *SVM*, support vector machine, *PPV and NPV*, Positive and Negative Predictive Value; *CI*, confidence interval.

When breaking down the true positive cases into the three subgroups of the AKI progression/persistence composite outcome, it can be seen that patients with AKI on admission were easier to discriminate when compared to those with no AKI on admission, likely due to the importance of the renal biomarkers in the model and the higher pre-test probability of those with ongoing AKI. Despite this, about two-thirds of patients who developed new AKI were correctly identified by all of the models.

The diagnostic performance of the MARS model with three alternative cut points is shown in table 4. These alternative cut points were determined using the 50th, 75th, and 90th percentiles of the MARS model probability score. These cut points were associated with a 5.7%, 9.7% and 27.5% probability, respectively, of having AKI progression/persistence. 28-day mortality also followed the same trend with patients below the 50th percentile cut point having a 2.5% mortality; those in the 50th to 75th group having a 4.1% mortality; those in the 75th to 90th percentile group having 8.5% mortality; and those above the 90th percentile having an 18.8% mortality. As expected, the MARS (50th) model had the highest sensitivity and lowest negative likelihood ratio, whereas the MARS (90th) had the highest specificity and positive likelihood ratio across all models.

MARS score cut points	50th	75th	90th
Test + (%)	1414 (50)	707 (25)	282 (10)
Accuracy % (95% CI)	55 (54, 55)	78 (78, 79)	90 (90, 91)
Sensitivity % (95% CI)	88 (82, 92)	77 (70, 83)	53 (46, 60)
Specificity % (95% CI)	53 (52, 53)	79 (78, 79)	93 (93, 93)
PPV % (95% CI)	11 (10, 12)	20 (18, 21)	34 (29, 38)
NPV % (95% CI)	98 (98, 99)	98 (98, 99)	97 (96, 97)
LR+ (95% CI)	1.8 (1.7, 2)	3.6 (3.2, 3.9)	7.5 (6.2, 9)
LR- (95% CI)	0.23 (0.15, 0.35)	0.29 (0.22, 0.4)	0.51 (0.43, 0.58)

Table 4. Performance of the MARS model with three alternative cut points. *MARS*, multivariate adaptive regression splines; *PPV and NPV*, Positive and Negative Predictive Value; *CI*, confidence interval.

The diagnostic performance of the two ensemble models is shown in table 5. The ensemble *Any model +* had the highest sensitivity and lowest negative likelihood ratio, whereas *All models +* had the highest specificity and positive likelihood ratio. Almost 20% of the true positive patients did not overlap between the two ensembles, which is an indication that the three prediction models are discriminating cases in different ways and gives the ensemble approach more value.

	Any model +	All models +
Test + (%)	887 (31.4)	454 (16)
Accuracy % (95% CI)	73 (72, 74)	86 (85, 87)
Sensitivity % (95% CI)	83 (77, 88)	67 (60, 73)
Specificity % (95% CI)	72 (72, 73)	87 (87, 88)
PPV % (95% CI)	17 (16, 18)	26 (23, 29)
NPV % (95% CI)	99 (98, 99)	98 (97, 98)
LR+ (95% CI)	3 (2.7, 3.2)	5.3 (4.5, 6)
LR- (95% CI)	0.23 (0.16, 0.32)	0.38 (0.31, 0.47)

Table 5. Performance of the two ensemble models in the validation set. In the *Any model +*, any patient that was test + in any of the three prediction models in Table 3 was considered test +. In the *All models +*, only patients that were test + in all of the three prediction models in Table 3 were considered test +. *PPV and NPV*, Positive and Negative Predictive Value; *CI*, confidence interval.

b. Net Benefit

Table 6 presents the Net Benefit results for the three different proposed scenarios.

	W_1 ($500)	W_2 ($5,000)	W_3 ($50,000)	Sum*
Random Forest	–15.6	2.9	4.7	–8
MARS (Youden's)	–13.8	2.9	4.6	–6.3
MARS (50^{th}%)	–38.9	1.1	5.1	–32.7
MARS (75^{th}%)	–15.2	2.9	4.7	–7.6
MARS (90^{th}%)	–3.3	2.7	3.3	**2.7**
SVM	–11.4	2.9	4.4	–4.1
Any model +	–20.8	2.7	5.0	–13.1
All models +	–7.6	**3.0**	4.1	–0.5
No model, test all	–8.7	–3.0	**5.4**	–0.3
No model, test none	**0**	0	0	

Table 6. Net benefit percentage in three different cost-based scenarios for a theoretical new urine AKI biomarker. In this Net Benefit analysis we assign a plausible cost of $500 to a new urine AKI biomarker test that will be performed only on test + patients in the different models and a test all or test none situation. The W_1, W_2 and W_3 represent a $500, $5,000 and $50,000 gain, respectively, when a true positive patient is detected. This example assumes the new urine AKI

biomarker accurately discriminates true positives and true negatives. *The sum of the different net benefits is assuming an equal weight of 1/3 for each scenario.

In this proposed example of a new urine AKI biomarker with three different gain scenarios, it is clear that in the low gain scenario (W_1) the best approach would be to not implement the test. If the gain for identifying true positives was \$5,000 (W_2) the *All models* + ensemble model would produce the highest gain. With substantial gains of \$50,000 per true positive identified (W_3), testing all patients might become beneficial. Using this Net Benefit approach, once the likelihood of each of the three scenarios is ascertained, one can assign probabilities and determine the approach with the highest chance of being beneficial. If there was no *a priori* information on the potential economic benefit of the test and an equal probability is assigned to each scenario, then the only model with a positive net benefit would be the MARS (90^{th}) in this example.

To further illustrate this example from an financial standpoint, a pediatric ICU that admits 1000 patient per year that implements the new urine AKI biomarker only in patients who are test + in the *All models* + ensemble would have the following combinations of loses and gains based on the different scenarios: W_1 = –\$59,000; W_1 = \$130,000; W_1 = \$2,020,000.

Discussion

We developed and internally validated three different prediction models and two ensemble models of AKI progression/persistence in the pediatric ICU. These data-driven prediction models have a good diagnostic performance in the validation dataset, dividing patients into groups with a 13 to 17-fold difference in the incidence of AKI progression/persistence. These models were developed using only EHR data that is objective, available in real time during the first 12 hours of ICU care and generalizable across pediatric ICUs.

Many risk factors have been associated with AKI in the pediatric and adult literature (7,11,17,19, 20-27). As previously mentioned, none of these studies have attempted to use these risk factors to develop data-driven prediction models of AKI in children. Our data-driven feature selection process identified many of these risk factors as predictors, but also found many of these same risk factors to have no independent effect on the outcome when analyzed in a multivariate model.

We have developed our AKI progression/persistence prediction models to take advantage of the EHR-based digital infrastructure in healthcare to facilitate deployment in the form of CDS systems (29, 45, 46). CDS systems are considered by most public and private national health organizations as one of the key instruments in the effort to improve the quality of patient care (47,48). Even though the data on outcomes remains sparse, CDS systems have shown to be effective at improving healthcare processes (49). We attempted to stay consistent with current recommendations from the medical informatics community in our AKI progression/persistence score, by mining existing clinical databases to develop data-driven CDS systems, but ensure our variables are generalizable and able to be implemented across different organizations (50). Bates and colleagues recommend integrating the CDS systems into the workflow of clinicians, making it available in real time, and limiting the amount of information required to be entered by the physician into the CDS system to a minimum (51). Our AKI progression/persistence models were designed with these principles in mind. By using real-time EHR data from the first 12 hours of ICU stay, a CDS system using this model would require no direct input from the clinician. In regards to the integration with the clinical workflow, our AKI progression/persistence models were designed to provide valuable information at 12 hours of ICU care, which is a time point when very active medical decision-making is still taking place.

Several intervention, preventive strategies, and research initiatives could be designed around a CDS tool powered by our AKI progression/persistence models. Relatively low risk strategies such as nephrotoxic medication avoidance and strict intake and output control could be easily implemented for those patients who fall in the high-risk group. Following the Renal Angina Index example, our AKI progression/persistence models could be also used as a screening tool for the general pediatric ICU population to obtain a novel AKI biomarker in the first 24 hours of ICU admission (16). To that effect we present a Net Benefit analysis of the implementation of a new urine AKI biomarker and show the different levels of benefit that could be achieved based on three different gain scenarios. A more accurate clinical utility analysis could be performed as the value of novel AKI biomarkers are analyzed.

Our study has several limitations. First, our prediction model was derived using data from a single institution and has not been validated in a different institution. Second, the data used to derive the model spans nine years of ICU care, which could be associated with significant variability in care. Third, variables that were not routinely recorded at the

beginning of the study period (e.g. serum lactate levels) were not included in the study. Fourth, only data recorded during routine patient care was used, which helps with the generalizability of the model and the integration with the clinical workflow, but also results in the problem of unknown values for variables not collected in certain patients. Lastly, we used a composite outcome, AKI progression/persistence, which can introduce bias in the model training process given the higher pre-test probability of patients with ongoing AKI on admission. Having said that, AKI progression/persistence represents in theory all patients with a worsening renal function trajectory who are identified at different time points in their course and the model still maintains a acceptable degree of clinical utility as a screening tool in a general pediatric ICU population.

Conclusion

We developed and validated several data-driven AKI progression/persistence prediction models in a general pediatric ICU population using only EHR data that is objective, available in real time during the first 12 hours of ICU care and generalizable across pediatric ICUs. A cost-based Net Benefit analysis shows that these models could have high clinical utility in different plausible scenarios.

Based on the different sensitivity of the models to patients with and without AKI on admission, the next steps of this project will be to explore what threshold levels might be more appropriate for the different types of patients at risk (i.e. those with and without AKI on admission). In addition, the Net Benefit analysis should be performed with real use cases to determine the potential clinical utility prior to implementation of an AKI prediction model-based diagnostic test (e.g. a novel AKI urine biomarker), preventive bundle (e.g. avoidance of nephrotoxic drugs and renal dosing), or interventions (e.g. fenoldopam).

References

(1) Tangri N, Kent DM. Toward a Modern Era in Clinical Prediction: The TRIPOD Statement for Reporting Prediction Models. Am J Kidney Dis 2015 Jan 15.

(2) Anthony Celi L, Mark RG, Stone DJ, Montgomery RA. "Big data" in the intensive care unit. Closing the data loop. American journal of respiratory and critical care medicine 2013;187(11):1157-1160.

(3) Moons KG, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. Heart 2012 May;98(9):683-690.

(4) Badawi O, Brennan T, Celi LA, Feng M, Ghassemi M, Ippolito A, et al. Making big data useful for health care: a summary of the inaugural mit critical data conference. JMIR medical informatics 2014;2(2).

(5) Labarère J, Bertrand R, Fine MJ. How to derive and validate clinical prediction models for use in intensive care medicine. Intensive Care Med 2014;40(4):513-527.

(6) Sanchez-Pinto LN, Goldstein SL, Schneider JB, Khemani RG. Association Between Progression and Improvement of Acute Kidney Injury and Mortality in Critically Ill Children. Pediatr Crit Care Med 2015 Jun 30.

(7) Akcan-Arikan A, Zappitelli M, Loftis LL, Washburn KK, Jefferson LS, Goldstein SL. Modified RIFLE criteria in critically ill children with acute kidney injury. Kidney Int 2007 May;71(10):1028-1035.

(8) de Geus HR, Betjes MG, Bakker J. Biomarkers for the prediction of acute kidney injury: a narrative review on current status and future challenges. Clinical kidney journal 2012;5(2):102-108.

(9) Chawla LS, Abell L, Mazhari R, Egan M, Kadambi N, Burke HB, et al. Identifying critically ill patients at high risk for developing acute renal failure: a pilot study. Kidney Int 2005 Nov;68(5):2274-2280.

(10) Malhotra RM, Macedo E, Bouchard J, Wynn S, Mehta R. Prediction of acute kidney injury (AKI) by risk factors classification [Abstract]. J Am Soc Nephrol 2009;20: 000.

(11) Li S, Krawczeski CD, Zappitelli M, Devarajan P, Thiessen-Philbrook H, Coca SG, et al. Incidence, risk factors, and outcomes of acute kidney injury after pediatric cardiac surgery: a prospective multicenter study. Crit Care Med 2011 Jun;39(6):1493-1499.

(12) Alkandari O, Eddington KA, Hyder A, Gauvin F, Ducruet T, Gottesman R, et al. Acute kidney injury is an independent risk factor for pediatric intensive care unit mortality, longer length of stay and prolonged mechanical ventilation in critically ill children: a two-center retrospective cohort study. Crit Care 2011 Jun 10;15(3):R146.

(13) Zappitelli M, Bernier PL, Saczkowski RS, Tchervenkov CI, Gottesman R, Dancea A, et al. A small post-operative rise in serum creatinine predicts acute kidney injury in children undergoing cardiac surgery. Kidney Int 2009 Oct;76(8):885-892.

(14) Goldstein SL, Chawla LS. Renal angina. Clin J Am Soc Nephrol 2010 May;5(5):943-949.

(15) Basu RK, Zappitelli M, Brunner L, Wang Y, Wong HR, Chawla LS, et al. Derivation and validation of the renal angina index to improve the prediction of acute kidney injury in critically ill children. Kidney Int 2013 Sep 18.

(16) Basu RK, Wang Y, Wong HR, Chawla LS, Wheeler DS, Goldstein SL. Incorporation of biomarkers with the renal angina index for prediction of severe AKI in critically ill children. Clin J Am Soc Nephrol 2014 Apr;9(4):654-662.

(17) Duzova A, Bakkaloglu A, Kalyoncu M, Poyrazoglu H, Delibas A, Ozkaya O, et al. Etiology and outcome of acute kidney injury in children. Pediatr Nephrol 2010 Aug;25(8):1453-1461.

(18) Bailey D, Phan V, Litalien C, Ducruet T, Merouani A, Lacroix J, et al. Risk factors of acute renal failure in critically ill children: A prospective descriptive epidemiological study. Pediatr Crit Care Med 2007 Jan;8(1):29-35.

(19) Moffett BS, Goldstein SL. Acute kidney injury and increasing nephrotoxic-medication exposure in noncritically-ill children. Clin J Am Soc Nephrol 2011 Apr;6(4):856-863.

(20) Ileri T, Ertem M, Ozcakar ZB, Ince EU, Biyikli Z, Uysal Z, et al. Prospective evaluation of acute and chronic renal function in children following matched related donor hematopoietic stem cell transplantation. Pediatr Transplant 2010 Feb;14(1):138-144.

(21) Fisher BT, Zaoutis TE, Leckerman KH, Localio R, Aplenc R. Risk factors for renal failure in pediatric patients with acute myeloid leukemia: a retrospective cohort study. Pediatr Blood Cancer 2010 Oct;55(4):655-661.

(22) Detaille T, Anslot C, de Clety SC. Acute kidney injury in paediatric bone marrow patients. Acta Clin Belg Suppl 2007;(2)(2):401-404.

(23) Hui-Stickle S, Brewer ED, Goldstein SL. Pediatric ARF epidemiology at a tertiary care center from 1999 to 2001. Am J Kidney Dis 2005 Jan;45(1):96-101.

(24) Xie Q, Zhou Y, Xu Z, Yang Y, Kuang D, You H, et al. The ratio of CRP to prealbumin levels predict mortality in patients with hospital-acquired acute kidney injury. BMC Nephrol 2011 Jun 29;12:30.

(25) Liu YL, Prowle J, Licari E, Uchino S, Bellomo R. Changes in blood pressure before the development of nosocomial acute kidney injury. Nephrol Dial Transplant 2009 Feb;24(2):504-511.

(26) Shema-Didi L, Ore L, Geron R, Kristal B. Is anemia at hospital admission associated with in-hospital acute kidney injury occurrence? Nephron Clin Pract 2010;115(2):c168-76.

(27) Haase M, Bellomo R, Story D, Letis A, Klemz K, Matalanis G, et al. Effect of mean arterial pressure, haemoglobin and blood transfusion during cardiopulmonary bypass on post-operative acute kidney Injury. Nephrol Dial Transplant 2011 Jun 15.

(28) Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating. : Springer Science & Business Media; 2008.

(29) Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, et al. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. BMJ 2013 Feb 5;346:e5595.

(30) Bellazzi R, Zupan B. Predictive data mining in clinical medicine: current issues and guidelines. Int J Med Inf 2008;77(2):81-97.

(31) Yoo I, Alafaireet P, Marinov M, Pena-Hernandez K, Gopidi R, Chang J, et al. Data mining in healthcare and biomedicine: a survey of the literature. J Med Syst 2012;36(4):2431-2448.

(32) Jenny MA, Hertwig R, Ackermann S, Messmer AS, Karakoumis J, Nickel CH, et al. Are Mortality and Acute Morbidity in Patients Presenting With Nonspecific Complaints Predictable Using Routine Variables? Acad Emerg Med 2015;22(10):1155-1163.

(33) Austin PC, Tu JV, Ho JE, Levy D, Lee DS. Using methods from the data-mining and machine-learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes. J Clin Epidemiol 2013;66(4):398-407.

(34) Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. Computational and structural biotechnology journal 2015;13:8-17.

(35) Kerr KF, Meisner A, Thiessen-Philbrook H, Coca SG, Parikh CR. Developing risk prediction models for kidney injury and assessing incremental value for novel biomarkers. Clin J Am Soc Nephrol 2014 Aug 7;9(8):1488-1496.

(36) KDIGO A. Work Group: KDIGO clinical practice guideline for acute kidney injury. Kidney Int Suppl 2012;2(1):1-138.

(37) Genuer R, Poggi J, Tuleau-Malot C. Variable selection using random forests. Pattern Recog Lett 2010;31(14):2225-2236.

(38) Khalilia M, Chakraborty S, Popescu M. Predicting disease risks from highly imbalanced data using random forest. BMC Med Inform Decis Mak 2011 Jul 29;11:51-6947-11-51.

(39) Lee PH. Resampling methods improve the predictive power of modeling in class-imbalanced datasets. International journal of environmental research and public health 2014;11(9):9776-9789.

(40) Milborrow S, Hastie T, Tibshirani R. earth: Multivariate Adaptive Regression Spline Models. R package version 3.2-7 2014.

(41) Breiman L. randomForest: Breiman and Cutler's random forests for classification and regression. R package version 4.6-12 2015.

(42) Dimitriadou E, Hornik K, Leisch F, Meyer D, Weingessel A, Leisch MF. The e1071 package. Misc Functions of Department of Statistics (e1071), TU Wien. R package version 1.6-7 2015.

(43) Basu RK, Zappitelli M, Brunner L, Wang Y, Wong HR, Chawla LS, et al. Derivation and validation of the renal angina index to improve the prediction of acute kidney injury in critically ill children. Kidney Int 2014 Mar;85(3):659-667.

(44) Basu RK, Kaddourah A, Terrell T, Mottes T, Arnold P, Jacobs J, et al. Assessment of Worldwide Acute Kidney Injury, Renal Angina and Epidemiology in critically ill children (AWARE): study protocol for a prospective observational study. BMC Nephrol 2015 Feb 26;16:24-015-0016-6.

(45) Grossman C, McGinnis JM. Digital Infrastructure for the Learning Health System:: The Foundation for Continuous Improvement in Health and Health Care: Workshop Series Summary. : National Academies Press; 2011.

(46) Smith M, Saunders R, Stuckhardt L, McGinnis JM. Best care at lower cost: the path to continuously learning health care in America. : National Academies Press; 2013.

(47) Zuckerman AE. The role of health information technology in quality improvement in pediatrics. Pediatr Clin North Am 2009 Aug;56(4):965-973.

(48) Institute of Medicine . Committee on Quality of Health Care in America. Crossing the quality chasm: a new health system for the 21st century. Washington, D.C.: National Academy Press; 2001.

(49) Bright TJ, Wong A, Dhurjati R, Bristow E, Bastian L, Coeytaux RR, et al. Effect of clinical decision-support systemsa systematic review. Ann Intern Med 2012;157(1):29-43.

(50) Sittig DF, Wright A, Osheroff JA, Middleton B, Teich JM, Ash JS, et al. Grand challenges in clinical decision support. J Biomed Inform 2008;41(2):387-392.

(51) Bates DW, Kuperman GJ, Wang S, Gandhi T, Kittler A, Volk L, et al. Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. J Am Med Inform Assoc 2003 Nov-Dec;10(6):523-530.