ACTA BIOMEDICA LOVANIENSIA 812
KU Leuven
Biomedical Sciences Group
Faculty of Medicine
Department of Cellular and Molecular Medicine
Laboratory of Intensive Care Medicine

Chao-Yuan HUANG

Computerized Decision Support to Assess the Kidney Function in Critically Ill Patients

Leuven University Press

Thesis submitted in partial fulfillment of the requirements for the degree of «Doctor of Biomedical Sciences».

Jury:

Supervisor:	Prof. Dr. Geert Meyfroidt
Co-Supervisors:	Dr. Fabian Güiza Grandas
	Prof. Dr. Greet De Vlieger
Chair examining committee:	Prof. Dr. Koen Milisen
Chair public defense:	Prof. Dr. Frank Claessens
Internal jury:	Prof. Dr. Björn Meijers
	Prof. Dr. Celine Vens
External jury:	Prof. Dr. Eric Hoste
	Prof. Dr. Marlies Ostermann

©2023 by Leuven University Press / Presses Universitaires de Louvain / Universitaire Pers Leuven.
Minderbroedersstraat 4 - bus 5602, B-3000 Leuven (Belgium)

All rights reserved. Except in those cases expressly determined by law, no part of this publication may be multiplied, saved in an automated data file or made public in any way whatsoever without the express prior written consent of the publishers.

ISBN 978 94 6165 353 6
D/2023/1869/15

Contents

List of abbreviations

A	
AKI	Acute kidney injury
AKI-3	Acute kidney injury stage 3
APACHE	Acute Physiology and Chronic Health Evaluation
ARC	Augmented renal clearance
AUPRC	Area under the Precision-recall curve
AUROC	Area under the receiver operating characteristic curve
B	
BGA	Blood gas analysis
BiVAD	Biventricular assist device
BMI	Body mass index
C	
CART	Classification and regression tree
CG	Cockcroft-Gault
CI	Confidence interval
CITL	Calibration-in-the-large
CKD	Chronic kidney disease
CKD-EPI	Chronic Kidney Disease Epidemiology Collaboration
COVID-19	Coronavirus disease 19
CPU	Central processing unit
CrCl	Creatinine clearance
CS	Calibration slope
E	
ECG	Electrocardiogram
ECMO	Extracorporeal membrane oxygenation
EEG	Electroencephalogram
eGFR	estimated Glomerular filtration rate
EHR	Electronic health record
EPaNIC	Early versus Late Parenteral Nutrition in Critically Ill Adults

ESKD	End-stage kidney disease
F	
FFT	Fast-Fourier Transform
FN	False negative
FP	False positive
FPR	False positive rate
FNR	False negative rate
FWO	Fonds Wetenschappelijk Onderzoek
G	
GBM	Gradient boosting method
GFR	Glomerular filtration rate
H	
HDF5	Hierarchical Data Format—Version 5
I	
IABP	Intra-aortic balloon pump
ICU	Intensive care unit
IQR	Interquartile range
K	
KDIGO	Kidney Disease: Improving Global Outcomes
KRT	Kidney replacement therapy
KU Leuven	Katholieke Universiteit Leuven
L	
Lasso	Least absolute shrinkage and selection operator
LR+	Positive likelihood ratio
LR-	Negative likelihood ratio
LIME	Local interpretable model-agnostic explanations
LOS	Length of stay
LVAD	Left ventricular assist device
M	
MAE	Mean absolute error
MD5	Message Digest Method 5

MDRD	Modification of Diet in Renal Disease
MSSQL	Microsoft structured query language
MV	Mechanical ventilation
MySQL	My structured query language
N	
NGAL	Neutrophil gelatinase-associated lipocalin
NGAL_AKI3	NGAL measured on the first day of AKI stage 3
NPV	Negative predictive value
NYHA	New York Heart Association
P	
PCR	Polymerase chain reaction
PDMS	Patient data management system
PEEP	Positive end-expiratory pressure
PK	Pharmacokinetic
PPV	Positive predictive value
PRC	Precision-recall curve
R	
RCT	Randomized controlled trial
RMSE	Root-mean-square error
ROC	Receiver operating characteristic
S	
SCr	Serum creatinine
SHA-1	Secure Hash Algorithm 1
SHA-256	Secure Hash Algorithm 256
SHAP	Shapley Additive exPlanations
SMOTE	Synthetic minority oversampling method
SOFA	Sequential Organ Failure Assessment
SQL	Structured query language
T	
TN	True negative
TNR	True negative rate

TP	True positive
TPR	True positive rate
TRIPOD	Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis
U	
UCr	Urine creatinine
UO	Urine output
USA	United States of America
UZ Leuven	University Hospitals Leuven

1

Introduction

1.1 Critical illness

Patients who suffer from severe illness or life-threatening conditions that require intensive follow-up and/or specific treatment are admitted to intensive care units (ICUs). There are millions of patients admitted to ICUs every year worldwide. Conditions leading to ICU admission include sepsis, major surgery, severe burns, serious trauma, heart failure, etc. Given the deadly impact of these diseases, many critically ill patients do not survive, with variable rates of ICU mortality depending on the chronic health and acute disease [1]. In addition, patients who survive the acute phase often suffer from prolonged consequences of their critical episode, hallmarked by vulnerability to infections, prolonged organ dysfunction, muscle weakness, and cognitive impairment, all leading to a decreased quality of life. In the ICU, a group of professionals including intensivists, specialized nurses, physiotherapists, and pharmacists take care of these patients to optimize their organ functions, allow the patients to survive their critical condition, and improve long-term outcome. ICUs are equipped with monitors to assess and register vital organ functions on a continuous basis to facilitate the treatment of these patients. Support of vital organ systems through effective drugs or mechanical devices is often necessary, including inotropic drugs or vasopressors for cardiovascular support, mechanical ventilation, extracorporeal membrane oxygenation, or kidney replacement therapy.

1.2 Kidney function during critical illness

1.2.1 Acute kidney injury

Acute kidney injury (AKI) is defined by a sudden decrease of renal excretory function leading to an acute aggregation of wastes and disturbance of the fluid and electrolyte balance of the body. Patients with AKI show different symptoms depending on the underlying cause and may suffer from oliguria, swelling, tiredness, nausea, coma, vomiting and even cardiac arrhythmia. Elderly people with diabetes mellitus, heart failure, and/or anemia are at risk for developing AKI.

AKI is a disease that is caused by many different pathophysiological pathways. In the ICU, the most frequent etiologies for AKI are sepsis and hypovolemia [2]. Other possible causes include cardiac failure, hepatorenal syndrome, interstitial nephritis, glomerulonephritis or postrenal obstruction of the urinary tract.

In 2012, the Kidney Disease: Improving Global Outcomes (KDIGO) was proposed with an aim to unify AKI diagnostic criteria in 2012 [3]. Depending on the level of increased serum creatinine from its baseline and the amount of urinary output volume, KDIGO guidelines classify AKI into three stages of increasing severity (**Table 1.1**).

Table 1.1 AKI definition according to KDIGO criteria

Stage	Serum creatinine criteria	Urine output criteria
1	Increase by 1.5 – 1.9 times baseline within 7 days OR Increase by ≥ 0.3 mg/dL (26.5 μmol/L) within 48 hours	Less than 0.5 mL/kg/h for 6 – 12 hours
2	Increase by 2 – 2.9 times baseline	Less than 0.5 mL/kg/h for ≥ 12 hours
3	Increase by ≥ 3 times baseline OR Increase to ≥ 4 mg/dL (353.6 μmol/L) OR Kidney replacement therapy initiation OR In patients < 18 years, decrease in estimated GFR to <35 mL/min/1.73m^2	Less than 0.3 mL/kg/h for ≥ 24 hours OR Anuria for ≥ 12 hours

Abbreviations: KDIGO, Kidney Disease: Improving Global Outcomes; AKI, acute kidney injury; GFR, glomerular filtration rate.

AKI is important, since it is one of the most prevalent complications in critically ill patients, with prevalence varying between 20-57% depending on the patient population and use of AKI definition criteria [2, 4, 5]. Furthermore, AKI is associated with a worse outcome, with ICU mortality greater than 50% [4] and association of four-fold to six-fold increased mortality than the general inpatient population [6, 7]. Additionally, AKI is associated with a longer length of stay [7, 8], higher financial cost [7, 9], and increased risk of chronic kidney disease and kidney failure [10, 11].

Current treatments for AKI are mainly preventive and supportive with hemodynamic management, treatment of infections, avoiding hyperglycemia and nephrotoxic medications, and replacing the kidney function until recovery in case of kidney failure. As AKI is associated with a worse outcome and treatment is mainly preventive, early detection of AKI may avoid possible complications, increase the possibility of kidney function recovery, and ultimately improve patient outcome and quality of life.

For critically ill patients, a variety of prediction models for AKI have been proposed [12–20]. Based on the large multi-center randomized controlled trial EPaNIC database [21], Flechet et al. developed the AKIpredictor, a series of models to predict AKI using advanced machine learning techniques [20]. The different AKIpredictor models each use a different set of features, as they become available at different time points during the clinical course: before admission, upon admission, on the first morning after admission, and after 24 hours. The models can predict AKI at any stage (1-2-3) or only the most severe stages (2-3). The AKIpredictor is available as an online calculator [22]. The accuracy of AKIpredictor was first assessed in a validation cohort, where it demonstrated remarkable results and outperformed the commonly used AKI biomarker serum Neutrophil gelatinase-associated lipocalin (NGAL) [20]. In addition, the AKIpredictor was able to predict severe AKI (defined as AKI stage 2 or 3 based on the KDIGO serum creatinine criteria and/or urinary output criteria)

with a similar discrimination but a higher net benefit as compared to physicians [23]. Even while the AKIpredictor has demonstrated a robust performance in retrospective and prospective patient cohorts, further external validation in different clinical settings would still be useful to demonstrate external validity.

Once AKI has occurred, the evolution of AKI and its recovery is also of importance as it is strongly associated with mortality, progressive renal dysfunction [24] and end-stage kidney disease [25]. Therefore, accurate prediction of the course of AKI recovery may improve general ICU and post-ICU care.

The ability of several kidney biomarkers to predict AKI recovery has been disappointing [26–34], which may be partially explained by the heterogeneous patient population with different pathophysiological causes of AKI and the timing of recovery definition. With the growth of electronic health records, several clinical prediction models are proposed for patients with AKI. However, the majority of them prioritized mortality prediction rather than renal recovery prediction [35–42]. Although clinical prediction model may be a promising tool, large databases frequently lack a good AKI recovery evaluation as the recovery definition is not uniform, and baseline serum creatinine is frequently unknown [43], which poses difficulties for the development of AKI recovery prediction models.

Nevertheless, there were two studies with special focus on developing AKI recovery prediction models [44, 45]. In particular, Itenov et al. developed and validated an AKI recovery prediction model in critically ill patients with AKI on ICU admission by using the Cox regression models [44]. The model had a fair predictive performance which may partially be explained by the fact that AKI recovery was defined within a pre-specified timepoint of 28 days, while some patients might still be undergoing recovery from critical illness in the ICU. In addition, they only investigated patients with AKI on ICU admission, but AKI may happen unexpectedly at any time during the ICU stay. In the other study, Lee et al. proposed an AKI recovery prediction model for dialysis-requiring AKI patients by using logistic regression and classification and regression tree (CART) [45]. However, there were some limitations. First, only poor discrimination was demonstrated, which may not meet the high standard of clinical practice. Second, they defined kidney recovery as kidney replacement therapy independence, but even AKI patients without KRT are associated with unfavorable outcomes [2]. Finally, no external validation was performed, which might lead to over-optimistic results. Given the shortcomings of currently available AKI recovery prediction models, there is a need for a better AKI recovery prediction model.

1.2.2 Augmented renal clearance

In contrast to AKI, augmented renal clearance (ARC) is at the other end of the renal clearance spectrum, where patients may have high renal clearance which is frequently induced by a stress-induced hyperdynamic state. Although there is no universally accepted ARC definition, it is commonly described as having a measured urinary creatinine clearance (CrCl) greater than 130 ml/min/1.73m^2. Critically ill patients frequently have ARC, with prevalence ranging between 20 and 65 percent [46]. Even though many studies reported no major difference in the clinical outcome or mortality between patients with and without ARC [47–52], it has been demonstrated that ARC may result in less exposure to drugs that are often prescribed, including beta-lactams, vancomycin, and anticoagulants [53, 54]. Additionally, in order to increase exposure and lower the chance of treatment failure, it has also been suggested that antibiotic doses should be raised in patients with ARC [55].

It is currently unclear what physiological mechanism causes ARC in critically ill patients. Systemic inflammatory response syndrome (SIRS) and kidney functional reserve (KFR) are the two primary hypotheses

underlying the emergence of ARC [46, 56]. The first theory suggests that the cytokines and pro-inflammatory mediators released during SIRS, coupled with intensive fluid therapy and inotropic drugs, enhance cardiac output and renal blood flow and elevate glomerular filtration rate (GFR). The second theory proposes that KFR may play a role. KFR is the term used to describe the kidney's ability to enhance GFR in response to specific physiological or pathological stimuli, such as pregnancy and high-protein diet [57]. Instead of a single mechanism, the combination of SIRS and enhanced KFR is believed by many researchers to be a possible cause of ARC [46].

There are some strategies for managing ARC patients. First, to make up for the increased excretion, the dosage of renally excreted drugs might need to be raised, by giving a higher dose, shortening the dosage interval or increasing the total duration of therapy. In addition, therapeutic drug monitoring can be routinely performed to adjust regimens when needed. Finally, another option is to switch to a medication that is not primarily excreted by the kidneys.

Age, gender, surgery/trauma/neuro-related diagnoses, and illness severity score are some of the variables that can screen individuals for the risk of ARC. Several studies [47, 49–51, 55, 58–66] consistently demonstrated a strong relationship between age and ARC. In addition, a significant correlation between ARC and male gender was seen in a number of studies [47, 50, 55, 58, 60–62, 64, 66]. Some research [50, 58, 60, 62, 63] have identified an association between ARC and surgery/trauma/neuro-related diagnoses. Finally, some studies [51, 63, 66] reported that a lower severity score was a risk factor for ARC.

Due to the high prevalence and strong association with adverse consequences in critically ill patients, it is important to predict the onset of ARC. Therefore, several studies [61, 63, 67, 68] have been conducted to predict the onset of ARC, but most of them were based on a small and selected subgroup of critically ill patients and not externally validated. Additionally, they predicted the onset of ARC once for the entire ICU stay, while some patients may develop ARC intermittently during their ICU stay. To address the abovementioned shortcomings, using the large multi-center M@tric database [69], Gijsen et al. developed a tool to predict the onset of ARC on the next ICU day named “ARC predictor” [62], by employing a generalized estimating equation (GEE) logistic regression algorithm and backward feature selection method. Six routinely collected clinical variables are included in the ARC predictor: age, sex, day from ICU admission, serum creatinine of the previous day, trauma related diagnosis on ICU admission (True/False), and cardiac surgery related diagnosis on ICU admission (True/False). Based on the given six features, the ARC predictor generates the predicted probability of ARC, which is then converted into a prediction for ARC on the following ICU day with a predetermined classification threshold. Although the threshold is by default set at 20% since this value maximized sensitivity and specificity in the original study, it can be manually altered.

In the validation cohort, the ARC predictor outperformed the two previous reference models (ARC Score [68] and ARCTIC Score [61]) with good discrimination and calibration and a broad clinical usefulness range. The ARC predictor is now accessible to the general public as an online predictor (**Figure 1.1**) [70]. Although the ARC predictor demonstrated significant promise in the original development study, additional external validation in an independent study cohort is required before it can be used in clinical practice.

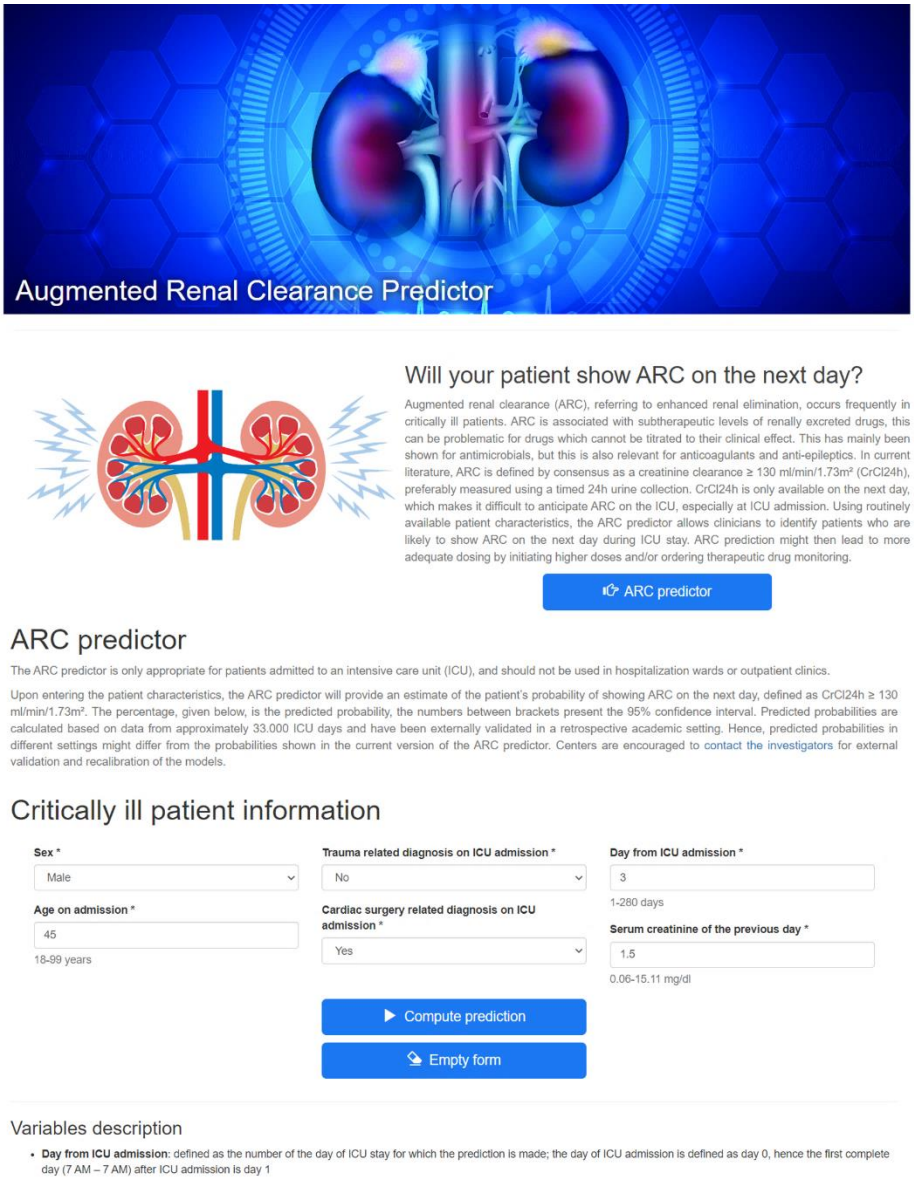


Figure 1.1 Screenshot of the ARC predictor website, www.arcpredictor.com

1.2.3 Creatinine clearance

Daily evaluation of kidney function in the ICU is routinely performed because of the high prevalence of AKI and ARC in critically ill patients. The glomerular filtration rate (GFR) is a quantification of the filtration function of kidney. It reflects the volume of fluid that is filtered through the kidneys per unit of time. In clinical practice, GFR is usually estimated based upon patients characteristics and serum creatinine by the Cockcroft-Gault equation [71], the Modification of Diet in Renal Disease (MDRD) study equation [72], or the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation [73] (**Table 1.2**). However, these frequently used formulas were not developed in ICU setting, and it is well known that they are not appropriate for accurate estimations of the kidney function of critically ill patients [52, 64, 74–76], especially in those patients with prolonged ICU stay [77].

Table 1.2 Equations kidney function estimation

measured CrCl (ml/min/1.73m^2) = $UCr(mg/dL) \times UO(ml/day)/SCr\ (mg/dL)/1440(min/day) \times 1.73/(0.007184 \times height(cm)^{0.725} \times weight(kg)^{0.425})$
$eCrCl_{CG}$ (ml/min) = $[(140 - age(years)) \times weight\ (kg)]/SCr(mg/dL) \times 72) \times 0.85(if\ female)$
$eGFR_{MDRD}$ (ml/min/1.73m^2) = $175 \times SCr(mg/dL)^{-1.154} \times age\ (years)^{-0.203} \times 0.742\ (if\ female) \times 1.212\ (if\ black)$
$eGFR_{CKD\text{-}EPI}$ (ml/min/1.73m^2) = $141 \times min(SCr(mg/dL)/\kappa, 1)^{\alpha} \times max(SCr(mg/dL)/\kappa, 1)^{-1.209} \times 0.993^{age(years)} \times 1.018\ (if\ female) \times 1.159\ (if\ black)$

CrCl, creatinine clearance; UCr, urine creatinine; UO, urine output; SCr, serum creatinine; $eCrCl_{CG}$, creatinine clearance estimated by the 'Cockcroft-Gault' equation; $eGFR_{MDRD}$, glomerular filtration rate estimated by the '4-variable Modification of Diet in Renal Disease' equation; $eGFR_{CKD\text{-}EPI}$, glomerular filtration rate estimated by the 'Chronic Kidney Disease Epidemiology Collaboration' equation; α = -0.329 if female and -0.411 if male; κ = 0.7 if female and 0.9 if male; min$(SCr(mg/dL)/\kappa, 1)$, minimum between $SCr(mg/dL)/\kappa$ and 1; max$(SCr(mg/dL)/\kappa, 1)$, maximum between $SCr(mg/dL)/\kappa$ and 1

Therefore, it is preferable to assess kidney function by measurement of GFR rather than using estimations. While adequate tools to measure real-time GFR are currently not available, GFR is best determined by urinary clearance of filtration markers. Inulin is the gold standard of exogeneous filtration marker, since inulin is not secreted, reabsorbed, nor metabolized by the kidney [78]. However, the usage of inulin is constrained by its price and lack of availability.

Creatinine is a suitable endogenous filtration marker, since it is neither reabsorbed nor metabolized by the kidney and only minimally excreted by the tubules. Creatinine clearance (CrCl) can be calculated with urine output, serum creatinine, and urinary creatinine (measured on 8- to 24-hour sample [79]). However, the need of several hours for CrCl calculation hinders quick decision-making on renally excreted drug administration. Additionally, by the time urinary CrCl is available, the calculation has already fallen behind the actual kidney function as it reflects the CrCl during the previous interval and is less reliable because the kidney function is known to vary quickly in critically ill patients [80]. Therefore, kidney function prediction may enable more appropriate treatment approaches by providing a more precise evaluation of kidney function.

Several studies developed kidney function prediction models with a focus on predicting AKI and ARC. However, both AKI and ARC are based on categorized definitions while the kidney function is a continuous parameter. Predicting CrCl for the entire kidney function spectrum is more in accordance with clinical and physiological reality. To the best of our knowledge, no prediction models for daily prediction of CrCl have been proposed.

1.3 Big data and machine-learning

Over the past decades, ICUs have been computerized through the implementation of electronic health records (EHR). Advancements in storage capacity, computation ability, and global internet, have definitely promoted the roll-out of these systems worldwide, and have gradually reduced the implementation cost [81].

In the first place, EHRs are used to optimize individual patient care as they can efficiently present the latest, accurate, and comprehensive information about patients to healthcare providers. In addition, EHRs make it easier to efficiently share data across multiple health care organizations in multi-center studies. Finally, EHRs generate sufficient amounts of patient data that are necessary to find associations with adequate statistical power.

The amount of clinical data gathered from ICU patients is huge. All vital organs are continuously monitored in a high-resolution fashion, and many EHRs also record the type and dosage of administered drugs, laboratory results, clinical notes. These result in large 'volume' of a wide 'variety' of data continuously accumulating at a high 'velocity' in ICUs. The three Vs comprise the key concepts of 'big data', which refers to datasets that are too large or complex to be handled by traditional data-processing software. Big data must add 'value' by generating novel insights and revealing undiscovered patterns in order to be useful. In addition, data 'veracity' is of equal importance. If the source of data is contaminated, research findings can be misleading. As a result, these last two Vs have been recently added to the big data definition [82]. In particular, these data appear to be extremely useful to build predictive models for a variety of conditions and outcomes [83].

Having big data is not enough; we also need advanced analytic capabilities and artificial intelligence to derive useful insights and valuable information from millions of data points. Artificial intelligence refers to techniques that aim to mimic a human's behavior regardless of the methods, including simple rule-based if-else conditions, advanced machine-learning algorithms, and powerful deep-learning techniques. Machine learning is a specific field of artificial intelligence where the computer system learns based upon previous examples. Next, it infers from what it has learned, and predicts for previously unseen conditions based on the learned mechanism. Machine learning is used in various fields such as physiologic waveform analysis, image analysis, and natural language processing.

In order to predict a certain event or outcome, supervised machine learning can be used. In this technique, a model learns the relationship between the given input data and one or more target outcomes. Target outcomes must be well indicated (often called labeled). In medicine, supervised learning can be used to identify people who are more likely to develop a disease and as such would benefit the most from specialized medical care. Examples of supervised algorithms include linear models (e.g., logistic regression, lasso, ridge) and tree-based algorithms (such as decision tree, random forest, and gradient boosting method).

According to the property of prediction target, supervised learning can be further divided into two subtypes: classification and regression. In a classification task, a model is trained to determine which category the new patient falls into. For example, if researchers want to know whether acute kidney injury (AKI) can be predicted by several routinely collected clinical data, classification algorithms can be applied to a dataset containing clinical data of interest and a label indicating the presence of AKI ('with AKI' or 'without AKI'). Regression can be used to construct prediction models for a continuous outcome. For instance, in a study to predict creatinine clearance (CrCl) of the next ICU day for each critically ill patient, regression algorithms can be utilized on a dataset including relevant clinical data and a label indicating the exact creatinine clearance on the next ICU day.

1.4 Development of machine-learning models in ICU

1.4.1 Feature engineering

Regardless of prediction tasks, it is necessary to include relevant data, and it is believed that meaningful data can produce better predictive performance than raw data. The process of extracting useful information from the raw data by using the domain knowledge is called feature engineering. The feature engineering process can involve basic derivatives or sophisticated data transformations. The slope of a linear regression model, for example, can be used to identify data trends, where different temporal window sizes could be chosen to provide a thorough insight about the patient's physiological stability. Complex transformation such as Fast-Fourier Transform (FFT) [84], which provides frequency information of the raw data by decomposing a signal into a combination of several spectral components, has also showed success in audio, video, and electrocardiogram (ECG) signal processing [85]. As an extension from the FFT, cepstrum analysis [86] describing the changing rate of information in different frequency bands has gained success in a variety of fields including speech recognition, radar and sonar applications, and analysis of electroencephalogram (EEG).

1.4.2 Feature selection

Having abundant features may not be helpful [87], and it is necessary to perform feature selection, which refers to a process to reduce the number of features by removing noisy, redundant, or less important features while keeping the most important features during the model development stage. Feature selection is one of the most important procedures in machine learning, since it significantly and directly influences the predictive performance. Feature selection is especially important in medicine when the incidence of a disease is low, and each disease presentation may have abundant features. This unique condition makes model development prone to overfitting, which refers to the situation where a prediction model performs well on the data used for model development and yields poor performance on unseen data. Additionally, feature selection has a profound impact on clinical utility. Specifically, compared with a powerful model depending on several monitoring data with difficult calculations, a relatively less powerful model that is only based on simple features with high-availability data and trivial calculations may be preferred as it is much easier to implement in the clinical practice. Furthermore, feature selection changes the way of interpreting prediction results. A successful feature selection sheds light on the most relevant features and bring novel insights, while a poor feature selection results in a prediction model containing unreasonable features that clinicians are reluctant to utilize.

Two types of feature selection methods are frequently used: intrinsic (or embedded) and wrapper methods [88, 89]. Intrinsic methods automatically select features during the model training phase. There are two types of regularization: L1 (Lasso) and L2 (Ridge) regularizations. L1 regularization has a stronger tendency to turn coefficients to zero and to create sparse models [90]. Tree-based models automatically search for the most important features via their contributions to predictive performance or the decrease in uncertainty (impurity) over all the trees. Therefore, the most important features can be chosen by looking at how frequently a feature is used. Intrinsic methods are preferred in cases where the dataset is small.

Wrapper methods pick the most important features by trying different combinations of feature subsets for a specific machine-learning algorithm on a given dataset. These methods are computationally expensive and prone to overfitting, but they usually demonstrate better performance than filter methods. Backward feature

selection is one of the wrapper methods, where it starts with a full feature subset. At each step, the feature that causes the least reduction in performance after removal is discarded from the feature subset. The removal process is also repeated until the stopping criteria are met. Wrapper methods typically show better performance given a large dataset [91].

Despite the efficiency and simplicity of automatic feature selection methods, one should not solely rely on them. Integration of the knowledge from experts in the field can largely reduce the computation time and increase the chance of building a robust model. For instance, there may be no logical justification to include a feature in the model, even when it is coincidentally very predictive for a prediction task. In this scenario, spurious correlation will successfully deceive automatic feature selection methods, while domain experts will recognize the lack of rationality.

1.4.3 Machine-learning algorithms

The method for feature selection has to be considered together with the choice of machine-learning algorithm. There are various machine-learning algorithms that all have their advantages and disadvantages. There is no single algorithm that consistently outperforms the others. Consequently, the choice of algorithm depends on the research question, input and output data characteristics, and training and running time requirement. The methods relevant to or employed in this thesis are listed below, and they were categorized into two subtypes: linear and tree-based algorithms.

1.4.3.1 Linear models

1.4.3.1.1 Logistic regression

Logistic regression is the most commonly used machine-learning algorithm for binary medical outcomes due to its simplicity [92]. As indicated in its name, it utilizes a 'logistic function' (or logit) to convert the probability to log odds (**Figure 1.2**). Logistic regression is used to estimate the relationship between a dependent categorical variable and one or several independent variable(s) (in nominal, ordinal, or continuous formats). The logistic regression can be represented in a mathematical way as follows:

Given N independent variables x_i, the predicted probability of having an event $p(y = 1)$ is

$$p(y = 1) = sigmoid(\sum_{i=1}^{N} \alpha_i \times x_i + \beta)$$

where α_i represents the coefficient for the corresponding independent x_i, β is known as the intercept, $sigmoid(x) = \frac{1}{1+e^{-x}}$, and e is the mathematical constant

Logistic regression can also be rewritten with logit function and log-odds instead of sigmoid function and probability: given N independent variables x_i, the log-odds for having an event is

$$\ln\left(\frac{p(y = 1)}{1 - p(y = 1)}\right) = \sum_{i=1}^{N} \alpha_i \times x_i + \beta$$

where α_i represents the coefficient for the corresponding independent x_i, β is known as the intercept, $\ln(x)$ is the natural logarithm with mathematical constant e as its base.

The α_i describes the effect on the prediction, when having a 1-unit increase in the corresponding independent variables x_i, while keeping the remaining variables constant. When there are multiple independent variables included in the analysis, we call this 'adjusted', and the corresponding coefficient α_i is the multivariate effect, given the effect of other independent variables. The odds-ratio of an independent variable x_i can be calculated by applying the exponential function to its corresponding coefficient (e^{α_i}).

Logistic regression has strengths. First, logistic regression is easy to implement, which can save resources including the human resources to maintain and to integrate the model into applications and reduce the time needed to make a prediction which plays a crucial role in real-time applications. Second, logistic regression is easy to interpret while other machine-learning models have been known for their 'black-box' characteristics.

Nevertheless, logistic regression has some limitations. First, as a linear model, logistic regression does not automatically account for the non-linear functions between the variables. Since it is the inherent limitation of logistic regression, the additional terms must be introduced explicitly, or more advanced machine-learning algorithms should be considered if non-linear relationship needs to be described. Second, logistic regression requires the absence of multicollinearity in the dataset. Multicollinearity means that there are highly correlated independent variables in the dataset. Violating this assumption gives wrong interpretation about the coefficients, weakens the statistical power, and results in less trustworthy p-values. Third, the independence of observations is still another crucial assumption that logistic regression requires. In other words, observations must be independent from one another and cannot be matched data or repeated measurements.

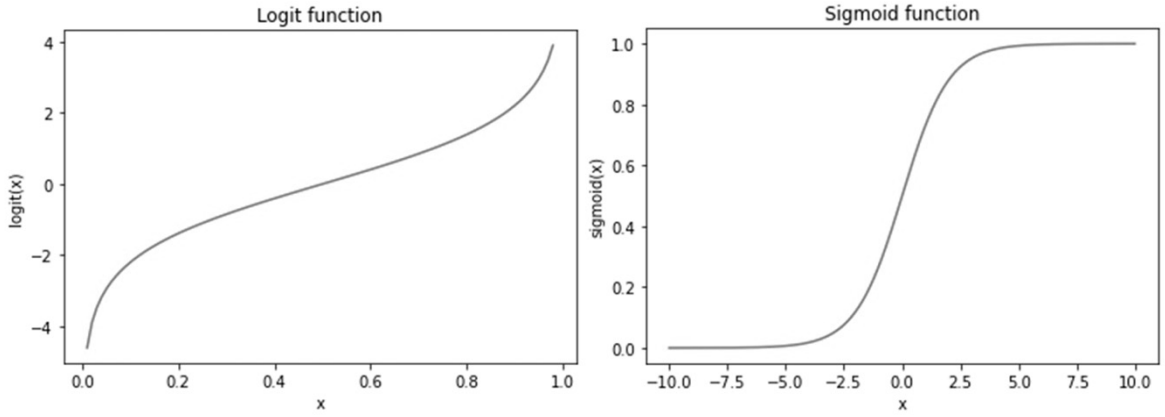


Figure 1.2 Visualization of the logit function (left) and sigmoid function (right)

1.4.3.2 Tree-based algorithms

Tree-based algorithms are based on a series of conditional rules (Yes/No) to approximate the outcome of interest. They can be used for both classification and regression. Additionally, tree-based algorithms can handle multi-output problems where the data of a patient is used as input and two or more prediction outputs are generated. Finally, by the virtue of a series of conditional rules, tree-based algorithms can depict a non-linear relationship. In this thesis, we focus on three algorithms: decision tree, random forest, and gradient boosting method.

1.4.3.2.1 Decision tree

Decision tree is an algorithm that has only one tree-and resembles a human decision-making process (**Figure 1.3**). A decision tree starts at a node (root) where a question needs to be answered before the data can be split into two branches (sizes are not necessarily the same). After the answer is given and the data is split, another

question is asked, and the already-split data is further decomposed into two parts. This procedure is repeated continuously until the stopping criteria are met. Finally, the input data will arrive at a specific terminal node (also called leaf) with corresponding predicted value(s), which could be predicted probability or continuous predicted outcome of interest for classification and regression tasks respectively.

Although a decision tree can be easily explained with visualization of the entire tree structure, it has some limitations. First, it frequently produces a tree that is extremely complex and cannot be generalized when tested on other datasets. To overcome this overfitting problem, some mitigations can be considered, such as selecting the most important features in advance, pruning the tree, limiting the minimum samples in each node, or specifying the maximum depth of the tree. Second, the fact that there is only one tree contributes to the instability of the decision tree. An entirely different decision tree could be produced even with a minor variation in the dataset. Third, a decision tree skewed toward the dominant class may be produced by a dataset with unequal numbers of cases from various classes. This issue can be solved by under-sampling the majority class, oversampling the minority class, or by penalizing minority-group errors more severely.

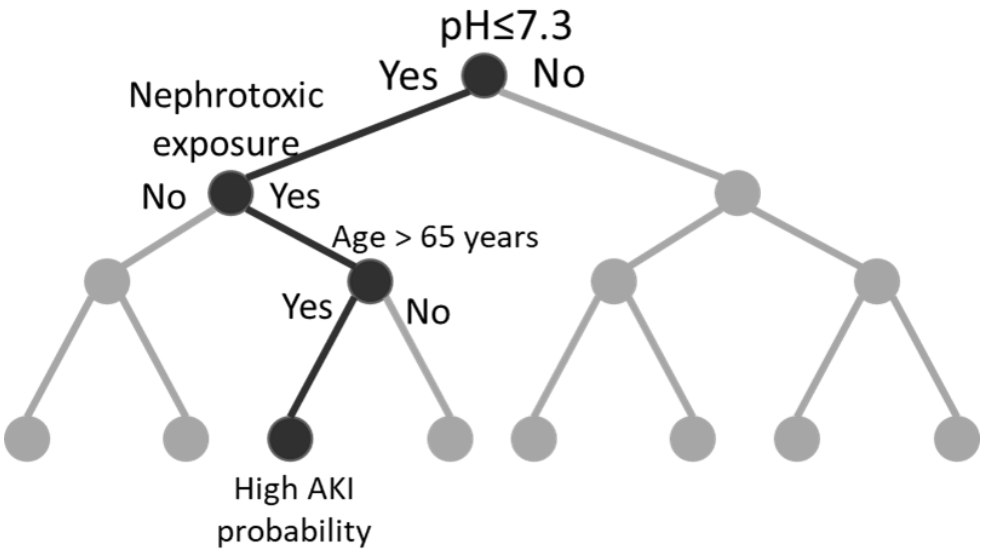


Figure 1.3 Example of a decision tree to predict the risk of acute kidney injury. AKI, acute kidney injury

1.4.3.2.2 Random forest

Random forest [93, 94] is a tree-based ensemble method. Ensemble methods are based on a concept that better performance can be achieved by combining the results from multiple models (often called base learners, as indicated in **Figure 1.4**) [95]. Given the tree-based nature, random forests can handle both classification and regression tasks. Instead of building multiple decision trees sequentially, random forest builds all trees simultaneously with bagging method. Bagging, bootstrapping or bootstrap aggregating, is a technique that randomly samples with replacement from the training set [96]. While some data may be sampled more than once using this technique, it is guaranteed that one sampling will not affect the others (mathematically, the covariance between any two samplings is zero), allowing each tree to grow independently from the others. Additionally,

since the sample size is maintained by recycling the utilized examples, the bagging approach reduces the issue of a dataset with a small population, which is a common circumstance in most clinical investigations.

Random forest has many strengths. First, it is resistant to outliers because it averages out the effects of extreme values locally. Second, it is effective at representing non-linear relationships. Third, it has low risk of overfitting because each tree overfit data in a different way, therefore the mistakes of one tree are compensated by the other trees. Finally, given that there are no intricate mathematical computations involved in the prediction procedure, a random forest can perform efficiently during the testing phase (only a series of conditional questions needs to be answered).

Random forest has some limitations. First, a random forest is computationally expensive and requires a lot of calculation and training time because each tree can grow deeply and there are many of them. This is especially true when there are lots of high dimensional training data. Second, continuous variables and categorical variables with a large number of distinct values tend to be preferred by random forest. Third, random forest is less interpretable than decision tree.

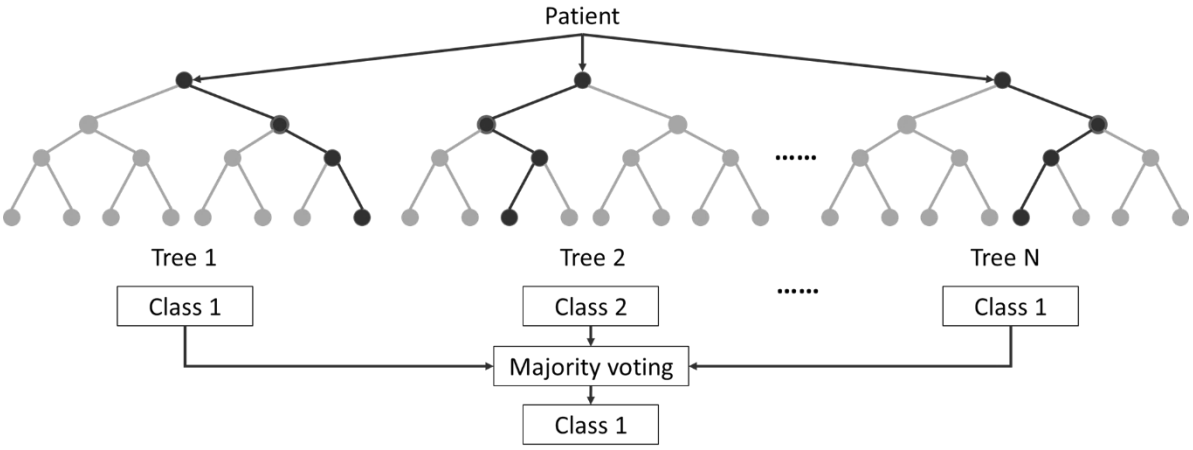


Figure 1.4 Example of a random forest to classify the new patient

1.4.3.2.3 Gradient boosting trees

As opposed to random forest, gradient-boosting trees [97] grow trees sequentially (**Figure 1.5**). The boosting approach is founded on the notion that each new tree is constructed to remedy the errors of its predecessors [98]. Based on the gradient descent procedure, which is an optimization technique that iteratively proceeds in the direction of steepest descent when reducing the loss function, the gradient-boosting trees algorithm corrects the errors when adding new trees. In the end, heavy weights (more focus) are given on difficult examples, and examples that are easy to predict receive small weights (less focus).

The gradient-boosting trees algorithm has many strengths. First, it is a tree-based ensemble method that can handle both classification and regression problems and can describe the non-linear relationship between independent and dependent variables. Second, gradient-boosting trees method usually outperforms random forest [97], which may be partly attributed to the ability to capture complex relationships by iteratively correcting the errors of predecessors. Third, no data pre-processing techniques are needed as the technique can work with data containing missing values, and neither standardization nor normalization is necessary.

Gradient-boosting trees method has some limitations. First, gradient boosting trees method is more prone to overfitting, when the model parameters are not properly fine-tuned. However, model generalizability can be increased by optimization of the model's parameters and regularization. Second, gradient-boosting trees method is less interpretable than decision tree. Although the relevance of each feature may be quantified, it is not always evident how one feature interacts with the others and influences the outcome of the prediction. Third, training gradient-boosting trees can take a long time, since each tree grows sequentially instead of parallelly.

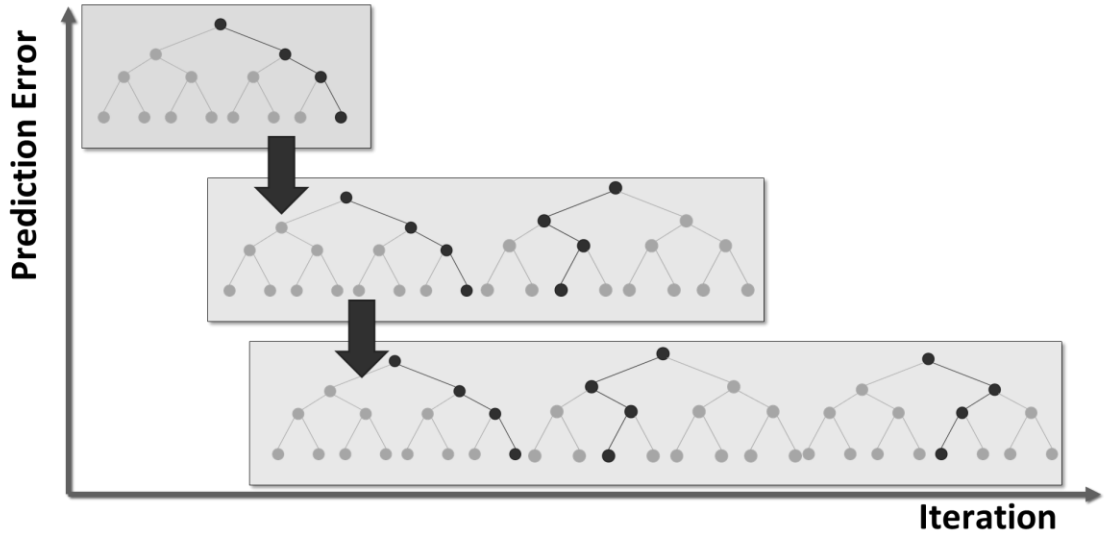


Figure 1.5 Example of a gradient boosting trees method where each new tree is built to correct the errors of its predecessors in each new iteration.

1.4.4 Evaluation metrics

After the development of prediction models, it is necessary to assess the predictive performance. Numerous evaluation measures exist; each provides different perspectives, and when reported altogether, they can provide a thorough picture of the model performance. Due to the inherent difference between classification and regression prediction targets, they have to be evaluated differently.

1.4.4.1 Classification

1.4.4.1.1 Confusion matrix

A confusion matrix can be created using the number of patients actually having the event (ground truths) and the number of patients that were predicted to have the event (prediction outcomes). The number of patients that are (in)correctly predicted as positive or negative is shown in a confusion matrix as true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), as shown in the **Table 1.3**. True positive rate, true negative rate, positive predictive value, and negative predictive value can be further defined as follows.

Table 1.3 Confusion matrix for binary classification

		Prediction outcomes	
		Positive	Negative
Ground truths	Positive	True positive (TP)	False negative (FN)
	Negative	False positive (FP)	True negative (TN)

True positive rate (TPR, sensitivity, recall): the probability of a positive predicted event, given the condition that the event is truly positive.

$$TPR = \frac{TP}{TP + FN}$$

True negative rate (TNR, specificity): the probability of a negative predicted event, given the condition that the event is truly negative.

$$TNR = \frac{TN}{TN + FP}$$

False positive rate (FPR, fall-out): the probability of a positive predicted event, given the condition that the event is truly negative.

$$FPR = \frac{FP}{FP + TN}$$

False negative rate (FNR, miss rate): the probability of a negative predicted event, given the condition that the event is truly positive.

$$FNR = \frac{FN}{FN + TP}$$

Positive likelihood ratio (LR+): the probability of a positive event predicted to be positive, divided by the probability of a negative event predicted to be positive.

$$LR+= \frac{TPR}{FPR}$$

Negative likelihood ratio (LR−): the probability of a positive event predicted to be negative, divided by the probability of a negative event predicted to be negative.

$$LR-= \frac{FNR}{TNR}$$

Positive predictive value (PPV, precision): the proportion of true positive results to all the positive prediction results.

$$PPV = \frac{TP}{TP + FP}$$

Negative predictive value (NPV): the proportion of true negative results to all the negative prediction results.

$$NPV = \frac{TN}{TN + FN}$$

1.4.4.1.2 Accuracy

The simplest and most popular method for determining whether or not a prediction model is reliable is to directly examine the model accuracy, which can be defined with the confusion matrix as follows: $\frac{TP+TN}{TP+TN+FP+FN}$.

However, the occurrence of an event or disease may be extremely rare in a medical context, making it inappropriate to evaluate model performance only on accuracy. In the case of a disease having a 5% prevalence in a dataset, for example, a simple estimate that the patient would be free from disease can obtain a very high accuracy of 95%. As a consequence, judging model performance solely on accuracy is clearly insufficient. The clinical usefulness of the evaluation should also be considered, along with other evaluation metrics like discrimination and calibration.

1.4.4.1.3 Discrimination

Discrimination evaluates a model's ability to discriminate patients with a specific condition (or event) from those without it. It is commonly measured with the area under the receiver operating characteristic (ROC) curve.

1.4.4.1.3.1 Receiver operating characteristic (ROC) curve and area under the ROC curve (AUROC)

The receiver operating characteristic (ROC) curve is a graph showing the sensitivity and the corresponding 1-specificity at all classification thresholds (**Figure 1.6**) [99]. Since the output of a classifier can be continuous predicted probability of an event, a classification threshold is needed to decide whether the final prediction result belongs to the positive or negative class. Instead of using a fixed threshold, the classification threshold can be tailored to the intended usage of the classifier and the acceptable risk for a positive or negative misclassification, based on the corresponding sensitivity and specificity at each classification threshold. It is noteworthy that there is a trade-off between the sensitivity and specificity. For example, a low classification threshold contributes to a high sensitivity at the cost of specificity, and it would be preferred if the cost of a false positive case is lower than the cost of a false negative, and vice versa.

Although the choice of optimal classification threshold depends on the intended usage, common ways of deciding the optimal classification threshold include (i) using the threshold that corresponding to the top-left corner of the ROC curve, since a perfect model should have 100% sensitivity and 100% specificity, and (ii) using the threshold that maximizes the Youden index, which is the sensitivity difference between a model and the diagonal axis for a specific specificity. These two methods both try to simultaneously maximize the sensitivity and specificity. The only difference is that the first method has a quadratic term in calculation of Euclidean distance to the top-left corner, whose clinical meaning remains unknown [100].

To estimate the overall performance of a model at all classification thresholds, the area under the ROC curve (AUROC) can be calculated. AUROC is a continuous measurement ranging from zero to one, where an AUROC of 0.5 represents a random model that is not different from flipping a coin, and an AUROC of one depicts a perfect model that makes no mistakes. The closer the AUROC is to one, the better the model's discrimination. There is no consensus about the adequate AUROC, which depends on the research question, intended usage, and reference for comparison. A prediction model with moderate or even poor discrimination may still be beneficial by saving resources (e.g., money, time, and human resources) and thus have a high clinical usefulness.

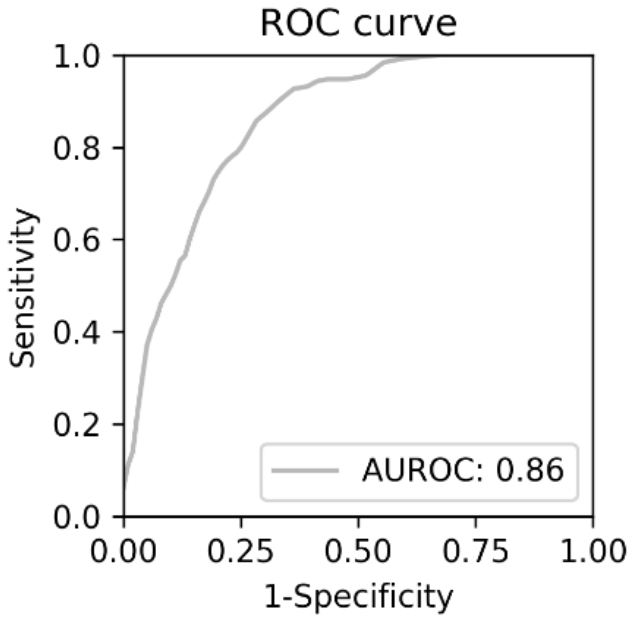


Figure 1.6 Example of a receiver operating characteristic curve. AUROC, area under the ROC curve; ROC curve, receiver operating characteristic curve

1.4.4.1.3.2 Precision-recall curve and area under the PR curve (AUPRC)

Precision-recall curve is a curve for evaluating the precision (positive predictive value) and corresponding recall (sensitivity) at all classification thresholds (**Figure 1.7**). Similar to the ROC curve, the classification threshold can be chosen based on the desired recall or precision. The major difference from ROC curve is that the precision recall curve considers the skewness in class distributions, since it evaluates the proportion of true positive cases in all positive predictions. If a dataset is imbalanced, the ROC curve may demonstrate over-optimistic results, while the precision recall curve allows for an unbiased interpretation of model performance [101, 102]. Unlike the ROC curve where only one curve is present, a 'baseline' model that predicts all patients as positive is usually added for comparison purposes in a precision recall curve. The further away that a model's curve is from the 'baseline' model's curve, the better its performance.

It is possible to utilize the threshold that maximizes the F1 score to select the best classification threshold that balances the importance of precision and recall. The F1 score is defined as follows.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Like the ROC curve, calculating the area under the precision recall curve (AUPRC) allows one to assess model performance. A well-performing model should have an AUPRC close to one, and a poor-performing model would have an AUPRC close to the event prevalence.

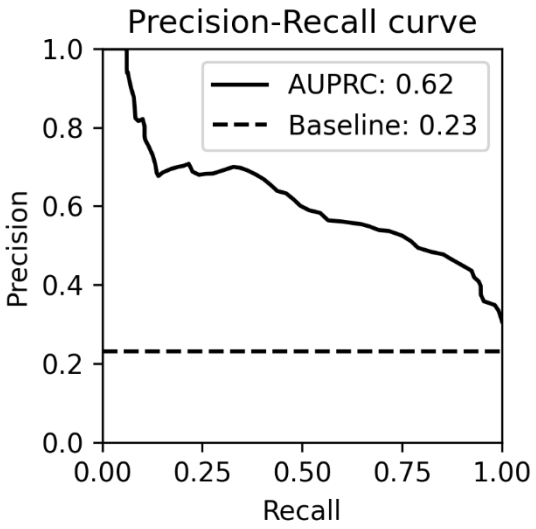


Figure 1.7 Example of a precision-recall curve, where the solid line indicates the model's precision-recall curve, while the dashed line represents the number of positive cases over the total number of patient-days. AUPRC, area under the precision recall curve; Baseline, the number of positive cases over the total number of patient-days

1.4.4.1.4 Calibration

Calibration assesses the degree of agreement between the predicted probabilities and observed outcome proportions [103]. For example, for a well calibrated model, out of the patients predicted to have a 0.4 probability of having the outcome, close to 40% actually do have the outcome. It is suggested to present the calibration in a graphical manner as a calibration plot (**Figure 1.8**), which is a curve with predicted probability on the x-axis and observed outcome proportion on the y-axis. In classification tasks, a locally weighted least squares regression smoother (LOESS) method should be applied to transform the patients' binary outcomes into continuous outcome proportions between zero and one (smoothing) by combining patients with similar predicted probabilities [104]. Since the diagonal axis of a calibration plot represents a perfectly calibrated prediction model, a model is considered well-calibrated if its calibration plot is not significantly different from the diagonal axis in a test of statistical significance. If a systematic deviation of a calibration plot from the diagonal axis is observed, the model is considered not well-calibrated. For instance, if the calibration plot is systematically above the diagonal axis, it means that the model has under-estimated the predicted probability of an event. And the opposite holds for an over-estimation situation. When miscalibration happens, a calibration of the uncalibrated model may be considered (re-calibration) by using Platt Scaling or Isotonic Regression [105, 106].

Calibration can be additionally characterized by the calibration slope and the calibration-in-the-large [103]. The calibration slope is calculated by the slope of a linear regression model, and it examines the spread of predicted probability of being positive. A calibration slope smaller than one represents that the predicted probability is too extreme (e.g. the predictions are either too high or too low), which is commonly observed in external validation studies where extreme values are present due to overfitting. The opposite is true for a prediction range that is too small. Calibration-in-the-large measures the overall calibration between predicted

probability and observed outcome, and it is calculated by the difference between average predicted probability and average observed outcome proportion. A positive calibration-in-the-large suggests underestimation, while a negative one represents over-estimation. In conclusion, a well-calibrated model should have calibration slope close to one and calibration-in-the-large close to zero.

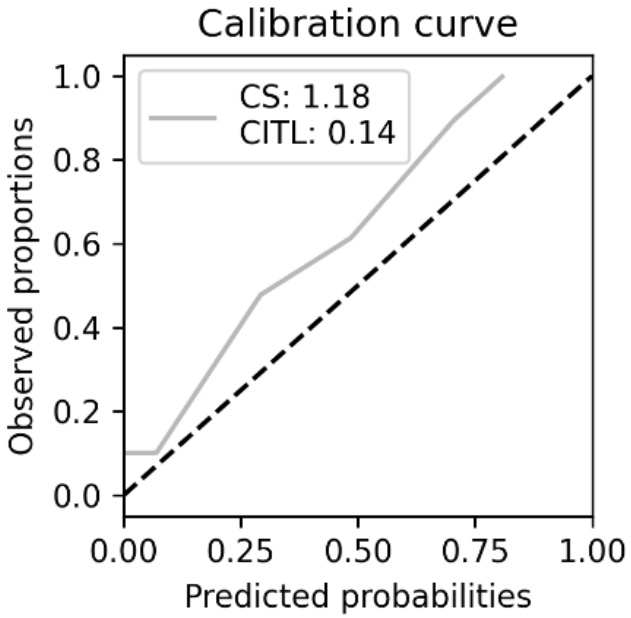


Figure 1.8 Example of a calibration curve, where the solid line indicates the model's calibration curve, and the black dashed line represents the calibration curve for a perfectly calibrated model. CS, calibration slope; CITLs, calibration-in-the-large

1.4.4.1.5 The decision curve

While discrimination and calibration are commonly reported in general prediction models, the decision curve (**Figure 1.9**), which provides information of the clinical usefulness of a model, is not commonly reported. As aforementioned, the output of a classifier is a continuous predicted probability, so a classification threshold needs to be set to classify the predictions as positive or negative. In a general prediction task, a default classification threshold of 0.5 is commonly used, since the importance of correctly classifying new inputs as positive or negative is equal. However, this is rarely the case in clinical practice. Missing one patient with the disease (FN) may have more adverse impact than misclassifying a patient as positive (FP), or vice versa. Therefore, an optimal classification threshold must be decided from not only a statistical perspective but also a clinical viewpoint. Once a classification threshold is decided, we can evaluate the clinical usefulness by comparing the potential benefits of utilizing the prediction model versus the default policy without the prediction model [107]. We may further study the clinical utility for all potential classification thresholds and visually display their relationship as a "decision curve" [108], similar to the ROC curve.

For a specific classification threshold p_t, the clinical usefulness can be quantified by the net benefit with the following equation, which considers the potential benefit and harm of true-positive and false-positive with a weighting factor.

$$Net\ benefit = \frac{TP}{N} - \frac{FP}{N} \times \frac{p_t}{1 - p_t}$$

where N is the sample size of the dataset, TP is the number of true-positive predictions, FP is the number of false-positive predictions, and $\frac{p_t}{1-p_t}$ is the odds of the classification threshold p_t. $\frac{p_t}{1-p_t}$ works as a weight to give different importance to FP from TP. For instance, if the classification threshold is set to 0.3, the harm of the FP is considered to be $\frac{3}{7}$ as import as the benefit of the TP.

As indicated above, the net benefit of a model has to be compared with that of two default policies: treat-all and treat-none. A model is considered clinically useful if the model's net benefit is larger than that of both two default policies. The net benefit of treat-all varies with disease prevalence and classification threshold. Treat-none always classifies patient as negative (TP and FP are always zero), so the net benefit of treat-none is zero for all possible classification thresholds.

We can further investigate the range of classification threshold where a model is considered clinically useful. The wider the clinical usefulness range, the more situations a model can be applied to. Since the importance of potential harms to potential benefits is different from patient to patient, a personalized classification threshold may be necessary. If a model has a wide clinical usefulness range, it can be potentially helpful for more patients in more situations.

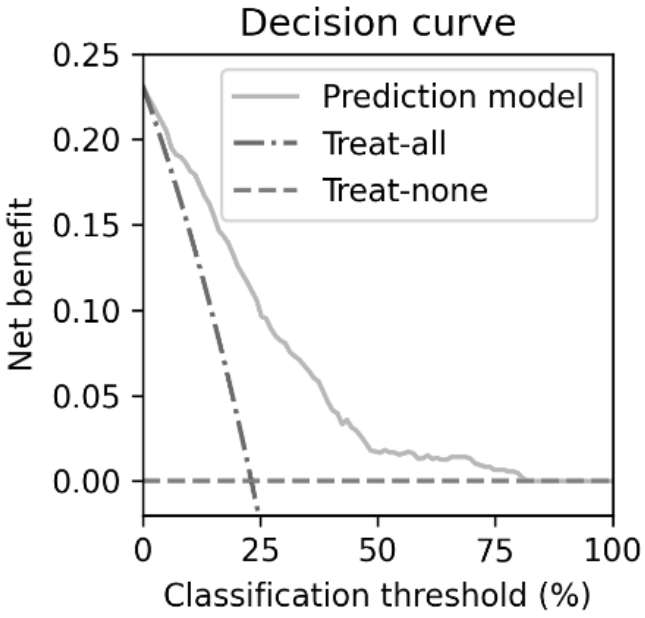


Figure 1.9 Example of a decision curve, where the solid line, dashed-dotted line, and dashed line indicate the net benefit of the prediction model, the default policy "treat-all", and the default policy "treat-none" respectively

1.4.4.2 Regression

1.4.4.2.1 Error, mean absolute error, and root mean square error

In a regression task we are interested in how closely our predictions match the ground truth, which is best indicated via direct comparison between the predicted and expected outcome of interest (often called as error). The commonly used error evaluation metrics include mean absolute error (MAE) and root mean square error (RMSE). Both summarize how much the predictions on average deviated from the real numerical outcome values, with RMSE more sensitive to large errors due to its square function. Since they both represent deviations, the

smaller the error measurements, the better the performance. Importantly, neither metric allows for the interpretation of error direction (overestimation or underestimation).

1.4.5 Internal validation and external validation

A prediction model is supposed to learn an underlying relationship from previous examples, to extrapolate from what it has learned, and to provide reliable predictions for previously unseen patients. Consequently, model performance should not be evaluated only on the patients used for model development but also on independent patient groups [109]. The procedure to evaluate performance outside the patients used for model development is called validation. There are two forms of validation: internal validation and external validation, depending on whether the development and validation datasets are independent.

1.4.5.1 Internal validation

Internal validation is a process to examine internal validity of a model in the same setting where the model development is performed. Internal validation is a crucial and indispensable step for the development of a prediction model, since it provides less biased measurement of model performance and helps reducing the risk of overfitting. In addition, a proper internal validation ensures the reproducibility. The most common internal validation methods include split-sample validation and K-fold cross validation [110].

1.4.5.1.1 Split-sample validation

Split-sample validation simply divides the dataset into two parts: one for model development and the other for validation of model performance. The dataset can be divided in any way, but certain details, such as the sample size ratio between the development set and validation set, should be focused on. Additionally, event incidence should be examined to ensure that it is high enough for both development and validation sets. A decent prediction model cannot be trained if the incidence in the development set is too low, and the performance cannot be trusted if the incidence in the validation set is not sufficient. One way to ensure the event incidence in both parts is to use stratified sampling, which forces the different partitions to have the same proportion of the target outcome.

Despite the simple computation and efficiency of split-sample validation, there is a limitation that the model and its performance may not be stable and reliable, since only part of the dataset is used for model development and the rest for model validation. Hence, more advanced internal validation methods such as K-fold cross validation are preferrable.

1.4.5.1.2 K-fold cross validation

Based on the concept of split-sample validation, K-fold cross validation is a validation method with a higher stability (**Figure 1.10**). First, the dataset is split into K folds. Subsequently, a model is trained on all but one fold and tested on the remaining fold. This training and testing procedure is repeated K times until all the folds have been used independently for testing. In the end, there are K testing measurements, and a corresponding confidence interval can be drawn.

The choice of the integer K is arguable. On the one hand, a large K is time-consuming and computationally demanding, since the training and testing procedures have to be repeated more times to evaluate model performance. The most extreme case is leave-one-out cross validation, where all but one patients are used for development, and validation is performed on one patient in each fold. On the other hand, a small K has the

drawback of not having a large sample size for training, and thus the testing performance is likely to be poor. Additionally, since the number of repetitions is lower for a small K, the testing performance is less stable. Commonly chosen Ks are 5, 10, or 20 folds.

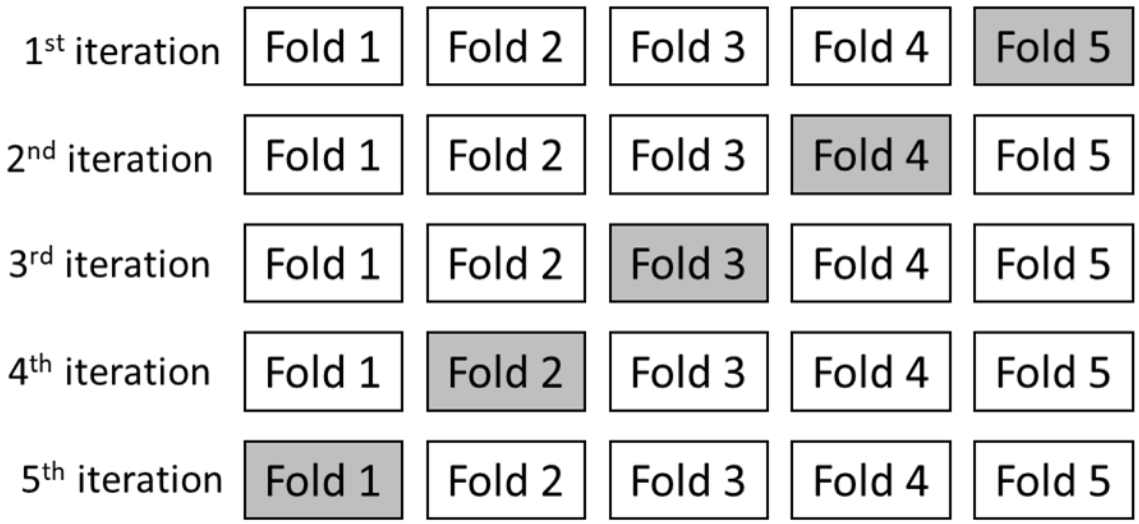


Figure 1.10 Example of K-fold cross validation, where k=5. The data are first split into five sets. Consequently, the model is trained and tested for five iterations. For each iteration, model is trained on four folds (i.e., the training sets, indicated in white) and tested on the reserved one fold (i.e., the testing test, indicated in grey). Finally, the model performance is summarized by the measurements from the testing set in each iteration.

1.4.5.2 External validation

As a result of the same setting, internal validation is still partially similar to development set and not systematically different, and thus internal validation can still provide over-optimistic results. Therefore, the optimal approach is to investigate a model's generalizability (or transportability) in an independent dataset [111]. There are many external validation methods with different settings, e.g., temporal validation, geographical validation, and validation in different patient populations. Temporal validation refers to model validation in a period that is different from the time the model was developed, e.g., ten years after model development. Geographical validation evaluates model performance in different locations from the place the model was built, e.g., other ICU units, other hospitals, and other countries. Validation in different patient populations may yield radically different results even at the same place during the same period, such as patients undergoing sepsis or cardiac surgery.

When external validation is performed, care should be taken with the sample size and disease prevalence. If the sample size of a study is too small, only low statistical power can be obtained, and model performance may be easily distorted by random effects. However, a large sample size also increases the chance of detecting differences that are not clinically relevant. Disease prevalence also matters. For instance, the classical rules of thumb for classification tasks indicate that at least 100 samples (preferably 200) should be in each group to draw concrete conclusions. Therefore, the acquired external validation results have to be carefully interpreted along with the sample size and disease prevalence.

1.5 Application of prediction models to clinical practice

Powerful prediction models with high accuracy are desirable, but simply having them on computers is insufficient. In the ideal situation, predictions run automatically at the patient's bedside rather than having clinicians spend time gathering patient data, manually entering data into models, and then turning prediction results into useful treatment plans. This way, the predictions could free up doctors to concentrate more on treating patients.

The general interest in machine learning predictions in the ICU population is reflected in a huge increase in the number of publications reporting machine learning models, with 48% of all reports published after 2015[112]. Nevertheless, few examples were successfully brought into the clinical practice. In a systematic review where Fleuren et al. introduced a clinically applicable scale (from level 1 to level 9) to assess the readiness of prediction models to be brought into clinical practice [113]. They showed that 93% of the developed models were not externally validated (level 4 or below), 5% were with external validation (level 5), 1% were integrated into workflow without exposure to clinicians (level 6), 1% were compared against related clinical outcomes (level 8), and none of them were integrated in the clinical workflow and evaluated at different centers (level 9).

The low clinical readiness of these prediction models may be explained by the numerous challenges. The first problem has to do with gathering well-structured patient data. It is preferable to use large amounts of high-quality data from several centers in order to create robust prediction models. However, the format and labeling of data from various centers may be radically different in their EHR systems, which makes it difficult to be integrated. Differences in data description and abbreviation may exist even in the same hospital. Therefore, researchers proposed Hierarchical Data Format—Version 5 (HDF5), a standardized format for transferring clinical data across various ICUs, as a solution to this problem [114]. The use of standardized formats would reduce the barrier to integration for large amounts of existing data.

The second issue relates to the lack of robustness in prediction models. A model that performs astonishingly only on the original dataset but not anywhere else is of limited usage. Preventing overly optimistic performance by boosting effective reporting from the start is one strategy to mitigate this. As a result, the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) Initiative developed a TRIPOD statement consisting of a checklist of 22 items, to increase the transparency of the reporting of the prediction model study [115]. The enhanced reporting is anticipated to improve the quality of the research and increase attention and resources for promising models.

The third issue is the black-box characteristics of prediction models while clinicians prefer strong evidence-based scientific support to deliver meaningful intervention and to adapt their treatment strategies to vulnerable ICU patients. However, many complex models were built with good accuracy at the expense of poor interpretability. In order to increase interpretability, a variety of techniques have been proposed, such as permutation importance [116] and Shapley Additive exPlanations (SHAP) [117], both of which showed consistency within the literature and clinical interpretation [118–120]. More reliable models are anticipated to be delivered at the patient's bedside with the use of these technologies to improve interpretability.

1.6 Conclusion

As a result of the expansion of electronic health records and the development of machine-learning algorithms, numerous models have been proposed for the prediction of kidney function in critically ill patients. The current

focus of prediction models for kidney function is mainly on the prediction of AKI and ARC. Although some models were shown to have good predictive performance, more external validations in large independent datasets are still needed, before the models can be implemented into clinical practice. In addition, AKI recovery prediction may also be helpful to guide therapeutic management as persisting acute kidney injury is associated with many unfavorable consequences. However, currently available AKI recovery prediction models have their limitations, and there is a need for a better AKI recovery prediction model. Finally, as a surrogate of the entire kidney function spectrum, CrCl may be more relevant than AKI and ARC. Since CrCl is measured based on 8- to 24-hour samples, prediction models for CrCl may enable more appropriate treatment approaches by providing a more accurate assessment of kidney function. Nevertheless, to the best of our knowledge, no prediction models for the daily prediction of CrCl have been proposed.

BIBLIOGRAPHY

1. Vincent J-L, Marshall JC, Namendys-Silva SA, et al (2014) Assessment of the worldwide burden of critical illness: the intensive care over nations (ICON) audit. Lancet Respir Med 2:380–386. https://doi.org/10.1016/S2213-2600(14)70061-X
2. Hoste EAJJ, Bagshaw SM, Bellomo R, et al (2015) Epidemiology of acute kidney injury in critically ill patients: the multinational AKI-EPI study. Intensive Care Med 41:1411–1423. https://doi.org/10.1007/s00134-015-3934-7
3. Kellum JA, Lameire N, Aspelin P, et al (2012) Kidney disease: Improving global outcomes (KDIGO) acute kidney injury work group. KDIGO clinical practice guideline for acute kidney injury. Kidney Int Suppl 2:1–138. https://doi.org/10.1038/kisup.2012.1
4. Case J, Khan S, Khalid R, Khan A (2013) Epidemiology of acute kidney injury in the intensive care unit. Crit Care Res Pract 2013:479730. https://doi.org/10.1155/2013/479730
5. Jiang L, Zhu Y, Luo X, et al (2019) Epidemiology of acute kidney injury in intensive care units in Beijing: The multi-center BAKIT study. BMC Nephrol 20:468. https://doi.org/10.1186/s12882-019-1660-z
6. Wang HE, Muntner P, Chertow GM, Warnock DG (2012) Acute kidney injury and mortality in hospitalized patients. Am J Nephrol 35:349–355. https://doi.org/10.1159/000337487
7. Chertow GM, Burdick E, Honour M, et al (2005) Acute kidney injury, mortality, length of stay, and costs in hospitalized patients. J Am Soc Nephrol 16:3365–3370. https://doi.org/10.1681/ASN.2004090740
8. Federspiel CK, Itenov TS, Mehta K, et al (2018) Duration of acute kidney injury in critically ill patients. Ann Intensive Care 8:30. https://doi.org/10.1186/s13613-018-0374-x
9. Silver SA, Chertow GM (2017) The Economic Consequences of Acute Kidney Injury. Nephron 137:297–301
10. See EJ, Jayasinghe K, Glassford N, et al (2019) Long-term risk of adverse outcomes after acute kidney injury: a systematic review and meta-analysis of cohort studies using consensus definitions of exposure. Kidney Int 95:160–172. https://doi.org/10.1016/j.kint.2018.08.036
11. Chawla LS, Kimmel PL (2012) Acute kidney injury and chronic kidney disease: an integrated clinical syndrome. Kidney Int 82:516–524. https://doi.org/10.1038/ki.2012.208
12. Le S, Allen A, Calvert J, et al (2021) Convolutional Neural Network Model for Intensive Care Unit Acute Kidney Injury Prediction. Kidney Int Reports 6:1289–1298. https://doi.org/https://doi.org/10.1016/j.ekir.2021.02.031
13. Malhotra R, Kashani KB, Macedo E, et al (2017) A risk prediction score for acute kidney injury in the intensive care unit. Nephrol Dial Transplant 32:814–822. https://doi.org/10.1093/ndt/gfx026
14. Wang Q, Tang Y, Zhou J, Qin W (2019) A prospective study of acute kidney injury in the intensive care unit: development and validation of a risk prediction model. J Transl Med 17:359. https://doi.org/10.1186/s12967-019-2118-6
15. Chiofolo C, Chbat N, Ghosh E, et al (2019) Automated Continuous Acute Kidney Injury Prediction and

Surveillance: A Random Forest Model. Mayo Clin Proc 94:783–792. https://doi.org/10.1016/j.mayocp.2019.02.009

16. Sato N, Uchino E, Kojima R, et al (2021) Prediction and visualization of acute kidney injury in intensive care unit using one-dimensional convolutional neural networks based on routinely collected data. Comput Methods Programs Biomed 206:106129. https://doi.org/https://doi.org/10.1016/j.cmpb.2021.106129

17. Li Y, Yao L, Mao C, et al (2018) Early Prediction of Acute Kidney Injury in Critical Care Setting Using Clinical Notes. In: 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp 683–686

18. Gao W, Wang J, Zhou L, et al (2021) Prediction of acute kidney injury in ICU with gradient boosting decision tree algorithms. Comput Biol Med 140:105097. https://doi.org/10.1016/j.compbiomed.2021.105097

19. Chen Y, Feng F, Li M, et al (2019) Development of a risk stratification-based model for prediction of acute kidney injury in critically ill patients. Med (United States) 98:e16867. https://doi.org/10.1097/MD.0000000000016867

20. Flechet M, Güiza F, Schetz M, et al (2017) AKIpredictor, an online prognostic calculator for acute kidney injury in adult critically ill patients: development, validation and comparison to serum neutrophil gelatinase-associated lipocalin. Intensive Care Med 43:764–773. https://doi.org/10.1007/s00134-017-4678-3

21. Casaer MP, Mesotten D, Hermans G, et al (2011) Early versus Late Parenteral Nutrition in Critically Ill Adults. N Engl J Med 365:506–517. https://doi.org/10.1056/NEJMoa1102662

22. AKI prediction. https://www.akipredictor.com/en/. Accessed 10 Feb 2022

23. Flechet M, Falini S, Bonetti C, et al (2019) Machine learning versus physicians' prediction of acute kidney injury in critically ill adults: A prospective evaluation of the AKIpredictor. Crit Care 23:. https://doi.org/10.1186/s13054-019-2563-x

24. Belayev LY, Palevsky PM (2014) The link between acute kidney injury and chronic kidney disease. Curr Opin Nephrol Hypertens 23:149–154. https://doi.org/10.1097/01.mnh.0000441051.36783.f3

25. Ishani A, Xue JL, Himmelfarb J, et al (2009) Acute kidney injury increases risk of ESRD among elderly. J Am Soc Nephrol 20:223–228. https://doi.org/10.1681/ASN.2007080837

26. Srisawat N, Murugan R, Lee M, et al (2011) Plasma neutrophil gelatinase-associated lipocalin predicts recovery from acute kidney injury following community-acquired pneumonia. Kidney Int 80:545–552. https://doi.org/10.1038/ki.2011.160

27. Srisawat N, Wen X, Lee MJ, et al (2011) Urinary biomarkers and renal recovery in critically ill patients with renal support. Clin J Am Soc Nephrol 6:1815–1823. https://doi.org/10.2215/CJN.11261210

28. Moon SJ, Park HB, Yoon SY, Lee SC (2013) Urinary biomarkers for early detection of recovery in patients with acute kidney injury. J Korean Med Sci 28:1181–1186. https://doi.org/10.3346/jkms.2013.28.8.1181

29. Aregger F, Uehlinger DE, Witowski J, et al (2014) Identification of IGFBP-7 by urinary proteomics as a novel prognostic marker in early acute kidney injury. Kidney Int 85:909–919. https://doi.org/10.1038/ki.2013.363

30. Meersch M, Schmidt C, Van Aken H, et al (2014) Urinary TIMP-2 and IGFBP7 as early biomarkers of acute kidney injury and renal recovery following cardiac surgery. PLoS One 9:e93460. https://doi.org/10.1371/journal.pone.0093460

31. Dewitte A, Joannès-Boyau O, Sidobre C, et al (2015) Kinetic eGFR and novel AKI biomarkers to predict renal recovery. Clin J Am Soc Nephrol 10:1900–1910. https://doi.org/10.2215/CJN.12651214

32. Zeng X-F, Li J-M, Tan Y, et al (2014) Performance of urinary NGAL and L-FABP in predicting acute kidney injury and subsequent renal recovery: a cohort study based on major surgeries. Clin Chem Lab Med 52:671–678. https://doi.org/10.1515/cclm-2013-0823

33. Gaião SM, Paiva JAO de C (2017) Biomarkers of renal recovery after acute kidney injury. Rev Bras Ter intensiva 29:373–381. https://doi.org/10.5935/0103-507X.20170051

34. Fiorentino M, Tohme FA, Murugan R, Kellum JA (2019) Plasma Biomarkers in Predicting Renal Recovery from Acute Kidney Injury in Critically Ill Patients. Blood Purif 48:253–261. https://doi.org/10.1159/000500423

35. Dharan KS, John GT, Antonisamy B, et al (2005) Prediction of mortality in acute renal failure in the tropics. Ren Fail 27:289–296. https://doi.org/10.1081/JDI-200056612

36. Chertow GM, Soroko SH, Paganini EP, et al (2006) Mortality after acute renal failure: Models for prognostic stratification and risk adjustment. Kidney Int 70:1120–1126. https://doi.org/10.1038/sj.ki.5001579

37. Demirjian S, Chertow GM, Zhang JH, et al (2011) Model to predict mortality in critically ill adults with acute kidney injury. Clin J Am Soc Nephrol 6:2114–2120. https://doi.org/10.2215/CJN.02900311

38. da Hora Passos R, Ramos JGR, Mendonça EJB, et al (2017) A clinical score to predict mortality in septic acute kidney injury patients requiring continuous renal replacement therapy: The HELENICC score. BMC Anesthesiol 17:21. https://doi.org/10.1186/s12871-017-0312-8

39. Kim Y, Park N, Kim J, et al (2019) Development of a new mortality scoring system for acute kidney injury with continuous renal replacement therapy. Nephrology 24:1233–1240. https://doi.org/10.1111/nep.13661

40. Kang MW, Kim J, Kim DK, et al (2020) Machine learning algorithm to predict mortality in patients undergoing continuous renal replacement therapy. Crit Care 24:42. https://doi.org/10.1186/s13054-020-2752-7

41. Lin K, Hu Y, Kong G (2019) Predicting in-hospital mortality of patients with acute kidney injury in the ICU using random forest model. Int J Med Inform 125:55–61. https://doi.org/10.1016/j.ijmedinf.2019.02.002

42. Xu Z, Luo Y, Adekkanattu P, et al (2019) Stratified mortality prediction of patients with acute kidney injury in critical care. In: Studies in Health Technology and Informatics. IOS Press, pp 462–466

43. Forni LG, Darmon M, Ostermann M, et al (2017) Renal recovery after acute kidney injury. Intensive Care Med 43:855–866. https://doi.org/10.1007/s00134-017-4809-x

44. Itenov TS, Berthelsen RE, Jensen JU, et al (2018) Predicting recovery from acute kidney injury in critically ill patients: Development and validation of a prediction model. Crit Care Resusc 20:54–60

45. Lee BJ, Hsu C yuan, Parikh R, et al (2019) Predicting Renal Recovery After Dialysis-Requiring Acute Kidney Injury. Kidney Int Reports 4:571–581. https://doi.org/10.1016/j.ekir.2019.01.015

46. Bilbao-Meseguer I, Rodríguez-Gascón A, Barrasa H, et al (2018) Augmented Renal Clearance in Critically Ill Patients: A Systematic Review. Clin Pharmacokinet 57:1107–1121. https://doi.org/10.1007/s40262-018-0636-7

47. Udy AA, Dulhunty JM, Roberts JA, et al (2017) Association between augmented renal clearance and clinical outcomes in patients receiving β-lactam antibiotic therapy by continuous or intermittent infusion: a nested cohort study of the BLING-II randomised, placebo-controlled, clinical trial. Int J Antimicrob Agents 49:624–630. https://doi.org/10.1016/j.ijantimicag.2016.12.022

48. Huttner A, Von Dach E, Renzoni A, et al (2015) Augmented renal clearance, low β-lactam concentrations and clinical outcomes in the critically ill: An observational prospective cohort study. Int J Antimicrob Agents 45:385–392. https://doi.org/10.1016/j.ijantimicag.2014.12.017

49. Kawano Y, Maruyama J, Hokama R, et al (2018) Outcomes in patients with infections and augmented renal clearance: A multicenter retrospective study. PLoS One 13:1–11. https://doi.org/10.1371/journal.pone.0208742

50. Dickerson RN, Crawford CN, Tsiu MK, et al (2021) Augmented renal clearance following traumatic injury in critically ill patients requiring nutrition therapy. Nutrients 13:. https://doi.org/10.3390/nu13051681

51. Wu CC, Tai CH, Liao WY, et al (2019) Augmented renal clearance is associated with inadequate antibiotic pharmacokinetic/ pharmacodynamic target in Asian ICU population: a prospective observational study. Infect Drug Resist 12:2531–2541. https://doi.org/10.2147/IDR.S213183

52. Adnan S, Ratnam S, Kumar S, et al (2014) Select critically ill patients at risk of augmented renal clearance: experience in a Malaysian intensive care unit. Anaesth Intensive Care 42:715–722. https://doi.org/10.1177/0310057x1404200606

53. He J, Yang Z-TT, Qian X, et al (2020) A higher dose of vancomycin is needed in critically ill patients with augmented renal clearance. Transl Androl Urol 9:2166–2171. https://doi.org/10.21037/tau-20-1048

54. Abdel El Naeem HEM, Abdelhamid MHE, Atteya DAM (2017) Impact of augmented renal clearance on enoxaparin therapy in critically ill patients. Egypt J Anaesth 33:113–117. https://doi.org/https://doi.org/10.1016/j.egja.2016.11.001

55. Claus BOMM, Hoste EA, Colpaert K, et al (2013) Augmented renal clearance is a common finding with worse clinical outcome in critically ill patients receiving antimicrobial therapy. J Crit Care 28:695–700. https://doi.org/10.1016/j.jcrc.2013.03.003

56. Luo Y, Wang Y, Ma Y, et al (2021) Augmented Renal Clearance: What Have We Known and What Will

We Do? Front Pharmacol 12:1–8. https://doi.org/10.3389/fphar.2021.723731

57. Sharma A, Mucino MJ, Ronco C (2014) Renal functional reserve and renal recovery after acute kidney injury. Nephron Clin Pract 127:94–100. https://doi.org/10.1159/000363721

58. Baptista JP, Martins PJ, Marques M, Pimentel JM (2020) Prevalence and Risk Factors for Augmented Renal Clearance in a Population of Critically Ill Patients. J Intensive Care Med 35:1044–1052. https://doi.org/10.1177/0885066618809688

59. De Waele JJ, Dumoulin A, Janssen A, Hoste EAJ (2015) Epidemiology of augmented renal clearance in mixed ICU patients. Minerva Anestesiol 81:1079–1085

60. Udy AA, Baptista JP, Lim NL, et al (2014) Augmented renal clearance in the ICU: Results of a Multicenter observational study of renal function in critically ill patients with normal plasma Creatinine concentrations. Crit Care Med 42:520–527. https://doi.org/10.1097/CCM.0000000000000029

61. Barletta JF, Mangram AJ, Byrne M, et al (2017) Identifying augmented renal clearance in trauma patients: Validation of the Augmented Renal Clearance in Trauma Intensive Care scoring system. In: The journal of trauma and acute care surgery. United States, pp 665–671

62. Gijsen M, Huang CY, Flechet M, et al (2020) Development and External Validation of an Online Clinical Prediction Model for Augmented Renal Clearance in Adult Mixed Critically Ill Patients: The Augmented Renal Clearance Predictor. Crit Care Med 48:e1260–e1268. https://doi.org/10.1097/CCM.0000000000004667

63. Nei AM, Kashani KB, Dierkhising R, Barreto EF (2020) Predictors of Augmented Renal Clearance in a Heterogeneous ICU Population as Defined by Creatinine and Cystatin C. Nephron 55905:313–320. https://doi.org/10.1159/000507255

64. Mulder MB, Eidelson SA, Sussman MS, et al (2019) Risk Factors and Clinical Outcomes Associated With Augmented Renal Clearance in Trauma Patients. J Surg Res 244:477–483. https://doi.org/10.1016/j.jss.2019.06.087

65. Nazer LH, AbuSara AK, Kamal Y (2021) Augmented renal clearance in critically ill patients with cancer (ARCCAN Study): A prospective observational study evaluating prevalence and risk factors. Pharmacol Res Perspect 9:2–7. https://doi.org/10.1002/prp2.747

66. Johnston BW, Perry D, Habgood M, et al (2021) Augmented renal clearance: a retrospective, cohort study of urinary creatinine clearance in critically ill patients in the United Kingdom. J Int Med Res 49:. https://doi.org/10.1177/03000605211015573

67. Tang L, Ding XY, Duan LF, et al (2021) A Regression Model to Predict Augmented Renal Clearance in Critically Ill Obstetric Patients and Effects on Vancomycin Treatment. Front Pharmacol 12:1–12. https://doi.org/10.3389/fphar.2021.622948

68. Udy AA, Roberts JA, Shorr AF, et al (2013) Augmented renal clearance in septic and traumatized patients with normal plasma creatinine concentrations: Identifying at-risk patients. Crit Care 17:. https://doi.org/10.1186/cc12544

69. M@tric project. https://www.matric.be/. Accessed 5 Oct 2021

70. ARC prediction. https://www.arcpredictor.com/. Accessed 13 Apr 2022

71. Cockcroft DW, Gault MH (1976) Prediction of creatinine clearance from serum creatinine. Nephron 16:31–41. https://doi.org/10.1159/000180580

72. Levey AS, Bosch JP, Lewis JB, et al (1999) A more accurate method to estimate glomerular filtration rate from serum creatinine: A new prediction equation. Ann Intern Med 130:461–470. https://doi.org/10.7326/0003-4819-130-6-199903160-00002

73. Levey AS, Stevens LA, Schmid CH, et al (2009) A new equation to estimate glomerular filtration rate. Ann Intern Med 150:604–612. https://doi.org/10.7326/0003-4819-150-9-200905050-00006

74. Al-Dorzi HM, Alsadhan AA, Almozaini AS, et al (2021) The Performance of Equations That Estimate Glomerular Filtration Rate against Measured Urinary Creatinine Clearance in Critically Ill Patients. Crit Care Res Pract 2021:. https://doi.org/10.1155/2021/5520653

75. Grootaert V, Willems L, Debaveye Y, et al (2012) Augmented Renal Clearance in the Critically Ill: How to Assess Kidney Function. Ann Pharmacother 46:925–959. https://doi.org/10.1345/aph.1Q708

76. Wongpraphairot S, Thongrueang A, Bhurayanontachai R (2022) Glomerular filtration rate correlation and agreement between common predictive equations and standard 24-hour urinary creatinine clearance in medical critically ill patients. PeerJ 10:e13556. https://doi.org/10.7717/peerj.13556

77. Schetz M, Gunst J, Van den Berghe G (2014) The impact of using estimated GFR versus creatinine clearance on the evaluation of recovery from acute kidney injury in the ICU. Intensive Care Med 40:1709–1717. https://doi.org/10.1007/s00134-014-3487-1

78. Rahn KH, Heidenreich S, Brückner D (1999) How to assess glomerular function and damage in humans. J Hypertens 17:309–317. https://doi.org/10.1097/00004872-199917030-00002

79. Cherry RA, Eachempati SR, Hydo L, Barie PS (2002) Accuracy of short-duration creatinine clearance determinations in predicting 24-hour creatinine clearance in critically ill and injured patients. J Trauma 53:267–271. https://doi.org/10.1097/00005373-200208000-00013

80. Huang C-Y, Güiza F, De Vlieger G, Meyfroidt G (2022) Daily fluctuations in kidney function in critically ill adults. Crit Care 26:347. https://doi.org/10.1186/s13054-022-04226-3

81. Evans RS (2016) Electronic Health Records: Then, Now, and in the Future. Yearb Med Inform Suppl 1:S48-61. https://doi.org/10.15265/IYS-2016-s006

82. Hadi HJ, Shnain AH, Hadishaheed S, Ahmad AH (2015) Big data and five V's characteristics. Int J Adv Electron Comput Sci 2:16–23

83. Kubben P (2019) Data sources. Fundam Clin data Sci 3–9

84. Heideman MT, Johnson DH, Burrus CS (1985) Gauss and the history of the fast Fourier transform. Arch Hist Exact Sci 34:265–277. https://doi.org/10.1007/BF00348431

85. Gothwal H, Kedawat S, Kumar R, others (2011) Cardiac arrhythmias detection in an ECG beat signal using fast fourier transform and artificial neural network. J Biomed Sci Eng 4:289

86. Childers DG, Skinner DP, Kemerait RC (1977) The cepstrum: A guide to processing. Proc IEEE

65:1428–1443. https://doi.org/10.1109/PROC.1977.10747

87. Berisha V, Krantsevich C, Hahn PR, et al (2021) Digital medicine and the curse of dimensionality. npj Digit Med 4:153. https://doi.org/10.1038/s41746-021-00521-5

88. Remeseiro B, Bolon-Canedo V (2019) A review of feature selection methods in medical applications. Comput Biol Med 112:103375. https://doi.org/https://doi.org/10.1016/j.compbiomed.2019.103375

89. Jović A, Brkić K, Bogunović N (2015) A review of feature selection methods with applications. In: 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). pp 1200–1205

90. Tibshirani R (1996) Regression shrinkage and selection via the lasso. J R Stat Soc Ser B 58:267–288

91. Hua J, Tembe WD, Dougherty ER (2009) Performance of feature-selection methods in the classification of high-dimension data. Pattern Recognit 42:409–424. https://doi.org/https://doi.org/10.1016/j.patcog.2008.08.001

92. Cramer JS (2002) The origins of logistic regression

93. Ho TK (1995) Random decision forests. In: Proceedings of 3rd international conference on document analysis and recognition. pp 278–282

94. Ho TK (1998) The random subspace method for constructing decision forests. IEEE Trans Pattern Anal Mach Intell 20:832–844

95. Opitz D, Maclin R (1999) Popular ensemble methods: An empirical study. J Artif Intell Res 11:169–198

96. Breiman L (1996) Stacked regressions. Mach Learn 24:49–64. https://doi.org/10.1007/BF00117832

97. Hastie T, Tibshirani R, Friedman J (2009) Boosting and Additive Trees. In: The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer New York, New York, NY, pp 337–387

98. Freund Y, Schapire RE, others (1996) Experiments with a new boosting algorithm. In: icml. pp 148–156

99. Kumar R, Indrayan A (2011) Receiver operating characteristic (ROC) curve for medical researchers. Indian Pediatr 48:277–287

100. Unal I (2017) Defining an Optimal Cut-Point Value in ROC Analysis: An Alternative Approach. Comput Math Methods Med 2017:3762651. https://doi.org/10.1155/2017/3762651

101. Ozenne B, Subtil F, Maucort-Boulch D (2015) The precision–recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases. J Clin Epidemiol 68:855–859. https://doi.org/https://doi.org/10.1016/j.jclinepi.2015.02.010

102. Sofaer HR, Hoeting JA, Jarnevich CS (2019) The area under the precision-recall curve as a performance metric for rare binary events. Methods Ecol Evol 10:565–577. https://doi.org/https://doi.org/10.1111/2041-210X.13140

103. Steyerberg EW (2009) Evaluation of performance. In: Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. Springer New York, New York, NY, pp 255–280

104. Austin PC, Steyerberg EW (2014) Graphical assessment of internal and external calibration of logistic

regression models by using loess smoothers. Stat Med 33:517–535. https://doi.org/10.1002/sim.5941

105. Platt J, others (1999) Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. Adv large margin Classif 10:61–74

106. Zadrozny B, Elkan C (2002) Transforming classifier scores into accurate multiclass probability estimates. In: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. pp 694–699

107. Steyerberg EW (2009) Clinical Usefulness. In: Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. Springer New York, New York, NY, pp 281–297

108. Vickers AJ, Elkin EB (2006) Decision curve analysis: a novel method for evaluating prediction models. Med Decis Mak an Int J Soc Med Decis Mak 26:565–574. https://doi.org/10.1177/0272989X06295361

109. Steyerberg EW, Harrell FEJ (2016) Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol 69:245–247. https://doi.org/10.1016/j.jclinepi.2015.04.005

110. Steyerberg EW, Harrell FE, Borsboom GJJM, et al (2001) Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. J Clin Epidemiol 54:774–781. https://doi.org/10.1016/S0895-4356(01)00341-9

111. Ramspek CL, Jager KJ, Dekker FW, et al (2020) External validation of prognostic models: what, why, how, when and where? Clin Kidney J 14:49–58. https://doi.org/10.1093/ckj/sfaa188

112. Shillan D, Sterne JAC, Champneys A, Gibbison B (2019) Use of machine learning to analyse routinely collected intensive care unit data: a systematic review. Crit Care 23:284. https://doi.org/10.1186/s13054-019-2564-9

113. Fleuren LM, Thoral P, Shillan D, et al (2020) Machine learning in intensive care medicine: ready for take-off? Intensive Care Med. 46:1486–1488

114. Laird P, Wertz A, Welter G, et al (2021) The critical care data exchange format: a proposed flexible data standard for combining clinical and high-frequency physiologic data in critical care. Physiol Meas 42:

115. Moons KGM, Altman DG, Reitsma JB, et al (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. Ann Intern Med 162:W1–W73. https://doi.org/10.7326/M14-0698

116. Random Forests. https://www.stat.berkeley.edu/~breiman/RandomForests/. Accessed 19 May 2022

117. Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems. pp 4766–4775

118. Cava W La, Bauer C, Moore JH, Pendergrass SA (2019) Interpretation of machine learning predictions for patient outcomes in electronic health records. AMIA . Annu Symp proceedings AMIA Symp 2019:572–581

119. Lundberg SM, Nair B, Vavilala MS, et al (2018) Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat Biomed Eng 2:749–760. https://doi.org/10.1038/s41551-018-0304-0

120. Janssen A, Hoogendoorn M, Cnossen MH, et al (2022) Application of SHAP values for inferring the optimal functional form of covariates in pharmacokinetic modeling. CPT Pharmacometrics \& Syst Pharmacol 11:1100–1110. https://doi.org/https://doi.org/10.1002/psp4.12828

2

Objectives

2.1 General aim

The general aim of this thesis is to develop and validate prediction models for kidney function in critically ill adults by applying machine-learning methods to routinely collected clinical data and translate the developed models into decision support applications for future usage in the clinical practice.

2.2 Research objectives

This thesis is divided into 3 primary sections, with each concentrating on a common medical condition or measurement in the intensive care unit (ICU). The first part deals with prediction of development and recovery of acute kidney injury (AKI) (Chapter 3, Chapter4). The second part covers daily prediction of augmented renal clearance (ARC) (Chapter 5). The third part focuses on daily creatinine clearance (CrCl) (Chapter 6, Chapter 7, Chapter 8, and Chapter 9).

Objective 1: the first objective of this thesis includes **(I)** the external validation of an existing machine-learning prediction model for the onset of AKI during the first week of ICU stay in critically ill adults from a large independent dataset. (**Chapter 3**), **(II)** development and validation of machine-learning prediction models for AKI recovery at hospital discharge in critically ill adults with ICU-acquired acute kidney injury stage 3, and to compare the performance with the most studied biomarker for acute kidney injury. (**Chapter 4**)

Objective 2: the second objective of this thesis is to externally validate an existing machine-learning prediction model for the presence of ARC on the next ICU day in critically ill coronavirus disease 19 patients. (**Chapter 5**)

Objective 3: the third objective of this thesis includes **(I)** to investigate the daily kidney function fluctuations in critically ill adults (**Chapter 6**), **(II)** to develop and validate machine-learning models for daily prediction of short-term CrCl in critically ill adults, with a comparison with the reference assuming CrCl remains unchanged (**Chapter 7**), **(III)** to perform an observational prospective study to compare the predictive performance of the developed CrCl prediction models with ICU physicians (**Chapter 8**), and **(IV)** to develop a prototype software with integration of the developed CrCl models that can visualize the prediction results along with explanations for CrCl of the next ICU day (**Chapter 9**).

3

External validation of the AKIpredictor in critically ill adults

Adapted from: **Chao-Yuan Huang**, Fabian Güiza, Greet De Vlieger, Geert Meyfroidt. “External validation of the AKIpredictor in critically ill adults”. *Intensive Care Medicine*. 2022;48(7):952-953. doi:10.1007/s00134-022-06746-6

3.1 Letter to the Editor

The early detection of acute kidney injury (AKI) in the intensive care unit (ICU) remains challenging. AKI is defined by a rise in serum creatinine and/or a reduced urine output, both late markers of the potential underlying kidney damage. There is a need to identify patients with high risk of developing AKI, so that early preventive or therapeutic interventions could be employed or studied, such as reducing or avoiding nephrotoxic drugs and optimizing blood pressure and fluid balance.

The AKIpredictor includes a series of models that predict AKI development within one week after ICU admission by using clinical data available at different time points. The AKIpredictor models demonstrated good performance, outperformed commonly used AKI biomarker neutrophil gelatinase-associated lipocalin (NGAL) in a validation cohort [1], and are available as an online calculator (www.akipredictor.com) [2]. When compared against ICU physicians in a prospective clinical trial, the AKIpredictor achieved similar discrimination and higher net benefit, thus outperforming physicians [3]. In the present study, we validated the AKIpredictor in a more recent ICU setting on a large heterogeneous cohort from the University Hospitals Leuven, included in the M@tric database [4], containing high-quality and complexly interrelated data from all adult patients annually admitted to the ICU from 2013 to 2018. Approval for the use of these patient data was obtained from the Ethics Committee of University Hospitals Leuven (S61364).

Kidney Disease: Improving Global Outcomes (KDIGO) serum creatinine and not urine output criteria were used to classify AKI stage 1 or higher. For better interpretation and comparison with earlier studies, predictive performance was examined by using the same evaluation metrics: area under the receiving operating characteristics (AUROC) curve (including sensitivity and specificity), calibration plot (including calibration slope (CS), and calibration-in-the-large (CITL)), and decision curve analysis.

Of the 20930 patients in the validation cohort, 11290 patients met the inclusion criteria (**Figure 3.1**). In total, 1239 patients (10.97%) developed AKI within one week after ICU admission. Complete descriptive statistics are available in **Table 3.1**. As shown in **Figure 3.2**, the AKIpredictor demonstrates good discrimination (AUROC: 0.75; CS: 0.75; CITL: -0.04), comparable to (but slightly lower than) the original study (AUROC: 0.80; CS: 0.78; CITL: -0.01) [1]. At the classification threshold of 14.5% that maximized sensitivity and specificity of 64% and 82% in the original study, the sensitivity and specificity are 55.53% and 79.64% respectively. Decision curve analysis demonstrates potential clinical usefulness across a broad range of classification thresholds (7.07% – 38.38%). The original classification threshold of 14.5% can still retain clinical utility.

The results demonstrate robustness of the AKIpredictor models, even while the original database where the models have been learned dates back ten years [5], during which clinical environment and health care processes evolved considerably. Based on these findings, the AKIpredictor can be a promising tool to identify AKI patients at an early stage. In the future, AKIpredictor could be combined with biomarkers to enhance performance. Whether improved stratification of patients with higher risk of AKI can benefit their outcomes, requires future prospective studies.

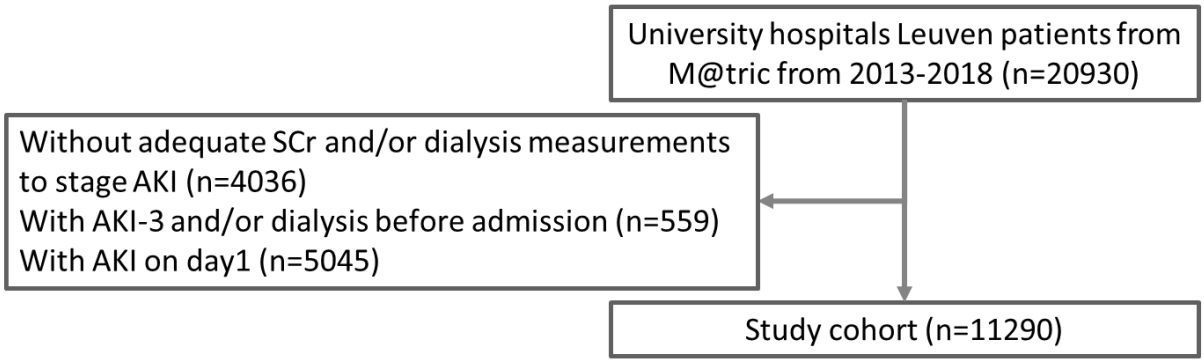


Figure 3.1 Study cohort. SCr, serum creatinine; AKI, acute kidney injury; AKI-3, acute kidney injury stage 3

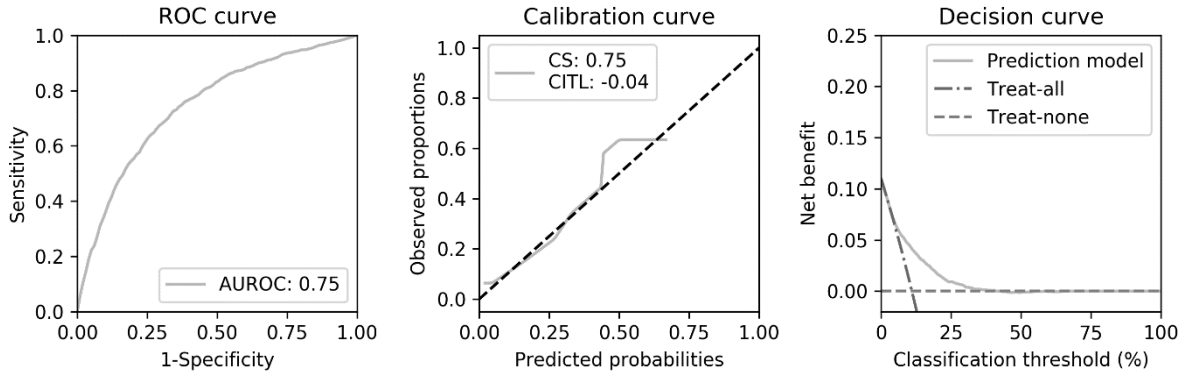


Figure 3.2 Predictive performance (left) ROC curve (middle) calibration curve (right) decision curve. ROC, receiver operating characteristic; CS, calibration slope; CITL, calibration-in-the-large

Table 3.1. Descriptive statistics

	All patients (n=11290)	AKI (n=1239)	Not AKI (n=10051)	p-value
Age, years, median (IQR)	62.88 (51.42 – 72.44)	67.75 (58.85 – 77.25)	62.25 (50.55 – 71.56)	<0.01
Gender male, number (%)	6719 (59.51)	708 (57.14)	6011 (59.80)	0.08
Height, cm, median (IQR)	170 (164 – 176)	169 (162 – 175)	170 (164 – 177)	<0.01
Weight, kg, median (IQR)	71 (60 – 82)	72 (61 – 83)	71 (60 – 82)	0.15
BMI, median (IQR)	24.57 (21.67 – 27.76)	24.97 (22.16 – 28.66)	24.49 (21.60 – 27.70)	<0.01
APACHE II score, median (IQR)	15 (12 – 19)	16 (15 – 23)	15 (12 – 18)	<0.01
Diabetes, number (%)	655 (5.80)	114 (9.20)	541 (5.38)	<0.01
Baseline serum creatinine, mg/dl, median (IQR)	0.97 (0.79 – 1.01)	0.96 (0.78 – 1.00)	0.97 (0.79 – 1.02)	<0.01
Elective admission, number (%)	3577 (31.68)	391 (31.56)	3186 (31.70)	0.95
Sepsis on ICU admission, number (%)	1523 (13.49)	222 (17.92)	1301 (12.94)	<0.01
Mechanical ventilation on day1, number (%)	5996 (53.11)	831 (67.07)	5165 (51.39)	<0.01
Mechanical hemodynamic support on ICU admission, number (%)	33 (0.29)	17 (1.37)	16 (0.16)	<0.01
Pharmacological hemodynamic support on ICU admission, number (%)	5216 (46.20)	740 (59.73)	4476 (44.53)	<0.01
Deceased at ICU discharge, number (%)	665 (5.89)	233 (18.81)	432 (4.30)	<0.01
Length of stay in ICU, days, median (IQR)	3.00 (2.00 – 7.00)	6.00 (3.00 – 12.50)	3.00 (2.00 – 6.00)	<0.01
Deceased at hospital discharge, number (%)	1326 (11.74)	332 (26.80)	994 (9.89)	<0.01

Length of stay in hospital, days, median (IQR)	18.00 (10.00 – 35.00)	24.00 (13.50 – 47.00)	17.00 (10.00 – 34.00)	<0.01

AKI, acute kidney injury; BMI, body mass index; APACHE II score, Acute Physiology And Chronic Health Evaluation score; ICU, intensive care unit; P-values were calculated by Mann–Whitney U test and Fisher's Exact Test for continuous and categorical data respectively. Baseline serum creatinine was back calculated with the Modification of Diet in Renal Disease (MDRD) equation, with an assumed normal baseline estimated Glomerular filtration rate (eGFR) of 75 mL/min/1.73m2. Mechanical hemodynamic support was defined as the presence of LVAD, BiVAD, ECMO, IABP. Pharmacological hemodynamic support was defined as any vasopressor or inotropic medication. Mechanical ventilation (MV) was defined as any form of MV/assisted ventilation with or without PEEP, with or without muscle relaxants, spontaneous breathing with positive end-expiratory pressure (PEEP)

3.A. Appendix

3.A.1 Supplementary methods

Baseline serum creatinine was estimated by Modification of Diet in Renal Disease (MDRD) formula with an assumed normal glomerular filtration rate (GFR) of 75 mL/min per 1.73 m2.

3.A.2 Supplementary figures

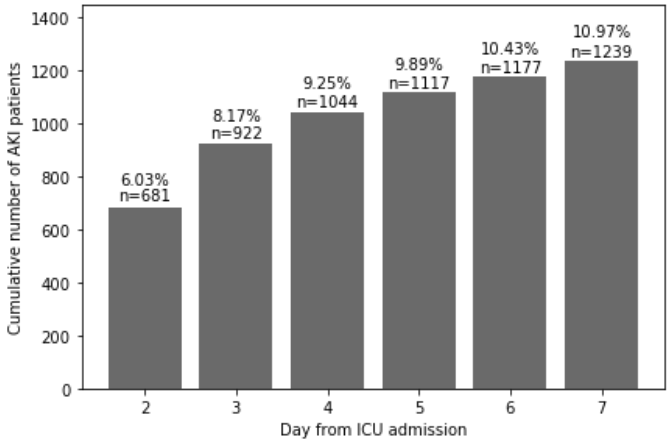


Figure 3.A.1 The cumulative number of AKI patients over time.

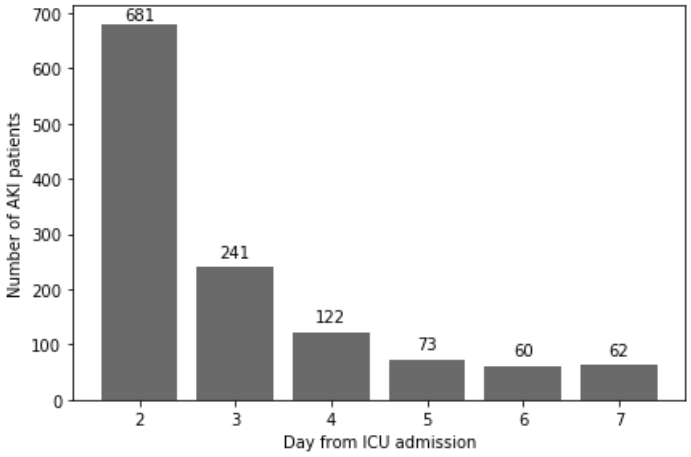


Figure 3.A.2 The number of patients with a new diagnosis of AKI based upon a rise in serum creatinine on each ICU day.

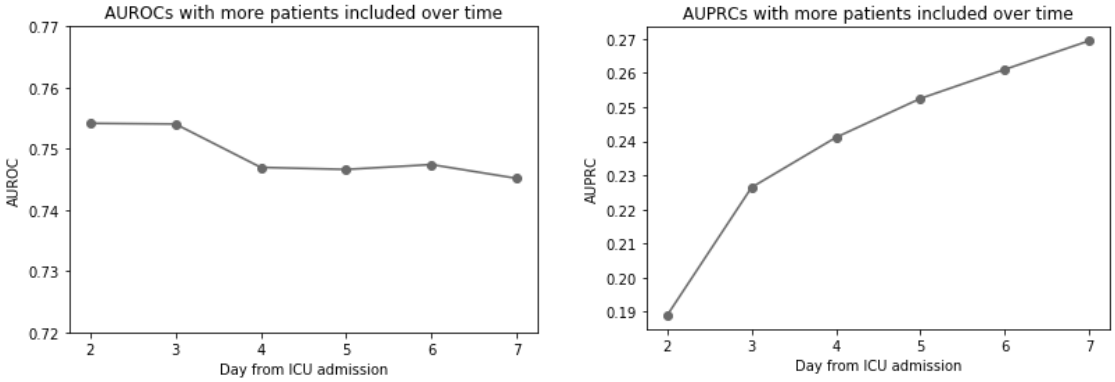


Figure 3.A.3 AUROC (left) and AUPRC (right) with more patients included over time. There was no statistically significant difference (DeLong's test p-value=0.54) between the ROC curve until day3 and the ROC curve until day4.

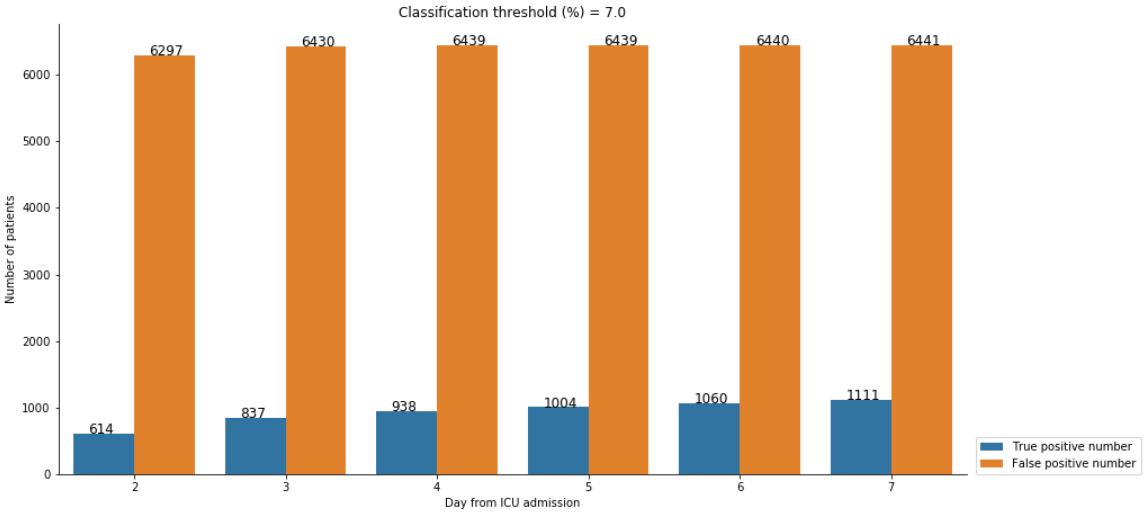


Figure 3.A.4 The cumulative number of true positive and false positive patients with classification threshold of 7.0 %.

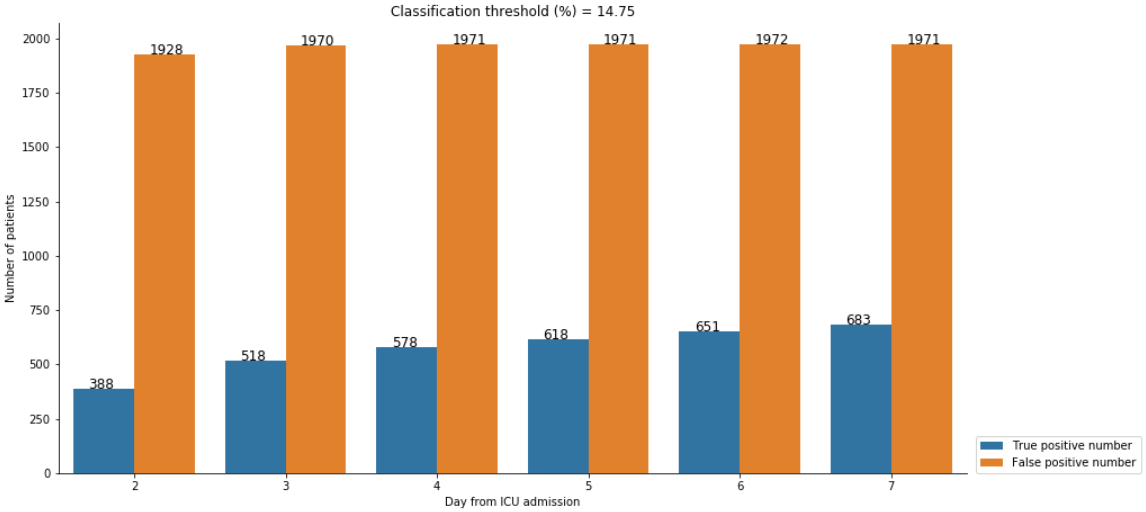


Figure 3.A.5 The cumulative number of true positive and false positive patients with classification threshold of 14.75 %.

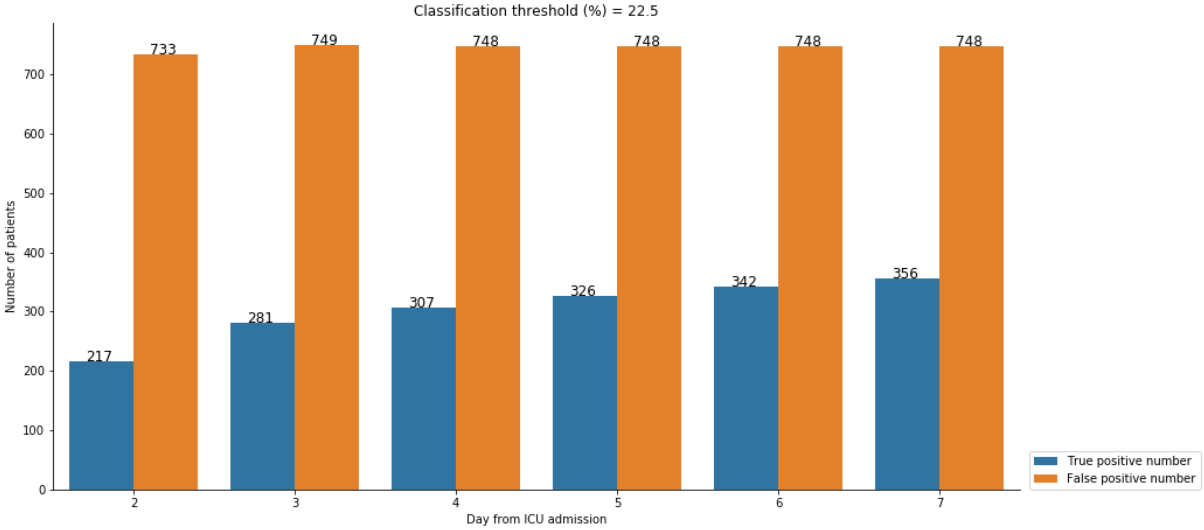


Figure 3.A.6 The cumulative number of true positive and false positive patients with classification threshold of 22.5 %.

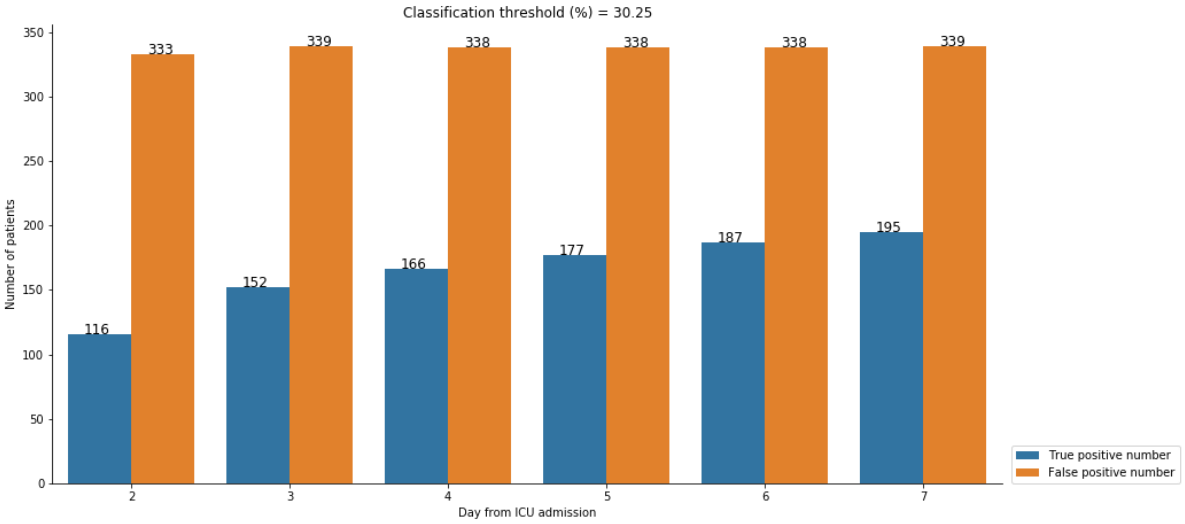


Figure 3.A.7 The cumulative number of true positive and false positive patients with classification threshold of 30.25 %.

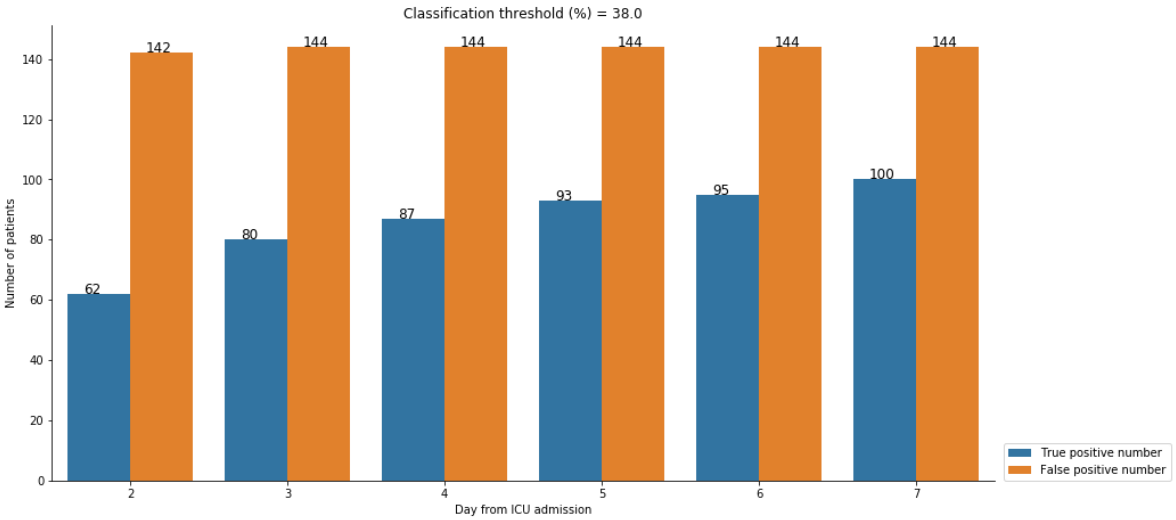


Figure 3.A.8 The cumulative number of true positive and false positive patients with classification threshold of 38.0 %.

GRANT SUPPORT AND CONFLICT OF INTEREST

C-Y.Huang receives funding from the Taiwanese Ministry of Education through a Taiwan-KU Leuven scholarship. G.Meyfroidt is funded by the Research Foundation, Flanders (FWO) as senior clinical investigator (1843118N).

The authors declare that they have no conflict of interest.

ACKNOWLEDGMENT AND PERSONAL CONTRIBUTION

Study concept and design	Meyfroidt, Güiza, De Vlieger
Data acquisition	**Huang**
Statistical analysis	**Huang**
Interpretation of results	**All authors**
Drafting of the manuscript	**Huang**
Manuscript revision	**All authors**
Principal investigator	Meyfroidt

The authors would like to thank Dominiek Cottem for the help in exporting data from the University Hospitals Leuven, included in the M@tric database.

BIBLIOGRAPHY

1. Flechet M, Güiza F, Schetz M, et al (2017) AKIpredictor, an online prognostic calculator for acute kidney injury in adult critically ill patients: development, validation and comparison to serum neutrophil gelatinase-associated lipocalin. Intensive Care Med 43:764–773. https://doi.org/10.1007/s00134-017-4678-3
2. AKI prediction. https://www.akipredictor.com/en/. Accessed 10 Feb 2022
3. Flechet M, Falini S, Bonetti C, et al (2019) Machine learning versus physicians' prediction of acute kidney injury in critically ill adults: A prospective evaluation of the AKIpredictor. Crit Care 23:. https://doi.org/10.1186/s13054-019-2563-x
4. M@tric project. https://www.matric.be/. Accessed 5 Oct 2021
5. Casaer MP, Mesotten D, Hermans G, et al (2011) Early versus Late Parenteral Nutrition in Critically Ill Adults. N Engl J Med 365:506–517. https://doi.org/10.1056/NEJMoa1102662

4

Development and validation of clinical prediction models for acute kidney injury recovery at hospital discharge in critically ill adults

Adapted from: **Chao-Yuan Huang**, Fabian Güiza, Greet De Vlieger, Pieter Wouters, Jan Gunst, Michael Casaer, Ilse Vanhorebeek, Inge Derese, Greet Van den Berghe, Geert Meyfroidt. "Development and validation of clinical prediction models for acute kidney injury recovery at hospital discharge in critically ill adults". *Journal of Clinical Monitoring and Computing*. Published online 2022. doi:10.1007/s10877-022-00865-7

Presented as:

- Poster presentation at the 32nd Annual Congress of the European Society of Intensive Care Medicine (ESICM), Berlin, Germany, October, 2019
- Poster presentation at the 40th Annual Congress of the International Symposium on Intensive Care and Emergency Medicine (ISICEM), Online, March, 2020

Abstract

Purpose: Acute kidney injury (AKI) recovery prediction remains challenging. The purpose of the present study is to develop and validate prediction models for AKI recovery at hospital discharge in critically ill patients with ICU-acquired AKI stage 3 (AKI-3).

Methods: Models were developed and validated in a development cohort (n=229) and a matched validation cohort (n=244) from the multicenter EPaNIC database to create prediction models with the least absolute shrinkage and selection operator (Lasso) machine-learning algorithm. We evaluated the discrimination and calibration of the models and compared their performance with plasma neutrophil gelatinase-associated lipocalin (NGAL) measured on first AKI-3 day (NGAL_AKI3) and reference model that only based on age.

Results: Complete recovery and complete or partial recovery occurred in 33.20% and 51.23% of the validation cohort patients respectively. The prediction model for complete recovery based on age, need for kidney replacement therapy (KRT), diagnostic group (cardiac/surgical/trauma/others), and sepsis on admission had an area under the receiver operating characteristics curve (AUROC) of 0.53. The prediction model for complete or partial recovery based on age, need for KRT, platelet count, urea, and white blood cell count had an AUROC of 0.61. NGAL_AKI3 showed AUROCs of 0.55 and 0.53 respectively. In cardiac patients, the models had higher AUROCs of 0.60 and 0.71 than NGAL_AKI3's AUROCs of 0.52 and 0.54. The developed models demonstrated a better performance over the reference models (only based on age) for cardiac surgery patients, but not for patients with sepsis and for a general ICU population.

Conclusion: Models to predict AKI recovery upon hospital discharge in critically ill patients with AKI-3 showed poor performance in the general ICU population, similar to the biomarker NGAL. In cardiac surgery patients, discrimination was acceptable, and better than NGAL. These findings demonstrate the difficulty of predicting non-reversible AKI early.

4.1 Introduction

Acute kidney injury (AKI) reflects an abrupt decline in renal function, resulting in a reduced capacity to regulate extracellular volume and to clear circulating substances. It is one of the most prevalent complications of critical illness and has a strong association with short- and long-term mortality. Patients with AKI suffer from longer lengths of stay, lower quality of life and heavier financial burden [1–5] The reported incidence of AKI in the intensive care unit (ICU) varies widely, depending on the population, and the use of different AKI definitions [6–10]. AKI criteria were first defined by the RIFLE criteria in 2004 and later modified by the Kidney Disease: Improving Global Outcomes (KDIGO) AKI criteria proposed in 2012 [11]. Current strategies for established AKI are mostly supportive, such as kidney replacement therapy (KRT) in case of metabolic complications or fluid overload, while the early initiation of KRT does not appear to be beneficial [12]. Predicting AKI recovery for critically ill patients could be useful so that clinicians can (I) identify patients with non-reversible AKI early, (II) provide patient and family counseling, and (III) make early post-discharge renal care plans.

Biomarkers have been studied to predict AKI development, but limited data exist on the possibility to predict AKI recovery [13–17]. Recently, the growth of electronic health records (EHR) created the possibility to use a large amount of data for clinical prediction models. Several models have been built to predict AKI development [18–25]. Outcome prediction tools in patients with AKI, have focused on predicting mortality instead of renal recovery [26–33].

Biomarkers and EHR-based models to predict AKI recovery [13–17, 34, 35] often show poor discrimination in the development cohort [35] and lack internal and/or external validation [13–16, 35] which could overestimate the performance and result in low generalizability. Most studies included only patients with AKI stage 3 or on KRT [13, 15–17, 35], and only two studies investigated all stages of AKI with AKI based upon the KDIGO criteria [14, 34]. The mortality and risk to develop end-stage kidney disease increases with AKI stages and thus predicting AKI recovery in AKI-3 patients is more clinically meaningful than prediction in less severe forms of AKI.

Here we develop and validate prediction models for AKI-3 recovery at hospital discharge by using routinely collected clinical data up to the first day of AKI-3 in general adult ICU patients. In addition, we evaluated the performance of the biomarker plasma neutrophil gelatinase associated lipocalin (NGAL) measured on the first day of AKI-3 (NGAL_AKI3) to predict AKI recovery and combined the clinical data model and NGAL model. All models were also evaluated in subgroups of patients with cardiac surgery and patients with sepsis on admission due to the high incidence of AKI [36, 37].

4.2 Methods

4.2.1 Prediction tasks and AKI definition

There are two prediction tasks in this study: 1) complete recovery and 2) complete or partial recovery at hospital discharge. We investigated both prediction tasks in ICU patients with AKI-3. AKI was diagnosed and staged based on the KDIGO serum creatinine criteria [11]. Baseline SCr was defined as the lowest SCr value three months up to one week before ICU admission for emergency admissions, and the lowest SCr value three months before ICU admission for elective admissions. If no pre-ICU SCr values were available, baseline SCr was

backcalculated with the Modification of Diet in Renal Disease (MDRD) equation with an assumption of a normal baseline estimated Glomerular filtration rate (eGFR) of 75 mL/min/1.73m^2 [38]. Complete recovery was defined as the absence of any stage of AKI and being alive without KRT at hospital discharge. Partial recovery was defined as AKI stage 1 or 2 and being alive without KRT at hospital discharge.

4.2.2 Study database

This study was a secondary analysis of the EPaNIC multicenter randomized controlled trial (RCT). The EPaNIC study compared early and late initiation of parenteral nutrition in 4640 adults admitted to seven ICUs between August 2007 and November 2010 [39]. Written informed consent was collected from all patients or their designated representatives. The institutional review board at each participating center and Belgium authorities approved the protocol and the consent forms (file number S50404). The present study is a preplanned continuation of a first study on AKI development [18] and makes use of the same predefined matched development and validation cohorts. As detailed in Appendix 4.A.1.7, the original EPaNIC database was split into cohorts matched for demographics, severity of illness, feeding strategy and relevant clinical outcomes. Patients were excluded if they had 1) history of end-stage kidney disease, 2) baseline SCr ≥ 4 mg/dL, 3) no available SCr measurements to stage AKI, 4) no monitoring or no medication data. For the present study, only patients with AKI-3 during their ICU stay were included.

4.2.3 Features for AKI recovery prediction

Only data up to the first day of AKI-3 in the ICU were used to predict AKI-3 recovery (**Figure 4.1**). The following data were used: 1) Admission data (only one value available during ICU stay): demographics, diagnosis, comorbidities, 2) Time-series data: heart rate, blood pressure, blood temperature, blood gas analysis data, laboratory test data, interventions, illness severity scores, 3) With the following medication data on the previous day of the first AKI-3 (true/false): non-steroid anti-inflammatory drugs, antiviral drugs, antifungal drugs, diuretics, vancomycin, β-lactam antibiotic, radiocontrast, aminoglycosides, ciclosporin/tacrolimus, ace inhibitors, vasopressors/inotropes, 4) Time-relevant data: Number of days with aforementioned medication data, first day of AKI-3. A list of all considered data is available at Appendix 4.A.1.1. Static data were retrieved from the EPaNIC study database (Filemaker Pro®; FileMaker Inc, FileMaker International), while the remaining data were retrieved from the patient data management system database (Microsoft SQL Server®; Microsoft®, Redmond, Washington, USA).

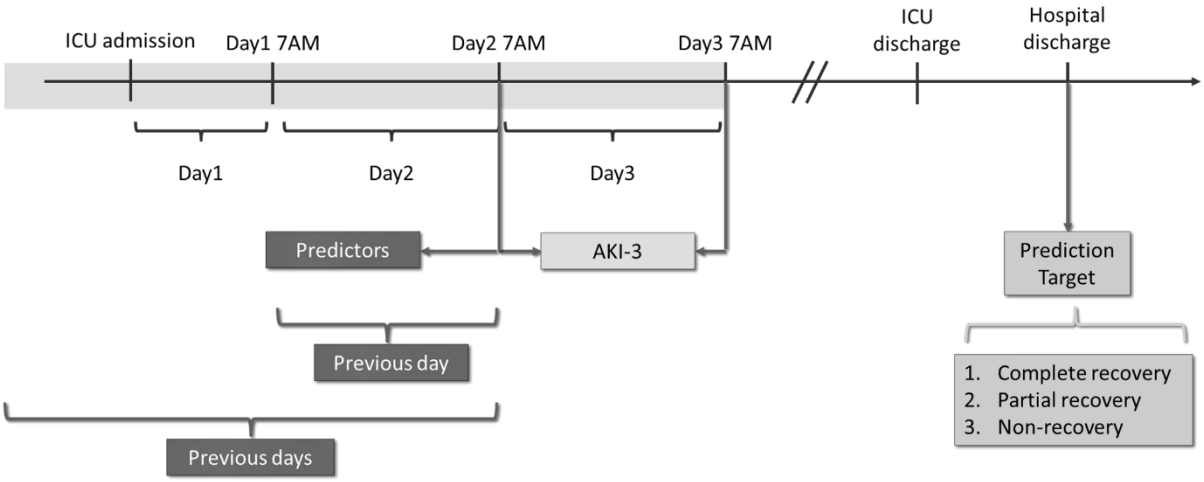


Figure 4.1 Example indicating the temporal relationship between the predictors and the prediction target. In the example, the patient developed AKI-3 on day3, and only data up to day3 were used as predictors. The prediction target of the model was the AKI recovery status (complete recovery, partial recovery, or non-recovery) at hospital discharge.

The minimum, maximum, mean, standard deviation, and linear regression slope were used to create derived characteristics from the time-series data. All the features with more than 10% missing values were excluded. For the remaining features, missing values were imputed with the mean for continuous data and the mode for categorical data. Finally, continuous data were standardized to zero mean and unit variance, and categorical data without order relation were converted into a form with binary data for each category.

4.2.4 Machine-learning algorithm and feature selection methods

The prediction models were trained with least absolute shrinkage and selection operator (Lasso), a machine-learning algorithm (Appendix 4.A.1.2), with features selected based on their known associations with renal function, correlation-based feature selection method (CfsSubsetEval) from Weka 3.8.4 system [40], and permutation importance measurements [41] (Appendix 4.A.1.3). In addition, the clinical meaningfulness of the selected features was confirmed with thorough discussion with 2 experienced ICU physicians.

4.2.5 Biomarker NGAL alone and in combination with prediction models

Plasma NGAL was measured on the first day of AKI-3 using the Human Lipocalin-2/NGAL Quantikine ELISA Kit (R&D Systems, Inc., Abingdon, UK). This kit uses a quantitative sandwich enzyme immunoassay technique with a measurable range from 0.2 to 10 ng/mL. We evaluated the predictive performance of plasma NGAL alone and in combination with the developed prediction models. NGAL_AKI3 was also assessed in the subgroups of patients with cardiac surgery and patients diagnosed with sepsis on ICU admission using the criteria of the American College of Chest Physicians–Society of Critical Care Medicine [42].

4.2.6 Metrics for predictive performance

Predictive performance was evaluated in terms of discrimination and calibration. Discrimination was reported by visualizing the receiver operating characteristics (ROC) curve with indication of the area under the receiver operating characteristics curve (AUROC). The closer the AUROC is to one, the better the discrimination. Calibration was evaluated visually and with the calibration slope, and calibration in the large [43]. A well-

calibrated model should have a calibration plot close to the diagonal axis, a calibration slope close to one, and a calibration in the large close to zero.

4.2.7 Model validation

At the model development stage, we internally validated the models by 100 repetitions of stratified 10-fold cross validation. At the model validation stage, models trained on the development cohort were applied on the previously unseen matched validation cohort to assess the generalizability. To examine the model usefulness, the developed models were compared with the reference models that were only based on age. Models were also evaluated in the predefined subgroups.

4.2.8 Descriptive statistical analysis and software used

All analyses were done in python 3.7.4 with the scikit-learn library 0.23.1. Continuous data were presented as means and interquartile ranges (IQR) while categorical data were expressed as numbers and percentages (%). To evaluate statistical significance of differences, Mann–Whitney U test and Fisher's Exact Test were used for continuous and categorical data respectively. A two-tailed P-value less than or equal to 0.05 was considered statistically significant.

4.3 Results

4.3.1 Study cohort: development and validation cohort

In total, 473 patients were included in this study (**Figure 4.2**). No significant difference between development (n=229) and validation cohorts (n=244) was observed in terms of baseline characteristics, AKI-relevant variables, and patient outcomes (**Table 4.A.1**) except for the surgery/trauma ($p<0.01$), transplant diagnostic groups ($p=0.01$), and early parenteral nutrition strategies ($p=0.02$). In the development and validation cohorts, 43.67% (100/229) and 43.03% (105/244) died before hospital discharge. Additionally, 37.55% (86/229) and 33.20% (81/244) patients completely recovered from AKI-3, while 51.09% (117/229) and 51.23% (125/244) patients completely or partially recovered from AKI-3. Age was the only patient characteristic consistently found significantly different between the recovery and non-recovery groups in both development and validation cohorts for both prediction tasks (**Table 4.A.2** and **Table 4.1**).

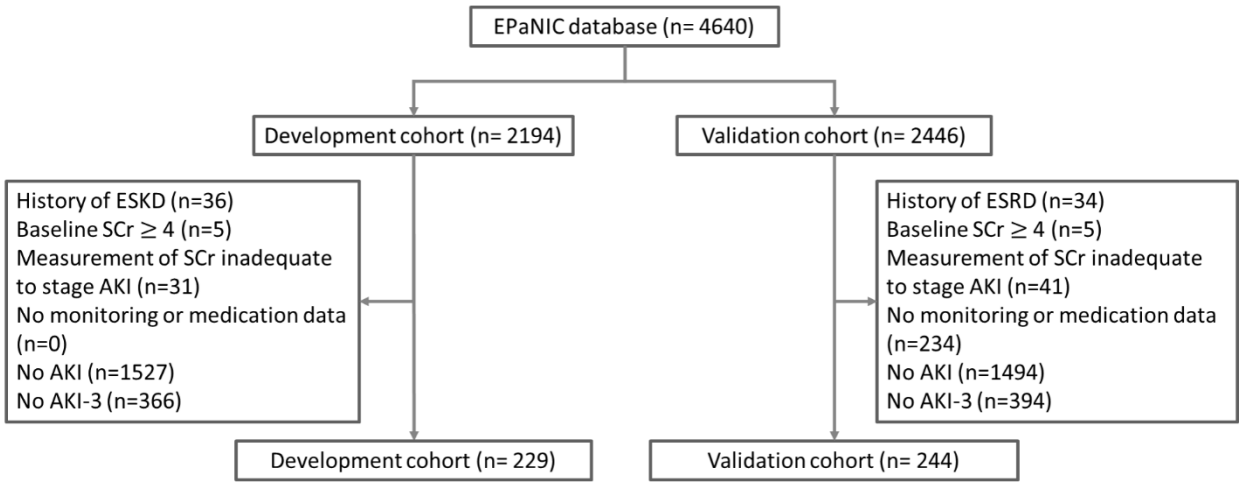


Figure 4.2 Consort diagram. ESKD, end-stage kidney disease; SCr, serum creatinine; AKI, acute kidney injury; AKI-3, acute kidney injury stage 3

Table 4.1 Patient characteristics and clinical outcomes of the two groups stratified by AKI recovery status in the validation cohort

	Validation cohort (n=244)	**Complete recovery (n=81)**	**Not complete recovery (n=163)**	**p-value**	**Complete or partial recovery (n=125)**	**Not complete or partial recovery (n=119)**	**p-value**
Age, years, median (IQR)	69.24 (57.87 – 76.45)	64.28 (53.87 – 73.23)	70.94 (60.05 – 78.22)	<0.01	65.02 (56.00 – 73.48)	71.86 (59.64 – 78.85)	0.01
Gender male, number (%)	148 (60.66)	48 (59.26)	100 (61.35)	0.78	72 (57.60)	76 (63.87)	0.36
BMI, median (IQR)	25.95 (23.03 – 29.39)	26.30 (23.41 – 30.37)	25.51 (22.70 – 29.26)	0.09	26.30 (23.39 – 30.37)	25.39 (22.70 – 28.82)	0.03
With true baseline, number (%)	133 (54.51)	46 (56.79)	87 (53.37)	0.68	60 (48.00)	73 (61.34)	0.04
Baseline serum creatinine, mg/dl, median (IQR)	0.97 (0.76 – 1.09)	0.98 (0.80 – 1.13)	0.96 (0.75 – 1.06)	0.12	0.96 (0.77 – 1.07)	0.97 (0.75 – 1.16)	0.53
Malignancy, number (%)	54 (22.13)	15 (18.52)	39 (23.93)	0.41	22 (17.60)	32 (26.89)	0.09
Chronic Kidney	60 (24.59)	22 (27.16)	38 (23.31)	0.53	25 (20.00)	35 (29.41)	0.10

Disease, number (%)							
Diabetes, number (%)	63 (25.82)	15 (18.52)	48 (29.45)	0.09	28 (22.40)	35 (29.41)	0.24
Elective admission, number (%)	49 (20.08)	17 (20.99)	32 (19.63)	0.87	22 (17.60)	27 (22.69)	0.34
Early parenteral nutrition strategies, number (%)	108 (44.26)	37 (45.68)	71 (43.56)	0.79	56 (44.80)	52 (43.70)	0.90
Sepsis on ICU admission, number (%)	146 (59.84)	43 (53.09)	103 (63.19)	0.17	68 (54.40)	78 (65.55)	0.09
Mechanical hemodynamic support on ICU admission, number (%)	20 (8.20)	7 (8.64)	13 (7.98)	1.00	7 (5.60)	13 (10.92)	0.16
Pharmacological hemodynamic	228 (93.44)	79 (97.53)	149 (91.41)	0.10	114 (91.20)	114 (95.80)	0.20

support on ICU admission, number (%)							
Cardiac surgery diagnostic group, number (%)	65 (26.64)	25 (30.86)	40 (24.54)	0.36	34 (27.20)	31 (26.05)	0.89
Medical diagnostic group, number (%)	76 (31.15)	28 (34.57)	48 (29.45)	0.46	39 (31.20)	37 (31.09)	1.00
Neuro diagnostic group, number (%)	5 (2.05)	2 (2.47)	3 (1.84)	1.00	2 (1.60)	3 (2.52)	0.68
Surgery/Trauma diagnostic group, number (%)	89 (36.48)	23 (28.40)	66 (40.49)	0.07	42 (33.60)	47 (39.50)	0.35
Transplant diagnostic	9 (3.69)	3 (3.70)	6 (3.68)	1.00	8 (6.40)	1 (0.84)	0.04

group, number (%)							
First AKI-3 day, days, median (IQR)	2.00 (1.00 – 4.00)	2.00 (1.00 – 3.00)	2.00 (1.00 – 4.00)	0.75	2.00 (1.00 – 3.00)	2.00 (1.00 – 4.00)	0.18
With KRT when first AKI-3 in ICU, number (%)	161 (65.98)	51 (62.96)	110 (67.48)	0.57	74 (59.20)	87 (73.11)	0.03
Deceased at ICU discharge, number (%)	81 (33.20)	0 (0.00)	81 (49.69)	<0.01	0 (0.00)	81 (68.07)	<0.01
LOS in ICU, days, median (IQR)	15.00 (8.00 – 30.00)	20.00 (11.00 – 32.00)	14.00 (7.00 – 29.00)	0.02	17.00 (10.00 – 29.00)	14.00 (8.00 – 30.50)	0.58
With KRT at ICU discharge, number (%)	94 (38.52)	9 (11.11)	85 (52.15)	<0.01	17 (13.60)	77 (64.71)	<0.01
Deceased at hospital discharge, number (%)	105 (43.03)	0 (0.00)	105 (64.42)	<0.01	0 (0.00)	105 (88.24)	<0.01

LOS in hospital, days, median (IQR)	36.00 (16.00 – 63.00)	52.00 (30.00 – 75.00)	26.00 (12.00 – 54.00)	<0.01	44.00 (27.00 – 67.00)	20.00 (9.00 – 48.00)	<0.01
With KRT at hospital discharge, number (%)	77 (31.56)	0 (0.00)	77 (47.24)	<0.01	0 (0.00)	77 (64.71)	<0.01

Abbreviations: AKI-3, acute kidney injury stage 3; ICU, intensive care unit; IQR, interquartile range; LOS, length of stay; KRT, kidney replacement therapy.

4.3.2 Features selected for AKI recovery

We removed 36.67% (154/420) features with 10% missingness. The remaining 266 features were subjected to the feature selection methods, as shown in **Figure 4.A.1**. The majority of the excluded features were generated with feature engineering techniques used to capture the dynamics of the signals. Based on known associations with renal function, the correlation-based feature selection method, and permutation importance, five features were chosen for complete recovery prediction: age, sepsis on admission, surgery/trauma diagnostic group, KRT on the first AKI-3 day in ICU, and cardiac surgery diagnostic group. Following the same strategy, five features were selected for complete or partial recovery: age, KRT on the first AKI-3 day in ICU, minimum platelet count before first AKI-3, maximum urea before first AKI-3, and maximum white blood cell count before first AKI-3. Univariate analyses of the selected features and the two prediction targets in the development and validation cohorts are presented in **Table 4.A.3** and **Table 4.A.4**. Comparisons of features between development and validation cohorts are presented in **Table 4.A.5** and **Table 4.A.6**. There were no missing features for complete recovery prediction and a maximum of 5.24% and 1.64% missing features respectively in the development and validation cohorts for complete or partial recovery prediction (**Table 4.A.7**).

4.3.3 Model performance: complete recovery prediction

In the validation cohort (**Table 4.2**, **Figure 4.3**, and **Table 4.A.8**), the model demonstrated an AUROC of 0.53, a calibration slope of 0.27, and a calibration in the large of -0.07. In the subgroup analyses (**Figure 4.A.2** and **Figure 4.A.3**), the model had an AUROC of 0.60 and 0.56 for cardiac and septic patients respectively. As illustrated in **Figure 4.4**, the predicted probabilities of complete recovery were not significantly different in the patients with recovery as compared to those without recovery in the validation cohort nor in the cardiac and septic patients separately (p-values for general ICU patients, cardiac patients, and septic patients were 0.43, 0.17, and 0.27). The reference model that was only based on age resulted in higher AUROCs of 0.62 for both general ICU patients and septic patients and a lower AUROC of 0.56 in cardiac patients, as shown in **Figure 4.5**.

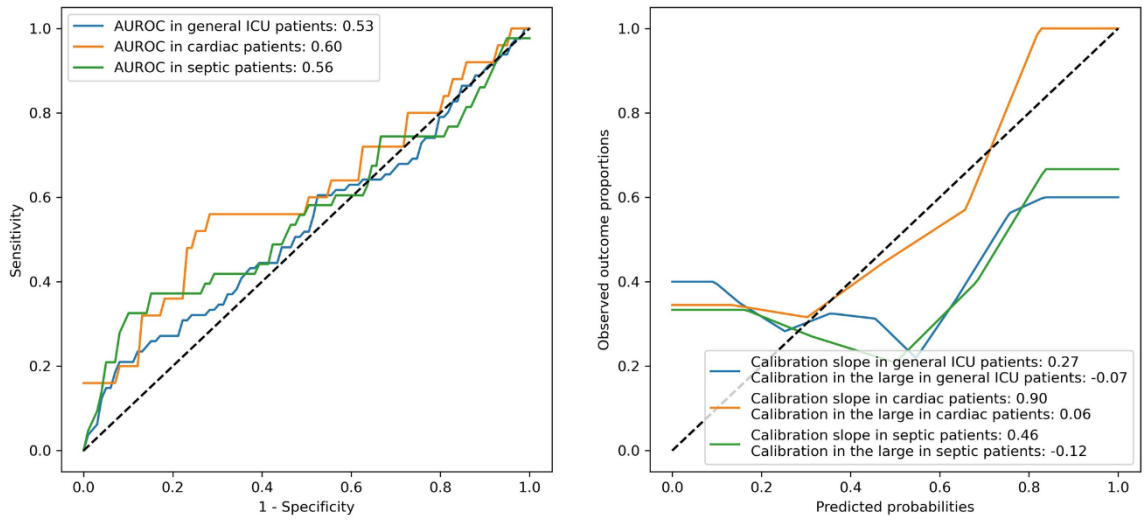


Figure 4.3 Model performance for complete recovery prediction in the validation cohort. (left) ROC curve (right) calibration curve (blue) in general ICU patients (orange) in cardiac patients (green) in septic patients

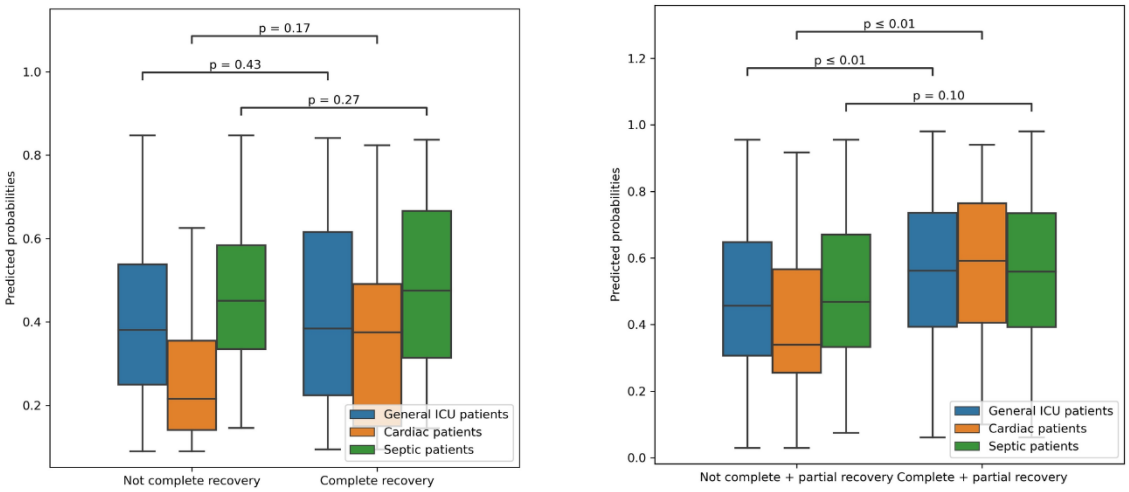


Figure 4.4 Predicted probabilities of prediction model for (left) complete recovery and (right) complete or partial recovery in the validation cohort. Reported p values are calculated based on Mann–Whitney U test. (blue) in general ICU patients (orange) in cardiac patients (green) in septic patients

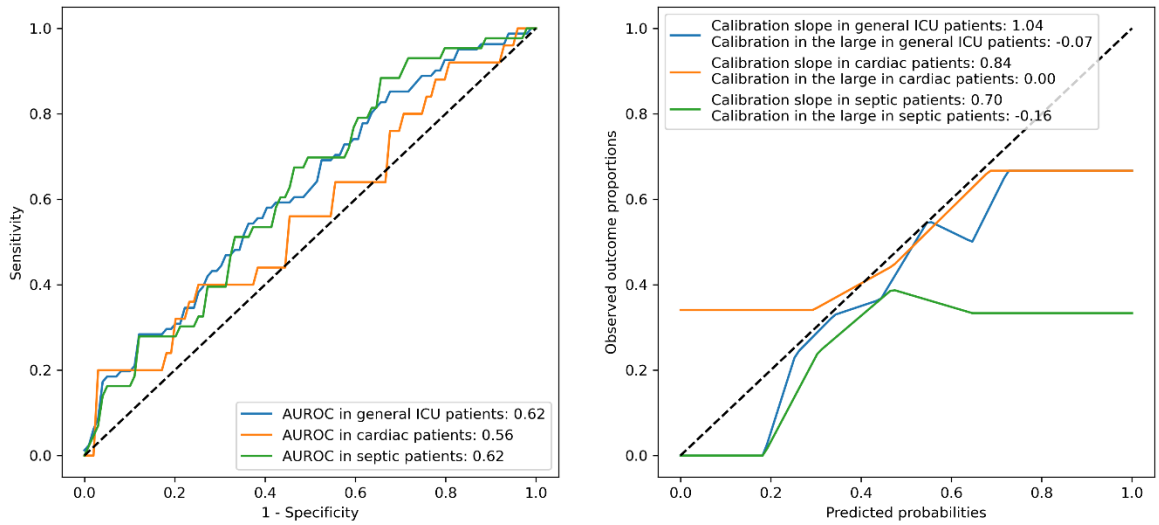


Figure 4.5 The performance of the reference model only based on age for complete recovery prediction in the validation cohort. (left) ROC curve (right) calibration curve (blue) in general ICU patients (orange) in cardiac patients (green) in septic patients

4.3.4 Model performance: complete or partial recovery prediction

In the validation cohort (**Table 4.2**, **Figure 4.6**, and **Table 4.A.8**), the model demonstrated an AUROC of 0.61, a calibration slope of 0.32, and calibration in the large close to zero. In the subgroup analysis (**Figure 4.A.4** and **Figure 4.A.5**), the AUROCs were 0.71 and 0.58 for cardiac and septic patients respectively. As depicted in **Figure 4.4**, a significant difference of predicted probabilities between the patients with recovery as compared to those without recovery was only found for general ICU patients and cardiac patients in the validation cohort (p-values for general ICU patients, cardiac patients, and septic patients were <0.01, <0.01, and 0.10). The reference model that was only based on age showed similar AUROCs of 0.60 and 0.57 for general ICU patients and septic patients and a lower AUROC of 0.57 in cardiac patients, as indicated in **Figure 4.7**.

Table 4.2 Summary of discrimination for prediction model, biomarker NGAL measured on first AKI-3 day, and prediction model with biomarker NGAL measured on first AKI-3 day

			AUROC	
			Development cohort	**Validation cohort**
Complete recovery	**Prediction model**	**General ICU patients**	0.73 ± 0.01	0.53
		Cardiac patients	0.67 ± 0.02	0.60
		Septic patients	0.69 ± 0.01	0.56
	Biomarker NGAL_AKI3	**General ICU patients**	0.53	0.55
		Cardiac patients	0.51	0.52
		Septic patients	0.51	0.63
	Prediction model with biomarker NGAL_AKI3	**General ICU patients**	0.72 ± 0.01	0.54
		Cardiac patients	0.65 ± 0.02	0.6
		Septic patients	0.69 ± 0.01	0.58
Complete or partial recovery	**Prediction model**	**General ICU patients**	0.74 ± 0.01	0.61
		Cardiac patients	0.75 ± 0.01	0.71
		Septic patients	0.70 ± 0.01	0.58
	Biomarker NGAL_AKI3	**General ICU patients**	0.52	0.53
		Cardiac patients	0.51	0.54
		Septic patients	0.53	0.61

	Prediction model with biomarker NGAL_AKI3	**General ICU patients**	0.73 ± 0.01	0.61
		Cardiac patients	0.76 ± 0.02	0.71
		Septic patients	0.69 ± 0.02	0.59

Mean ± standard deviation was obtained from 100 repetitions of stratified 10-fold cross validation. AUROC, area under the receiver operating characteristic curve; NGAL_AKI3, neutrophil gelatinase-associated lipocalin measured on the first AKI-3 day.

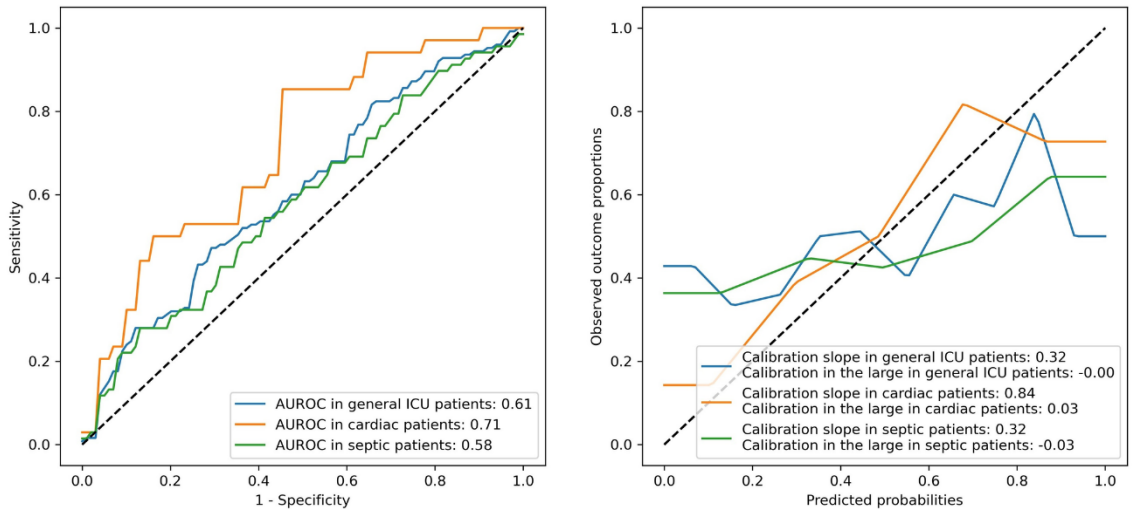


Figure 4.6 Model performance for complete or partial recovery prediction in the validation cohort. (left) ROC curve (right) calibration curve (blue) in general ICU patients (orange) in cardiac patients (green) in septic patients

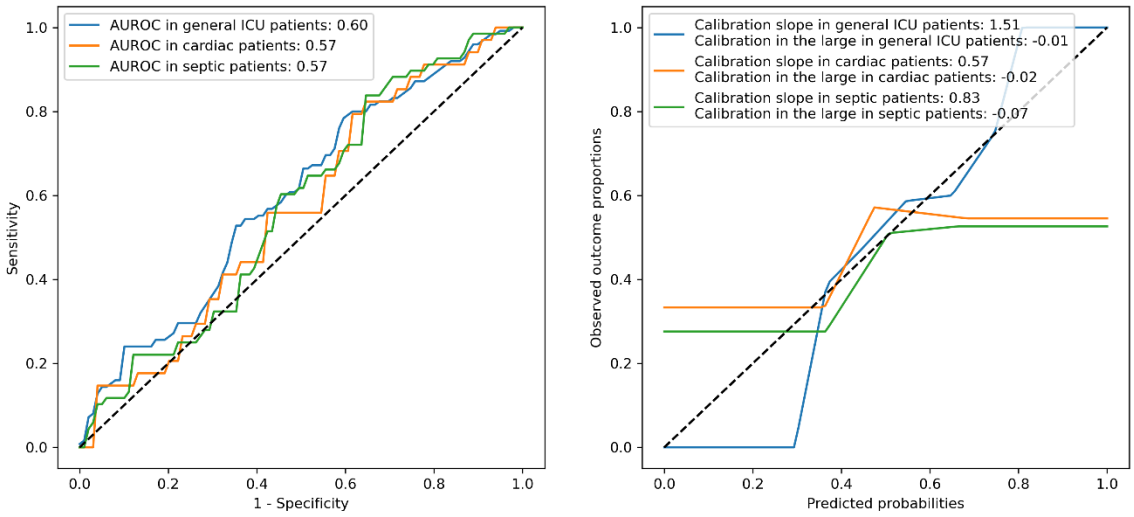


Figure 4.7 The performance of the reference model only based on age for complete or partial recovery prediction in the validation cohort. (left) ROC curve (right) calibration curve (blue) in general ICU patients (orange) in cardiac patients (green) in septic patients

4.3.5 Performance of NGAL alone and in combination with prediction models

There were 3.93% and 2.46% missing NGAL_AKI3 values in the development and validation cohorts. In the subgroup analysis (**Table 4.A.9** and **Table 4.A.10**), non-cardiac patients and septic patients had significantly higher NGAL_AKI3 levels compared to cardiac patients and non-septic patients (p-values<0.01). Table 2 summarizes the AUROCs of the prediction models, NGAL_AKI3, and their combination for complete recovery prediction and complete or partial recovery prediction. For complete recovery prediction in the validation cohort, the respective AUROCs were 0.53, 0.55, and 0.54 for the general ICU population. The AUROCs were 0.60, 0.52, and 0.60 for cardiac patients, and 0.56, 0.63, and 0.58 for septic patients. For complete or partial recovery prediction, the respective AUROCs were 0.61, 0.53, and 0.61 for the general ICU population. The AUROCs were 0.71, 0.54, and 0.71 for cardiac patients, and 0.58, 0.61, and 0.59 for septic patients. More details about the predictive performance are available at **Table 4.A.11**.

4.3.6 Lasso coefficients and permutation importance

Judging from the magnitude of the lasso coefficients and permutation importance, for both complete recovery and complete or partial recovery prediction tasks, age and KRT on the first AKI-3 day were consistently among the three most important features not only in the development cohort but also in the validation cohort (**Figure 4.A.6**, **Figure 4.A.7**, **Figure 4.A.8**, and **Figure 4.A.9**).

4.4 Discussion

In this retrospective study of critically ill adults with AKI-3 in the ICU, we developed and validated prediction models for complete recovery and complete or partial recovery at hospital discharge, with mean AUROCs of 0.53 and 0.61 in the validation cohort. The models' performance dropped substantially in the validation cohort. This finding highlights the need for validation in previously unseen patients to evaluate a model's generalizability and to prevent overoptimistic results.

In the subgroup analyses of the validation cohort, compared to the entire validation cohort, better model performance was observed in the cardiac patients, likely because one third of the EPaNIC database consisted of cardiac patients which is a more homogeneous subpopulation. Although the developed models demonstrated better performance for both prediction tasks in cardiac patients than general ICU patients of validation cohort, these findings still need prospective validation in independent cohorts. The features selected in this study were in line with prior work. To be more specific, age has been consistently considered an independent feature for renal recovery prediction [27, 28, 31, 32, 34, 35, 44]. Platelet count, shown to be associated with AKI, mortality, and renal recovery [45–49] was used to predict KRT dependence and mortality [50]. Surgical patients also have a higher probability of renal recovery, compared to medical patients [29, 44]. White blood cell count was strongly associated with renal recovery [30, 31, 33, 51]. Sepsis was indicated to be relevant to survival and renal recovery [52–54]. Finally, blood urea level, reflecting the underlying kidney function, has also been used for renal recovery prediction [31].

The observed significantly higher NGAL levels in septic patients confirms previous findings [55]. In the general ICU population and in cardiac patients, the developed complete or partial recovery prediction model achieved better discrimination than NGAL, which was in line with previous studies evaluating biomarkers for

AKI recovery [13–17]. Overall, including NGAL into the prediction models did not improve performance of the models in this study, as opposed to findings of previous studies [13–17]. However, four of these studies lacked an external validation cohort [13–16], and it is questionable whether the reported small performance improvements are clinically relevant and worth the expensive biomarker costs.

Our study confirms the low incidence of complete recovery from AKI-3 at hospital discharge, with the majority of AKI-3 patients deceased at hospital discharge, which is in line with the previous studies where mortality was found to be high [54, 56] in AKI patients with need of KRT. Although clinical prediction models have been widely studied for other research questions, there are limited number of prediction models for AKI recovery [13–17, 34, 35]. Among the studies with at least one AKI recovery prediction model, five studies reported fair discrimination but were specifically designed to examine predictive performance of biomarkers instead of clinical prediction models [13–17]. Itenov et al. [34] developed and validated an AKI recovery prediction model with fair performance, but they defined AKI recovery within a pre-defined timepoint of 28 days, when some patients are still recovering from critical illness in the ICU. Similarly, Lee et al. obtained poor discrimination with a 0.64 AUROC without external validation, which as discussed, might be an optimistic result [35].

Although we included granular clinical data accompanied with advanced feature selection methods and a machine learning algorithm, poor performance was found after validation in previously unseen patients, especially for complete recovery prediction, where the developed model did not outperform the reference model that was only based on age. This could be partially explained by the fact that although development and validation cohorts were matched in the beginning, significant difference between the cohorts was observed for surgical/trauma diagnostic group, one of the features used for complete recovery prediction, and for early parental nutrition strategies, which was associated with late recovery and more complications, compared to late parenteral nutrition strategies [39]. A systematic review reported similar performance drops with AUROCs below 0.7 at external validation in previous AKI mortality prediction studies [57]. This lack of accurate and generalizable prediction models for AKI outcome could indicate that conditions and events after diagnosis of AKI might be more important for AKI recovery than those preceding the diagnosis. Therefore, by integrating information after AKI-3 onset, the prediction accuracy might be improved [58], but at the expense of shortening the time window for clinical intervention.

Our study has many strengths. First, our prediction models are the first proposed for AKI-3 patients that predict AKI-3 recovery at hospital discharge. Second, both AKI-3 recovery and AKI definitions followed KDIGO consensus criteria with baseline SCr and patient outcome examined carefully. Third, validation in previously unseen patients were conducted to fairly report model performance and examine model generalizability. Fourth, not only static data but also time-series data with feature engineering techniques were considered. Fifth, the sparse models with only five features in each prediction task increase the usability. Sixth, transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) statement was used as a reporting benchmark [59]. Seventh, we compared the performance of the prediction models against the NGAL biomarker and evaluated the predictive ability of NGAL when being added to the developed models. Finally, both discrimination and calibration were reported to give complete understanding of model performance.

There are several limitations in our study. First, the small dataset prohibited us from using more advanced machine-learning algorithms such as random forest and neural network models that typically require larger sets of examples, but the lasso models with sparse features are easier to interpret and more robust when used in different patient populations. Second, due to the retrospective nature of this study, there was no impact analysis, and no causal relation conclusions can be drawn. Third, the study is based on a database from a RCT dating back to 2010, which limits its generalizability. Fourth, not all patients had true baseline SCr available, where it was substituted by back-calculation based on the MDRD equation recommended by KDIGO. However, our results remained unchanged with or without true baseline SCr (Appendix 4.A.5.2). Finally, development and validation cohort were not perfectly matched, in particular with regards to surgery/trauma and transplant diagnostic groups and timing of initiation of parenteral nutrition. Since the EPaNIC database was divided into matched development and validation cohorts before applying the inclusion and exclusion criteria of the present study, some imbalances could not be avoided in this subgroup of patients with AKI-3. In addition, the EPaNIC trial demonstrated that late initiation of parenteral nutrition resulted in a lower incidence of AKI, less need of KRT and shorter duration of KRT in the ICU. Therefore, it cannot be excluded that these imbalances might have contributed to the performance decrease in the validation cohort.

4.5 Conclusion

Using clinical data collected until the first day of AKI-3, we have built prediction models for AKI recovery at hospital discharge with AUROC of 0.61. NGAL measured on the first AKI-3 day demonstrated similar-to-worse performance than the developed models. Although the proposed models were developed and validated carefully by cross validation and in previously unseen validation cohort, these findings demonstrate the difficulty of predicting non-reversible AKI at an early stage. Larger studies with inclusion of more data after AKI-3 are needed to examine whether the machine-learning models can make better predictions.

4.A Appendix

4.A.1 Supplementary methods

4.A.1.1 Features for AKI recovery prediction

- Demographics: Age, Baseline serum creatinine, Gender (male/female), Height, Weight, Pre-admission hemoglobin, With true baseline (True/False)
- Diagnosis: Diagnostic group on admission (Cardiac surgery, Medical, neuro, Non-cardiac non-transplant surgery/Trauma, Transplant), Sepsis on admission
- Comorbidities: Chronic kidney disease (yes/no), Diabetes (yes/no), History of malignancy (yes/no)
- Blood gas analysis data: Partial pressure of carbon dioxide, Partial pressure of oxygen, Bicarbonate, Calculated oxygen saturation, Lactate, Glucose, Sodium, Potassium, pH value
- Lab test data: Bilirubin, Serum creatinine, C-reactive protein level, Lactate, Plasma urea level, Hematocrit, White blood cell count, Red blood cell count, Platelet count, Chloride
- Interventions: Mechanical hemodynamic support upon ICU admission (True/False), Pharmacological hemodynamic support upon ICU admission (True/False), With renal replacement therapy on first AKI-3 day (True/False), Respiratory support (True/False), Early initiation of parenteral nutrition strategies (True/False)
- Illness severity scores: Child-Pugh Score, New York Heart Association (NYHA) Score, Sequential Organ Failure Assessment (SOFA) score, Acute Physiology and Chronic Health Evaluation (APACHE) II score
- Medication data: non-steroid anti-inflammatory drugs, antiviral drugs, antifungal drugs, diuretics, vancomycin, β-lactam antibiotic, radiocontrast, aminoglycosides, ciclosporin/tacrolimus, ace inhibitors, vasopressors/inotropes

4.A.1.2 Machine-learning algorithm: least absolute shrinkage and selection operator (Lasso)

Lasso is a linear regression algorithm where the feature coefficients are fitted under the constraint that their magnitude should be as small as possible. Compared to unconstrainted linear regression algorithms, lasso models are more prone to turning coefficients to exactly zeros. Keeping only the most important features makes the lasso models easier to interpret and more generalizable to different datasets.

4.A.1.3 Feature selection methods: correlation-based feature selection method and permutation importance

The correlation-based feature selection method iteratively adds features to the subset. The features in the subset are highly correlated to the prediction target but have low correlation among themselves. Permutation importance measures the decrease in the model score when the feature's values are randomly shuffled. The bigger the drop in model score, the more dependent the model is on the feature and the more predictive the feature is for the outcome parameter. Unless stated otherwise, the area under the receiver operating characteristics curve (AUROC) was the score used to evaluate the permutation importance.

4.A.1.4 Features selected for complete recovery prediction

With renal replacement therapy on first day of AKI-3 was added based on known associations from the literatures (**Figure 4.A.1**). Cardiac diagnostic group was added for better interpretability. Age, surgery/trauma diagnostic group, with anti-inflammatory medication on the previous day of AKI-3, and sepsis on ICU admission were

selected by correlation-based feature selection method. With anti-inflammatory medication on the previous day of AKI-3 was removed due to unbalanced data, where only two patients were found to have anti-inflammatory medication on the previous day of AKI-3.

4.A.1.5 Features selected for complete or partial recovery prediction

Age was added based on literature (**Figure 4.A.1**). Eight features selected by correlation-based feature selection method: 1) With renal replacement therapy on first day of AKI-3, 2) Maximum of SOFA score before first AKI-3, 3) Number of days with vasopressors/inotropes medication before AKI-3, 4) Maximum of hemoglobin before first AKI-3, 5) Minimum of calculated oxygen saturation of the previous day of first AKI-3, 6) Linear regression slope of serum creatinine before first AKI-3, 7) Maximum of urea before first AKI-3, 8) Standard deviation of white blood cell count before first AKI-3. Four features were removed due to low permutation importance: 1) Minimum of calculated oxygen saturation of the previous day of first AKI-3, 2) Number of days with vasopressors/inotropes medication before AKI-3, 3) Linear regression slope of serum creatinine before first AKI-3, 4) Maximum of hemoglobin before first AKI-3. Finally, standard deviation of white blood cell count before first AKI-3 was replaced with maximum of white blood cell count before first AKI-3 due to the strong correlation (spearman correlation coefficient=0.73) and the easier calculation. Maximum of SOFA score before first AKI-3 was replaced with maximum of SOFA coagulation score before first AKI-3 due to relatively higher correlation (spearman correlation coefficient=0.60) than other SOFA sub scores, and it was further replaced with minimum of platelet count before first AKI-3 to have a broader continuous range with more details.

4.A.1.6 Early initiation of parenteral nutrition strategies

Since early initiation of parenteral nutrition has been shown to lead to late recovery and more complications in the EPaNIC study, we hypothesized that AKI recovery incidence might be influenced by the randomization of feeding strategies and that forcing the randomization of feeding strategies into the models would improve the prediction. To test the hypothesis, prediction models with and without forcing early initiation of parenteral nutrition strategies were built and their predictive performance was compared.

4.A.1.7 The database split including the differences in cohort sizes

The present study is a preplanned continuation of a first study on AKI development [18] and makes use of the same predefined matched development and validation cohorts. To perform the matching, the entire dataset was randomly permuted and split into two sets of equal size and the patient characteristics listed in **Table 4.A.12** were statistically compared. This procedure was repeated until two sets of 2194 and 2195 patients were obtained such that no statistical difference at the 0.05 level was observed between any of the characteristics. One set was thereafter referred to as the development cohort and the other as the validation cohort. However, as the majority of the data required for modeling was also available for the patients without monitoring information, it was later decided to include these extra 251 patients in the validation cohort.

4.A.2 Supplementary results

4.A.2.1 Results of forcing early initiation of parenteral nutrition strategies

No significantly differences were observed for early initiation of parenteral nutrition strategies for both prediction tasks in both cohorts (p-values were all above 0.05 as indicated in **Table 4.A.2** and **Table 4.1**). Forcing early initiation of parenteral nutrition strategies into the models did not influence the prediction performance, except

for the small changes in the subgroups of complete prediction task (**Table 4.A.13**). For the complete or partial complete prediction, since lasso algorithm turned the coefficient of early initiation of parenteral nutrition strategies to zero, no performance changes were found.

4.A.2.2 Model performance: complete recovery prediction

The internal validation in the development cohort of the complete recovery at hospital discharge models, resulted in 0.73 ± 0.01 AUROC, 1.16 ± 0.03 calibration slope, and 0.01 ± 0.01 calibration in the large (**Figure 4.A.10** and **Table 4.A.14**). As illustrated in **Figure 4.A.11**, a significant difference of predicted probabilities between the recovery and non-recovery groups was consistently observed in the development cohort (p-values for general ICU patients, cardiac patients, and septic patients were 0.02, 0.03, and <0.001).

4.A.2.3 Model performance in the development cohort: complete or partial recovery prediction

Upon internal validation in the development cohort, the models for complete or partial recovery at hospital discharge, obtained 0.74 ± 0.01 AUROC, 1.00 ± 0.03 calibration slope, and 0.00 ± 0.01 calibration in the large (**Figure 4.A.12** and **Table 4.A.14**). As depicted in **Figure 4.A.11**, a significant difference of predicted probabilities between the recovery and non-recovery groups was consistently observed in the development cohort (p-values for general ICU patients, cardiac patients, and septic patients were 0.01, <0.001, and <0.001)

4.A.3 Supplementary tables

Table 4.A.1 Patient characteristics and clinical outcomes stratified by development and validation cohorts

	All patients (n=473)	Development cohort (n=229)	Validation cohort (n=244)	p-value
Age, years, median (IQR)	67.68 (57.53 – 75.70)	66.26 (56.56 – 75.38)	69.24 (57.87 – 76.45)	0.16
Gender male, number (%)	289 (61.10)	141 (61.57)	148 (60.66)	0.85
BMI, median (IQR)	25.95 (23.03 – 29.39)	26.12 (22.86 – 29.41)	25.95 (23.03 – 29.39)	0.64
With true baseline, number (%)	269 (56.87)	136 (59.39)	133 (54.51)	0.31
Baseline serum creatinine, mg/dl, median (IQR)	0.96 (0.76 – 1.08)	0.96 (0.75 – 1.07)	0.97 (0.76 – 1.09)	0.39
Malignancy, number (%)	93 (19.66)	39 (17.03)	54 (22.13)	0.17
Chronic Kidney Disease, number (%)	102 (21.56)	42 (18.34)	60 (24.59)	0.12
Diabetes, number (%)	114 (24.10)	51 (22.27)	63 (25.82)	0.39
Elective admission, number (%)	98 (20.72)	49 (21.40)	49 (20.08)	0.73
Early parenteral nutrition strategies, number (%)	234 (49.47)	126 (55.02)	108 (44.26)	0.02
Sepsis on ICU admission, number (%)	266 (56.24)	120 (52.40)	146 (59.84)	0.12
Mechanical hemodynamic support on ICU admission, number (%)	45 (9.51)	25 (10.92)	20 (8.20)	0.35
Pharmacological hemodynamic support on ICU admission, number (%)	443 (93.66)	215 (93.89)	228 (93.44)	0.85
Cardiac surgery diagnostic group, number (%)	145 (30.66)	80 (34.93)	65 (26.64)	0.06
Medical diagnostic group, number (%)	142 (30.02)	66 (28.82)	76 (31.15)	0.62
Neuro diagnostic group, number (%)	11 (2.33)	6 (2.62)	5 (2.05)	0.77

Surgery/Trauma diagnostic group, number (%)	143 (30.23)	54 (23.58)	89 (36.48)	<0.01
Transplant diagnostic group, number (%)	32 (6.77)	23 (10.04)	9 (3.69)	0.01
First AKI-3 day, days, median (IQR)	2.00 (1.00 – 4.00)	2.00 (1.00 – 4.00)	2.00 (1.00 – 4.00)	0.13
With KRT when first AKI-3 in ICU, number (%)	314 (66.38)	153 (66.81)	161 (65.98)	0.92
Deceased at ICU discharge, number (%)	160 (33.83)	79 (34.50)	81 (33.20)	0.77
LOS in ICU, days, median (IQR)	16.00 (8.00 – 31.00)	16.00 (8.00 – 32.00)	15.00 (8.00 – 30.00)	0.66
With KRT at ICU discharge, number (%)	182 (38.48)	88 (38.43)	94 (38.52)	1.00
Deceased at hospital discharge, number (%)	205 (43.34)	100 (43.67)	105 (43.03)	0.93
LOS in hospital, days, median (IQR)	34.00 (16.00 – 62.00)	32.00 (17.00 – 59.00)	36.00 (16.00 – 63.00)	0.72
With KRT at hospital discharge, number (%)	153 (32.35)	76 (33.19)	77 (31.56)	0.77
Complete recovery, number (%)	167 (35.31)	86 (37.55)	81 (33.20)	0.34
Complete or partial recovery, number (%)	242 (51.16)	117 (51.09)	125 (51.23)	1.00

Abbreviations: AKI-3, acute kidney injury stage 3; ICU, intensive care unit; IQR, interquartile range; LOS, length of stay; KRT, kidney replacement therapy; BMI, body mass index

Table 4.A.2 Patient characteristics and clinical outcomes of the two groups stratified by AKI recovery status in the development cohort

	Development cohort (n=229)	Complete recovery (n=86)	Not complete recovery (n=143)	p-value	Complete or partial recovery (n=117)	Not complete or partial recovery (n=112)	p-value
Age, years, median (IQR)	66.26 (56.56 – 75.38)	61.98 (50.83 – 72.70)	68.56 (60.19 – 77.29)	<0.01	65.27 (51.34 – 73.18)	68.47 (60.95 – 78.00)	<0.01
Gender male, number (%)	141 (61.57)	50 (58.14)	91 (63.64)	0.48	69 (58.97)	72 (64.29)	0.42
BMI, median (IQR)	26.12 (22.86 – 29.41)	25.34 (22.81 – 29.34)	26.42 (23.61 – 29.46)	0.29	26.12 (23.23 – 29.38)	26.04 (22.67 – 29.76)	0.91
With true baseline, number (%)	136 (59.39)	51 (59.30)	85 (59.44)	1.00	63 (53.85)	73 (65.18)	0.11
Baseline serum creatinine, mg/dl, median (IQR)	0.96 (0.75 – 1.07)	0.96 (0.74 – 1.05)	0.96 (0.76 – 1.10)	0.68	0.93 (0.74 – 1.03)	0.96 (0.77 – 1.15)	0.11
Malignancy, number (%)	39 (17.03)	15 (17.44)	24 (16.78)	1.00	19 (16.24)	20 (17.86)	0.86
Chronic Kidney Disease, number (%)	42 (18.34)	15 (17.44)	27 (18.88)	0.86	17 (14.53)	25 (22.32)	0.17
Diabetes, number (%)	51 (22.27)	19 (22.09)	32 (22.38)	1.00	23 (19.66)	28 (25.00)	0.35
Elective admission, number (%)	49 (21.40)	20 (23.26)	29 (20.28)	0.62	25 (21.37)	24 (21.43)	1.00
Early parenteral nutrition strategies, number (%)	126 (55.02)	43 (50.00)	83 (58.04)	0.27	62 (52.99)	64 (57.14)	0.60
Sepsis on ICU admission, number (%)	120 (52.40)	54 (62.79)	66 (46.15)	0.02	66 (56.41)	54 (48.21)	0.24
Mechanical hemodynamic support on ICU admission, number (%)	25 (10.92)	5 (5.81)	20 (13.99)	0.08	6 (5.13)	19 (16.96)	0.01
Pharmacological hemodynamic support on ICU admission, number (%)	215 (93.89)	80 (93.02)	135 (94.41)	0.78	108 (92.31)	107 (95.54)	0.41

Cardiac surgery diagnostic group, number (%)	80 (34.93)	24 (27.91)	56 (39.16)	0.09	35 (29.91)	45 (40.18)	0.13
Medical diagnostic group, number (%)	66 (28.82)	22 (25.58)	44 (30.77)	0.45	30 (25.64)	36 (32.14)	0.31
Neuro diagnostic group, number (%)	6 (2.62)	1 (1.16)	5 (3.50)	0.41	2 (1.71)	4 (3.57)	0.44
Surgery/Trauma diagnostic group, number (%)	54 (23.58)	31 (36.05)	23 (16.08)	<0.01	37 (31.62)	17 (15.18)	<0.01
Transplant diagnostic group, number (%)	23 (10.04)	8 (9.30)	15 (10.49)	0.82	13 (11.11)	10 (8.93)	0.66
First AKI-3 day, days, median (IQR)	2.00 (1.00 – 4.00)	2.00 (1.00 – 3.00)	2.00 (1.00 – 5.00)	0.27	2.00 (1.00 – 3.00)	2.00 (1.00 – 7.00)	0.05
With KRT when first AKI-3 in ICU, number (%)	153 (66.81)	46 (53.49)	107 (74.83)	<0.01	63 (53.85)	90 (80.36)	<0.01
LOS in ICU, days, median (IQR)	16.00 (8.00 – 32.00)	15.50 (8.00 – 30.00)	16.00 (9.00 – 32.00)	0.83	14.00 (8.00 – 27.00)	18.00 (9.50 – 35.00)	0.10
With KRT at ICU discharge, number (%)	88 (38.43)	6 (6.98)	82 (57.34)	<0.01	11 (9.40)	77 (68.75)	<0.01
LOS in hospital, days, median (IQR)	32.00 (17.00 – 59.00)	41.50 (25.50 – 71.50)	27.00 (15.50 – 54.00)	<0.01	36.00 (22.00 – 67.00)	24.00 (13.50 – 51.50)	<0.01

Abbreviations: AKI-3, acute kidney injury stage 3; ICU, intensive care unit; IQR, interquartile range; LOS, length of stay; KRT, kidney replacement therapy; BMI, body mass index

Table 4.A.3 Complete recovery prediction feature list and univariate association in the development and validation cohorts

	Development cohort (n=229)	Complete recovery (n=86)	Not complete recovery (n=143)	p-value	Validation cohort (n=244)	Complete recovery (n=81)	Not complete recovery (n=163)	p-value
Age, years, median (IQR)	66.26 (56.56 – 75.38)	61.98 (50.83 – 72.70)	68.56 (60.19 – 77.29)	<0.01	69.24 (57.87 – 76.45)	64.28 (53.87 – 73.23)	70.94 (60.05 – 78.22)	<0.01
Surgery/Trauma diagnostic group, number (%)	54 (23.58)	31 (36.05)	23 (16.08)	<0.01	89 (36.48)	23 (28.40)	66 (40.49)	0.07
Sepsis on ICU admission, number (%)	120 (52.40)	54 (62.79)	66 (46.15)	0.02	146 (59.84)	43 (53.09)	103 (63.19)	0.17
With KRT when first AKI-3 in ICU, number (%)	153 (66.81)	46 (53.49)	107 (74.83)	<0.01	161 (65.98)	51 (62.96)	110 (67.48)	0.57
Cardiac surgery diagnostic group, number (%)	80 (34.93)	24 (27.91)	56 (39.16)	0.09	65 (26.64)	25 (30.86)	40 (24.54)	0.36

A significant difference between the recovery and non-recovery group was found in all features except for cardiac diagnostic group in the development cohort, while only age showed a significant difference in the validation cohort. Abbreviations: AKI-3, acute kidney injury stage 3; ICU, intensive care unit; IQR, interquartile range; KRT, kidney replacement therapy.

Table 4.A.4 Complete or partial recovery prediction feature list and univariate association in the development and validation cohorts

	Development cohort (n=229)	Complete or partial recovery (n=117)	Not complete or partial recovery (n=112)	p-value	Validation cohort (n=244)	Complete or partial recovery (n=125)	Not complete or partial recovery (n=119)	p-value
Age, years, median (IQR)	66.26 (56.56 – 75.38)	65.27 (51.34 – 73.18)	68.47 (60.95 – 78.00)	<0.01	69.24 (57.87 – 76.45)	65.02 (56.00 – 73.48)	71.86 (59.64 – 78.85)	0.01
Maximum of white blood cell count before first AKI-3, × 10^9/L, median (IQR)	14.12 (9.89 – 19.45)	13.40 (9.63 – 18.04)	14.98 (11.32 – 21.99)	0.04	14.29 (9.35 – 19.80)	14.65 (8.79 – 19.80)	14.10 (9.47 – 19.68)	0.52
Maximum of urea before first AKI-3, mg/dL, median (IQR)	120.00 (84.00 – 175.50)	106.00 (76.50 – 144.00)	139.00 (94.00 – 210.00)	<0.01	116.00 (84.00 – 167.00)	114.00 (84.00 – 158.00)	122.00 (84.00 – 177.00)	0.73
Minimum of platelet before first AKI-3, × 10^9/L, median (IQR)	89.00 (61.00 – 138.00)	99.50 (74.00 – 156.00)	74.00 (50.00 – 109.50)	<0.01	95.00 (51.00 – 158.00)	96.00 (55.00 – 167.50)	93.00 (45.00 – 140.00)	0.23
With KRT when first AKI-3 in ICU, number (%)	153 (66.81)	63 (53.85)	90 (80.36)	<0.01	161 (65.98)	74 (59.20)	87 (73.11)	0.03

A significant difference was found in all features in the development cohort, while only age and KRT on the first day of AKI-3 showed significant difference in the validation cohort. Abbreviations: AKI-3, acute kidney injury stage 3; ICU, intensive care unit; IQR, interquartile range; KRT, kidney replacement therapy.

Table 4.A.5 Comparison of complete recovery prediction features between the development and validation cohorts

	All patients (n=473)	Development cohort (n=229)	Validation cohort (n=244)	p-value
Age, years, median (IQR)	67.68 (57.53 – 75.70)	66.26 (56.56 – 75.38)	69.24 (57.87 – 76.45)	0.16
Cardiac surgery diagnostic group, number (%)	145 (30.66)	80 (34.93)	65 (26.64)	0.06
Surgery/Trauma diagnostic group, number (%)	143 (30.23)	54 (23.58)	89 (36.48)	<0.01
Sepsis on ICU admission, number (%)	266 (56.24)	120 (52.40)	146 (59.84)	0.12
With KRT when first AKI-3 in ICU, number (%)	314 (66.38)	153 (66.81)	161 (65.98)	0.92

A Significant difference between development and validation cohorts was found in surgical/trauma diagnostic group for complete recovery prediction. Abbreviations: AKI-3, acute kidney injury stage 3; ICU, intensive care unit; IQR, interquartile range; KRT, kidney replacement therapy.

Table 4.A.6 Comparison of complete or partial recovery prediction features between the development and validation cohorts

	All patients (n=473)	Development cohort (n=229)	Validation cohort (n=244)	p-value
Age, years, median (IQR)	67.68 (57.53 – 75.70)	66.26 (56.56 – 75.38)	69.24 (57.87 – 76.45)	0.16
Maximum of white blood cell count before first AKI-3, × 10^9/L, median (IQR)	14.21 (9.63 – 19.63)	14.12 (9.89 – 19.45)	14.29 (9.35 – 19.80)	0.60
Maximum of urea before first AKI-3, mg/dL, median (IQR)	118.00 (84.00 – 172.00)	120.00 (84.00 – 175.50)	116.00 (84.00 – 167.00)	0.48
Minimum of platelet before first AKI-3, × 10^9/L, median (IQR)	93.00 (56.00 – 143.00)	89.00 (61.00 – 138.00)	95.00 (51.00 – 158.00)	0.83
With KRT when first AKI-3 in ICU, number (%)	314 (66.38)	153 (66.81)	161 (65.98)	0.92

No features revealed significant difference between development and validation cohorts for complete or partial recovery prediction. Abbreviations: AKI-3, acute kidney injury stage 3; ICU, intensive care unit; IQR, interquartile range; KRT, kidney replacement therapy.

Table 4.A.7 Number of missing complete or partial recovery prediction features in the development and validation cohorts

	Development cohort (n=229)	**Validation cohort (n=244)**
Age	0	0
Maximum of white blood cell count before first AKI-3	6 (2.62%)	1 (0.41%)
Maximum of urea before first AKI-3	6 (2.62%)	0
Minimum of platelet before first AKI-3	12 (5.24%)	4 (1.64%)
With KRT when first AKI-3 in ICU	0	0

Abbreviations: AKI-3, acute kidney injury stage 3; ICU, intensive care unit; IQR, interquartile range; KRT, kidney replacement therapy;

Table 4.A.8 Performance of prediction model for complete recovery and complete or partial recovery in the validation cohort

Measurement	Complete recovery prediction Mean ± standard deviation	Complete or partial recovery prediction Mean ± standard deviation
Number of patients	244	244
Recovery incidence	33.20%	51.23%
AUROC	0.53	0.61
Sensitivity[1]	0.48	0.55
Specificity[1]	0.56	0.57
Calibration slope	0.27	0.32
Calibration in the large	-0.07	-0.00
Classification Threshold (%)	40.55	50.71

Abbreviations: AUROC, area under the receiver operating characteristic curve; Mean ± standard deviation was obtained from 100 repetitions of stratified 10-fold cross validation. [1]Evaluated at the threshold that maximized sensitivity and specificity in the development cohort

Table 4.A.9 Comparison of NGAL measured on the first AKI3 day between cardiac and non-cardiac subgroups

	Development cohort (n=229)	Cardiac AKI-3 patients (n=80)	Non-cardiac AKI-3 patients (n=149)	p-value	Validation cohort (n=244)	Cardiac AKI-3 patients (n=65)	Non-cardiac AKI-3 patients (n=179)	p-value
NGAL measured on first AKI-3 day, median (IQR)	653.64 (407.10 – 1147.17)	575.40 (380.50 – 967.96)	719.79 (414.43 – 1311.09)	<0.01	720.04 (450.77 – 1241.83)	594.94 (386.66 – 785.31)	856.03 (480.47 – 1374.62)	<0.01

Abbreviation: NGAL, neutrophil gelatinase-associated lipocalin; AKI3, acute kidney injury stage 3

Table 4.A.10 Comparison of NGAL measured on the first AKI3 day between septic and non-septic subgroups

	Development cohort (n=229)	Septic AKI-3 patients (n=120)	Non-septic AKI-3 patients (n=109)	p-value	Validation cohort (n=244)	Septic AKI-3 patients (n=146)	Non-septic AKI-3 patients (n=98)	p-value
NGAL measured on first AKI-3 day, median (IQR)	653.64 (407.10 – 1147.17)	819.88 (485.14 – 1365.58)	553.89 (377.44 – 919.94)	<0.01	720.04 (450.77 – 1241.83)	881.43 (497.20 – 1488.16)	602.44 (432.24 – 965.98)	<0.01

Abbreviation: NGAL, neutrophil gelatinase-associated lipocalin; AKI3, acute kidney injury stage 3

Table 4.A.11 Summary of prediction performance for prediction model, biomarker NGAL_AKI3, and prediction model with biomarker NGAL_AKI3

			Development cohort			Validation cohort		
			AUROC	Calibration slope	Calibration in the large	AUROC	Calibration slope	Calibration in the large
Complete recovery	Prediction model	General ICU patients	0.73 ± 0.01	1.16 ± 0.03	0.01 ± 0.01	0.53	0.27	-0.07
		Cardiac patients	0.67 ± 0.02	0.76 ± 0.26	-0.03 ± 0.04	0.6	0.9	0.06
		Septic patients	0.69 ± 0.01	1.05 ± 0.11	0.03 ± 0.02	0.56	0.46	-0.12
	Biomarker NGAL_AKI3	General ICU patients	0.53	0.82	0. 16	0.55	0.5	-0.02
		Cardiac patients	0.51	-0.17	-0.24	0.52	-0.52	-0.34
		Septic patients	0.51	0.42	-0.03	0.63	0.66	-0.08
	Prediction model with biomarker NGAL_AKI3	General ICU patients	0.72 ± 0.01	1.15 ± 0.03	0.01 ± 0.01	0.54	0.31	-0.08
		Cardiac patients	0.65 ± 0.02	0.32 ± 0.40	-0.11 ± 0.08	0.6	0.83	0.05
		Septic patients	0.69 ± 0.01	1.01 ± 0.12	0.04 ± 0.02	0.58	0.51	-0.12
Complete or partial recovery	Prediction model	General ICU patients	0.74 ± 0.01	1.00 ± 0.03	0.00 ± 0.01	0.61	0.32	0
		Cardiac patients	0.75 ± 0.01	1.03 ± 0.09	-0.01 ± 0.01	0.71	0.84	0.03
		Septic patients	0.70 ± 0.01	0.89 ± 0.10	0.03 ± 0.01	0.58	0.32	-0.03
	Biomarker NGAL_AKI3	General ICU patients	0.52	0.67	0.22	0.53	0.32	0.11
		Cardiac patients	0.51	-0.12	-0.08	0.54	-0.72	-0.28
		Septic patients	0.53	0.28	0.03	0.61	0.45	0.04

	Prediction model with biomarker NGAL_AKI3	**General ICU patients**	0.73 ± 0.01	0.99 ± 0.03	0.00 ± 0.01	0.61	0.42	-0.02
		Cardiac patients	0.76 ± 0.02	0.93 ± 0.09	-0.00 ± 0.01	0.71	0.8	0.03
		Septic patients	0.69 ± 0.02	0.85 ± 0.10	0.02 ± 0.01	0.59	0.37	-0.04

Mean ± standard deviation was obtained from 100 repetitions of stratified 10-fold cross validation. AUROC, area under the receiver operating characteristic curve; NGAL_AKI3, neutrophil gelatinase-associated lipocalin measured on the first AKI-3 day

Table 4.A.12 Parameters that were matched between the development and validation cohorts in the matching.

Age
Gender
APACHE II score
Cardiac surgery diagnostic group
Sepsis
AKI stage
AKI-3
Renal replacement therapy
LOS in ICU
LOS in hospital
Hospital mortality
Day-90 mortality
Randomization for parenteral nutrition strategies

Table 4.A.13 Comparison of prediction performance in the validation cohort, before and after forcing the early initiation of parenteral nutrition strategies into the models

			Before forcing the early initiation of parenteral nutrition strategies into the models			After forcing the early initiation of parenteral nutrition strategies into the models		
			AUROC	Calibration slope	Calibration in the large	AUROC	Calibration slope	Calibration in the large
Complete recovery	Prediction model	General ICU patients	0.53	0.27	-0.07	0.53	0.17	-0.06
		Cardiac patients	0.6	0.9	0.06	0.63	0.93	0.05
		Septic patients	0.56	0.46	-0.12	0.54	0.28	-0.13
	Prediction model with biomarker NGAL_AKI3	General ICU patients	0.54	0.31	-0.08	0.54	0.18	-0.06
		Cardiac patients	0.6	0.83	0.05	0.63	0.92	0.05
		Septic patients	0.58	0.51	-0.12	0.56	0.24	-0.14
Complete or partial recovery	Prediction model	General ICU patients	0.61	0.32	0	0.61	0.32	0
		Cardiac patients	0.71	0.84	0.03	0.71	0.84	0.03
		Septic patients	0.58	0.32	-0.03	0.58	0.32	-0.03
	Prediction model with biomarker NGAL_AKI3	General ICU patients	0.61	0.42	-0.02	0.61	0.42	-0.02
		Cardiac patients	0.71	0.8	0.03	0.71	0.8	0.03
		Septic patients	0.59	0.37	-0.04	0.59	0.37	-0.04

AUROC, area under the receiver operating characteristic curve; NGAL_AKI3, neutrophil gelatinase-associated lipocalin measured on the first AKI-3

Table 4.A.14 Performance of prediction model for complete recovery and complete or partial recovery in the development cohort

Measurement	Complete recovery prediction Mean ± standard deviation	Complete or partial recovery prediction Mean ± standard deviation
Number of patients	229	229
Recovery incidence	37.55%	51.09%
AUROC	0.73 ± 0.01	0.74 ± 0.01
Sensitivity[1]	0.63 ± 0.03	0.71 ± 0.03
Specificity[1]	0.73 ± 0.03	0.69 ± 0.02
Calibration slope	1.16 ± 0.03	1.00 ± 0.03
Calibration in the large	0.01 ± 0.01	0.00 ± 0.01
Classification Threshold (%)	40.82	50.43

Abbreviations: AUROC, area under the receiver operating characteristic curve; Mean ± standard deviation was obtained from 100 repetitions of stratified 10-fold cross validation. [1]Evaluated at the threshold that maximized sensitivity and specificity in the development cohort

4.A.4 Supplementary figures

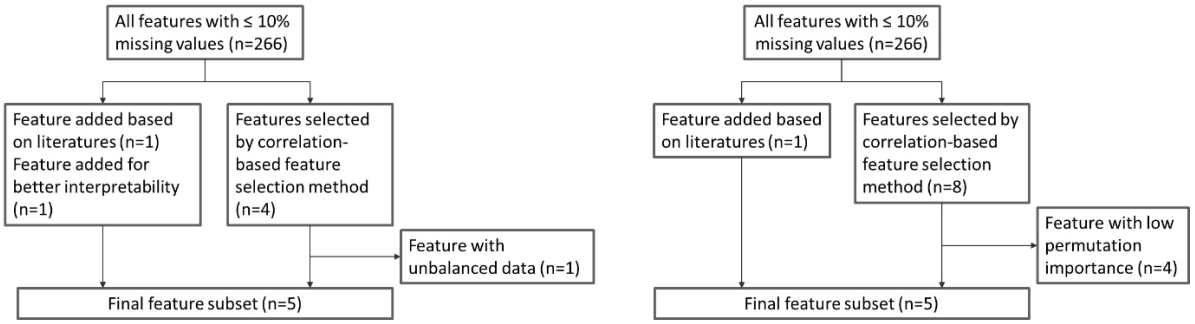


Figure 4.A.1 Feature selection flow diagram for (left) complete recovery (right) complete or partial recovery using the development cohort

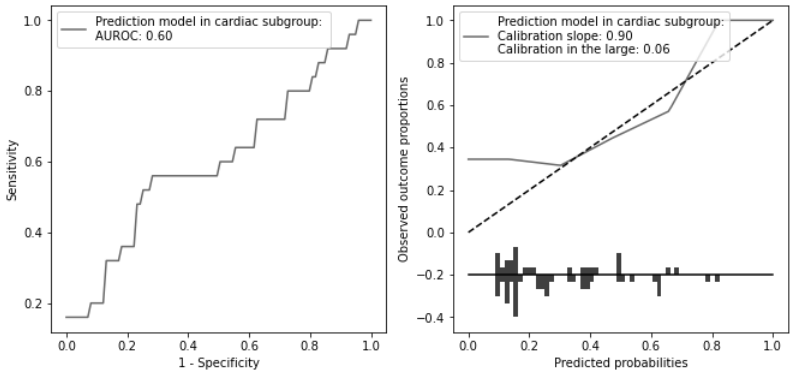


Figure 4.A.2 Cardiac subgroup analysis of the validation cohort for complete recovery prediction

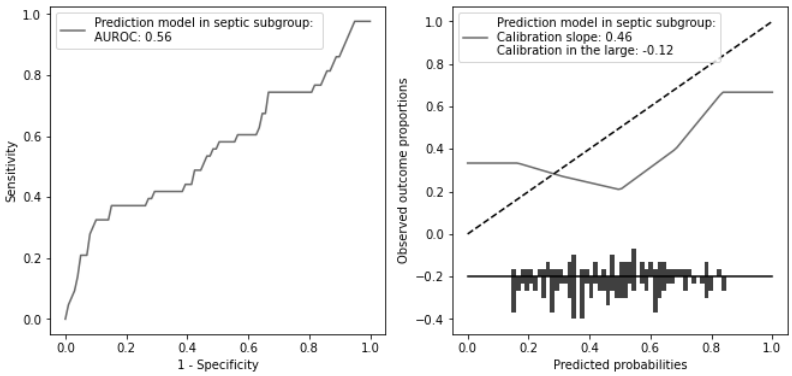


Figure 4.A.3 Septic subgroup analysis of the validation cohort for complete recovery prediction

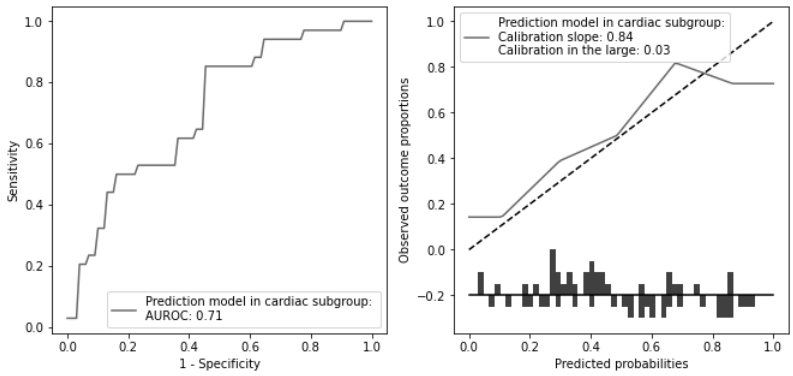


Figure 4.A.4 Cardiac subgroup analysis of the validation cohort for complete or partial recovery prediction

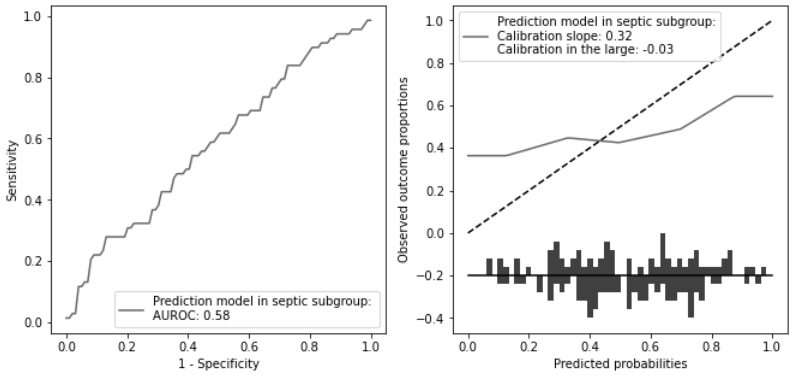


Figure 4.A.5 Septic subgroup analysis of the validation cohort for complete or partial recovery prediction

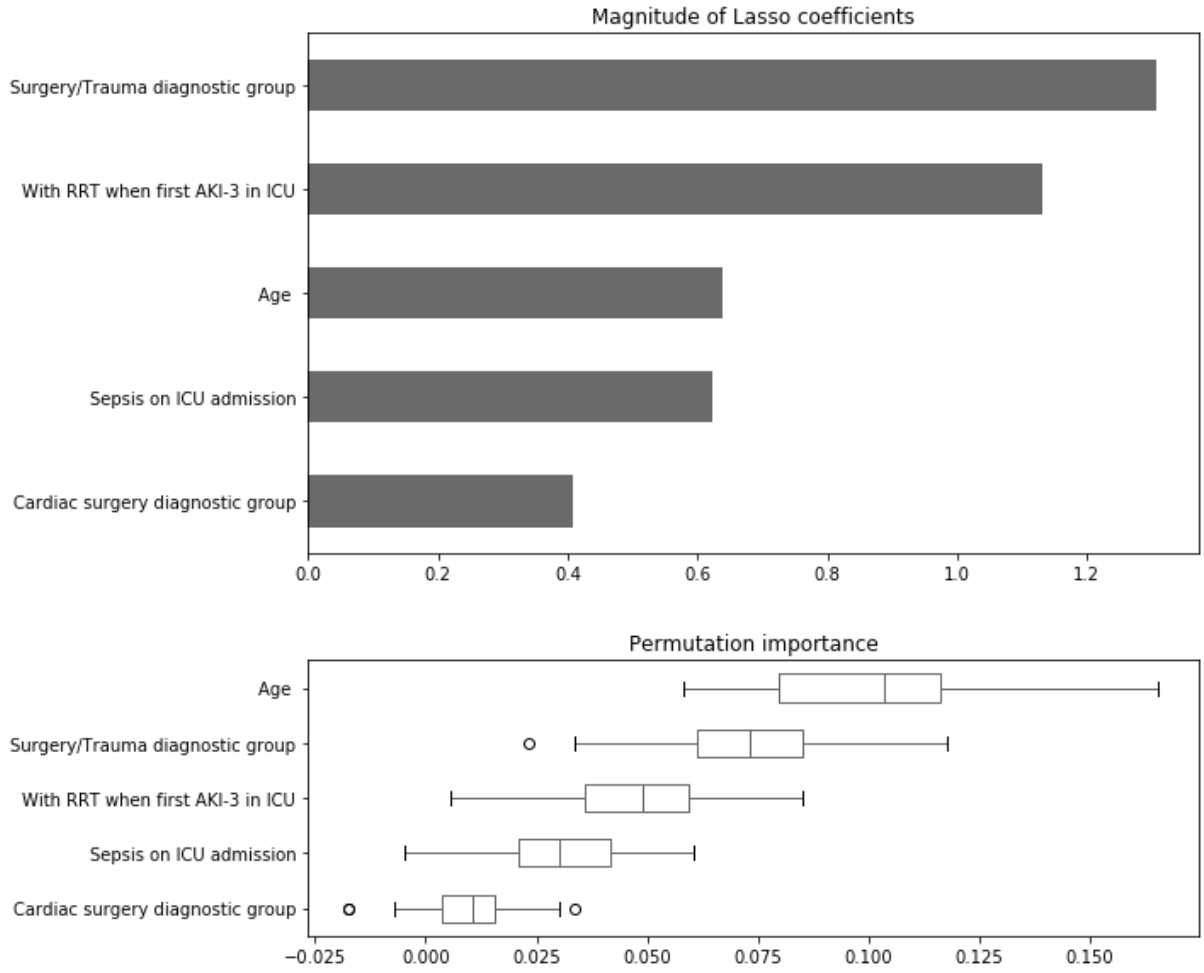


Figure 4.A.6 Magnitude of Lasso coefficients and permutation importance for complete recovery prediction in the development cohort

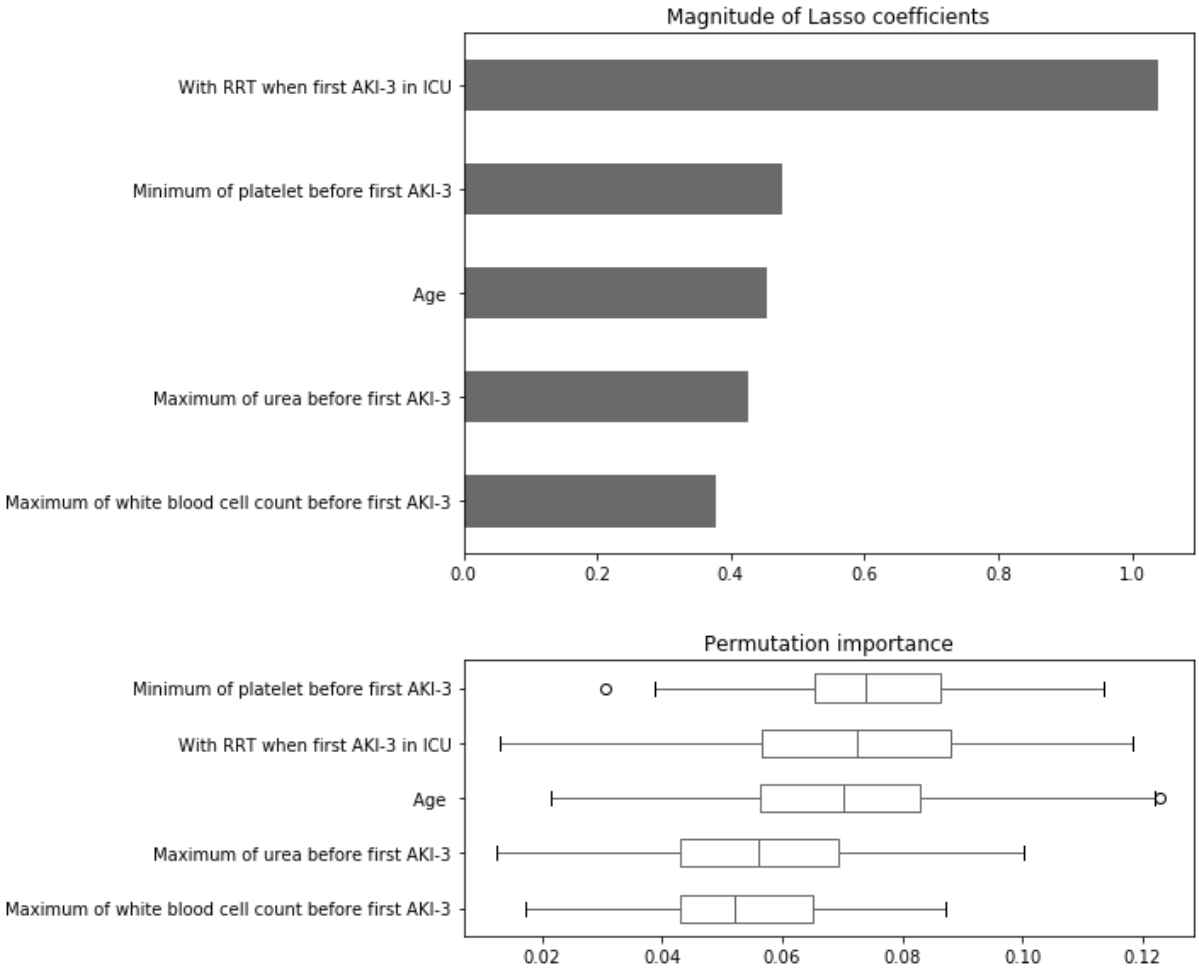


Figure 4.A.7 Magnitude of Lasso coefficients and permutation importance for complete or partial recovery prediction in the development cohort

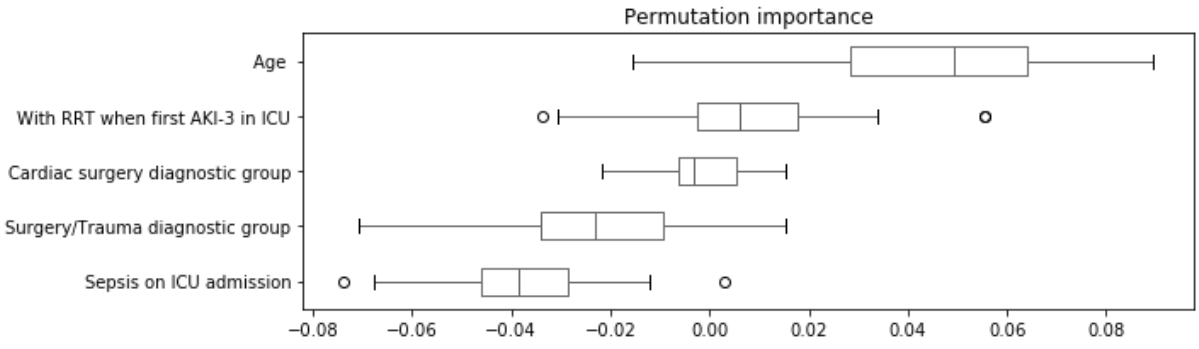


Figure 4.A.8 Permutation importance for complete recovery prediction in the validation cohort

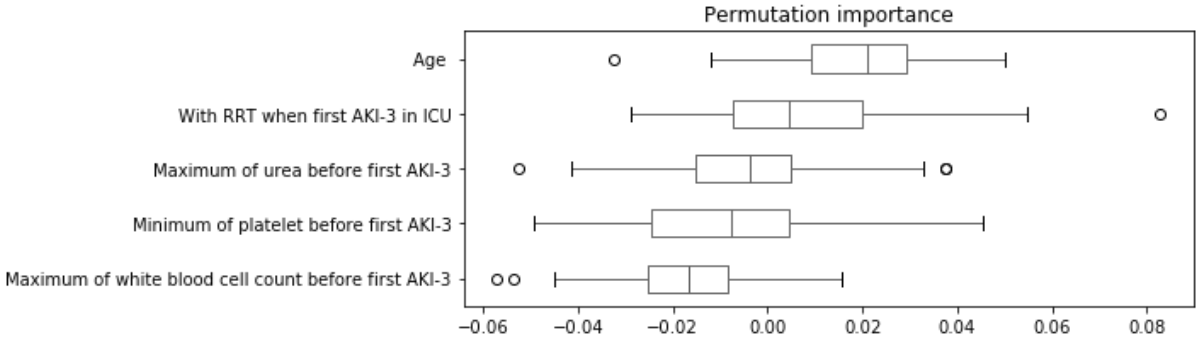


Figure 4.A.9 Permutation importance for complete or partial recovery prediction in the validation cohort

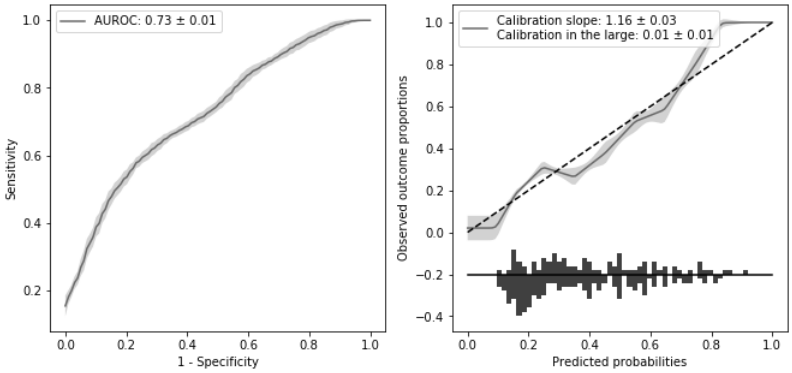


Figure 4.A.10 Model performance for complete recovery prediction in the development cohort. (left) ROC curve (right) calibration curve.

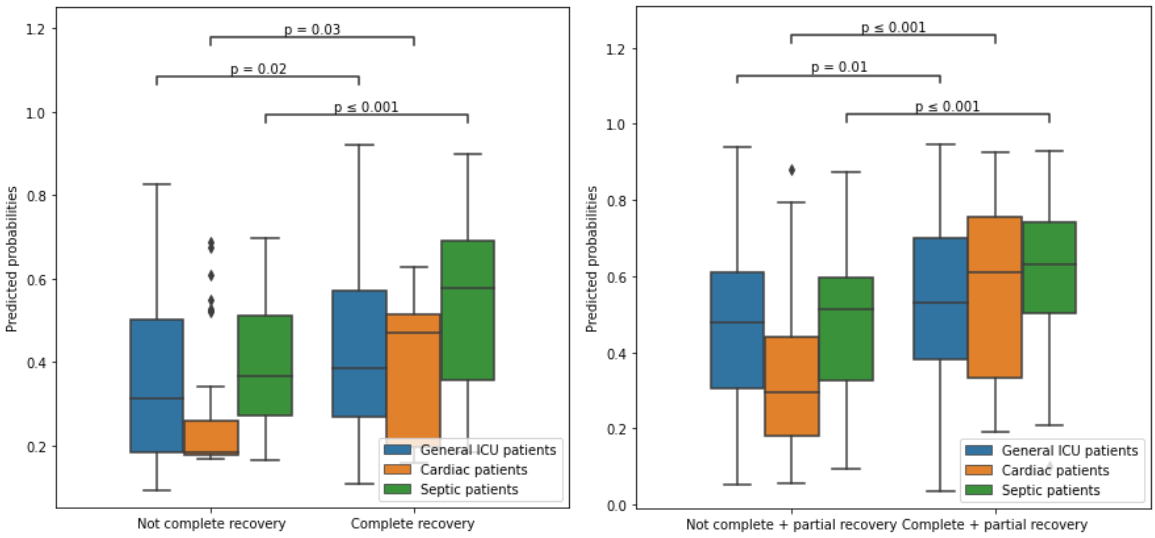


Figure 4.A.11 Predicted probabilities of prediction model for (left) complete recovery and (right) complete or partial recovery in the development cohort.

Reported p values are calculated based on Mann–Whitney U test. Predicted probabilities were the averages of 100 repetitions of stratified 10-fold cross validation. (blue) in general ICU patients (orange) in cardiac patients (green) in septic patients

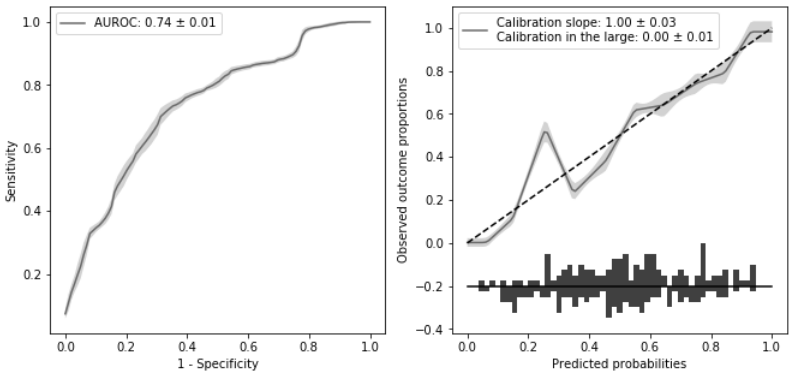


Figure 4.A.12 Model performance for complete or partial recovery prediction in the development cohort. (left) ROC curve (right) calibration curve.

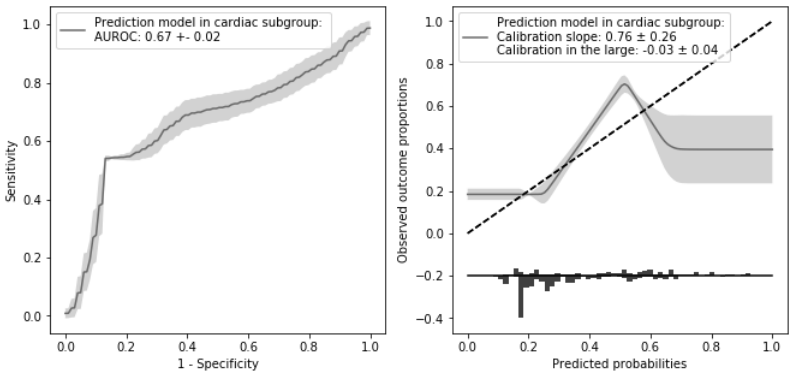


Figure 4.A.13 Cardiac subgroup analysis of the development cohort for complete recovery prediction

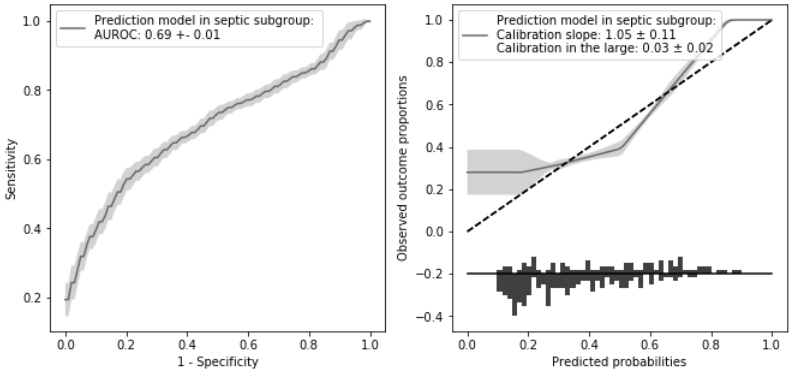


Figure 4.A.14 Septic subgroup analysis of the development cohort for complete recovery prediction

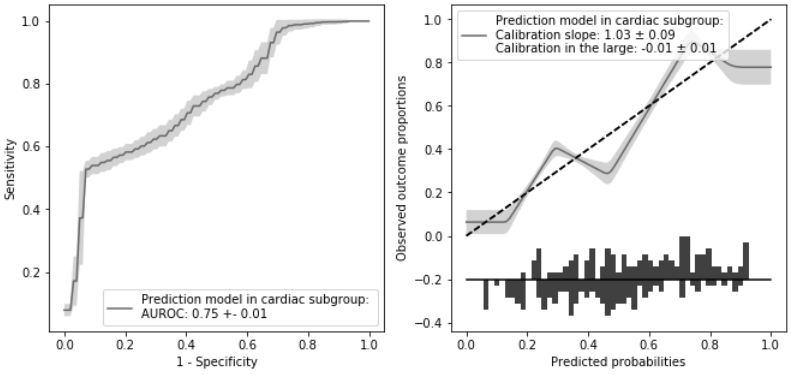


Figure 4.A.15 Cardiac subgroup analysis of the development cohort for complete or partial recovery prediction

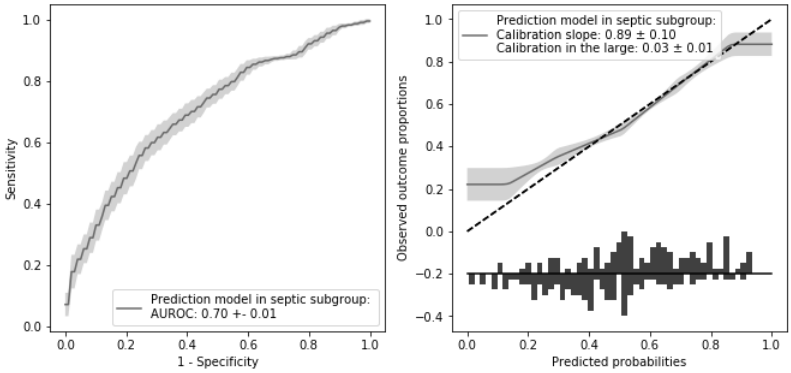


Figure 4.A.16 Septic subgroup analysis of the development

4.A.5 Supplementary discussion

4.A.5.1 The role of early initiation of parenteral nutrition strategies in AKI recovery prediction

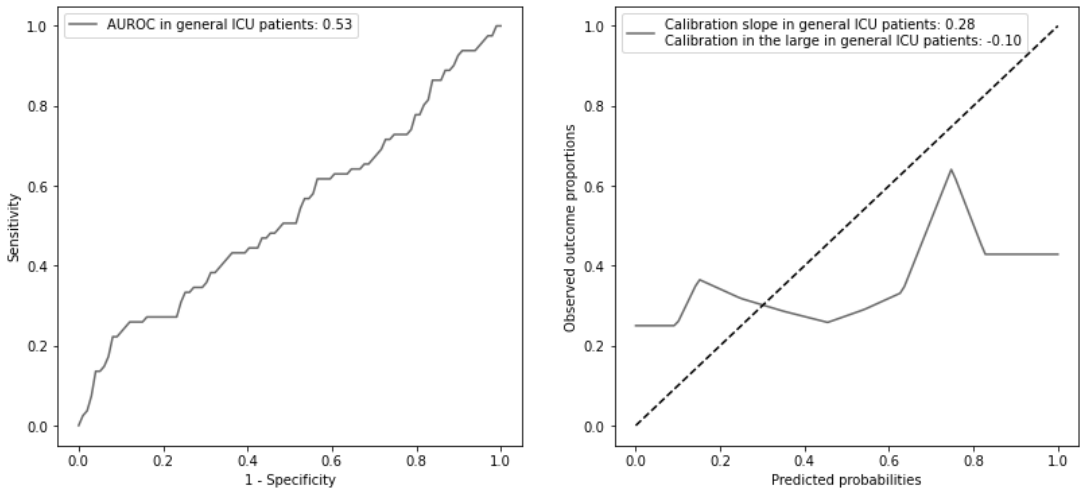
Although early initiation of parenteral nutrition strategies was shown to lead to late recovery and more complications in the EPaNIC study, no significant differences were observed in both AKI prediction tasks in both cohorts of this study. In addition, after the addition of early initiation of parenteral nutrition strategies into the prediction models, only small changes in the predictive performance were observed in the subgroups for complete recovery prediction. Therefore, we rejected the hypothesis that forcing the randomization of feeding strategies into the models would improve the prediction.

4.A.5.2 The results with "True baseline serum creatinine (True/False)" in the developed model

Please see below the results with "True baseline serum creatinine (True/False)" added to the developed models, where the AUROC remains unchanged in both prediction tasks.

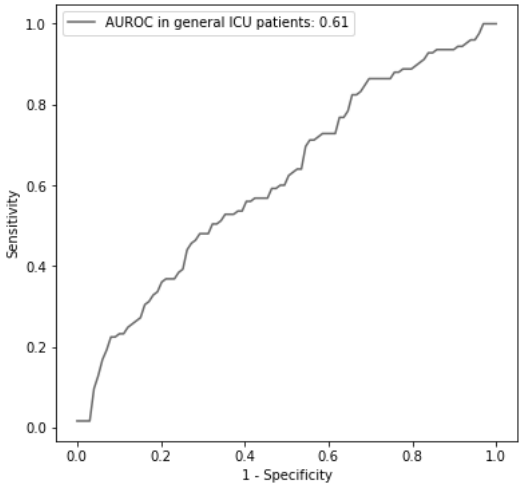
Complete recovery:

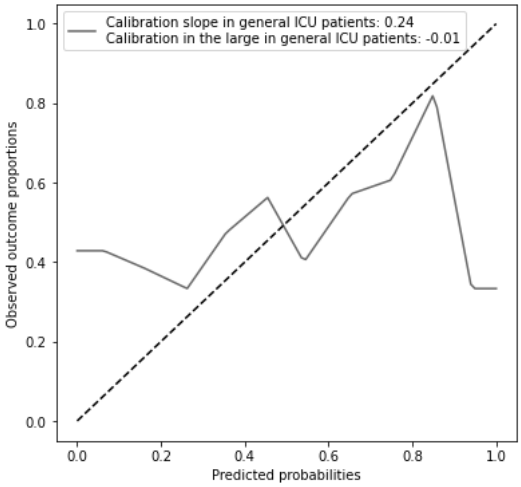
The developed model with "True baseline serum creatinine (True/False)"



Complete or partial recovery:

The developed model with “True baseline serum creatinine (True/False)”

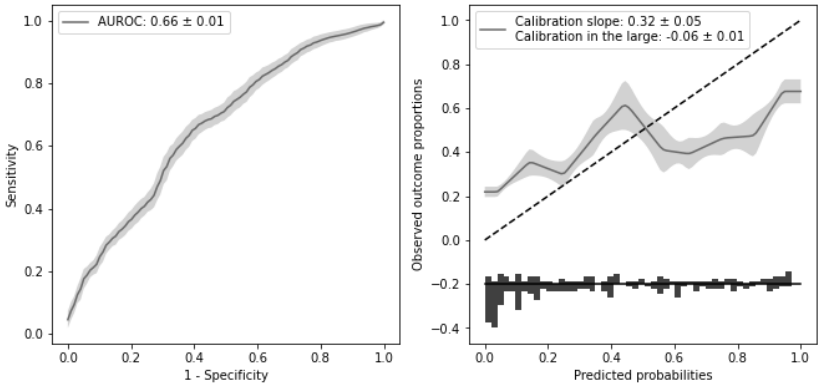




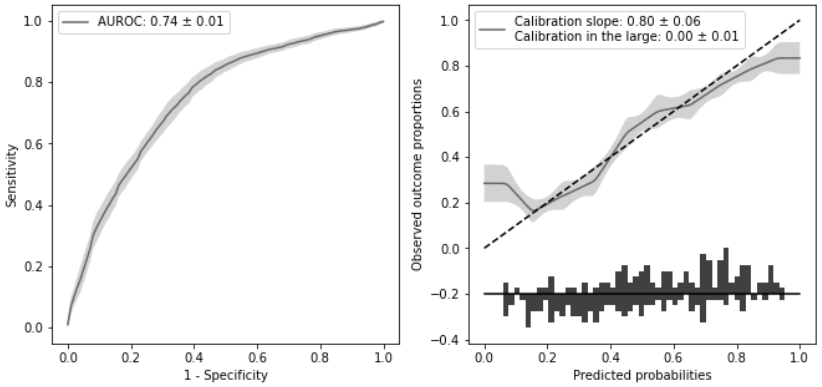
4.A.5.3 The comparison of prediction results with different machine learning algorithms: random forest, gradient boosting, and support vector machine

Random forest:

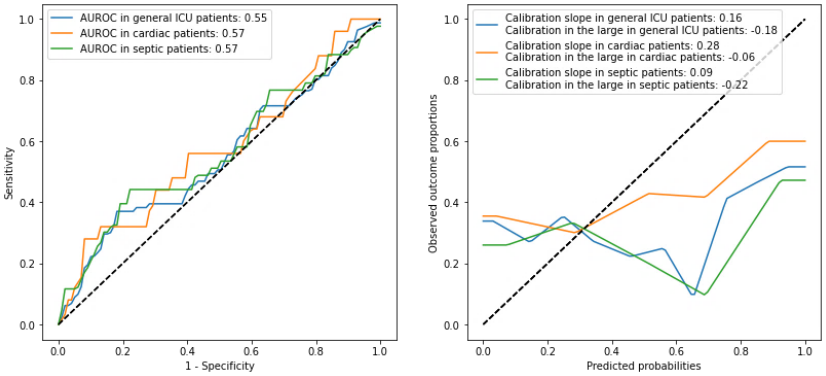
Ten-fold cross validation for complete recovery prediction:



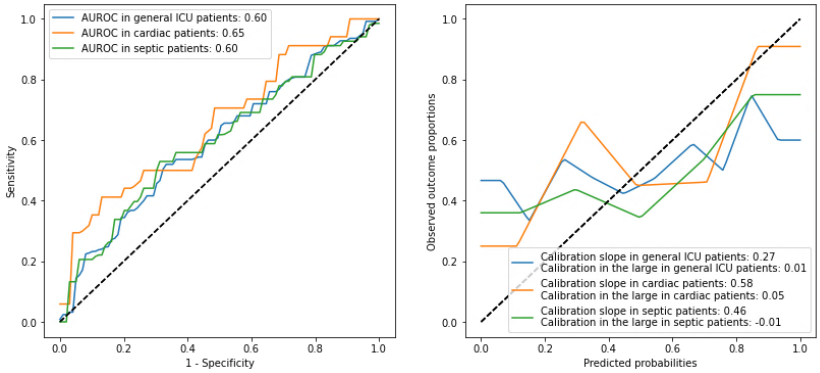
Ten-fold cross validation for complete or partial recovery prediction:



Validation cohort results for complete recovery prediction:

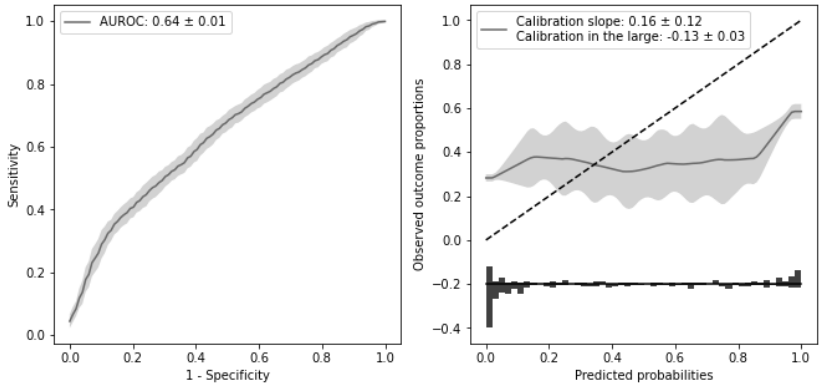


Validation cohort results for complete or partial recovery prediction:

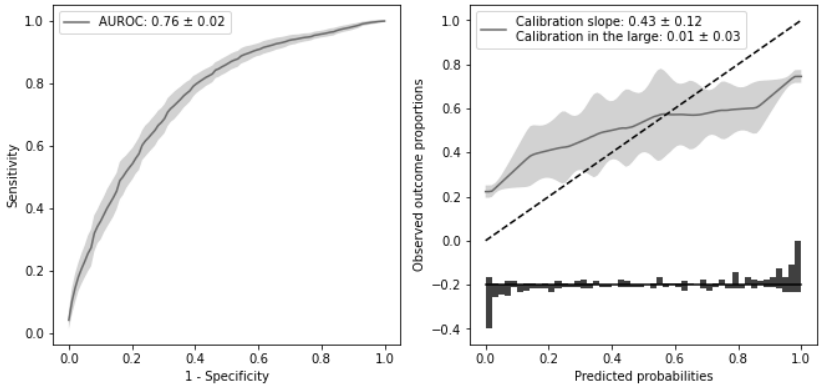


Gradient boosting:

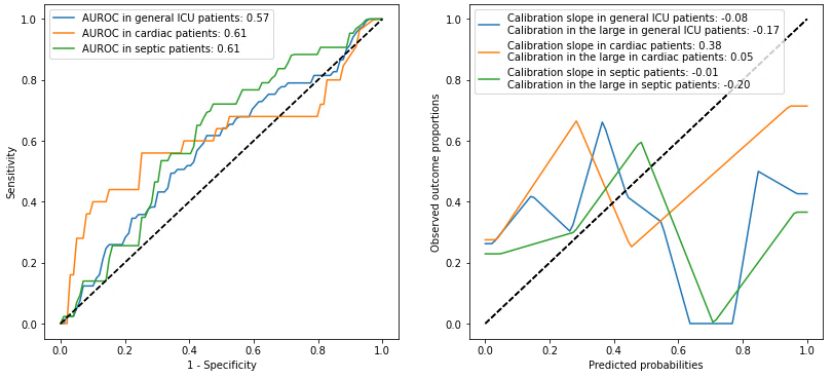
Ten-fold cross validation for complete recovery prediction:



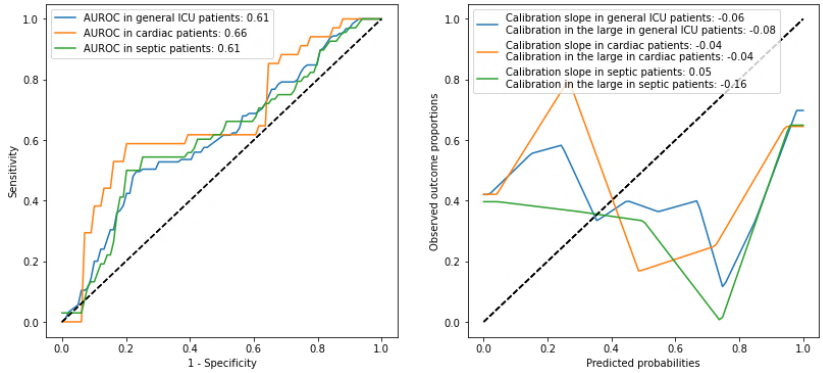
Ten-fold cross validation for complete or partial recovery prediction:



Validation cohort results for complete recovery prediction:

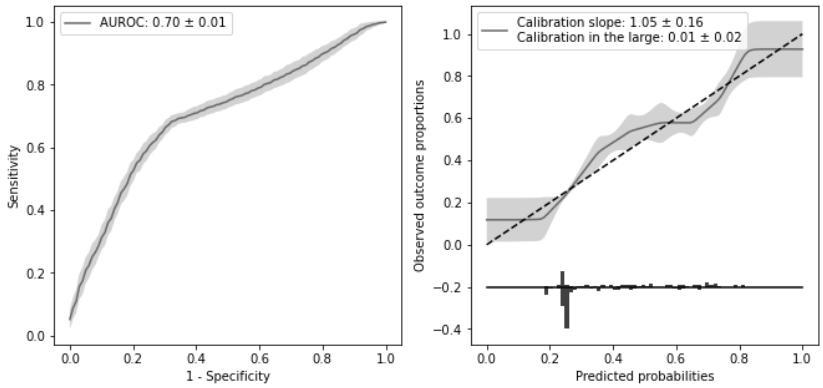


Validation cohort results for complete or partial recovery prediction:

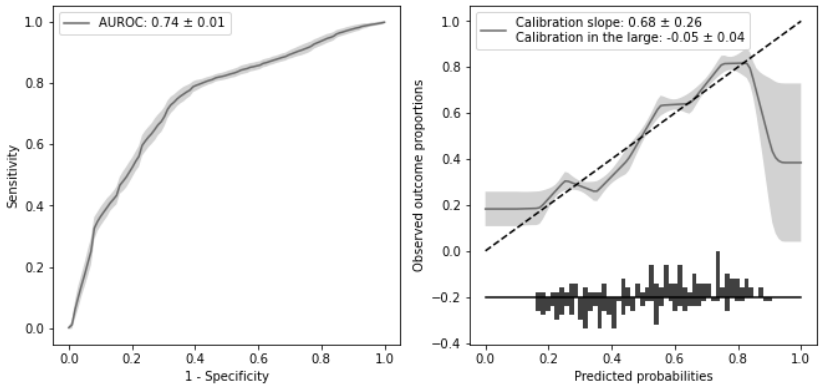


Support vector machine (SVM):

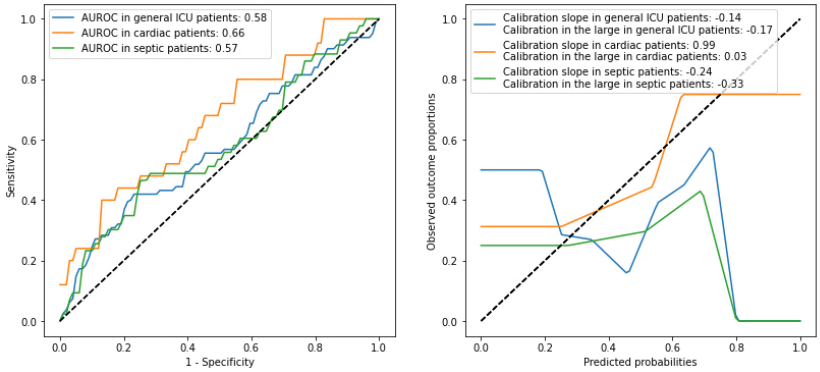
Ten-fold cross validation for complete recovery prediction:



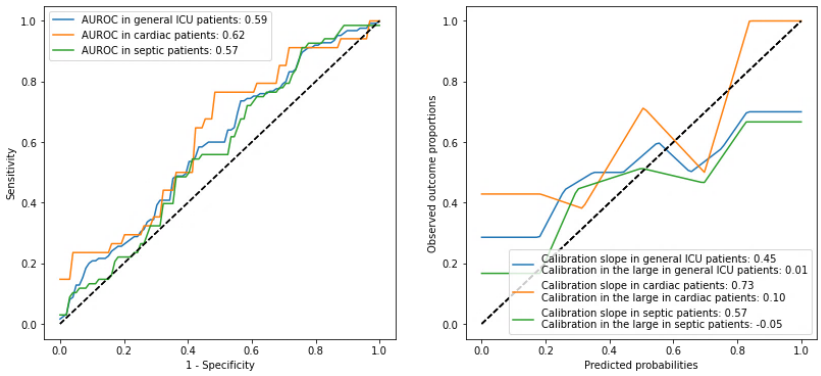
Ten-fold cross validation for complete or partial recovery prediction:



Validation cohort results for complete recovery prediction:



Validation cohort results for complete or partial recovery prediction:



For random forest, gradient boosting, and support vector machine, the validation cohort results demonstrated slightly better AUROCs of 0.55, 0.57, and 0.58 in the complete recovery prediction, and similar AUROCs of 0.60, 0.61, and 0.59 in the complete or partial recovery prediction. Despite the slightly better discrimination of other algorithms for complete recovery prediction, before moving to the validation cohort, Lasso algorithm showed similar to better discrimination and decent calibration in the ten-fold cross validation, so we believe the choice of Lasso algorithm was appropriate.

GRANT SUPPORT AND CONFLICT OF INTEREST

C-Y.Huang receives a grant from the Taiwan-KU Leuven scholarship. J.Gunst holds a postdoctoral research fellowship supported by the University Hospitals Leuven. G. De Vlieger receives a clinical fellowship grant of the Flanders Research Foundation (1701719 N). G.Meyfroidt received a grant of the Flanders Research Foundation as senior clinical investigator. This work was supported by the Methusalem program of the Flemish government (through the University of Leuven to G. Van den Berghe, METH/08/07 and to G. Van den Berghe and I.Vanhorebeek, METH14/06).

All the authors declare no competing interests that are relevant to the content of this article.

ACKNOWLEDGMENT AND PERSONAL CONTRIBUTION

Study concept and design	Meyfroidt, Güiza
Data acquisition	**Huang,** Güiza, Wouters, Gunst, Casaer, Vanhorebeek, Derese, Van den Berghe
Statistical analysis	**Huang**
Interpretation of results	**All authors**
Drafting of the manuscript	**Huang,** Güiza, De Vlieger, Meyfroidt
Manuscript revision	**All authors**
Principal investigator	Meyfroidt

Chao-Yuan Huang and Fabian Güiza have contributed equally to this work. The authors would like to thank the members of the EPaNIC research group for helpfully providing the EPaNIC database, which was used to develop the AKI recovery prediction models.

BIBLIOGRAPHY

1. Chertow GM, Burdick E, Honour M, et al (2005) Acute kidney injury, mortality, length of stay, and costs in hospitalized patients. J Am Soc Nephrol 16:3365–3370. https://doi.org/10.1681/ASN.2004090740

2. Lysak N, Bihorac A, Hobson C (2017) Mortality and cost of acute and chronic kidney disease after cardiac surgery. Curr. Opin. Anaesthesiol. 30:113–117

3. Silver SA, Chertow GM (2017) The Economic Consequences of Acute Kidney Injury. Nephron 137:297–301

4. Dasta JF, Kane-Gill S (2019) Review of the Literature on the Costs Associated With Acute Kidney Injury. J. Pharm. Pract. 32:292–302

5. Alshaikh HN, Katz NM, Gani F, et al (2018) Financial Impact of Acute Kidney Injury After Cardiac Operations in the United States. Ann Thorac Surg 105:469–475. https://doi.org/10.1016/j.athoracsur.2017.10.053

6. Jiang L, Zhu Y, Luo X, et al (2019) Epidemiology of acute kidney injury in intensive care units in Beijing: The multi-center BAKIT study. BMC Nephrol 20:468. https://doi.org/10.1186/s12882-019-1660-z

7. Gammelager H, Christiansen CF, Johansen MB, et al (2012) One-year mortality among Danish intensive care patients with acute kidney injury: a cohort study. Crit Care 16:R124. https://doi.org/10.1186/cc11420

8. Fujii T, Uchino S, Takinami M, Bellomo R (2014) Validation of the kidney disease improving global outcomes criteria for AKI and comparison of three criteria in hospitalized patients. Clin J Am Soc Nephrol 9:848–854. https://doi.org/10.2215/CJN.09530913

9. Koeze J, Keus F, Dieperink W, et al (2017) Incidence, timing and outcome of AKI in critically ill patients varies with the definition used and the addition of urine output criteria. BMC Nephrol 18:70. https://doi.org/10.1186/s12882-017-0487-8

10. Ronco C, Bellomo R, Kellum JA (2019) Acute kidney injury. Lancet 394:1949–1964

11. Khwaja A (2012) KDIGO clinical practice guidelines for acute kidney injury. Nephron - Clin. Pract. 120:c179-84

12. Gaudry S, Hajage D, Benichou N, et al (2020) Delayed versus early initiation of renal replacement therapy for severe acute kidney injury: a systematic review and individual patient data meta-analysis of randomised clinical trials. Lancet (London, England) 395:1506–1515. https://doi.org/10.1016/S0140-6736(20)30531-6

13. Srisawat N, Murugan R, Lee M, et al (2011) Plasma neutrophil gelatinase-associated lipocalin predicts recovery from acute kidney injury following community-acquired pneumonia. Kidney Int 80:545–552. https://doi.org/10.1038/ki.2011.160

14. Dewitte A, Joannès-Boyau O, Sidobre C, et al (2015) Kinetic eGFR and novel AKI biomarkers to predict renal recovery. Clin J Am Soc Nephrol 10:1900–1910. https://doi.org/10.2215/CJN.12651214

15. Fiorentino M, Tohme FA, Murugan R, Kellum JA (2019) Plasma Biomarkers in Predicting Renal Recovery from Acute Kidney Injury in Critically Ill Patients. Blood Purif 48:253–261. https://doi.org/10.1159/000500423

16. Srisawat N, Wen X, Lee MJ, et al (2011) Urinary biomarkers and renal recovery in critically ill patients with renal support. Clin J Am Soc Nephrol 6:1815–1823. https://doi.org/10.2215/CJN.11261210

17. Pike F, Murugan R, Keener C, et al (2015) Biomarker enhanced risk prediction for adverse outcomes in critically Ill patients receiving RRT. Clin J Am Soc Nephrol 10:1332–1339. https://doi.org/10.2215/CJN.09911014

18. Flechet M, Güiza F, Schetz M, et al (2017) AKIpredictor, an online prognostic calculator for acute kidney injury in adult critically ill patients: development, validation and comparison to serum neutrophil gelatinase-associated lipocalin. Intensive Care Med 43:764–773. https://doi.org/10.1007/s00134-017-4678-3

19. Malhotra R, Kashani KB, Macedo E, et al (2017) A risk prediction score for acute kidney injury in the intensive care unit. Nephrol Dial Transplant 32:814–822. https://doi.org/10.1093/ndt/gfx026

20. Chiofolo C, Chbat N, Ghosh E, et al (2019) Automated Continuous Acute Kidney Injury Prediction and Surveillance: A Random Forest Model. Mayo Clin Proc 94:783–792. https://doi.org/10.1016/j.mayocp.2019.02.009

21. Zimmerman LP, Reyfman PA, Smith ADR, et al (2019) Early prediction of acute kidney injury following ICU admission using a multivariate panel of physiological measurements. BMC Med Inform Decis Mak 19:16. https://doi.org/10.1186/s12911-019-0733-z

22. Tomašev N, Glorot X, Rae JW, et al (2019) A clinically applicable approach to continuous prediction of future acute kidney injury. Nature 572:116–119. https://doi.org/10.1038/s41586-019-1390-1

23. Li Y, Yao L, Mao C, et al (2018) Early Prediction of Acute Kidney Injury in Critical Care Setting Using Clinical Notes. In: 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp 683–686

24. Chen Y, Feng F, Li M, et al (2019) Development of a risk stratification-based model for prediction of acute kidney injury in critically ill patients. Med (United States) 98:e16867. https://doi.org/10.1097/MD.0000000000016867

25. Parreco J, Soe-Lin H, Parks JJ, et al (2019) Comparing machine learning algorithms for predicting acute kidney injury. Am Surg 85:725–729. https://doi.org/10.1177/000313481908500731

26. Dharan KS, John GT, Antonisamy B, et al (2005) Prediction of mortality in acute renal failure in the tropics. Ren Fail 27:289–296. https://doi.org/10.1081/JDI-200056612

27. Chertow GM, Soroko SH, Paganini EP, et al (2006) Mortality after acute renal failure: Models for prognostic stratification and risk adjustment. Kidney Int 70:1120–1126. https://doi.org/10.1038/sj.ki.5001579

28. Demirjian S, Chertow GM, Zhang JH, et al (2011) Model to predict mortality in critically ill adults with acute kidney injury. Clin J Am Soc Nephrol 6:2114–2120. https://doi.org/10.2215/CJN.02900311

29. da Hora Passos R, Ramos JGR, Mendonça EJB, et al (2017) A clinical score to predict mortality in septic acute kidney injury patients requiring continuous renal replacement therapy: The HELENICC score. BMC Anesthesiol 17:21. https://doi.org/10.1186/s12871-017-0312-8

30. Kim Y, Park N, Kim J, et al (2019) Development of a new mortality scoring system for acute kidney injury with continuous renal replacement therapy. Nephrology 24:1233–1240. https://doi.org/10.1111/nep.13661

31. Kang MW, Kim J, Kim DK, et al (2020) Machine learning algorithm to predict mortality in patients undergoing continuous renal replacement therapy. Crit Care 24:42. https://doi.org/10.1186/s13054-020-2752-7

32. Lin K, Hu Y, Kong G (2019) Predicting in-hospital mortality of patients with acute kidney injury in the ICU using random forest model. Int J Med Inform 125:55–61. https://doi.org/10.1016/j.ijmedinf.2019.02.002

33. Xu Z, Luo Y, Adekkanattu P, et al (2019) Stratified mortality prediction of patients with acute kidney injury in critical care. In: Studies in Health Technology and Informatics. IOS Press, pp 462–466

34. Itenov TS, Berthelsen RE, Jensen JU, et al (2018) Predicting recovery from acute kidney injury in critically ill patients: Development and validation of a prediction model. Crit Care Resusc 20:54–60

35. Lee BJ, Hsu C yuan, Parikh R, et al (2019) Predicting Renal Recovery After Dialysis-Requiring Acute Kidney Injury. Kidney Int Reports 4:571–581. https://doi.org/10.1016/j.ekir.2019.01.015

36. Thiele RH, Isbell JM, Rosner MH (2015) AKI associated with cardiac surgery. Clin. J. Am. Soc. Nephrol. 10:500–514

37. Peerapornratana S, Manrique-Caballero CL, Gómez H, Kellum JA (2019) Acute kidney injury from sepsis: current concepts, epidemiology, pathophysiology, prevention and treatment. Kidney Int. 96:1083–1099

38. Kellum JA (2014) How can we define recovery after acute kidney injury? Considerations from epidemiology and clinical trial design. Nephron - Clin Pract 127:81–88. https://doi.org/10.1159/000363681

39. Casaer MP, Mesotten D, Hermans G, et al (2011) Early versus Late Parenteral Nutrition in Critically Ill Adults. N Engl J Med 365:506–517. https://doi.org/10.1056/NEJMoa1102662

40. Witten IH, Frank E, Hall MA, Pal CJ (2016) Data Mining: Practical Machine Learning Tools and Techniques. Elsevier

41. Strobl C, Boulesteix AL, Zeileis A, Hothorn T (2007) Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC Bioinformatics 8:25. https://doi.org/10.1186/1471-2105-8-25

42. Bone RC, Balk RA, Cerra FB, et al (1992) Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. In: Chest. pp 1644–1655

43. Steyerberg EW (2009) Evaluation of performance. In: Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. Springer New York, New York, NY, pp 255–280

44. Kellum JA, Sileanu FE, Bihorac A, et al (2017) Recovery after acute kidney injury. Am J Respir Crit Care Med 195:784–791. https://doi.org/10.1164/rccm.201604-0799OC

45. Fan H, Zhao Y, Zhu JH, et al (2015) Thrombocytopenia as a predictor of severe acute kidney injury in patients with heat stroke. Ren Fail 37:877–881. https://doi.org/10.3109/0886022X.2015.1022851

46. Kertai MD, Zhou S, Karhausen JA, et al (2016) Platelet Counts, Acute Kidney Injury, and Mortality after Coronary Artery Bypass Grafting Surgery. Anesthesiology 124:339–352. https://doi.org/10.1097/ALN.0000000000000959

47. Koo CH, Eun Jung D, Park YS, et al (2018) Neutrophil, Lymphocyte, and Platelet Counts and Acute Kidney Injury After Cardiovascular Surgery. J Cardiothorac Vasc Anesth 32:212–222. https://doi.org/10.1053/j.jvca.2017.08.033

48. Griffin BR, Jovanovich A, You Z, et al (2019) Effects of Baseline Thrombocytopenia and Platelet Decrease Following Renal Replacement Therapy Initiation in Patients With Severe Acute Kidney Injury. Crit Care Med 47:e325–e331. https://doi.org/10.1097/CCM.0000000000003598

49. Wu M, Luan YY, Lu JF, et al (2020) Platelet count as a new biomarker for acute kidney injury induced by hemorrhagic shock. Platelets 31:94–102. https://doi.org/10.1080/09537104.2019.1581921

50. Li DH, Wald R, Blum D, et al (2020) Predicting mortality among critically ill patients with acute kidney injury treated with renal replacement therapy: Development and validation of new prediction models. J Crit Care 56:113–119. https://doi.org/10.1016/j.jcrc.2019.12.015

51. Han SS, Ahn SY, Ryu J, et al (2014) U-shape relationship of white blood cells with acute kidney injury and mortality in critically Ill patients. Tohoku J Exp Med 232:177–185. https://doi.org/10.1620/tjem.232.177

52. Uchino S, Kellum JA, Bellomo R, et al (2005) Acute renal failure in critically ill patients: A multinational, multicenter study. J Am Med Assoc 294:813–818. https://doi.org/10.1001/jama.294.7.813

53. Bagshaw SM (2006) Epidemiology of renal recovery after acute renal failure. Curr. Opin. Crit. Care 12:544–550

54. Bagshaw SM, Laupland KB, Doig CJ, et al (2005) Prognosis for long-term survival and renal recovery in critically ill patients with severe acute renal failure: a population-based study. Crit Care 9:R700. https://doi.org/10.1186/cc3879

55. Bagshaw SM, Bennett M, Haase M, et al (2010) Plasma and urine neutrophil gelatinase-associated lipocalin in septic versus non-septic acute kidney injury in critical illness. Intensive Care Med 36:452–461. https://doi.org/10.1007/s00134-009-1724-9

56. Czempik P, Cieśla D, Knapik P, Krzych Ł (2018) Mortality of patients with acute kidney injury requiring renal replacement therapy. Adv Clin Exp Med 27:327–333. https://doi.org/10.17219/acem/65066

57. Ohnuma T, Uchino S (2017) Prediction models and their external validation studies for mortality of patients with acute kidney injury: A systematic review. PLoS One 12:e0169341. https://doi.org/10.1371/journal.pone.0169341

58. Beaubien-Souligny W, Wald R (2019) Predicting Outcomes in Acute Kidney Injury Survivors: Searching for the Crystal Ball. Kidney Int. Reports 4:520–521

59. Moons KGM, Altman DG, Reitsma JB, et al (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. Ann Intern Med 162:W1–W73. https://doi.org/10.7326/M14-0698

5

External Validation of the Augmented Renal Clearance Predictor in Critically Ill COVID-19 Patients

Adapted from: **Chao-Yuan Huang**, Fabian Güiza, Matthias Gijsen, Isabel Spriet, Dieter Dauwe, Yves Debaveye, Marijke Peetermans, Joost Wauters, Greet Van den Berghe, Geert Meyfroidt, Greet De Vlieger. "External Validation of the Augmented Renal Clearance Predictor in Critically Ill COVID-19 Patients". Submitted for publication.

Presented as:

- Poster presentation at the 42nd Annual Congress of the Belgian Society of Intensive Care Medicine (SIZ), Brussels, Belgium, June, 2022

Abstract

Aim: The ARC predictor, a prediction model developed to predict onset of augmented renal clearance (ARC) on the next intensive care unit (ICU) day, showed good performance and outperformed two existing models in a general ICU setting. The aim is to externally validate the ARC predictor in critically ill coronavirus disease 19 (COVID-19) patients.

Methods: This is a retrospective, single-center, observational study in critically ill COVID-19 adults admitted to the ICU of the University Hospitals Leuven from February 2020 to January 2021. Patient-days with available serum creatinine (SCr) on any given ICU day and measured creatinine clearance (CrCl) on the next ICU day were enrolled. ARC prediction was based on six clinical data: age, sex, day from admission, SCr, trauma, and cardiac surgery. Model was evaluated by discrimination, calibration, and decision curve.

Results: A total of 120 patients (1064 patient-days) were included, and ARC was found in 57 (47.5%) patients, corresponding to 246 (23.1%) patient-days. The ARC predictor demonstrated good discrimination and calibration (area under the receiver operating characteristics curve of 0.86, calibration slope of 1.18, and calibration-in-the-large of 0.14) and wide clinical usefulness range (4.04–81.82%). At the default classification threshold of 20% in the original study, the sensitivity and specificity were 72% and 81%.

Conclusions: The ARC predictor is able to accurately predict ARC in the critically ill COVID-19 patients, based on six routinely collected clinical data. These results imply the potential of ARC predictor to optimize renally cleared drug dosages in this specific ICU population.

5.1 Introduction

Augmented renal clearance (ARC) is common in critically ill patients, with prevalence varying between 20-65% [1]. Currently, there is no generally accepted ARC definition, but it is commonly defined by urinary creatinine clearance (CrCl) being larger than 130 ml/min/1.73m^2 [1], measured from the past 8- to 24-hour window depending on the urine collection time [2]. It has been shown that ARC leads to a decreased exposure to commonly used antibiotics such as beta-lactams and vancomycin [3–5], as well as anticoagulants [6]. Consequently, it has been suggested that increased antibiotic doses are necessary in patients with ARC to increase exposure and decrease the risk of treatment failure [7].

Given the adverse consequences of ARC, the ARC predictor was developed to predict the presence of ARC on the next intensive care unit (ICU) day. The ARC predictor outperformed two existing models (ARC Score [8] and ARCTIC Score [9]) in the validation cohort [10], and has been made publicly available as an online calculator [11]. Despite the success of the ARC predictor in the general ICU setting, and despite its potential as a tool to optimize drug dosing, its performance in other independent patient populations remains to be investigated before it can be recommended for broad clinical use [12].

Since the beginning of 2020, ICUs worldwide have been overwhelmed by a large number of critically ill coronavirus disease 19 (COVID-19) patients. More than 525 million people have been infected, and over six million people have died from COVID-19 [13]. Patients with COVID-19 are mainly characterized by respiratory illness, but many critically ill COVID-19 patients may also suffer from reduced kidney function with an acute kidney injury (AKI) prevalence of 18-81% [14]. In addition, it was recently reported that ARC, the other extreme of renal function spectrum, also occurs frequently in COVID-19 population with prevalence of 25-72% [15].

To the best of our knowledge, only the epidemiology of ARC has been described in critically ill COVID-19 patients [15–19], and there are currently no studies investigating ARC prediction model performance in this patient population. Therefore, we aim to externally validate the ARC predictor in previously unseen critically ill COVID-19 patients.

5.2 Methods

5.2.1 Study databases with inclusion and exclusion criteria

Model validation was performed in adult COVID-19 pneumonia patients admitted to critical care in the University Hospitals Leuven from February 2020 to January 2021, with a positive polymerase chain reaction (PCR) for SARS-coronavirus-2 on a respiratory sample. Ethical approval was obtained from the Ethics Committee (EC) Research UZ/KU Leuven (S66365) with study title of "Machine learning tools in critically ill COVID-19 patients: external validation of the Acute Kidney Injury and Augmented Renal Clearance predictors" on 8th April 2022, and the need for informed consent form was waived because of the non-interventional nature of the study. The study was conducted in compliance with the principles of the Declaration of Helsinki and its later revisions. Patients were excluded if they had end-stage kidney disease defined as chronic hemodialysis and/or kidney transplant upon ICU admission. Patient-days were excluded if they had 1) no available serum creatinine (SCr) measured on the ICU day prior to the day for which the prediction is made, 2) no measured CrCl on the ICU day for which the prediction is made, 3) kidney replacement therapy (KRT) on the day for which the

prediction is made, 4) onset of intermittent dialysis during the previous ICU days, 5) incomplete ICU day (day0), and/or 6) KRT on the day prior to the day for which the prediction is made.

5.2.2 ARC definition

Daily CrCl was measured for each ICU day based on the daily 24h urine output (UO), urinary creatinine (UCr), and SCr with correction for an average body surface area: CrCl (ml/min/1.73m^2) = UCr (mg/dL) × 24h UO (ml/day) / SCr (mg/dL) / 1440 (min/day) × 1.73 / (0.007184 × height (cm)$^{0.725}$ × weight (kg)$^{0.425}$). If more than one value was available on the same ICU day, mean was applied for UCr and SCr, and summation was applied to UO. ARC was defined as a measured CrCl larger than 130 ml/min/1.73m². Data were retrieved from the patient data management system database (Microsoft SQL Server®; Microsoft®, Redmond, Washington, USA). After application of the exclusion criteria, there were no patient-days with missing values for any ARC predictor feature.

5.2.3 ARC predictor

The ARC predictor is a model developed by Gijsen et al. [10] to predict ARC on the next ICU day based on six routinely collected clinical variables: age, sex, day from ICU admission, SCr of the previous day, trauma related diagnosis on ICU admission (True/False), and cardiac surgery related diagnosis on ICU admission (True/False), by using a generalized estimating equation (GEE) logistic regression with backward feature selection. The ARC predictor calculates the predicted probability with the provided six features, which can then be translated into a prediction for ARC on the next ICU day according to a prespecified classification threshold. This classification threshold is set by default at 20%, which maximized sensitivity and specificity in the original study, although the threshold can also be manually adapted.

5.2.4 Evaluation metrics for predictive performance

For better interpretation and comparison with the original study, model performance was evaluated by using the same evaluation metrics: receiving operating characteristics (ROC) curve (including area under the ROC curve (AUROC), sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio, and negative likelihood ratio), calibration plot (including calibration slope (CS), and calibration-in-the-large (CITL)) [20], and decision curve analysis [21]. Precision-recall (PR) curve (including area under the PR curve (AUPRC) was also examined. To further investigate the importance of each ARC predictor feature and to examine whether all features were still predictive for this study cohort, 100 repetitions of AUROC-based permutation importance were measured [22]. Boxplot was used to compare the predicted probabilities between (i) the patient-days with and without ARC during the entire ICU stay and on each ICU day within two weeks after ICU admission, and (ii) the patients with and without ARC where the predicted probabilities were averaged over their ICU stay, regardless of presence of ARC on that day. Additionally, percentage of ARC days and number of ARC days in ARC patients were investigated and visualized with boxplot.

5.2.5 Descriptive analyses and software used

All analyses were performed in Python 3.7.4 (Python Software Foundation, http://www.python.org) with SciPy version 1.7.3 (SciPy.org) and Scikit-learn library 1.0.2 (scikit-learn.org). Descriptive statistics were used to describe the study population, with continuous data presented as medians and interquartile ranges (IQR) and categorical data expressed as counts and percentages (%). To evaluate statistical significance of differences, GEE

model was used with patient identification number as grouping variable. A two-tailed P-value less than or equal to 0.05 was considered statistically significant.

5.3 Results

5.3.1 Study cohort

In total, 120 patients (1064 patient-days) were included, among which ARC was found on at least one ICU day in 57 (47.5%) patients, corresponding to 246 patient-days (23.1%) (**Figure 5.1**). The descriptive statistics per patient and per patient-day are shown in **Table 5.1** and **Table 5.A.1**. Seventy-two percent of the study cohort were male patients. Baseline SCr was missing in 178 (16.73%) patient-days, corresponding to 20 (16.67%) patients, and imputed with median baseline SCr of 0.86 mg/dL. The median (interquartile range, IQR) age was 67 (59 – 75) years, median (IQR) body mass index (BMI) was 28.7 (25.8 – 33.1) kg/m^2, and median (IQR) ICU length of stay (LOS) was 14 (9 – 24) days. Patients with ARC were significantly younger (61 (57 – 67) vs. 73 (65 – 78) years, $p < 0.01$), had lower baseline serum creatinine (0.9 (0.7 – 0.9) vs. 0.9 (0.9 – 1.1), $p < 0.01$), and lower Acute Physiology and Chronic Health Evaluation II (APACHE II) score (18 (13 – 23)) vs. 19 (17 – 27), $p < 0.01$). In ARC patients, the median (IQR) first day of ARC was 1 (1 – 2) day from ICU admission, median (IQR) percentage of ARC days was 61.54% (25 – 100%), and median (IQR) days with ARC was 2 (1 – 6) days (**Figure 5.A.1**). Patient-days with ARC had significantly higher CrCl (152.7 (138.3 – 175.6) vs. 74.8 (52.3 – 102.2) ml/min/1.73m^2, $p < 0.01$). In comparison to the original ARC predictor development cohort, this study cohort consisted of 10% more males (72.5 vs. 62.5%), had comparable ages (67 vs. 65 years), and showed an ICU LOS almost twice as long (14 vs. 8 days).

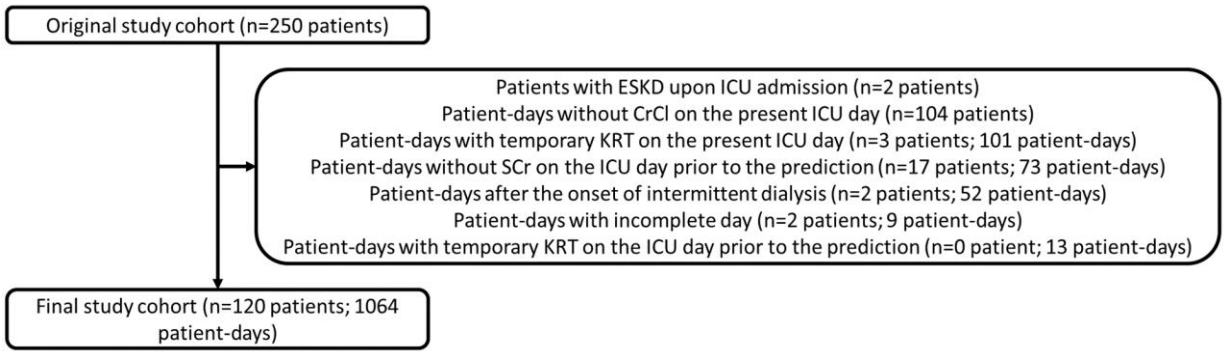


Figure 5.1. Consort diagram. ESKD, end-stage kidney disease; CrCl, creatinine clearance; KRT, kidney replacement therapy; SCr, serum creatinine; ICU, intensive care unit

Table 5.1 Patient characteristics and clinical outcomes

Variables	All patients (n=120)	ARC (n=57, 47.50%)	Not ARC (n=63, 52.50%)	p-value
Age, years, median (IQR)	67 (59 – 75)	61 (57 – 67)	73 (65 – 78)	<0.01
Gender male, number (%)	87 (72.5)	40 (70.2)	47 (74.6)	0.59
Height, m, median (IQR)	1.7 (1.6 – 1.8)	1.8 (1.7 – 1.8)	1.7 (1.6 – 1.8)	0.03
Weight, kg, median (IQR)	85.0 (71.5 – 104.0)	86.0 (70.0 – 104.0)	85.0 (73.5 – 101.5)	0.63
BMI, median (IQR)	28.7 (25.8 – 33.1)	28.1 (26.1 – 32.8)	29.0 (25.8 – 34.3)	0.71
Baseline serum creatinine, mg/dl, median (IQR)	0.9 (0.8 – 1.0)	0.9 (0.7 – 0.9)	0.9 (0.9 – 1.1)	<0.01
APACHE II score, median (IQR)	19 (15 – 25)	18 (13 – 23)	19 (17 – 27)	<0.01
Day from ICU admission, day, median (IQR)	6.0 (3.5 – 10.0)	6.5 (4.0 – 10.0)	6.0 (3.2 – 10.5)	0.73
Creatinine clearance, ml/min/1.73m^2, median (IQR)	91.3 (54.7 – 132.5)	133.8 (106.4 – 165.0)	55.9 (27.8 – 81.9)	<0.01
Length of stay in ICU, days, median (IQR)	14 (9 – 24)	15 (8 – 24)	14 (10 – 24)	0.11

BMI, body mass index; APACHE II score, Acute Physiology and Chronic Health Evaluation II score; IQR, interquartile range

5.3.2 ARC predictor external validation performance

The ARC predictor demonstrated good discrimination comparable to the original study (AUROC: 0.86 vs. 0.89) but slightly worse calibration (CS: 1.18 vs. 0.95; CITL: 0.14 vs. 0.12) (**Figure 5.2**). At the classification threshold of 20% that maximized sensitivity and specificity in the original study, the sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio, and negative likelihood ratio are 72.36%, 81.17%, 53.61%, 90.71%, 3.84, and 0.34 respectively (**Figure 5.3**), in comparison to 87.9%, 76.9%, 48.3%, 96.3%, 3.8, and 0.16 in the original study. Decision curve analysis demonstrated potential clinical usefulness across a broad range of classification thresholds (4.04 – 81.82%), similar to the original study (1 – 71%). The ARC predictor showed a higher AUPRC of 0.62 than the baseline AUPRC of 0.23 (**Figure 5.A.2**). As described in **Figure 5.4**, regardless of the day from ICU admission, the predicted probabilities were significantly higher in patients and patient-days with ARC, compared to patients and patient-days without ARC. On each ICU day

within the first two weeks of ICU admission, predicted probabilities were significantly higher in patient-days with ARC than patient-days without ARC (**Figure 5.A.3**).

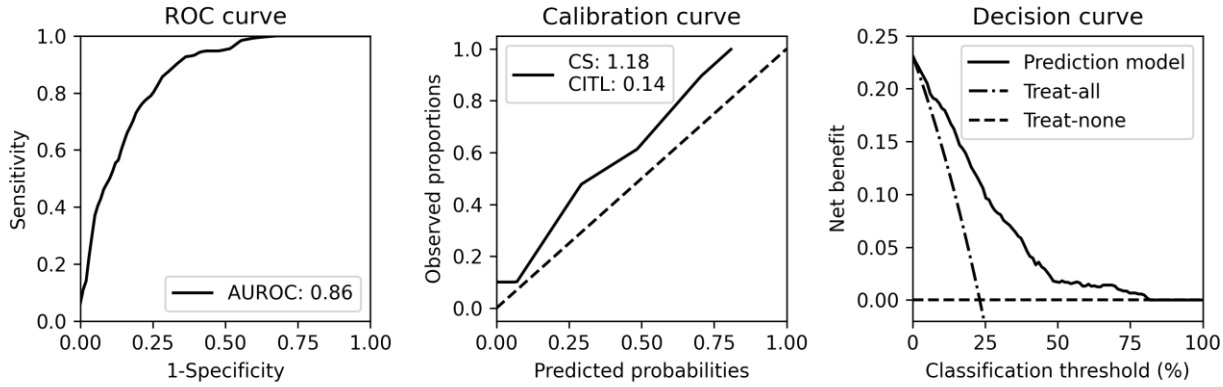


Figure 5.2. ARC predictor performance represented by (i) ROC curve (left) (ii) calibration curve (middle) (iii) decision curve (right). AUROC, area under the receiver operating characteristics curve; CS, calibration slope; CITL, calibration in the large

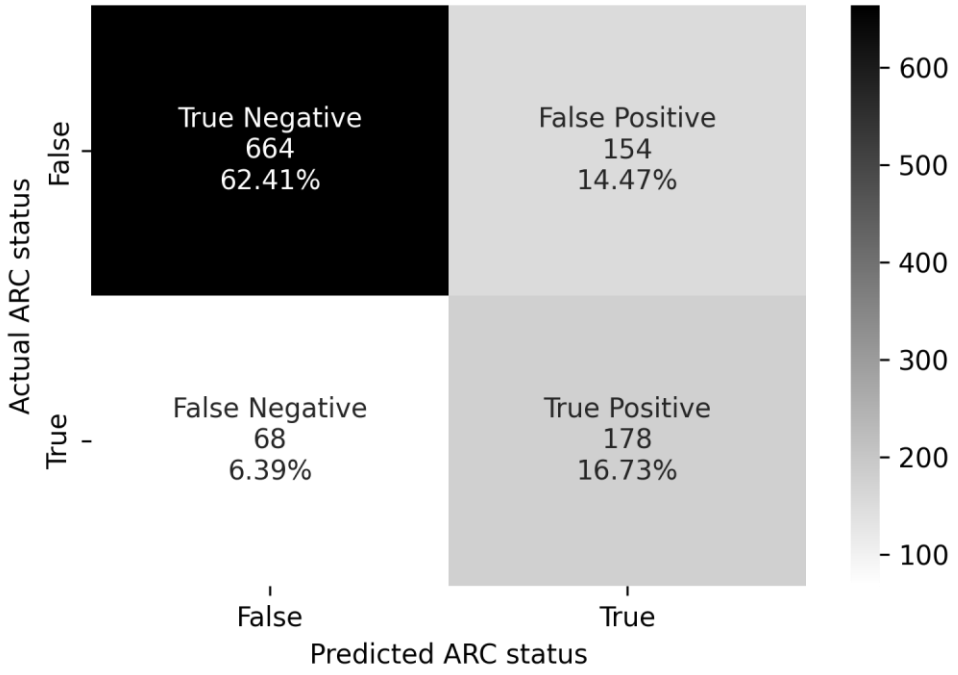


Figure 5.3. Confusion matrix with numbers and percentages of ICU days with true/false positives/negatives. ARC, augmented renal clearance

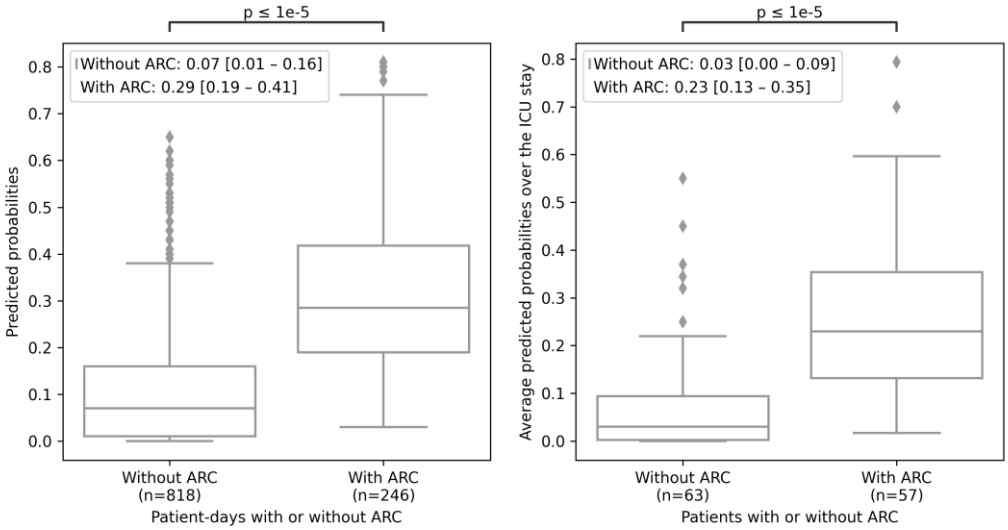


Figure 5.4. Comparison of predicted probabilities of ARC on the next ICU day between (left) patient-days with and without ARC and (right) patients with and without ARC during their ICU stay, with predicted probabilities average over their ICU stay, regardless of presence of ARC on that day. ARC, augmented renal clearance

5.3.3 ARC predictor feature importance

In the present study cohort, SCr of the previous day was the most important feature, followed by day from admission, age, and sex, with median permutation importance of 0.28, 0.10, 0.04, and 0.02 (**Figure 5.A.4**). Since all patients were admitted due to the respiratory insufficiency resulted from COVID-19, they were all without trauma and cardiac surgery related diagnosis on ICU admission and thus with zero permutation importance for these two features.

5.4 Discussion

In this external validation study, we found that ARC predictor had good performance in predicting the presence of ARC on the next ICU day in critically ill COVID-19 patients. Specifically, the robustness of the ARC predictor was confirmed by the comparable ARC predictor discrimination demonstrated in this study compared to the original study and by the significantly higher predicted probabilities in patients and patient-days with ARC. Nevertheless, the calibration plot expressed that the ARC predictor slightly underestimated the ARC risk in this population. Finally, the decision curve analysis showed a similar wide clinical usefulness range and that the default classification threshold of 20% that maximized the sensitivity and specificity in the original study was still able to attain clinical usefulness in this critically ill COVID-19 population. These results demonstrated the potential of the ARC predictor for risk stratification and drug dose adjustment in this specific population of critically ill patents.

Despite the fact that only one percent decrease in discrimination was found and that the clinical usefulness range was still wide, worse calibration was identified. This was expected and understandable, given the significantly different patient characteristics between this critically ill COVID-19 cohort and the original ARC predictor development cohort. The development cohort also consisted of trauma and cardiac surgery

patients and included no COVID-19 patients, which are known to display substantially longer ICU stays than general ICU patients [23]. In addition to the worse calibration, lower sensitivity compared to the original study was also detected, which both indicated that the ARC predictor underestimated the risk of ARC. This underestimation of ARC risk may be partially explained by the fact that the two features contributing to a higher predicted probability, trauma-related and cardiac-related diagnosis at admission, were zeros in our cohort. Nevertheless, COVID-19 patients might experience systemic inflammatory response syndrome, which could (in-)directly overlap with the mechanism of ARC [15] and could consequently increase the ARC risk.

Based on the permutation importance plot, we noticed that the SCr of the previous day was the most important feature, which was reasonable since it directly and timely reflected the time-variant renal function. The second most important feature was day from ICU admission. We found that ARC occurred relatively early, which was opposed to the findings of Beunders et al. who found that ARC occurred late on median (IQR) day 28 (21–42) following ICU admission during COVID-19 infections [16] but in line with the previous studies in general ICU patients (the highest ARC prevalence was observed on day 5) and critically ill COVID-19 patients (median (IQR) first day of ARC was day 2 (3–5) of ICU stay) [1, 18]. Afterwards, age was ranked the third most important feature, which was expected since age has consistently shown a significant association with ARC in many studies [7, 9, 31–34, 10, 24–30]. Age might be more relevant in this patient population, since it has been noticed that some COVID-19 variants are more prevalent in young patients [35, 36]. Also, the association between male gender and ARC is well-known [7, 9, 10, 25, 27, 29–31, 34]. Finally, the permutation importance plot revealed an important message that these four ARC predictor features were with positive permutation importance and thus were all effectively contributing to the final robust predictive performance.

Our study has many strengths. First, not only discrimination, but also calibration and clinical usefulness were investigated, which were considered the key measures to evaluate model performance [37]. Second, the reporting of this study was performed following the Transparent Reporting of a Multivariate Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guidelines [38]. Third, this study was based on a large high-quality COVID-19 cohort without any missing value in the ARC predictor features, and thus no imputation methods were applied, so the presented results are reliable and trustworthy. Fourth, the reported permutation importance helped understanding the contribution of each ARC predictor feature in this study cohort. Finally, the higher predicted probabilities in patients and patient-days with ARC were explicitly investigated, either as a whole in the entire ICU stay regardless of the ICU days, or on each ICU day within two weeks after ICU admission.

There are still several limitations in our study. First, by virtue of the retrospective study nature, there was no impact analysis about whether or not the ARC predictor could help attaining pharmacokinetic targets, optimizing renally cleared drug dosage, and/or improving patient outcome. Second, this is a single-center study in Belgium, while the optimum is to validate the ARC predictor performance in a larger multi-center international setting. Third, there might be a selection bias resulting from the exclusion criteria where patient-days with need for temporary KRT, unavailable SCr on the previous day, and/or unavailable CrCl on the present ICU day were removed, but it is an inherent bias resulting from the same inclusion and exclusion criteria as the original study, and these inclusion/exclusion criteria were necessary to ensure that only reliable CrCls were used for performance evaluation. Fourth, we did not evaluate whether ARC predictor might improve drug dosage. Future studies are

needed to assess whether the ARC predictor is able to improve the drug dosage of antibiotics and low molecular weight heparins.

5.5 Conclusion

We have demonstrated the robustness of ARC predictor in predicting the presence of ARC on the next ICU day in previously unseen critically ill COVID-19 patients, based on six routinely collected clinical variables in the ICU. Despite the promising performance, these findings should be prospectively validated in independent patient populations before the ARC predictor can be implemented for risk stratification or used to inform optimized dosing strategies in routine clinical ICU practice.

5.A Appendix

5.A.1 Supplementary tables

Table 5.A.1 Patient characteristics and clinical outcomes (per patient-day)

Variables	All patient-days (n=1064)	ARC (n=246, 23.12%)	Not ARC (n=818, 76.88%)	p-value
Age, years, median (IQR)	65 (60 – 72)	60 (52 – 65)	70 (60 – 73)	0.03
Gender male, number (%)	805 (75.7)	173 (70.3)	632 (77.3)	0.35
Height, m, median (IQR)	1.8 (1.6 – 1.8)	1.8 (1.7 – 1.8)	1.7 (1.6 – 1.8)	0.19
Weight, kg, median (IQR)	85.0 (75.0 – 104.0)	86.0 (70.0 – 105.0)	85.0 (76.5 – 103.0)	0.73
BMI, median (IQR)	28.7 (26.2 – 32.1)	28.4 (26.3 – 32.1)	28.7 (26.2 – 31.9)	0.19
Baseline serum creatinine, mg/dl, median (IQR)	0.9 (0.8 – 0.9)	0.9 (0.7 – 0.9)	0.9 (0.8 – 1.0)	<0.01
APACHE II score, median (IQR)	19 (15 – 25)	17 (14 – 22)	19 (16 – 25)	<0.01
Day from ICU admission, day, median (IQR)	9.0 (4.0 – 18.0)	8.0 (4.0 – 13.0)	10.0 (5.0 – 20.0)	<0.01
Creatinine clearance, ml/min/1.73m^2, median (IQR)	91.9 (60.4 – 127.9)	152.7 (138.3 – 175.6)	74.8 (52.3 – 102.2)	<0.01
Length of stay in ICU, days, median (IQR)	22 (13 – 38)	18 (11 – 23)	24 (15 – 44)	<0.01

Statistically significant difference was examined by using univariable generalized estimating equation (GEE) model, with ICU admission number as clustering variable, and days with ARC as reference group. BMI, body mass index; APACHE II score, Acute Physiology and Chronic Health Evaluation II score; IQR, interquartile range

5.A.2 Supplementary figures

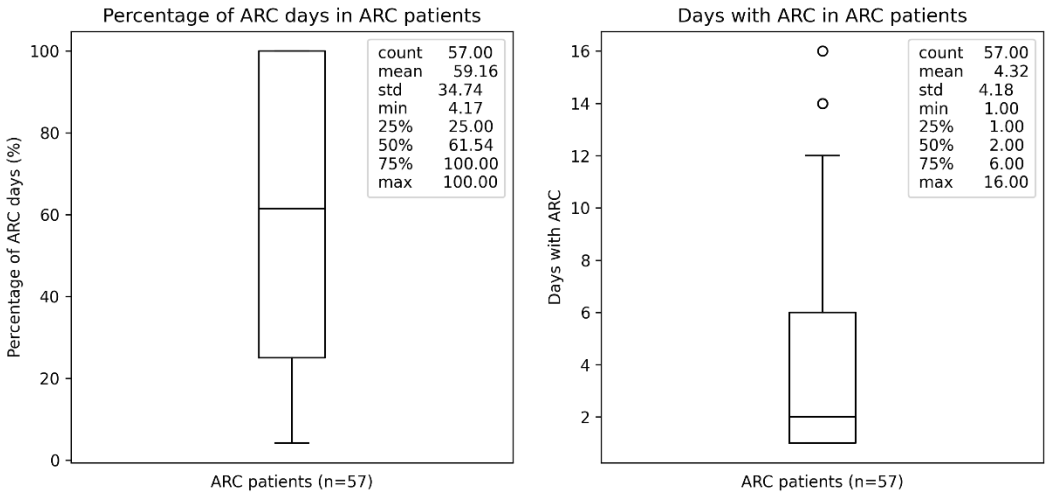


Figure 5.A.1 Percentage of ARC days (left) and number of days with ARC (right) in ARC patients; ARC, augmented renal clearance

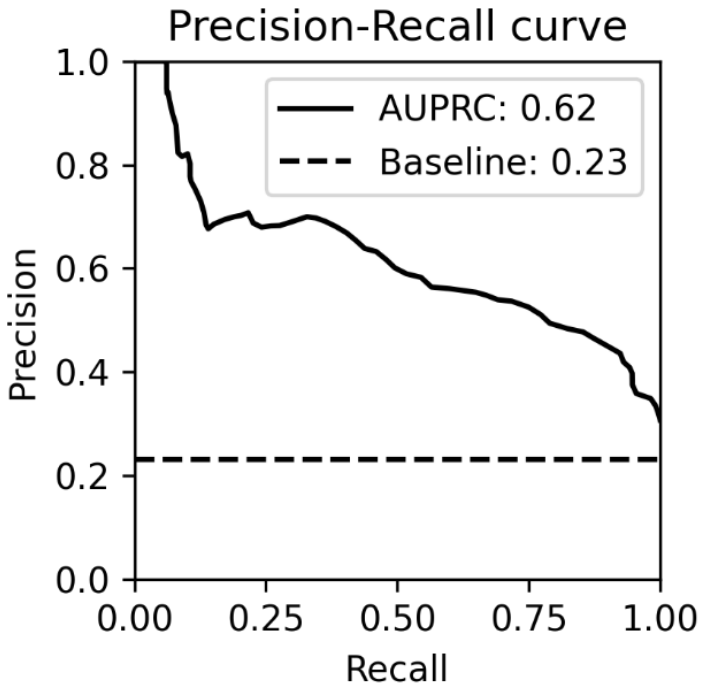


Figure 5.A.2 Precision recall curve. AUPRC, area under the precision recall curve. Baseline, the number of positive cases (patient-days with presence of augmented renal clearance) over the total number of patient-day

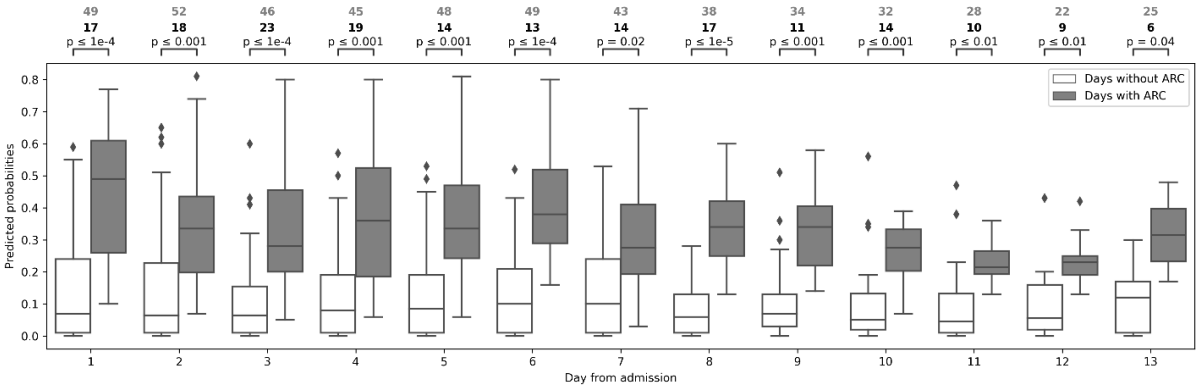


Figure 5.A.3 Comparison of predicted probabilities of ARC on each ICU day between patient-days with and without ARC, within the first two weeks of ICU admission. The black and grey numbers above the figure indicated the numbers of patient-days with and without ARC on each ICU day. ARC, augmented renal clearance

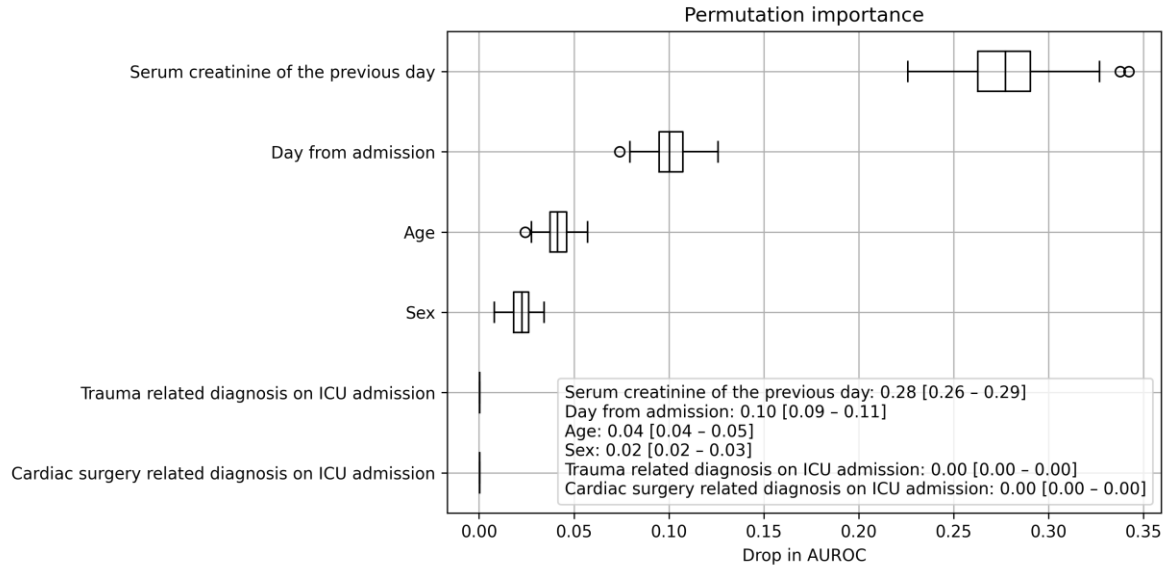


Figure 5.A.4 Boxplots of permutation importance of all augmented renal clearance predictor features with 100 repetitions. AUROC, area under the receiver operating characteristics curve

GRANT SUPPORT AND CONFLICT OF INTEREST

C-Y.Huang receives funding from the Taiwan-KU Leuven scholarship. M.Gijsen is a postdoctoral research fellow at KU Leuven (file number PDMT2/21/077). I.Spriet is funded by the Clinical Research Fund, University Hospitals Leuven. G.Meyfroidt is funded by the research foundation, Flanders (FWO) as senior clinical investigator.

All the authors declare no competing interests that are relevant to the content of this article.

ACKNOWLEDGMENT AND PERSONAL CONTRIBUTION

Study concept and design	**Huang**, Güiza, De Vlieger, Meyfroidt
Data acquisition	**Huang**, De Vlieger
Statistical analysis	**Huang**
Interpretation of results	**All authors**
Drafting of the manuscript	**Huang**, Güiza, De Vlieger, Meyfroidt
Manuscript revision	**All authors**
Principal investigator	Meyfroidt

The authors would like to thank the members of the University Hospitals Leuven COVID-19 units for providing the necessary data used to validate the ARC predictor. The authors would like to mention that some authors (C-Y.**Huang**, M.Gijsen, Y.Debaveye, G.Meyfroidt, J.Wauters, I.Spriet) participated in the development of the ARC predictor.

BIBLIOGRAPHY

1. Bilbao-Meseguer I, Rodríguez-Gascón A, Barrasa H, et al (2018) Augmented Renal Clearance in Critically Ill Patients: A Systematic Review. Clin Pharmacokinet 57:1107–1121. https://doi.org/10.1007/s40262-018-0636-7

2. Cherry RA, Eachempati SR, Hydo L, Barie PS (2002) Accuracy of short-duration creatinine clearance determinations in predicting 24-hour creatinine clearance in critically ill and injured patients. J Trauma 53:267–271. https://doi.org/10.1097/00005373-200208000-00013

3. He J, Yang Z-TT, Qian X, et al (2020) A higher dose of vancomycin is needed in critically ill patients with augmented renal clearance. Transl Androl Urol 9:2166–2171. https://doi.org/10.21037/tau-20-1048

4. Huttner A, Von Dach E, Renzoni A, et al (2015) Augmented renal clearance, low β-lactam concentrations and clinical outcomes in the critically ill: An observational prospective cohort study. Int J Antimicrob Agents 45:385–392. https://doi.org/10.1016/j.ijantimicag.2014.12.017

5. Udy AA, Varghese JM, Altukroni M, et al (2012) Subtherapeutic initial β-lactam concentrations in select critically ill patients: association between augmented renal clearance and low trough drug concentrations. Chest 142:30–39. https://doi.org/10.1378/chest.11-1671

6. Abdel El Naeem HEM, Abdelhamid MHE, Atteya DAM (2017) Impact of augmented renal clearance on enoxaparin therapy in critically ill patients. Egypt J Anaesth 33:113–117. https://doi.org/https://doi.org/10.1016/j.egja.2016.11.001

7. Claus BOMM, Hoste EA, Colpaert K, et al (2013) Augmented renal clearance is a common finding with worse clinical outcome in critically ill patients receiving antimicrobial therapy. J Crit Care 28:695–700. https://doi.org/10.1016/j.jcrc.2013.03.003

8. Udy AA, Roberts JA, Shorr AF, et al (2013) Augmented renal clearance in septic and traumatized patients with normal plasma creatinine concentrations: Identifying at-risk patients. Crit Care 17:. https://doi.org/10.1186/cc12544

9. Barletta JF, Mangram AJ, Byrne M, et al (2017) Identifying augmented renal clearance in trauma patients: Validation of the Augmented Renal Clearance in Trauma Intensive Care scoring system. In: The journal of trauma and acute care surgery. United States, pp 665–671

10. Gijsen M, Huang CY, Flechet M, et al (2020) Development and External Validation of an Online Clinical Prediction Model for Augmented Renal Clearance in Adult Mixed Critically Ill Patients: The Augmented Renal Clearance Predictor. Crit Care Med 48:e1260–e1268. https://doi.org/10.1097/CCM.0000000000004667

11. ARC prediction. https://www.arcpredictor.com/. Accessed 13 Apr 2022

12. Steyerberg EW, Harrell FEJ (2016) Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol 69:245–247. https://doi.org/10.1016/j.jclinepi.2015.04.005

13. COVID-19 Map - Johns Hopkins Coronavirus Resource Center. https://coronavirus.jhu.edu/map.html. Accessed 19 May 2022

14. Silver SA, Beaubien-Souligny W, Shah PS, et al (2021) The Prevalence of Acute Kidney Injury in Patients Hospitalized With COVID-19 Infection: A Systematic Review and Meta-analysis. Kidney Med 3:83-98.e1. https://doi.org/https://doi.org/10.1016/j.xkme.2020.11.008

15. Rhoney DH, Brooks AB, Nelson NR (2022) Augmented Renal Clearance: An Under-Recognized Phenomenon Associated With COVID-19. Crit care Explor 4:e0617. https://doi.org/10.1097/CCE.0000000000000617

16. Beunders R, van de Wijgert IH, van den Berg M, et al (2021) Late augmented renal clearance in patients with COVID-19 in the intensive care unit. A prospective observational study. J Crit Care 64:7–9. https://doi.org/10.1016/j.jcrc.2021.02.009

17. Nelson N, Brooks A, Rhoney D (2022) 250: A RAPID REVIEW OF AUGMENTED RENAL CLEARANCE IN PATIENTS WITH COVID-19. Crit Care Med 50:

18. Dhaese S, Peperstraete H, Hoste E, et al (2021) Augmented renal clearance in critically ill COVID-19 patients: Forewarned is forearmed. J Crit Care 66:93–95. https://doi.org/10.1016/j.jcrc.2021.08.004

19. Murt A, Dincer MT, Karaca C (2021) Augmented Renal Clearance in COVID-19. Nephron 145:386–387. https://doi.org/10.1159/000515423

20. Steyerberg EW (2009) Evaluation of performance. In: Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. Springer New York, New York, NY, pp 255–280

21. Vickers AJ, Elkin EB (2006) Decision curve analysis: a novel method for evaluating prediction models. Med Decis Mak an Int J Soc Med Decis Mak 26:565–574. https://doi.org/10.1177/0272989X06295361

22. Random Forests. https://www.stat.berkeley.edu/~breiman/RandomForests/. Accessed 19 May 2022

23. Higgins TL, Stark MM, Henson KN, Freeseman-Freeman L (2021) Coronavirus Disease 2019 ICU Patients Have Higher-Than-Expected Acute Physiology and Chronic Health Evaluation-Adjusted Mortality and Length of Stay Than Viral Pneumonia ICU Patients. Crit Care Med 49:e701–e706. https://doi.org/10.1097/CCM.0000000000005012

24. Nei AM, Kashani KB, Dierkhising R, Barreto EF (2020) Predictors of Augmented Renal Clearance in a Heterogeneous ICU Population as Defined by Creatinine and Cystatin C. Nephron 55905:313–320. https://doi.org/10.1159/000507255

25. Udy AA, Dulhunty JM, Roberts JA, et al (2017) Association between augmented renal clearance and clinical outcomes in patients receiving β-lactam antibiotic therapy by continuous or intermittent infusion: a nested cohort study of the BLING-II randomised, placebo-controlled, clinical trial. Int J Antimicrob Agents 49:624–630. https://doi.org/10.1016/j.ijantimicag.2016.12.022

26. Kawano Y, Maruyama J, Hokama R, et al (2018) Outcomes in patients with infections and augmented renal clearance: A multicenter retrospective study. PLoS One 13:1–11. https://doi.org/10.1371/journal.pone.0208742

27. Mulder MB, Eidelson SA, Sussman MS, et al (2019) Risk Factors and Clinical Outcomes Associated With Augmented Renal Clearance in Trauma Patients. J Surg Res 244:477–483. https://doi.org/10.1016/j.jss.2019.06.087

28. Nazer LH, AbuSara AK, Kamal Y (2021) Augmented renal clearance in critically ill patients with cancer (ARCCAN Study): A prospective observational study evaluating prevalence and risk factors. Pharmacol Res Perspect 9:2–7. https://doi.org/10.1002/prp2.747

29. Johnston BW, Perry D, Habgood M, et al (2021) Augmented renal clearance: a retrospective, cohort study of urinary creatinine clearance in critically ill patients in the United Kingdom. J Int Med Res 49:. https://doi.org/10.1177/03000605211015573

30. Baptista JP, Martins PJ, Marques M, Pimentel JM (2020) Prevalence and Risk Factors for Augmented Renal Clearance in a Population of Critically Ill Patients. J Intensive Care Med 35:1044–1052. https://doi.org/10.1177/0885066618809688

31. Dickerson RN, Crawford CN, Tsiu MK, et al (2021) Augmented renal clearance following traumatic injury in critically ill patients requiring nutrition therapy. Nutrients 13:. https://doi.org/10.3390/nu13051681

32. Wu CC, Tai CH, Liao WY, et al (2019) Augmented renal clearance is associated with inadequate antibiotic pharmacokinetic/ pharmacodynamic target in Asian ICU population: a prospective observational study. Infect Drug Resist 12:2531–2541. https://doi.org/10.2147/IDR.S213183

33. De Waele JJ, Dumoulin A, Janssen A, Hoste EAJ (2015) Epidemiology of augmented renal clearance in mixed ICU patients. Minerva Anestesiol 81:1079–1085

34. Udy AA, Baptista JP, Lim NL, et al (2014) Augmented renal clearance in the ICU: Results of a Multicenter observational study of renal function in critically ill patients with normal plasma Creatinine concentrations. Crit Care Med 42:520–527. https://doi.org/10.1097/CCM.0000000000000029

35. Twohig KA, Nyberg T, Zaidi A, et al (2022) Hospital admission and emergency care attendance risk for SARS-CoV-2 delta (B.1.617.2) compared with alpha (B.1.1.7) variants of concern: a cohort study. Lancet Infect Dis 22:35–42. https://doi.org/10.1016/S1473-3099(21)00475-8

36. Chadeau-Hyam M, Wang H, Eales O, et al (2022) SARS-CoV-2 infection and vaccine effectiveness in England (REACT-1): a series of cross-sectional random community surveys. Lancet Respir Med 10:355–366. https://doi.org/10.1016/S2213-2600(21)00542-7

37. Steyerberg EW, Vergouwe Y (2014) Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J 35:1925–1931. https://doi.org/10.1093/eurheartj/ehu207

38. Moons KGM, Altman DG, Reitsma JB, et al (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. Ann Intern Med 162:W1–W73. https://doi.org/10.7326/M14-0698

6

Daily fluctuations in kidney function in critically ill adults

Adapted from: **Chao-Yuan Huang**, Fabian Güiza, Greet De Vlieger, Geert Meyfroidt. “Daily fluctuations in kidney function in critically ill adults”. *Critical Care.* 2022;26(1):347. doi:10.1186/s13054-022-04226-3

6.1 Letter to the Editor

It is well known that the kidney function can change rapidly during critical illness, with either sudden increases [1] or decreases [2] in renal clearance that may have potentially important consequences for drug dosage adjustments involving renally excreted drugs. However, the true incidence and degree of these fluctuations have never been described systematically. In this study, we aim to investigate kidney function fluctuations as defined by the daily differences in creatinine clearance (CrCl) in critically ill adults.

For the present study, data were retrieved from patients included in the large multicenter EPaNIC randomized controlled trial [3] that compared two parenteral nutrition strategies in 4640 critically ill adults between 2007 and 2010. The Ethics Committee of University Hospitals Leuven approved the use of these patient data for additional analyses (S50404). Daily CrCl was calculated by multiplying the urinary creatinine (UCr) (measured on 24-hour urine collection) with the 24-hour urine output (UO), divided by the serum creatinine (SCr) times the collection time (1440 minutes), without correcting for body surface area (BSA).

$$CrCl\ (ml/min) = \frac{UCr \times UO}{SCr \times 1440}$$

As the University Hospitals Leuven were the only EPaNIC center where CrCl was calculated daily, patients from other centers were excluded. Laboratory results were exported from patient data management system database (Microsoft SQL Server®; Microsoft®, Redmond, Washington, USA), and the remaining data were retrieved from EPaNIC trial database (Filemaker Pro®; FileMaker Inc, FileMaker International). As a measure of daily fluctuations in CrCl, we calculated the difference between the CrCl of each pair of two consecutive days. In case the absolute difference in CrCl of 2 consecutive days was less than 20 ml/min, the daily fluctuation was labeled as 'stable'. A difference larger than 20 ml/min was defined as 'unstable'. We decided to use 20 ml/min as a cutoff as this is a meaningful difference for drug dosing and because the CrCl variability in healthy volunteers has been reported with mean differences of 21.7 ml/min/1.73m² and variations of 18.7% [4]. Depending on the direction of the fluctuation, we used the categories 'unstable-upward' if the CrCl on the next consecutive day was >20 ml/min higher than the previous day, or 'unstable-downward' if the CrCl on the next consecutive day was >20 ml/min lower compared to the day before. We also investigated the relative change in CrCl, defining a relative instable CrCl as a change of > 20% between two consecutive CrCl values. Finally, we examined CrCl fluctuations in the first week of ICU admission separately.

Of the 4389 patients in the study cohort, 2825 patients, corresponding to 18494 patient-days, met the inclusion criteria (**Figure 6.A.1**). Descriptive statistics are available in **Table 6.A.1**. In terms of absolute instability (**Figure 6.1a**), CrCl remained stable in 65% of days, while 19% were unstable-upward and 16% were unstable-downward. Across the CrCl range, the percentage of stable days decreased approximately linearly with increasing CrCl values, such that stability was above 70% for CrCl below 75 ml/min and around 30% for CrCl above 180 ml/min. Additionally, more than 50% of the days were unstable-downward for CrCl above 180 ml/min, and the percentage of unstable-upward days was 20% for CrCl in a range of 45–180 ml/min. When we used the relative definition of a 20% difference (**Figure 6.1b**), the percentages of stable, unstable-upward, and unstable-downward days were 58%, 25%, and 17%, respectively, with an overall CrCl stability ranging around 60% throughout the CrCl range. The percentage of unstable-upward cases declined roughly linearly across the CrCl range, ranging from more than 30% for CrCl below 60 ml/min to about 6% for CrCl beyond 180 ml/min.

Additionally, for CrCl below 180 ml/min, the percentage of unstable-downward cases remained low, around 15%, and it rose to 37% for CrCl above 180 ml/min. For absolute and relative differences alike, more instability was observed in the first week of ICU stay (**Figure 6.1c** and **Figure 6.1d**). Specifically, for patient-days within the first week of ICU admission, 39% were unstable, including 23% unstable-upward and 16% unstable-downward days for absolute difference, and 50% were unstable, including 31% unstable-upward and 19% unstable-downward days for relative difference.

To conclude, our findings confirm that potentially clinically significant changes in kidney function may occur on a daily basis in critically ill patients on approximately 35–40% of days, depending on the definition of instability. This instability mainly occurs in the first week of ICU admission and is more pronounced when the CrCl is higher. The measured CrCl is known to overestimate the glomerular filtration rate (GFR) as compared to inulin clearance, but it has been shown the most reliable and cheap method to assess the GFR on a daily basis in the ICU [5]. While these findings have to be confirmed in independent cohorts of critically ill patients, additional investigations are needed to determine the factors associated with fluctuations in renal clearance.

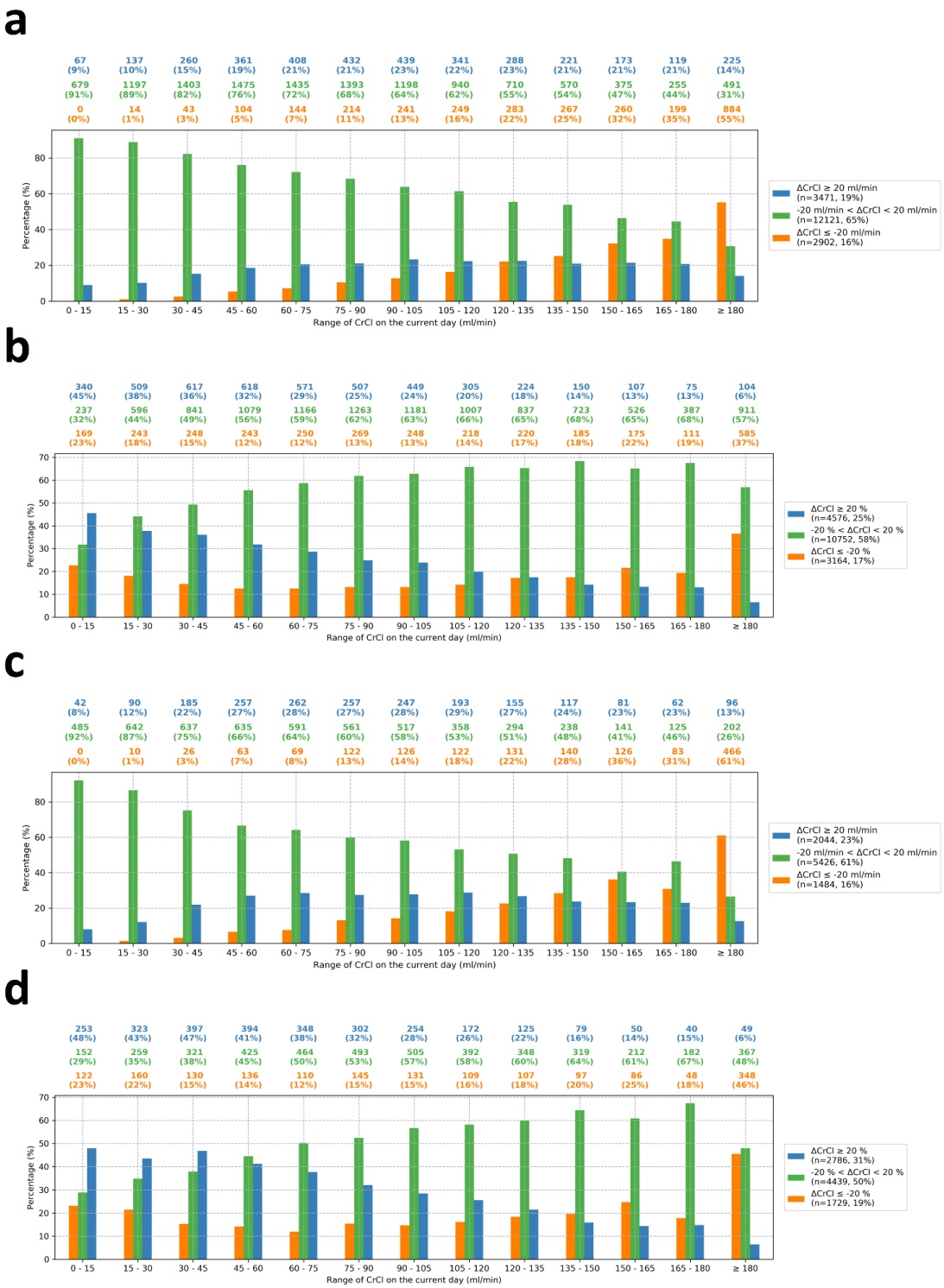


Figure 6.1 Percentage of patient-days during the entire ICU stay (**a**, **b**) and within the first week of ICU admission (**c**, **d**) with stable, unstable-upward, or unstable-downward CrCl for different CrCl ranges. The number and percentage of patient-days for each CrCl ranges are indicted above the figure. **a, c** The blue, orange, and green bars represent respectively: an increase larger than 20 ml/min in the CrCl of the next day compared to the current day (*unstable-upward*), a decrease larger than 20 ml/min in the CrCl of the next day compared to the current day (*unstable-downward*), and an absolute difference smaller than 20 ml/min between the CrCl of the next and the current day (*stable*). ΔCrCl, CrCl of the next day minus CrCl of the current day; **b, d** The blue, orange, and green bars represent respectively: an increase larger than 20% in the CrCl of the next day compared to the current day (*unstable-upward*), a decrease larger than 20% in the CrCl of the next day compared to the current day (*unstable-downward*), and an absolute difference smaller than 20% between the CrCl of the next and the current day (*stable*). ΔCrCl, (CrCl of the next day-CrCl of the current day)/(CrCl of the current day); CrCl, creatinine clearance.

6.A Appendix

6.A.1 Supplementary tables

Table 6.A.1 Patient characteristics and clinical outcomes

	Study cohort (n=2825)
Age, years, median (IQR)	67.59 (56.16 – 75.61)
Gender male, number (%)	1747 (61.84)
Emergency admission, number (%)	1272 (45.03)
APACHE II score, median (IQR)	22 (16 – 32)
Reason for admission	
Cardiac surgery, number (%)	1655 (58.58)
Medical disease, number (%)	114 (4.04)
Neurology and neurosurgery, number (%)	119 (4.21)
Trauma and other surgery, number (%)	662 (23.43)
Transplantation, number (%)	275 (9.73)
ICU mortality, number (%)	162 (5.73)
Length of stay in ICU, days, median (IQR)	5 (3 – 11)

APACHE II score, Acute Physiology and Chronic Health Evaluation II score; IQR, interquartile range; ICU, intensive care unit

6.A.2 Supplementary figures

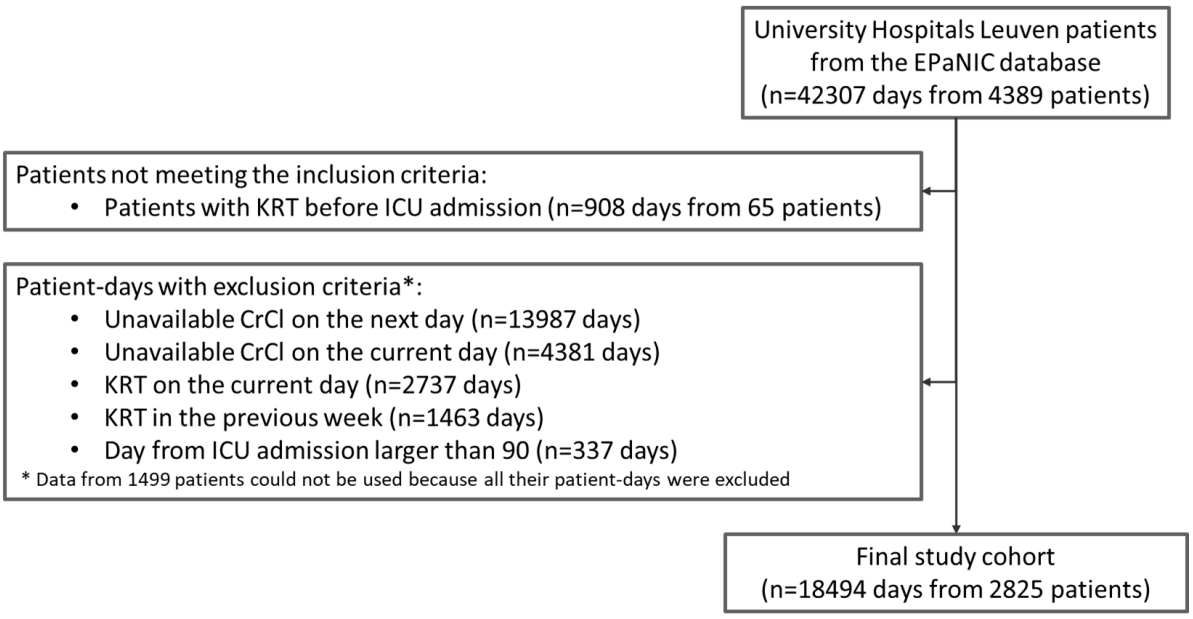


Figure 6.A.1 Study cohort. CrCl, creatinine clearance; KRT, kidney replacement therapy; ICU, intensive care unit.

GRANT SUPPORT AND CONFLICT OF INTEREST

C-Y.Huang receives funding from the Taiwanese Ministry of Education through a Taiwan-KU Leuven scholarship. G.Meyfroidt is funded by the Research Foundation, Flanders (FWO) as senior clinical investigator (1843123N).

The authors declare that they have no conflict of interest.

ACKNOWLEDGMENT AND PERSONAL CONTRIBUTION

Study concept and design	Meyfroidt, Güiza, De Vlieger
Data acquisition	**Huang,** Güiza
Statistical analysis	**Huang**
Interpretation of results	**All authors**
Drafting of the manuscript	**Huang**
Manuscript revision	**All authors**
Principal investigator	Meyfroidt

BIBLIOGRAPHY

1. Johnston BW, Perry D, Habgood M, et al (2021) Augmented renal clearance: a retrospective, cohort study of urinary creatinine clearance in critically ill patients in the United Kingdom. J Int Med Res 49:. https://doi.org/10.1177/03000605211015573
2. Jiang L, Zhu Y, Luo X, et al (2019) Epidemiology of acute kidney injury in intensive care units in Beijing: The multi-center BAKIT study. BMC Nephrol 20:468. https://doi.org/10.1186/s12882-019-1660-z
3. Casaer MP, Mesotten D, Hermans G, et al (2011) Early versus Late Parenteral Nutrition in Critically Ill Adults. N Engl J Med 365:506–517. https://doi.org/10.1056/NEJMoa1102662
4. Toffaletti JG, McDonnell EH (2008) Variation of serum creatinine, cystatin C, and creatinine clearance tests in persons with normal renal function. Clin Chim Acta 395:115—119. https://doi.org/10.1016/j.cca.2008.05.020
5. Carlier M, Dumoulin A, Janssen A, et al (2015) Comparison of different equations to assess glomerular filtration in critically ill patients. Intensive Care Med 41:427–435. https://doi.org/10.1007/s00134-014-3641-9

7

Development and validation of the creatinine clearance predictor machine learning models in critically ill adults

Adapted from: **Chao-Yuan Huang**, Fabian Güiza, Pieter Wouters, Liese Mebis, Giorgia Carra, Jan Gunst, Philippe Meersseman, Michael Casaer, Greet Van den Berghe, Greet De Vlieger, Geert Meyfroidt. "Development and validation of the creatinine clearance predictor machine learning models in critically ill adults". Ready for submission.

Presented as:

- Poster presentation at the 33rd Annual Congress of the European Society of Intensive Care Medicine (ESICM), Online, December, 2020
- Poster presentation at the 34th Annual Congress of the European Society of Intensive Care Medicine (ESICM), Online, October, 2021

Awarded the **Outstanding pitch** of the 2021 Janssen Pharmaceutica - VLIR/ULiège Interactive Career Day Data Sciences, Online, October 2021

Abstract

Purpose: In critically ill patients, measured creatinine clearance (CrCl) is the most reliable method to evaluate glomerular filtration rate, but may vary subsequently on a day-to-day basis. We developed and externally validated models to predict CrCl one day ahead, and compared them with current clinical practice assuming the CrCl remains stable.

Methods: A gradient boosting method (GBM) machine-learning algorithm was used to develop the models on data from 2825 patients from the EPaNIC multicenter randomized controlled trial database. Three models were developed: a "Core" model based on demographic, admission diagnosis, and daily laboratory results; a "Core+BGA" model adding blood gas analysis results; and a "Core+BGA+Monitoring" model also including high-resolution monitoring data. Model performance was evaluated against the true CrCl by absolute difference, and root-mean-square error (RMSE). We externally validated the models on 9576 patients from the University Hospitals Leuven, included in the M@tric database.

Results: All three developed models showed smaller RMSEs than the current clinical practice. Assuming the same CrCl of the day of prediction showed 20.6 (95% CI 20.3-20.9) ml/min absolute difference with true CrCl and 40.1 ml/min RMSE in the external validation cohort, while the developed model having the smallest RMSE (the Core+BGA+Monitoring model) had 18.1 (95% CI 17.9-18.3) ml/min absolute difference with true CrCl and 28.9 ml/min RMSE.

Conclusion: Prediction models based on routinely collected clinical data in the ICU were able to accurately predict next day CrCl. These models could be useful for hydrophilic drug dosage adjustment or stratification of patients at risk.

7.1 Introduction

Critical illness often affects the kidney function. Epidemiologic studies have shown that 40-60% of ICU patients have episodes of acute kidney injury (AKI) [1, 2], and in 20-65% of the patients, days with an augmented renal clearance (ARC) occur [3]. Accurate assessment of kidney function is crucial for patient risk stratification and drug dosage adjustment, especially for renally cleared drugs such as vancomycin and β-lactam anti-microbials. Most often, the renal function is evaluated through the serum creatinine, and renal clearance is estimated based upon the modification of diet in renal disease study (MDRD) [4], or chronic kidney disease epidemiology collaboration (CKD-EPI) [5] formulas. However, these commonly used estimation formulas were derived from non-critically ill patients and thus have their limitations to properly estimate the kidney function in the intensive care unit (ICU) setting [6–10]. Creatinine clearance (CrCl) measured from a 24-h urine collection reflects best the renal function in both reduced and augmented renal clearance in routine clinical practice [3, 11]. Especially in long-stay patients, there is a gap between the estimated renal clearance based upon the serum creatinine and the measured creatinine clearance [11]. As the kidney function may change rapidly in critically ill patients [12, 13], even a calculated CrCl based on the urinary CrCl of the past 24h may lag behind the true kidney function. Potentially, accurate prediction of the kidney function of the next 24h could allow for more suitable therapeutic interventions.

Existing machine-learning predictions for renal function have focused on predicting the onset of AKI [14–22], because of its high incidence [23] and strong associations with higher mortality, longer length of stay, and heavier financial burden [24, 25]. Other models [26–30] predict ARC, which is commonly defined as the presence of CrCl greater than 130 ml/min/1.73m^2 and has significant consequences concerning the pharmacokinetics of hydrophilic drugs. Studies have demonstrated that ARC patients need higher antibiotic doses [31], have more treatment failure [32], and a doubled risk of subtherapeutic vancomycin serum concentrations [33]. AKI and ARC prediction models were based on categorized definitions. As the glomerular filtration rate (GFR) is in fact continuous, being able to predict the entire kidney function spectrum corresponds better with the clinical and physiological reality.

Despite the importance and need of continuous kidney function prediction, to the best of our knowledge, no prediction models for daily prediction of CrCl in critically ill patients exist. Hence, this study aims to develop and validate prediction models that apply machine learning algorithm to routinely collected patient data to predict CrCl one day ahead. To evaluate the model usefulness, the developed models were also compared with the current clinical practice that uses the CrCl from the day of prediction.

7.2 Methods

7.2.1 Prediction tasks and CrCl definition

This study aims to predict the CrCl of the next patient day. CrCl was calculated by daily 24h urine output (UO), urinary creatinine (UCr), and serum creatinine (SCr) without correction for an average body surface area: CrCl (ml/min) = Urine creatinine (mg/dL) × 24h Urine output (ml/day) / Serum creatinine (mg/dL) / 1440 (min/day). In an additional analysis, the same methodology was applied to develop models to predict the average CrCl over the next two days ahead (Appendix 7.A.1).

7.2.2 Study databases with inclusion and exclusion criteria

The large multicenter EPaNIC randomized controlled trial (RCT) database [34], where two parenteral nutrition strategies were compared in 4640 critically ill adults between August 2007 and November 2010, was used for model development. This study was conducted on the basis of prior informed consent by all patients or their legal representatives, and the consent forms included the permission to use the data for additional research (S50404). For this secondary study, patients were eligible for inclusion if they had no kidney replacement therapy (KRT) before ICU. Patient-days were excluded if there were 1) no CrCl measurements on the next day, 2) no CrCl measurements on the day of prediction, 3) KRT on the day of prediction, 4) KRT in the previous week, and/or 5) all patient days beyond 90 days in the ICU were excluded.

External validation was performed on a dataset of 20930 patients of the University Hospitals Leuven who were included in the large multicenter M@tric database between 2013 and 2018 [35]. The M@tric database contains high-quality data from all adult patients admitted to the ICUs of the three largest university hospitals in Belgium. Ethical approval for the M@tric database collection was received from the Ethics Committee (EC) of University Hospitals Ghent. Approval for the use of these patient data in the present study was obtained from the EC of University Hospitals Leuven (S61364). The study was conducted in compliance with the principles of the Declaration of Helsinki and its later revisions. The same exclusion criteria as described above for the EPaNIC development dataset were applied.

7.2.3 Feature engineering

Only data up to the day of predicting CrCl were used as input to the models. The considered data included: 1) Admission data: demographics, diagnosis, and comorbidities, 2) Time-series data such as minute-by-minute monitoring data and daily or hourly laboratory results, 3) Medication-related data, 4) Time-related data: day of the week, and day from ICU admission. Data were retrieved from both the EPaNIC study database (Filemaker Pro®; FileMaker Inc, FileMaker International) and the patient data management system (PDMS) database (Microsoft SQL Server®; Microsoft®, Redmond, Washington, USA).

The minimum, maximum, mean, standard deviation, linear regression slope, Fast Fourier transform (FFT), cepstrum analysis, autoregressive analyses, and first-order derivative were applied to derive characteristics from the timestamped data. All the features with more than 10% missing values were excluded. For the remaining features, missing values were imputed with the mean and the mode from the development cohort for continuous data and categorical data respectively. Finally, continuous data were standardized to zero mean and unit variance, and categorical data without order relation were converted into a form with binary data for each category.

7.2.4 Machine-learning algorithm, feature selection methods, and clinical prediction models

The prediction models were trained with the gradient-boosting regressor method [36], with features selected based on data availability in the PDMS system, backward elimination method [37], and thorough discussions with two experienced ICU physicians (GDV and GM). Hyperparameters were fine-tuned with Optuna hyperparameter optimization software [38].

For each prediction task, three models with progressively more features were developed which are meant to be utilized sequentially, based on the data availability at the bedside.

- A "**Core model**" using only admission data and daily routine laboratory results.
- A "**Core+BGA model**" that adds to the above, blood gas analysis data.
- A "**Core+BGA+Monitoring model**" that adds to the above, monitoring data (heart rate, mean arterial blood pressure, and respiratory rate).

7.2.5 Internal and external validation

Models were developed and internally validated on the EPaNIC database with 10-fold cross validation. At the external validation stage, models trained on the entire EPaNIC database were applied on the previously unseen external validation cohort to assess generalizability. To examine the model usefulness, model performance was further compared against the current clinical practice: using as prediction for one-day ahead the same CrCl value of the day of prediction. This reference CrCl was henceforth referred to as **Prediction day's CrCl**. To assess daily fluctuations in CrCl, we calculated the difference between the CrCl of each pair of two consecutive days. It was labeled as stable if the absolute difference was less than 20 ml/min and unstable if more than 20 ml/min, as this is a meaningful difference for drug dosing and because the CrCl variability in healthy volunteers has been reported with mean differences of 21.7 ml/min/1.73m² and variations of 18.7% [39].

7.2.6 Evaluation metrics for predictive performance

Absolute difference between the predicted and actual CrCls, and root-mean-square error (RMSE) were computed for all available patient-days, stable days, and unstable days for each model in both cohorts. RMSE measures the errors between the model predictions and the target CrCl values and is sensitive to large errors. Predictive performance was also evaluated visually with scatter plots and plots of daily absolute difference and RMSE for all available patient-days, stable days, and unstable days during the first week of ICU stay. As multiple patient-days were available in many patients, no overall p-value can be calculated as this may be biased by repeated measures, but we compared the absolute difference on a day-by-day basis with the Diebold-Mariano test [40]. Count-based feature importance of the developed models was visualized with bar plots.

7.2.7 Descriptive analyses and software used

Python 3.7.4 (Python Software Foundation, http://www.python.org), SciPy version 1.3.1 (SciPy.org), and Scikit-learn library 0.24.2 (scikit-learn.org) were used for all analyses. The study population was described using descriptive statistics, with continuous data presented as medians and interquartile ranges (IQR) and categorical data expressed as counts and percentages (%). Mann–Whitney U test and Fisher's Exact Test were used to

evaluate statistical significance of differences for continuous and categorical data respectively. Significance levels were set at the 5% level.

7.3 Results

7.3.1 Study cohorts

7.3.1.1 Development cohort

For the model development, data were retrieved from 2825 patients, equivalent to 18494 patient-days (**Figure 7.A.1**). The descriptive statistics were shown in **Table 7.1**. Median (interquartile range, IQR) age was 67.6 (56.2–75.6) years, median (IQR) Acute Physiology and Chronic Health Evaluation II (APACHE II) score was 22 (16–32), the majority was with cardiac surgery (n=1655, 58.6%), and median (IQR) average CrCl over the entire ICU stay was 93.5 (58.2–131.8) ml/min. The median ICU length of stay (IQR) was 5 (3–11) days, and 162 (5.7%) patients died before ICU discharge. There were 6371 (34.5%) unstable days.

7.3.1.2 External validation cohort

For the external validation of the developed models, data from 9576 patients were used, corresponding to 53943 patient-days. The age was younger (median (IQR) 65.6 (54.6–75.0) years, $p<0.01$), emergency admission was less frequently (n=3231, 33.7%, $p<0.01$) , APACHE II score was lower (median (IQR) 17 (13–21), $p<0.01$), cardiac surgery was still the major admission diagnosis but occurred less often (n=3229, 33.7%, $p<0.01$), and average CrCl over the entire ICU stay was similar (median (IQR) 93.1 (56.2–133.6) ml/min, $p=0.7$) (**Table 7.1**). The ICU length of stay was shorter (median (IQR) 4 (2–9) days, $p<0.01$), and the ICU mortality was lower (n=200, 2.1%, $p<0.01$). There were 16514 (30.6%) unstable days.

Table 7.1 Patient characteristics and clinical outcomes

	Development cohort (n=2825)	Validation cohort (n=9576)	p-value
Age, years, median (IQR)	67.6 (56.2–75.6)	65.5 (54.6–75.0)	<0.01
Gender male, number (%)	1747 (61.8)	5816 (60.7)	0.3
Mean creatinine clearance over the entire ICU stay, ml/min, median (IQR)	93.5 (58.2–131.8)	93.1 (56.2–133.6)	0.7
Emergency admission, number (%)	1272 (45.0)	3231 (33.7)	<0.01
APACHE II score, median (IQR)	22 (16–32)	17 (13–21)	<0.01
Reason for admission			
Cardiac surgery, number (%)	1655 (58.6)	3229 (33.7)	<0.01
Medical disease, number (%)	114 (4.0)	2316 (24.2)	<0.01
Neurology, number (%)	119 (4.2)	363 (3.8)	0.3
Trauma and other surgery, number (%)	662 (23.4)	2859 (29.9)	<0.01
Transplantation, number (%)	275 (9.7)	809 (8.4)	0.04
ICU mortality, number (%)	162 (5.7)	200 (2.1)	<0.01
Length of stay in ICU, days, median (IQR)	5 (3–11)	4 (2–9)	<0.01

ICU, intensive care unit; APACHE II score, Acute Physiology and Chronic Health Evaluation II score

7.3.2 Features selected for CrCl prediction

Among the ten most predictive variables of the three models, seven were related to CrCl, one to urea level, and the remaining two were the baseline characteristics age and body mass index (BMI) (**Figure 7.1**). For the three models, the top ten most important features were features already available in the Core model. In other words, neither BGA nor monitoring data related features were among the top ten most important features of any model. The full set of features was presented in Appendix 7.A.2.

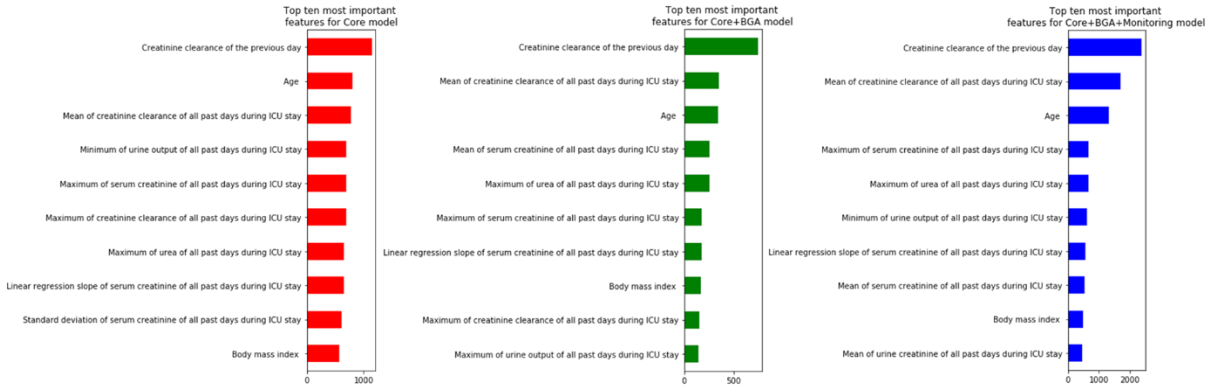


Figure 7.1 Top ten most important features of different models. The red, green, and blue bar plots are the results for the Core, Core+BGA, and Core+BGA+Monitoring models, respectively.

7.3.3 Externally validated model performance

The developed models performed well in both the internal validation (**Figure 7.A.2**, **Figure 7.A.3,** and **Table 7.A.1**) and the external validation (**Figure 7.2**, **Figure 7.3,** and **Table 7.2**). Specifically, for all patient-days, the model having the smallest RMSE in the validation cohort was the Core+BGA+Monitoring model, which exhibited 18.1 (95% CI 17.9-18.3) ml/min absolute difference and 28.9 ml/min RMSE, while the prediction day's CrCl assuming a constant CrCl led to a 20.6 (95% CI 20.3-20.9) ml/min absolute difference and 40.1 ml/min RMSE. The Core and Core+BGA models showed absolute difference of 18.5 (95% CI 18.4-18.7) and 18.5 (95% CI 18.3-18.7) ml/min, and RMSE of 29.0 and 29.2 ml/min.

Table 7.2 Summary of absolute difference between the predicted and actual creatinine clearances, and root-mean-square error for the developed models and the reference on all days, stable days, and unstable days in the validation cohort

	All days		Stable days		Unstable days	
	Absolute difference (ml/min) (95% CI)	Root-mean-square error (ml/min)	Absolute difference (ml/min) (95% CI)	Root-mean-square error (ml/min)	Absolute difference (ml/min) (95% CI)	Root-mean-square error (ml/min)
Prediction day's CrCl	20.6 (20.3-20.9)	40.1	7.6 (7.5-7.6)	9.3	50.1 (49.3-50.9)	71.1
Core model	18.5 (18.4-18.7)	29.0	11.4 (11.3-11.5)	15.1	34.8 (34.3-35.2)	47.3
Core+BGA model	18.5 (18.3-18.7)	29.2	11.0 (10.9-11.1)	14.6	35.4 (34.9-35.9)	47.9
Core+BGA+Monitoring model	18.1 (17.9-18.3)	28.9	10.5 (10.4-10.6)	13.9	35.5 (35.0-36.0)	47.9

During the stable days, the model having the smallest RMSE was the Core+BGA+Monitoring model, demonstrating 10.5 (95% CI 10.4-10.6) ml/min absolute difference and 13.9 ml/min RMSE, while the prediction day's CrCl showed 7.6 (95% CI 7.5-7.6) ml/min absolute difference and 9.3 ml/min RMSE. The model had a larger absolute difference of 3 ml/min on average than the prediction day's CrCl. The Core and Core+BGA models manifested 11.4 (95% CI 11.3-11.5) and 11.0 (95% CI 10.9-11.1) ml/min absolute difference, and 15.1 and 14.6 ml/min RMSE.

However, on the days when renal function was unstable, the model having the smallest RMSE was the Core model exhibiting 34.8 (95% CI 34.3-35.2) ml/min absolute difference and 47.3 ml/min RMSE, whereas the prediction day's CrCl showed 50.1 (95% CI 49.3-50.9) ml/min absolute difference and 71.1 ml/min RMSE. The model had a smaller absolute difference of 15 ml/min on average than the prediction day's CrCl. The Core+BGA and Core+BGA+Monitoring models demonstrated 35.4 (95% CI 34.9-35.9) and 35.5 (95% CI 35.0-36.0) ml/min absolute difference, and 47.9 and 47.9 ml/min RMSE.

These differences in prediction performance for stable and unstable days remained when analyzing the daily predictions during the first week of ICU stay only as evidenced in **Figure 7.2**, where there was significant difference in absolute difference between the Core+BGA+Monitoring model and the reference. The majority of the predictions was located near to the diagonal axis, denoting a good agreement between predicted and actual CrCls (**Figure 7.3**). The results of the two days ahead average CrCl predictions were discussed in Appendix 7.A.2.

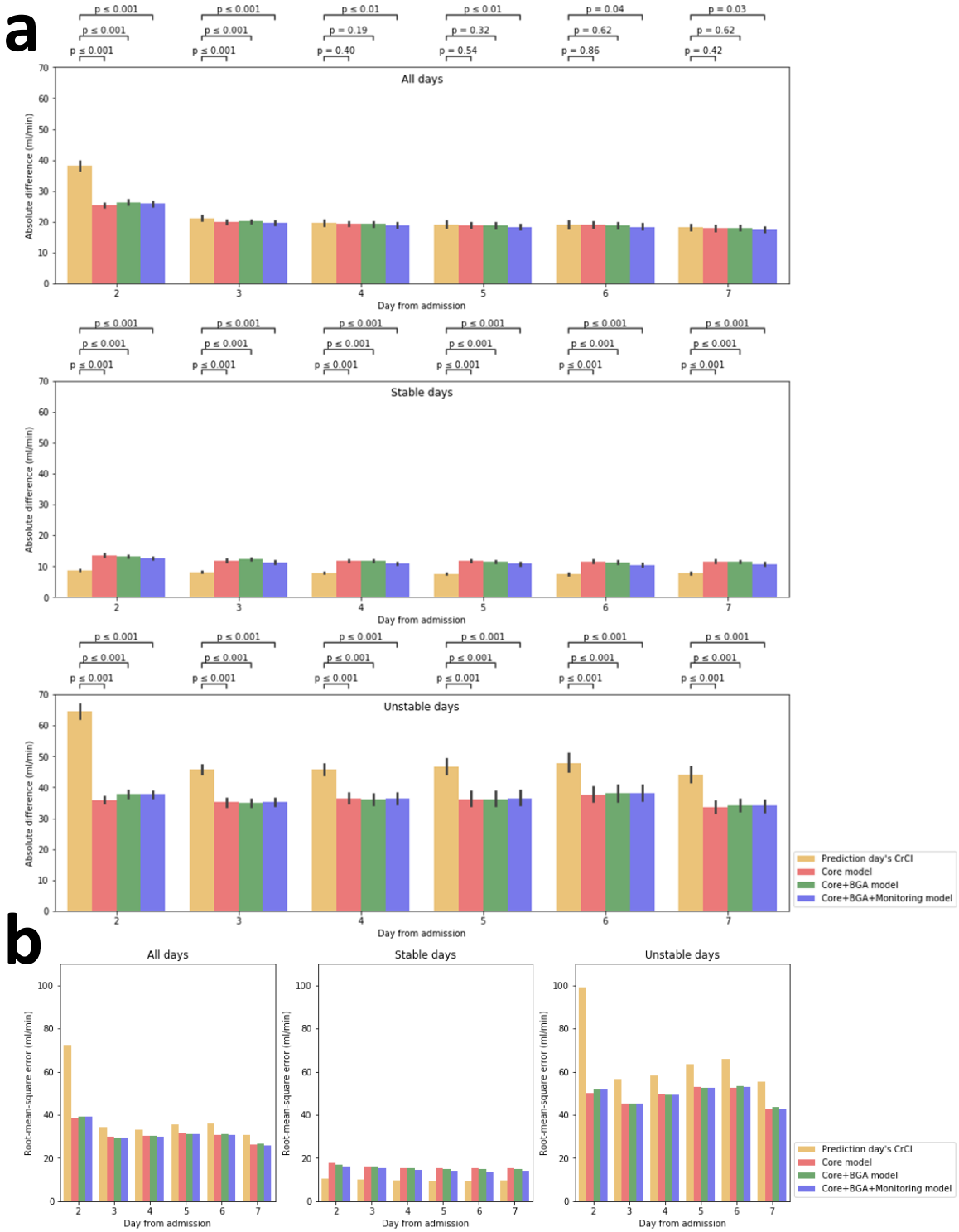


Figure 7.2 Temporal absolute difference (**a**) and root-mean-square error (**b**) of different models on all days, stable days, and unstable days within the first week of ICU admission in the validation cohort. The red, green, blue, and orange bars represent, respectively, the Core, Core+BGA, Core+BGA+Monitoring models, and the reference that assumes CrCl will remain the same compared to the day of prediction. Error bars represent 95% confidence intervals.

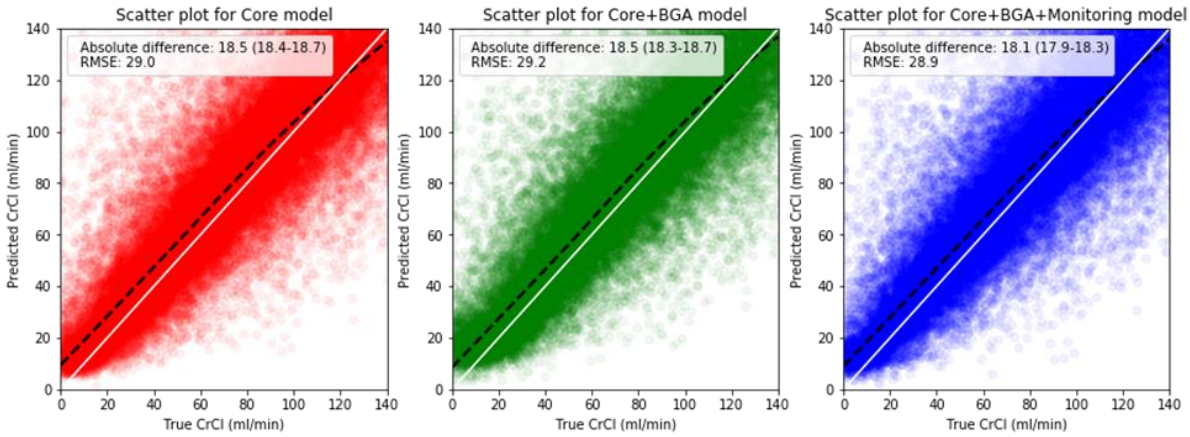


Figure 7.3 Relationships between predicted and actual CrCls for different models in the validation cohort. The red, green, and blue scatter plots show the results for the Core, Core+BGA, and Core+BGA+Monitoring models, respectively. The black dashed and white solid lines represent the lowess-based regression lines for the developed models and the diagonal axis. RMSE, root mean square error; CrCl, creatinine clearance

7.4 Discussion

In this study, we presented three models to predict daily CrCl in critically ill adults, based on information derived from routinely collected clinical data, and that the predictive performance remained similar when adding high-resolution data. The developed models were externally validated on previously unseen patients with good performance. Finally, the models demonstrated smaller RMSE than using the CrCl of the day of prediction (reflecting the current clinical practice), mainly during the days with high CrCl instability. The worse performance of the models than the reference during the days with low CrCl instability was not of clinical relevance, as the difference in absolute difference was only 3 ml/min on average. To the best of our knowledge, this study presents the first machine-learning algorithm for daily CrCl prediction in the ICU, by using routinely collected clinical data.

There are many reasons why there is a need for such daily prediction of CrCl during the entire ICU stay. First, measured urinary CrCl is currently considered the most suitable method to estimate the GFR [41], as many studies have shown the limited ability of estimation methods in proper assessment of kidney function in the ICU setting [6–9]. Second, a minimum of eight-hour time window of urine collection is necessary to ensure a reliable urinary CrCl measurement [42]. Consequently, the kidney function might have already changed by the time urine collection is complete. This delayed kidney function information could endanger patients by giving the physicians a false impression of renal function when prescribing drugs, as we observed here that the strategy that uses the measured CrCl of the past 24 hours leads to large estimation errors (RMSE of 40.1 ml/min). Third, several hydrophilic antibiotics are mainly eliminated by the kidneys, so dosage adjustment is necessary to prevent drug toxicity in reduced renal clearance patients [43] and treatment failure in ARC patients [32].

Having a reference to compare against helps understanding whether the models could have clinical usefulness. Compared to the current clinical practice of assuming the same CrCl as the day of prediction, our developed models reduced the RMSE from 40.1 to 28.9 ml/min. Importantly, in the subgroup of patient-days with stable renal function (comprising 60-70% of all patient-days), the developed models demonstrated a clinically insignificant larger absolute difference compared to the actual CrCl, around 3 ml/min on average, than

the reference assuming the same CrCl as the day of prediction. Noticeably, in the subgroup of patient-days with unstable renal function (comprising 30-40% of all patient-days), the developed models had clinically relevant smaller absolute differences compared to the actual CrCl, around 15 ml/min on average. This subgroup analysis of days with high CrCl instability clearly exhibited our models' capability of better capturing the dynamics of kidney function. Nevertheless, despite the large reduction in prediction errors during the unstable days, whether or not the models help in improving patient outcomes still needs to be investigated prospectively.

Our study has many strengths. First, the use of a general ICU population instead of specific subset of patients makes it more generally applicable, and the daily prediction truthfully reflected the fluctuating kidney function on each patient day, allowing for risk stratification and drug dose adjustment. Second, the reporting of this study was performed following the Transparent Reporting of a Multivariate Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guidelines [44]. Third, both internal and external validation were performed, and the developed models were compared against a reference to fairly report the model performance and robustness without overoptimism. Fourth, not only static data but also timestamped data applied with advanced feature engineering techniques were progressively included with increasing data resolution. Finally, the use of a very large validation dataset of approximately 54000 patient-days from over 9500 mixed critically ill patients attests to the robustness of the findings.

There are several limitations in our study. First, the development cohort was based on a RCT database in Belgium dating back to 2010, which might limit its generalizability in other settings. However, model performance remained unchanged when externally validated on a very large database with patient data collected up to 2018. Second, the use of high-resolution data might be difficult to implement in hospitals with limited resources, and some settings might even struggle to have the necessary data for the lower resolution Core model. Third, there might be a selection bias resulting from the exclusion from the analyses of patient-days with KRT on the day of prediction and in previous week, or of patient-days when less than 2 consecutive CrCls were available, or patient-days after the first 90 days in ICU. These exclusion criteria were necessary to ensure reliable CrCl prediction models could be developed. Fourth, this study was based on retrospective data, and the developed models still need prospective validation in independent cohorts. Fifth, the model performance was not compared against novel biomarkers such as cystatin C that may be less biased, but measured CrCl is a fast and cheap test, which are important characteristics as the measurements were performed on a daily basis. Sixth, the measurement of creatinine changed from the Jaffe method in the development cohort to the enzymatic method in the validation cohort, and it was found that the Jaffe method yielded higher creatinine values than the enzymatic method, particularly at low creatinine concentrations [45]. However, the Jaffe and enzymatic creatinine methods were shown with adequate overall agreement (r=.9994, r=.9998 in serum and urine respectively), and thus the influence of changed creatinine measurements was expected low. Finally, the developed models were not implemented as bedside tools, integrated into clinical practice, and transferred to other centers yet, but it was beyond the scope of this work and remains a challenging topic for future studies.

7.5 Conclusion

We have shown that CrCl can be accurately predicted one day in advance on a daily basis during ICU stay, with models developed based on routinely collected clinical data. We have also demonstrated the robustness of the developed models on previously unseen patients in external validation. The developed models' usefulness has

also been shown in comparisons with a reference reflecting current clinical practice, mainly on the patient-days with high renal function instability. Despite the promising performance, these findings should be prospectively validated in independent patient populations, before these prediction models can be further used for risk stratification or incorporated into a pharmacokinetic model to support a more optimized dose regimen.

7.A Appendix

7.A.1 Supplementary methods

For prediction of average CrCl of two days ahead, to assess daily fluctuations in CrCl, we calculated the difference between the CrCl of two consecutive days of three consecutive days. It was labeled as stable if either of the two absolute differences was less than 20 ml/min and unstable if both were more than 20 ml/min.

7.A.2 Supplementary results

For the average CrCl of two days ahead prediction task, the majority of the predictions was located near to the diagonal axis, denoting a good agreement between predicted and actual CrCls (**Figure 7.A.4**). When analyzing the daily predictions during the first week of ICU stay only as shown in **Figure 7.A.5**, the models were with lower RMSEs for all days and unstable days. The worse performance on stable days was not of clinical relevance as the difference was small, around 4 ml/min on average.

7.A.2.1 Complete feature list: One-day-ahead prediction

For the one-day-ahead Core model, a total of 26 features were used. 4 were derived from serum creatinine, 3 were derived from creatinine clearance, 3 were derived from medication/intervention (aminoglycosides medication on the day of prediction, number of days with respiratory support in all past days during ICU stay, and number of days with vasopressors/inotropes medication in all past days during ICU stay), 2 were derived from bilirubin, 2 were derived from urine output, 2 were derived from urine creatinine, 1 was derived from urea, 1 was derived from hematocrit, 1 was derived from c-reactive protein, 1 was derived from Sequential Organ Failure Assessment (SOFA) score, 1 was day of the week, 1 was gender, 1 was age, 1 was day from ICU admission, 1 was APACHE II score on the first day of ICU, and 1 was body mass index.

For the one-day-ahead Core+BGA model, a total of 65 features were used. 7 were derived from BGA pH value, 7 were derived from BGA bicarbonate, 6 were derived from BGA partial pressure of oxygen, 6 were derived from BGA lactate, 4 were derived from serum creatinine, 3 were derived from creatinine clearance, 3 were derived from medication/intervention (aminoglycosides medication on the day of prediction, number of days with respiratory support in all past days during ICU stay, and number of days with vasopressors/inotropes medication in all past days during ICU stay), 3 were derived from BGA sodium, 3 were derived from BGA potassium, 3 were derived from BGA partial pressure of carbon dioxide, 2 were derived from BGA glucose, 2 were derived from bilirubin, 2 were derived from urine output, 2 were derived from urine creatinine, 2 were derived from BGA hemoglobin, 1 was derived from urea, 1 was derived from hematocrit, 1 was derived from c-reactive protein, 1 was derived from SOFA score, 1 was day of the week, 1 was gender, 1 was age, 1 was day from ICU admission, 1 was APACHE II score on the first day of ICU, and 1 was body mass index.

For the one-day-ahead Core+BGA+Monitoring model, a total of 74 features were used. 7 were derived from BGA pH value, 7 were derived from BGA bicarbonate, 6 were derived from BGA partial pressure of oxygen, 6 were derived from BGA lactate, 5 were derived from heart rate, 4 were derived from serum creatinine, 3 were derived from creatinine clearance, 3 were derived from medication/intervention (aminoglycosides medication on the day of prediction, number of days with respiratory support in all past days during ICU stay, and number of days with vasopressors/inotropes medication in all past days during ICU stay), 3 were derived from BGA sodium, 3 were derived from BGA potassium, 3 were derived from BGA partial pressure of carbon dioxide, 2 were derived from BGA glucose, 2 were derived from bilirubin, 2 were derived from urine output, 2 were derived from urine creatinine, 2 were derived from respiratory rate, 2 were derived from mean arterial blood pressure, 2 were derived from BGA hemoglobin, 1 was derived from urea, 1 was derived from hematocrit, 1 was derived from c-reactive protein, 1 was derived from SOFA score, 1 was day of the week, 1 was gender, 1 was age, 1 was day from ICU admission, 1 was APACHE II score on the first day of ICU, and 1 was body mass index.

7.A.2.2 Complete feature list: Average-of-two-days-ahead prediction

For the prediction of average CrCl of two days ahead, a total of 67 features were employed in the Core model. 8 were derived from urea, 7 were derived from serum creatinine, 6 were derived from chloride, 6 were derived from platelet, 6 were derived from SOFA score, 4 were derived from urine output, 4 were derived from urine creatinine, 4 were derived from hematocrit, 4 were derived from creatinine clearance, 4 were derived from c-reactive protein, 3 were derived from white blood cell count, 2 were derived from medication/intervention, 2 were derived from bilirubin, 1 was age, 1 was weight, 1 was APACHE II score on the first day of ICU, 1 was transplant diagnostic group on ICU admission, 1 was gender, 1 was day from ICU admission, and 1 was day of the week.

For the prediction of average CrCl of two days ahead, a total of 130 features were employed in the Core+BGA model. 8 were derived from BGA partial pressure of oxygen, 8 were derived from urea, 8 were derived from BGA pH value, 7 were derived from serum creatinine, 7 were derived from BGA bicarbonate, 7 were derived from BGA sodium, 6 were derived from BGA hemoglobin, 6 were derived from chloride, 6 were derived from BGA potassium, 6 were derived from BGA lactate, 6 were derived from platelet, 6 were derived from SOFA score, 6 were derived from BGA glucose, 5 were derived from BGA partial pressure of carbon dioxide, 4 were derived from urine output, 4 were derived from BGA calculated oxygen saturation, 4 were derived from urine creatinine, 4 were derived from hematocrit, 4 were derived from creatinine clearance, 4 were derived from c-reactive protein, 3 were derived from white blood cell count, 2 were derived from medication/intervention, 2 were derived from bilirubin, 1 was age, 1 was weight, 1 was APACHE II score on the first day of ICU, 1 was transplant diagnostic group on ICU admission, 1 was gender, 1 was day from ICU admission, and 1 was day of the week.

For the prediction of average CrCl of two days ahead, a total of 149 features were employed in the Core+BGA+Monitoring model. 9 were derived from heart rate, 8 were derived from BGA partial pressure of oxygen, 8 were derived from urea, 8 were derived from BGA pH value, 7 were derived from serum creatinine, 7 were derived from BGA bicarbonate, 7 were derived from BGA sodium, 6 were derived from BGA hemoglobin, 6 were derived from chloride, 6 were derived from BGA potassium, 6 were derived from BGA lactate, 6 were derived from mean arterial blood pressure, 6 were derived from platelet, 6 were derived from SOFA score, 6

were derived from BGA glucose, 5 were derived from BGA partial pressure of carbon dioxide, 4 were derived from respiratory rate, 4 were derived from urine output, 4 were derived from BGA calculated oxygen saturation, 4 were derived from urine creatinine, 4 were derived from hematocrit, 4 were derived from creatinine clearance, 4 were derived from c-reactive protein, 3 were derived from white blood cell count, 2 were derived from medication/intervention, 2 were derived from bilirubin, 1 was age, 1 was weight, 1 was APACHE II score on the first day of ICU, 1 was transplant diagnostic group on ICU admission, 1 was gender, 1 was day from ICU admission, and 1 was day of the week.

7.A.3 Supplementary tables

Table 7.A.1 Summary of absolute difference between the predicted and actual creatinine clearances (CrCls), and root-mean-square error for different models and the reference on all days, stable days, and unstable days for the one-day-ahead prediction task in the development cohort

	All days		Stable days		Unstable days	
	Absolute difference (ml/min) (95% CI)	**Root-mean-square error (ml/min)**	**Absolute difference (ml/min) (95% CI)**	**Root-mean-square error (ml/min)**	**Absolute difference (ml/min) (95% CI)**	**Root-mean-square error (ml/min)**
Prediction day's CrCl	22.5 (22.0-23.0)	42.1	7.8 (7.7-7.9)	9.5	50.6 (49.4-51.8)	70.6
Core model	19.9 (19.5-20.2)	32.6	11.1 (10.9-11.3)	14.8	36.5 (35.6-37.4)	51.7
Core+BGA model	19.9 (19.5-20.3)	32.8	10.8 (10.6-11.0)	14.3	37.2 (36.3-38.1)	52.2
Core+BGA+Monitoring model	19.6 (19.2-19.9)	32.5	10.3 (10.2-10.5)	13.7	37.1 (36.2-38.0)	52.1

Table 7.A.2 Summary of absolute difference between the predicted and actual creatinine clearances (CrCls), and root-mean-square error for different models and the reference unstable-upward and unstable-downward days for the one-day-ahead prediction task in the development cohort

	Unstable-upward days		Unstable-downward days	
	Absolute difference (ml/min) (95% CI)	Root-mean-square error (ml/min)	Absolute difference (ml/min) (95% CI)	Root-mean-square error (ml/min)
Prediction day's CrCl	48.3 (46.9-49.8)	65.3	53.4 (51.4-55.3)	76.4
Core model	37.6 (36.1-39.1)	58.1	35.2 (34.3-36.1)	42.8
Core+BGA model	38.2 (36.7-39.7)	58.5	36.1 (35.2-37.0)	43.6
Core+BGA+Monitoring model	38.1 (36.7-39.6)	58.5	35.9 (35.0-36.8)	43.3

Unstable-upward days were the days of prediction where an increase larger than 20 ml/min in the CrCl of the next day was present; unstable-upward days were the days of prediction where a decrease larger than 20 ml/min in the CrCl of the next day was found; CrCl, creatinine clearance

Table 7.A.3 Summary of absolute difference between the predicted and actual creatinine clearances (CrCls), and root-mean-square error for different models and the reference unstable-upward and unstable-downward days for the one-day-ahead prediction task in the validation cohort

	Unstable-upward days		Unstable-downward days	
	Absolute difference (ml/min) (95% CI)	Root-mean-square error (ml/min)	Absolute difference (ml/min) (95% CI)	Root-mean-square error (ml/min)
Prediction day's CrCl	46.1 (45.3-46.8)	58.4	54.9 (53.5-56.4)	83.9
Core model	34.5 (33.8-35.3)	50.6	35.0 (34.4-35.6)	43.2
Core+BGA model	34.1 (33.3-34.9)	50.2	37.0 (36.5-37.6)	45.1
Core+BGA+Monitoring model	34.9 (34.1-35.6)	50.7	36.3 (35.7-36.8)	44.2

Unstable-upward days were the days of prediction where an increase larger than 20 ml/min in the CrCl of the next day was present; unstable-upward days were the days of prediction where a decrease larger than 20 ml/min in the CrCl of the next day was found; CrCl, creatinine clearance

7.A.4 Supplementary figures

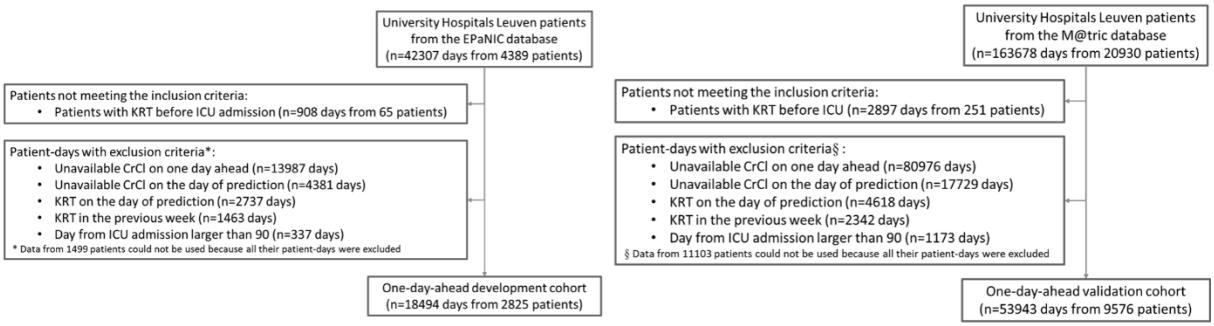


Figure 7.A.1 Consort diagram for prediction of one-day-ahead CrCl. CrCl, creatinine clearance; KRT, kidney replacement therapy; ICU, intensive care unit

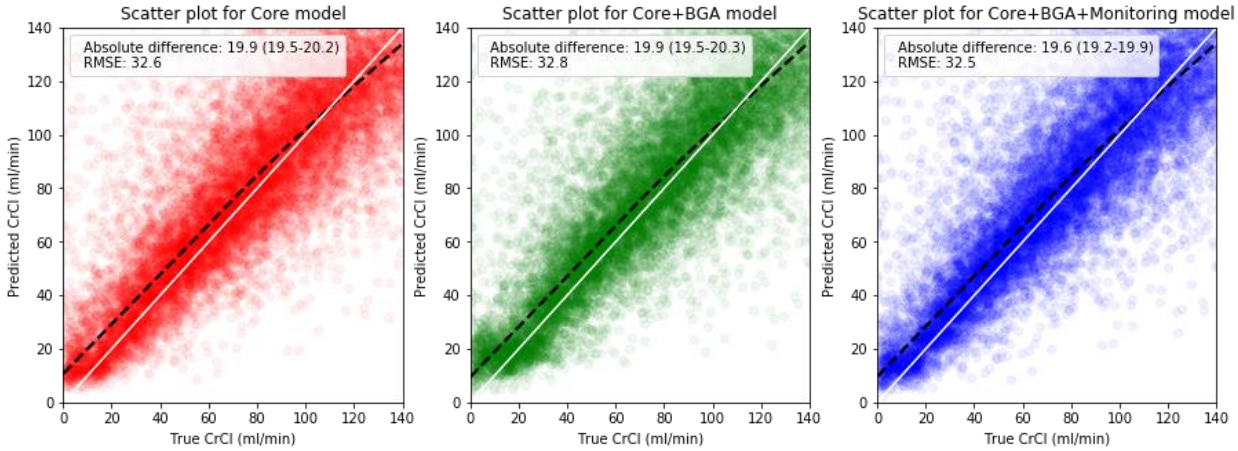


Figure 7.A.2 Relationships between predicted and actual CrCls for different models for the one-day-ahead prediction task in the development cohort. The red, green, and blue scatter plots show the results for the Core, Core+BGA, and Core+BGA+Monitoring models. The black dashed and white solid lines represent the lowess-based regression lines for the developed models and the diagonal axis. RMSE, root mean square error; CrCl, creatinine clearance

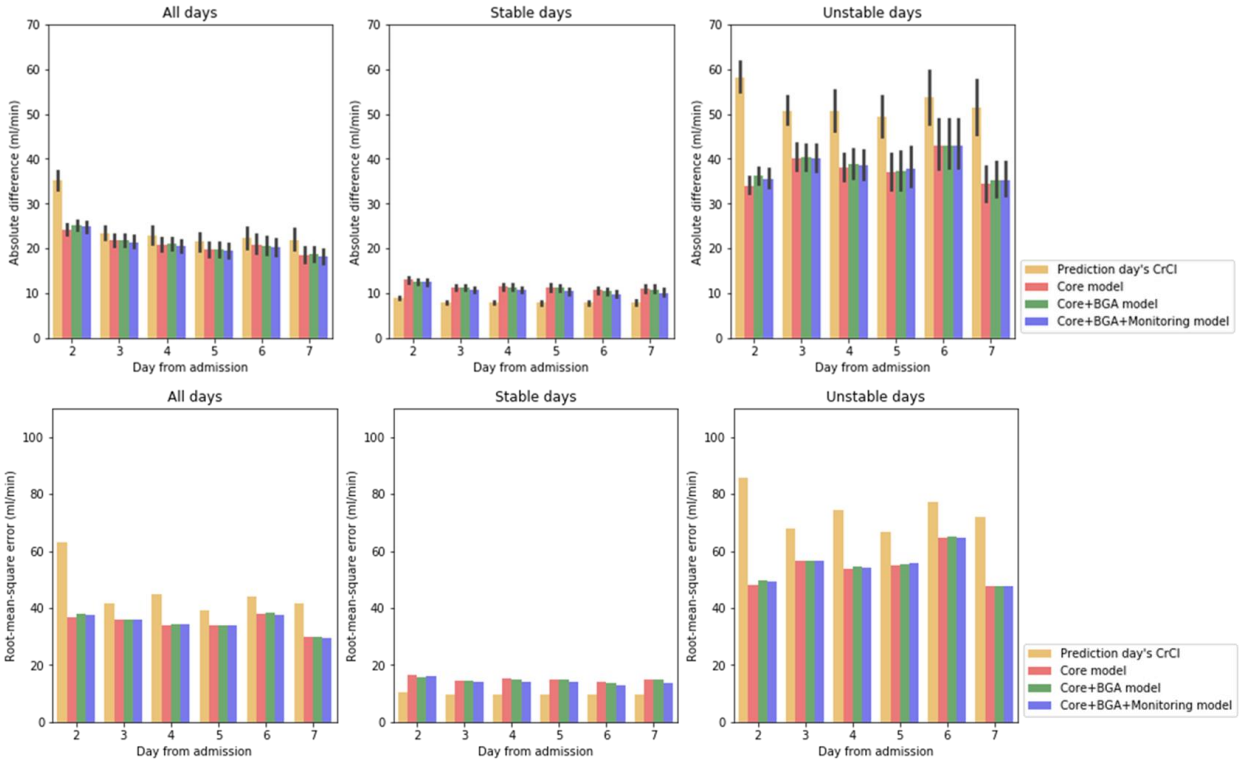


Figure 7.A.3 Temporal absolute difference (upper) and root-mean-square error (lower) of different models on all days, stable days, and unstable days within the first week of ICU admission for the one-day-ahead prediction task in the development cohort. The red, green, blue, and orange bars represent, respectively, the Core, Core+BGA, Core+BGA+Monitoring models, and the reference that assumes CrCl will remain the same compared to the day of prediction. Error bars represent 95% confidence intervals.

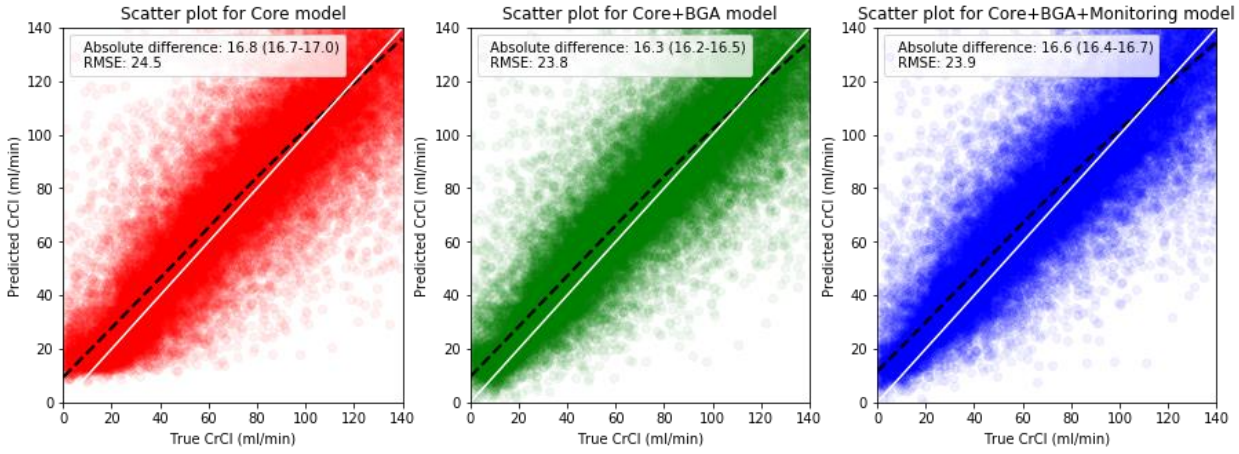


Figure 7.A.4 Relationships between predicted and actual CrCls for different models for the average of two days ahead prediction task in the validation cohort. The red, green, and blue scatter plots show the results for the Core, Core+BGA, and Core+BGA+Monitoring models. The black dashed and white solid lines represent the lowess-based regression lines for the developed models and the diagonal axis. RMSE, root mean square error; CrCl, creatinine clearance

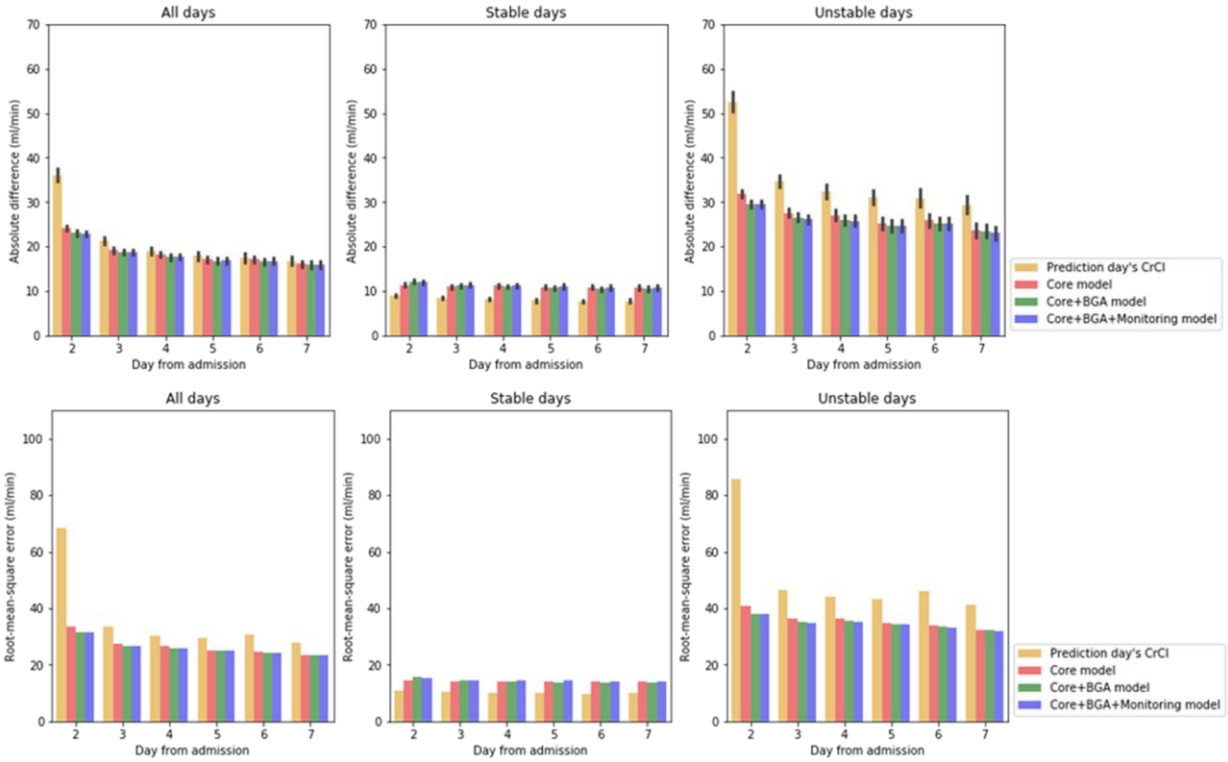


Figure 7.A.5 Temporal absolute difference (upper) and root-mean-square error (lower) of different models on all days, stable days, and unstable days within the first week of ICU admission for the average of two days ahead prediction task in the validation cohort. The red, green, blue, and orange bars represent, respectively, the Core, Core+BGA, Core+BGA+Monitoring models, and the reference that assumes CrCl will remain the same compared to the day of prediction. Error bars represent 95% confidence intervals.

GRANT SUPPORT AND CONFLICT OF INTEREST

C-Y.Huang receives funding from the Taiwan-KU Leuven scholarship. G.Meyfroidt is funded by the research foundation, Flanders (FWO) as senior clinical investigator (1843123N). J.Gunst is granted a postdoctoral research fellowship by the Clinical Research and Education Council of the University Hospitals Leuven. M.Casaer receives funding from the Research Foundation Flanders (FWO) (Grant No. 1832817N) and Onderzoeksraad, KU Leuven (Grant No. C24/17/070) and from the Private Charity Organization "Help Brandwonden Kids."

All the authors declare no competing interests that are relevant to the content of this article.

ACKNOWLEDGMENT AND PERSONAL CONTRIBUTION

Study concept and design	**Huang**, Güiza, De Vlieger, Meyfroidt
Data acquisition	**Huang**, Güiza, Wouters, Mebis, Meersseman, Casaer, Meyfroidt
Statistical analysis	**Huang**
Interpretation of results	**Huang**, Güiza, De Vlieger, Meyfroidt
Drafting of the manuscript	**Huang**, Güiza, De Vlieger, Meyfroidt
Manuscript revision	**All authors**
Principal investigator	Meyfroidt

The authors would like to thank Prof. Dr. Miet Schetz for her contribution to the EPaNIC database, and the members of the M@tric research group for designing the M@tric database, which were used to develop and validate the CrCl prediction models, respectively.

BIBLIOGRAPHY

1. Nisula S, Kaukonen KM, Vaara ST, et al (2013) Incidence, risk factors and 90-day mortality of patients with acute kidney injury in Finnish intensive care units: The FINNAKI study. Intensive Care Med 39:420–428. https://doi.org/10.1007/s00134-012-2796-5

2. Hoste EAJJ, Bagshaw SM, Bellomo R, et al (2015) Epidemiology of acute kidney injury in critically ill patients: the multinational AKI-EPI study. Intensive Care Med 41:1411–1423. https://doi.org/10.1007/s00134-015-3934-7

3. Bilbao-Meseguer I, Rodríguez-Gascón A, Barrasa H, et al (2018) Augmented Renal Clearance in Critically Ill Patients: A Systematic Review. Clin Pharmacokinet 57:1107–1121. https://doi.org/10.1007/s40262-018-0636-7

4. Levey AS, Bosch JP, Lewis JB, et al (1999) A more accurate method to estimate glomerular filtration rate from serum creatinine: A new prediction equation. Ann Intern Med 130:461–470. https://doi.org/10.7326/0003-4819-130-6-199903160-00002

5. Levey AS, Stevens LA, Schmid CH, et al (2009) A new equation to estimate glomerular filtration rate. Ann Intern Med 150:604–612. https://doi.org/10.7326/0003-4819-150-9-200905050-00006

6. Al-Dorzi HM, Alsadhan AA, Almozaini AS, et al (2021) The Performance of Equations That Estimate Glomerular Filtration Rate against Measured Urinary Creatinine Clearance in Critically Ill Patients. Crit Care Res Pract 2021:. https://doi.org/10.1155/2021/5520653

7. Mulder MB, Eidelson SA, Sussman MS, et al (2019) Risk Factors and Clinical Outcomes Associated With Augmented Renal Clearance in Trauma Patients. J Surg Res 244:477–483. https://doi.org/10.1016/j.jss.2019.06.087

8. Adnan S, Ratnam S, Kumar S, et al (2014) Select critically ill patients at risk of augmented renal clearance: experience in a Malaysian intensive care unit. Anaesth Intensive Care 42:715–722. https://doi.org/10.1177/0310057x1404200606

9. Grootaert V, Willems L, Debaveye Y, et al (2012) Augmented Renal Clearance in the Critically Ill: How to Assess Kidney Function. Ann Pharmacother 46:925–959. https://doi.org/10.1345/aph.1Q708

10. Wongpraphairot S, Thongrueang A, Bhurayanontachai R (2022) Glomerular filtration rate correlation and agreement between common predictive equations and standard 24-hour urinary creatinine clearance in medical critically ill patients. PeerJ 10:e13556. https://doi.org/10.7717/peerj.13556

11. Schetz M, Gunst J, Van den Berghe G (2014) The impact of using estimated GFR versus creatinine clearance on the evaluation of recovery from acute kidney injury in the ICU. Intensive Care Med 40:1709–1717. https://doi.org/10.1007/s00134-014-3487-1

12. Sunder S, Jayaraman R, Mahapatra HS, et al (2014) Estimation of renal function in the intensive care unit: The covert concepts brought to light. J Intensive Care 2:1–7. https://doi.org/10.1186/2052-0492-2-31

13. Huang C-Y, Güiza F, De Vlieger G, Meyfroidt G (2022) Daily fluctuations in kidney function in critically

ill adults. Crit Care 26:347. https://doi.org/10.1186/s13054-022-04226-3

14. Flechet M, Güiza F, Schetz M, et al (2017) AKIpredictor, an online prognostic calculator for acute kidney injury in adult critically ill patients: development, validation and comparison to serum neutrophil gelatinase-associated lipocalin. Intensive Care Med 43:764–773. https://doi.org/10.1007/s00134-017-4678-3

15. Malhotra R, Kashani KB, Macedo E, et al (2017) A risk prediction score for acute kidney injury in the intensive care unit. Nephrol Dial Transplant 32:814–822. https://doi.org/10.1093/ndt/gfx026

16. Chiofolo C, Chbat N, Ghosh E, et al (2019) Automated Continuous Acute Kidney Injury Prediction and Surveillance: A Random Forest Model. Mayo Clin Proc 94:783–792. https://doi.org/10.1016/j.mayocp.2019.02.009

17. Zimmerman LP, Reyfman PA, Smith ADR, et al (2019) Early prediction of acute kidney injury following ICU admission using a multivariate panel of physiological measurements. BMC Med Inform Decis Mak 19:16. https://doi.org/10.1186/s12911-019-0733-z

18. Tomašev N, Glorot X, Rae JW, et al (2019) A clinically applicable approach to continuous prediction of future acute kidney injury. Nature 572:116–119. https://doi.org/10.1038/s41586-019-1390-1

19. Li Y, Yao L, Mao C, et al (2018) Early Prediction of Acute Kidney Injury in Critical Care Setting Using Clinical Notes. In: 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp 683–686

20. Chen Y, Feng F, Li M, et al (2019) Development of a risk stratification-based model for prediction of acute kidney injury in critically ill patients. Med (United States) 98:e16867. https://doi.org/10.1097/MD.0000000000016867

21. Parreco J, Soe-Lin H, Parks JJ, et al (2019) Comparing machine learning algorithms for predicting acute kidney injury. Am Surg 85:725–729. https://doi.org/10.1177/000313481908500731

22. Rank N, Pfahringer B, Kempfert J, et al (2020) Deep-learning-based real-time prediction of acute kidney injury outperforms human predictive performance. npj Digit Med 3:1–12. https://doi.org/10.1038/s41746-020-00346-8

23. Ronco C, Bellomo R, Kellum JA (2019) Acute kidney injury. Lancet 394:1949–1964

24. Silver SA, Chertow GM (2017) The Economic Consequences of Acute Kidney Injury. Nephron 137:297–301

25. Chertow GM, Burdick E, Honour M, et al (2005) Acute kidney injury, mortality, length of stay, and costs in hospitalized patients. J Am Soc Nephrol 16:3365–3370. https://doi.org/10.1681/ASN.2004090740

26. Gijsen M, Huang CY, Flechet M, et al (2020) Development and External Validation of an Online Clinical Prediction Model for Augmented Renal Clearance in Adult Mixed Critically Ill Patients: The Augmented Renal Clearance Predictor. Crit Care Med 48:e1260–e1268. https://doi.org/10.1097/CCM.0000000000004667

27. Tang L, Ding XY, Duan LF, et al (2021) A Regression Model to Predict Augmented Renal Clearance in

Critically Ill Obstetric Patients and Effects on Vancomycin Treatment. Front Pharmacol 12:1–12. https://doi.org/10.3389/fphar.2021.622948

28. Nei AM, Kashani KB, Dierkhising R, Barreto EF (2020) Predictors of Augmented Renal Clearance in a Heterogeneous ICU Population as Defined by Creatinine and Cystatin C. Nephron 55905:313–320. https://doi.org/10.1159/000507255

29. Barletta JF, Mangram AJ, Byrne M, et al (2017) Identifying augmented renal clearance in trauma patients: Validation of the Augmented Renal Clearance in Trauma Intensive Care scoring system. In: The journal of trauma and acute care surgery. United States, pp 665–671

30. Udy AA, Roberts JA, Shorr AF, et al (2013) Augmented renal clearance in septic and traumatized patients with normal plasma creatinine concentrations: Identifying at-risk patients. Crit Care 17:. https://doi.org/10.1186/cc12544

31. He J, Yang Z-TT, Qian X, et al (2020) A higher dose of vancomycin is needed in critically ill patients with augmented renal clearance. Transl Androl Urol 9:2166–2171. https://doi.org/10.21037/tau-20-1048

32. Claus BOMM, Hoste EA, Colpaert K, et al (2013) Augmented renal clearance is a common finding with worse clinical outcome in critically ill patients receiving antimicrobial therapy. J Crit Care 28:695–700. https://doi.org/10.1016/j.jcrc.2013.03.003

33. Minkute R, Briedis V, Steponavičiute R, et al (2013) Augmented renal clearance - An evolving risk factor to consider during the treatment with vancomycin. J Clin Pharm Ther 38:462–467. https://doi.org/10.1111/jcpt.12088

34. Jones Q, Walden A (2011) Early versus late parenteral nutrition in critically ill adults. J Intensive Care Soc 12:338–339. https://doi.org/10.1177/175114371101200422

35. M@tric project. https://www.matric.be/. Accessed 5 Oct 2021

36. Ke G, Meng Q, Finley T, et al (2017) LightGBM: A highly efficient gradient boosting decision tree. In: Advances in Neural Information Processing Systems. pp 3147–3155

37. Guyon I, Weston J, Barnhill S, Vapnik V (2002) Gene selection for cancer classification using support vector machines. Mach Learn 46:389–422. https://doi.org/10.1023/A:1012487302797

38. Akiba T, Sano S, Yanase T, et al (2019) Optuna: A Next-generation Hyperparameter Optimization Framework. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery, New York, NY, USA, pp 2623–2631

39. Toffaletti JG, McDonnell EH (2008) Variation of serum creatinine, cystatin C, and creatinine clearance tests in persons with normal renal function. Clin Chim Acta 395:115—119. https://doi.org/10.1016/j.cca.2008.05.020

40. Diebold FX, Mariano RS (2002) Comparing Predictive Accuracy. J Bus \& Econ Stat 20:134–144. https://doi.org/10.1198/073500102753410444

41. De Waele JJ, Dumoulin A, Janssen A, Hoste EAJ (2015) Epidemiology of augmented renal clearance in mixed ICU patients. Minerva Anestesiol 81:1079–1085

42. Cherry RA, Eachempati SR, Hydo L, Barie PS (2002) Accuracy of short-duration creatinine clearance determinations in predicting 24-hour creatinine clearance in critically ill and injured patients. J Trauma 53:267–271. https://doi.org/10.1097/00005373-200208000-00013

43. Kane-Gill SL, Kirisci L, Verrico MM, Rothschild JM (2012) Analysis of risk factors for adverse drug events in critically ill patients*. Crit Care Med 40:823–828. https://doi.org/10.1097/CCM.0b013e318236f473

44. Moons KGM, Altman DG, Reitsma JB, et al (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. Ann Intern Med 162:W1–W73. https://doi.org/10.7326/M14-0698

45. Küme T, Sağlam B, Ergon C, Sisman AR (2018) Evaluation and comparison of Abbott Jaffe and enzymatic creatinine methods: Could the old method meet the new requirements? J Clin Lab Anal 32:. https://doi.org/10.1002/jcla.22168

8

A prospective comparison of ICU physicians and machine learning for the daily prediction of creatinine clearance in critically ill adults

Adapted from: **Chao-Yuan Huang**, Fabian Güiza, Pieter Wouters, Liese Mebis, Joachim Gidts, Lore Klewais, Greet Van den Berghe, Greet De Vlieger, Geert Meyfroidt. “A prospective comparison of ICU physicians and machine learning for the daily prediction of creatinine clearance in critically ill adults”. Ready for submission.

Abstract

Purpose: Critical illness often affects the kidney function with either an increase or decrease of the glomerular filtration rate on a day-to-day basis. The measured creatinine clearance (CrCl) is currently the best surrogate for the glomerular filtration rate. A set of machine-learning models to predict CrCl on a daily basis (the CrCl predictor) was recently developed. The aim of the present study was to prospectively validate the CrCl predictor and compare its accuracy with predictions made by ICU physicians.

Methods: We conducted a prospective study in a tertiary ICU. ICU physicians prospectively predicted patients' CrCl of the next day, which were collected via a questionnaire survey. Mean absolute error (MAE) and root mean square error (RMSE) with respect to the true CrCl were used to compare the predictive performance of the physicians and the CrCl predictor.

Results: A total of 54 physicians answered 1298 questionnaires for 197 patients corresponding to 704 patient-days. All three models of the CrCl predictor had lower RMSEs than the physicians' predictions. Specifically, ICU physicians showed 21.8 (95% CI 20.2-23.4) ml/min MAE and 36.4 ml/min RMSE, while the CrCl predictor model having the lowest RMSE (the Core+BGA+Monitoring model) demonstrated 19.0 (95% CI 17.8-20.3) ml/min MAE and 29.9 ml/min RMSE.

Conclusion: The CrCl predictor showed a good accuracy in a prospective validation and performed at least as well as the ICU physicians. These findings suggest that CrCl predictor can be a promising tool to adjust renally cleared drug dosage and stratify patients at risk.

8.1 Introduction

Critical illness often affects the kidney function as the glomerular filtration rate (GFR) may be increased or reduced depending on the premorbid stage and the clinical setting. A reduction of the GFR may lead to acute kidney injury (AKI), which occurs in 20-57% of the patients depending on the patient population and the use of different criteria of the AKI definition [1–4]. The presence of AKI associates with four-fold to six-fold increased mortality than the general inpatient population [5, 6], and prolongs the length of stay [6, 7]. On the other extreme of the kidney function is the augmented renal clearance (ARC), which is identified in 20-65% of critically ill patients [8]. ARC leads to an increased clearance of several drugs such as anticoagulants [9] and commonly used antibiotics such as beta-lactams and vancomycin [10–12], leading to a higher risk of treatment failure [13]. Predicting the renal clearance on the next day may guide physicians to optimize treatment. The measured creatinine clearance (CrCl) based on the urinary volume and the concentration of creatinine in serum and urine is currently the best surrogate for GFR during critical illness [14].

Recently, with the vast adoption of electronic health records and rapid growth of machine-learning algorithms, many models have been built to predict the onset of two extremes of kidney function spectrum: AKI [15–23] and ARC [24–28]. Recently we also developed a set of prediction models for the CrCl. This CrCl predictor predicts the CrCl of the next ICU day, based on routinely collected clinical information. The external validity and usefulness of the CrCl predictor models were confirmed through a comparison with the reference reflecting the current clinical practice (**Chapter 7**).

As a next step on the road to implementing these models into clinical practice, a prospective validation and comparison of the accuracy of the prediction models to the physicians' predictions is needed. Therefore, the aims of this study are to examine the external validity of the CrCl predictor, to investigate the accuracy of physicians in predicting short-term CrCl, and to compare the performance between them.

8.2 Methods

8.2.1 Study cohort

This prospective study was conducted in the surgical ICUs of the University Hospitals Leuven, Belgium from January to April 2022. Data collection and usage for the purpose of this study was approved by the Ethics Committee (EC) of University Hospitals Leuven (S65759 and S66669). Furthermore, this study was conducted in compliance with the principles of the Declaration of Helsinki laid down in the 1964 and its later amendments. We included all critically ill adults admitted to the surgical ICUs of the University Hospitals Leuven who did not receive kidney replacement therapy (KRT) before ICU admission. Patient-days were excluded if 1) no CrCl measurements were available on the next day, 2) no CrCl measurements were available on the day of prediction, 3) KRT was performed on the day of prediction, 4) KRT was performed in the previous week, 5) patients with shorter than 24-hour ICU length of stay, and/or 6) weekend or holidays (because of the workload of ICU physicians on call).

8.2.2 Study endpoints

The main objective of this study was to compare the predictive accuracy of ICU physicians and machine-learning models in predicting CrCl of the next day. The secondary objective was to examine whether the level of seniority and/or confidence level had an association with the predictive accuracy.

8.2.3 Prediction targets and CrCl definitions

CrCl was defined by serum creatinine (SCr), urine creatinine (Ucr), and 24-hour urine output (UO) without correction for body surface area:

$$CrCl\ (ml/min) = \frac{UCr \times UO}{SCr \times 1440}$$

8.2.4 CrCl predictor predictions

After the data collection at the end the study period of 3 months was finalized, one-day-ahead CrCl was predicted on each patient day, using the CrCl predictor models developed earlier (**Chapter 7**): the Core model using only admission data and daily routine laboratory results, the Core+BGA model that adds to the above, blood gas analysis data, and the Core+BGA+Monitoring model that adds to the above, monitoring data (heart rate, mean arterial blood pressure, and respiratory rate). More information about the features included in each model can be found in the original development study (**Chapter 7**). Missing values were imputed with the mean for continuous data and the mode for categorical data from the original study.

8.2.5 ICU physicians' predictions

All attending physicians of the surgical ICU team (junior resident, senior resident, and staff member) were asked to fill out questionnaires on a daily basis (**Figure 8.A.1**) in which they predicted the CrCl of the next day, and reported their levels of confidence (low, medium, or high confidence) with this prediction. The questionnaires were distributed to physician's office in each ICU unit, filled out by every physician for each patient under their care on each day, and collected two times a day (at noon and in the afternoon) to ensure a higher response rate.

ICU physicians were asked to fill in the questionnaire, which composed of the following questions.

- What is your prediction for CrCl (ml/min) on the next ICU day?
- How confident are you about your prediction? (low, medium, or high confidence)

The physicians' age, gender, time of ICU experience, and seniority level were also collected (**Figure 8.A.2**). The impact of seniority level and level of confidence on the predictive accuracy were investigated.

8.2.6 Evaluation metrics

Root-mean-square error (RMSE) and mean absolute error (MAE) between the predicted and actual CrCls were used to evaluate the predictive accuracy. Predictive performance was visualized and evaluated with scatter plots for the entire ICU stay, and plots of daily RMSE and MAE during the first week of ICU stay.

8.2.7 Statistical analysis

Data were presented as median with interquartile ranges (IQR) and counts with percentages (%) for continuous and categorical data in the descriptive statistics. All statistical analyses were conducted with Python version 3.7 (python.org), SciPy version 1.3.1 (scipy.org), and Scikit-learn version 0.24.2 (scikit-learn.org). Significance levels were set at the 5% level using the Mann–Whitney U test and Fisher's Exact Test for continuous data and categorical data respectively.

8.3 Results

8.3.1 Study cohort

Of the 459 patients who were included in this study, 262 patients were left out because all their patient-days were excluded so 197 patients remained for further analysis (**Figure 8.1**). Patients' characteristics were reported in **Table 8.1**. Median (interquartile range, IQR) age was 65 (55–75) years, 126 (64.0%) patients were male, median (IQR) Acute Physiology and Chronic Health Evaluation II (APACHE II) score was 16 (13–20), the major admission diagnosis was trauma and other surgery (n=81, 41.1%), and median (IQR) CrCl over the entire ICU stay was 97.0 (49.2–130.9) ml/min. The median (IQR) length of stay until the end of study were 7 (5–12) days, and 16 (8.1%) patients died before the end of study. In comparison to the original CrCl predictor development cohort, this study cohort consisted of comparable males (61.8% vs. 64.0%), had comparable medium ages (67 vs. 65 years), and showed a five-point lower median APACHE II score (22 vs. 16).

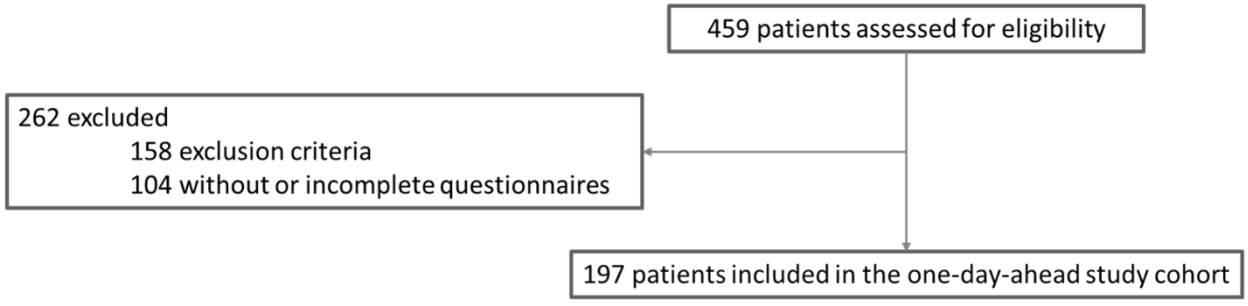


Figure 8.1 Consort diagram. CrCl, creatinine clearance; ICU, intensive care unit

Table 8.1 Patient characteristics and clinical outcomes

	All patients (n=197)
Age, years, median (IQR)	65 (55 – 75)
Gender male, number (%)	126 (64.0)
Day from ICU admission, median (IQR)	5.0 (3.5 – 6.6)
BMI, median (IQR)	24.7 (22.5 – 27.8)
Mean creatinine clearance over the ICU stay, ml/min, median (IQR)	97.0 (49.2 – 130.9)
Baseline serum creatinine, mg/dl, median (IQR)	0.9 (0.7 – 1.1)
APACHE II score, median (IQR)	16 (13 – 20)
Diabetes, number (%)*	
Emergency admission, number (%)*	
Sepsis on ICU admission, number (%)*	
Pharmacological hemodynamic support on ICU admission, number (%)*	
Reason for admission	
Cardiac surgery, number (%)	37 (18.8)
Medical disease, number (%)	44 (22.3)
Neurology, number (%)	14 (7.1)
Trauma and other surgery, number (%)	81 (41.1)
Transplantation, number (%)	21 (10.7)
Deceased during the study, number (%)	16 (8.1)
Length of stay until the end of study, median (IQR)	7 (5 – 12)

*The following variables are currently being retrieved: diabetes, emergency admission, sepsis on ICU admission, pharmacological hemodynamic support on ICU admission.

8.3.2 CrCl predictor predictions

All features were with less than 10% missing values, except for three features relating to respiratory rate and one feature derived from glucose level. The CrCl predictor was robust in this external validation cohort (**Table 8.2**): the Core+BGA+Monitoring model having the smallest RMSE showed 19.0 (95% CI 17.8-20.3) ml/min MAE and 29.9 ml/min RMSE, in comparison to 18.1 (95% CI 17.9-18.3) ml/min MAE and 28.9 ml/min RMSE in the validation cohort of the original study. The Core and Core+BGA models showed MAE of 19.7 (95% CI 18.4-20.9) and 19.4 (95% CI 18.1-20.7) ml/min, and RMSE of 30.5 and 30.6 ml/min, in contrast to MAE of 18.5 (95% CI 18.4-18.7) and 18.5 (95% CI 18.3-18.7) ml/min, and RMSE of 29.1 and 29.2 ml/min in the validation cohort of the original study.

Table 8.2 Summary of mean absolute error between the predicted and actual creatinine clearances, and root-mean-square error for the developed models and the physicians' predictions for the one-day-ahead prediction task

	Study cohort for the one-day-ahead CrCl prediction task	
	Mean absolute error (ml/min) (95% CI)	**Root-mean-square error (ml/min)**
Physicians' predictions	21.8 (20.2-23.4)	36.4
Core model	19.7 (18.4-20.9)	30.5
Core+BGA model	19.4 (18.1-20.7)	30.6
Core+BGA+Monitoring model	19.0 (17.8-20.3)	29.9

8.3.3 ICU physicians' predictions

Overall, 54 physicians completed 1298 questionnaires for 197 patients corresponding to 704 patient-days. From the 37 junior residents, 664 predictions were available in 182 patients. From the 8 senior residents 524 were available in 152 patients, and from the 9 staff members, 110 predictions were available in 63 patients (physicians' characteristics in **Table 8.A.1**). The median (IQR) hour to fill out the questionnaires was 2 (1–2) p.m. Clinicians completed the questionnaires for a median (IQR) of 8 (4–14) patients under their care.

The predictions made by ICU physicians demonstrated MAE of 21.8 (95% CI 20.2-23.4) ml/min and RMSE of 36.4 ml/min. In particular (**Table 8.A.2**), junior residents, senior residents, and staff members showed MAEs of 23.3 (95% CI 20.9-25.7), 20.9 (95% CI 18.5-23.3), and 16.9 (95% CI 13.5-20.4) ml/min, and RMSEs of 39.1, 34.8, and 25.0 ml/min respectively. Finally, the predictive performance of predictions made with low or medium confidence was clinically similar to that of predictions made with high confidence (MAE 22.4 (95% CI 20.6-24.2) vs. 20.0 (95% CI 16.6-23.3) ml/min, and RMSE 36.2 vs. 36.9 ml/min, for low or medium confidence and high confidence, respectively) (**Table 8.A.3**).

We evaluated the predictive performance during the first week of ICU stay separately. The CrCl predictor showed lower RMSE than the physicians, while the MAEs were not statistically significant (**Figure 8.2**). For both the CrCl predictor and ICU physicians alike, most predictions were close to the diagonal axis, indicating good agreement between predicted and true CrCls (**Figure 8.3**). In Appendix 8.A.1, the results of the two-day average CrCl predictions were presented, with the primary findings remaining the same.

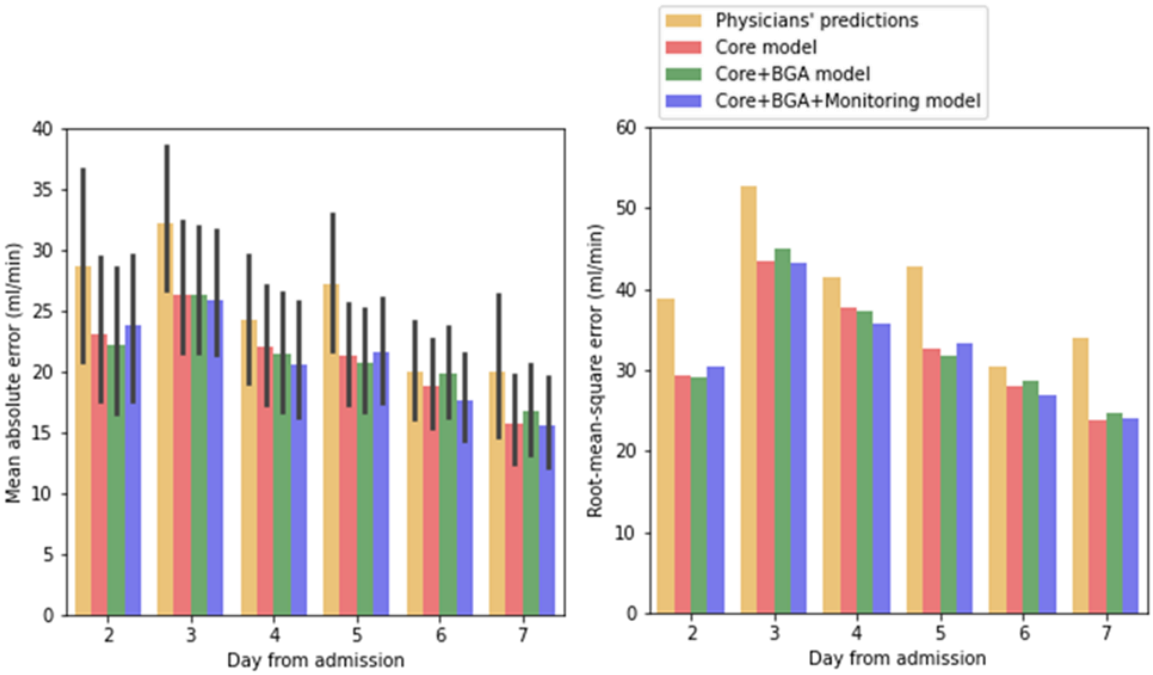


Figure 8.2 Temporal absolute difference (upper) and root-mean-square error (lower) of different within the first week of ICU admission. The orange, red, green, and blue bars represent, respectively, the ICU physicians' predictions, Core, Core+BGA, and Core+BGA+Monitoring models. Error bars represent 95% confidence intervals.

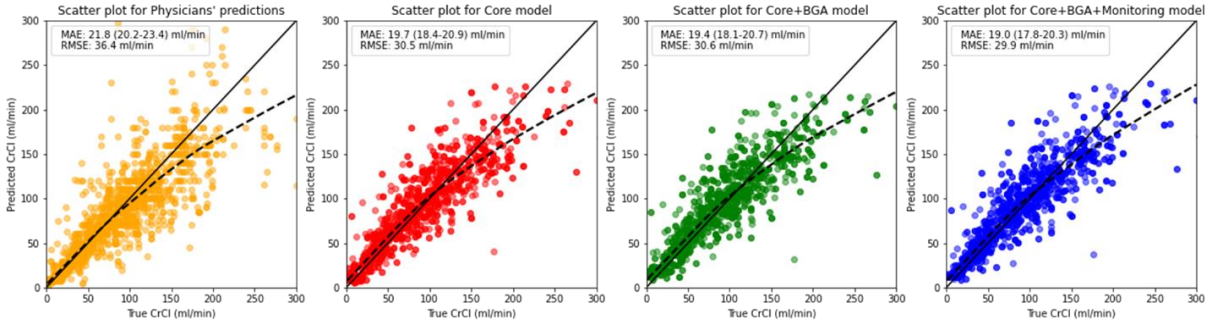


Figure 8.3 Relationships between predicted and actual CrCls for different models. The orange, red, green, and blue scatter plots show the results for the ICU physicians' predictions, Core, Core+BGA, Core+BGA+Monitoring models, respectively. The dashed and solid lines represent the lowess-based regression lines for the predictions and the diagonal axis. RMSE, root mean square error; CrCl, creatinine clearance

8.4 Discussion

In this study, we prospectively validated CrCl predictor in previously unseen critically ill adults and we found a comparable accuracy as in the original development study. Subsequently, we demonstrated that the CrCl predictor had lower RMSE than the ICU physicians' predictions. In particular, the CrCl predictor had slightly smaller RMSEs than the staff members and senior residents, and a much smaller RMSE than the junior residents. The absence of statistically significant MAEs within the first week of ICU admission indicated that the CrCl predictor performed at least as equally well as physicians. To the best of our knowledge, this study presents the first comparison between machine-learning algorithm and ICU physicians for daily CrCl prediction in the ICU.

The lower RMSE of the CrCl predictor and the absence of statistically significant MAEs within the first ICU week are noteworthy, given many factors favoring the physicians. First, physicians saw patients and thus had more information that were not available in the electronic health record [29], which might provide a more comprehensive understanding of patient condition. Second, the CrCl predictor based on the past information until 7AM, while most ICU physicians gave their predictions at 2PM, giving them seven hours of additional information about the patient circumstances. Besides, not all physicians succeeded in delivering a timely prediction before 2PM, regardless of the reason. As a result, ICU physicians may have anticipated better based on the treatment they provided on the day of prediction, which was not accessed by the CrCl predictor.

As our recently developed CrCl prediction model is the first model to predict CrCl, we did not find any study comparing the accuracy of ICU physicians with prediction models in predicting continuous kidney function. This underlines the importance of this work and future research. One prospective study investigating the onset of AKI is the nearest comparison [30]. Flechet et al. described a set of models for AKI, named AKIpredictor [23, 31], outperforming ICU physicians with similar discrimination and higher net benefit in an ICU population. Importantly, there are inherent differences between the present study from the previous work. First, the AKIpredictor was based on categorized AKI definitions, while the CrCl predictor focused on continuous kidney function. Consequently, the AKIpredictor could serve as a screening tool to identify patients with high risk of decreases in kidney function, and the CrCl predictor may help tailor the treatment to every patient with critical illness. Additionally, the AKIpredictor predicted AKI within the first week of ICU admission with an aim to identify the patients who may benefit the most with the focused care, and the CrCl predictor predicted on each ICU day in order to potentially facilitate daily renally cleared drug administration in the future.

The results of this study are promising for the future, showing many advantages of the CrCl predictor. First, the CrCl predictor's predictions are objective and time-invariant, as opposed to physicians' predictions, which are prone to be affected by emotion, stress, tiredness, and the limited time. Most importantly, physicians rely on a series of mental procedures, which cannot be quantified, reproduced, and studied for further improvement. Second, the performance not statistically significant different from the ICU physicians suggested that the CrCl predictor could at least perform equally well as physicians. In contrary to humans that cannot process more than three or four independent variables at the same time [32], machine-learning algorithms may identify hidden patterns based on large amounts of clinical data from multiple sources. Third, the CrCl predictor provides consistent predictions for all patients and could serve as an efficient, time-reducing and scalable screening tool to identify patients with high kidney function instability [33] based on the predictions generated for every patient on each day of their ICU stay.

However, there are several limitations in this study. First, this is a single-center Belgian study and the results may not be generalizable to other centers with different clinical settings. However, prospective validation of a machine-learning model in the same setting is needed before moving to other settings [34]. Second, the exclusion of patient-days with KRT on the day of prediction and in the previous week, or patient-days with less than two consecutive CrCls available precluded the short stayers and could have resulted in a selection bias. Nonetheless, the initiation and cessation of KRT has a huge effect on the SCr levels, resulting to the fact that CrCl in these circumstances is not a good surrogate of the GFR. Third, it remains unknown whether the use of the CrCl predictor can help improving the dosing adjustment of renally cleared drugs and stratification of high-risk patients, but it is out of the scope of this study and should be investigated in future studies.

8.5 Conclusion

We have demonstrated the robustness of the CrCl predictor on previously unseen patients in external validation. The ICU physicians' predictions had larger RMSEs than the CrCl predictor, with smaller RMSEs in higher seniority levels. The absence of statistically significant MAEs indicated that the CrCl predictor could perform at least as well as the ICU physicians. The findings suggest the potential added value of the CrCl predictor to physicians' predictions, especially for junior residents, and the possibility of implementation of the CrCl predictor into clinical decision support system to facilitate risk stratification and drug dosage adjustment involving renally cleared drugs in critically ill adults.

8.A Appendix

8.A.1 Supplementary results

8.A.1.1 For prediction of average CrCl of two days ahead: Study cohort

Data from 77 patients from the one-day-ahead study cohort were further left out because of the lack of prediction target, average CrCl of two days ahead, so 120 patients remained for further analysis (**Figure 8.A.3**). Patients' characteristics were reported in **Table 8.A.4**. Median (interquartile range, IQR) age was 65 (56 – 74) years, 77 (64.2%) patients were male, median (IQR) Acute Physiology and Chronic Health Evaluation II (APACHE II) score was 16 (13 – 21), the major admission diagnosis was trauma and other surgery (n=45, 37.5%), and median (IQR) CrCl over the entire ICU stay was 89.1 (47.4 – 120.1) ml/min. The median (IQR) length of stay until the end of study were 10 (8 – 16) days, and 15 (12.5%) patients died before the end of study. In comparison to the original CrCl predictor development cohort, this study cohort consisted of comparable males (64.2% vs. 60.9%), had comparable medium ages (65 vs. 67 years), and showed a nine-point lower median APACHE II score (16 vs. 25).

8.A.1.2 For prediction of average CrCl of two days ahead: CrClpredictor predictions

All features were with less than 10% missing values, except for four features relating to respiratory rate, one feature derived from Sequential Organ Failure Assessment (SOFA) score, one feature derived from respiratory support, and one feature derived from glucose level. The CrCl predictor was still robust in this external validation cohort (**Table 8.A.5**). In particular, the Core+BGA+Monitoring model having the smallest RMSE showed 17.8 (95% CI 16.3-19.3) ml/min MAE and 31.3 ml/min RMSE, in comparison to 16.6 (95% CI 16.4-16.7) ml/min MAE and 23.9 ml/min RMSE in the validation cohort of the original study. The Core and Core+BGA models showed MAE of 18.1 (95% CI 16.5-19.6) and 17.7 (95% CI 16.2-19.2) ml/min, and RMSE of 31.8 and 31.4 ml/min, in contrast to MAE of 16.8 (95% CI 16.7-17.0) and 16.35 (95% CI 16.2-16.5) ml/min, and RMSE of 24.4 and 23.8 ml/min in the validation cohort of the original study.

8.A.1.3 For prediction of average CrCl of two days ahead: ICU physicians' predictions

Overall, 51 physicians completed 796 questionnaires for 120 patients corresponding to 428 patient-days. The 35 junior residents made 402 predictions for 112 patients, The 8 senior residents made 334 predictions for 100 patients, and 8 staff members made 60 predictions for 34 patients (physicians' characteristics in **Table 8.A.6**).

Regardless of seniority level, the predictions made by ICU physicians demonstrated MAE of 21.1 (95% CI 19.3-23.0) ml/min and RMSE of 36.7 ml/min. In particular, junior residents, senior residents, and staff members showed MAEs of 22.6 (95% CI 19.8-25.3), 19.5 (95% CI 16.8-22.2), and 20.6 (95% CI 14.8-26.5) ml/min, and RMSEs of 36.1, 31.9, and 30.9 ml/min respectively (**Table 8.A.7**). Finally, the predictive performance of predictions made with low or medium confidence was clinically similar to that of predictions made with high confidence (MAE 21.2 (95% CI 19.3-23.2) vs. 20.6 (95% CI 14.8-26.4) ml/min, and RMSE 33.2 vs. 38.2 ml/min, for low or medium confidence and high confidence, respectively) (**Table 8.A.8)**.

In the subgroup of predictions during the first week of ICU stay, the CrCl predictor showed lower RMSE than the predictions of physicians, but the MAE was not significant (**Figure 8.A.4**). For both the CrCl predictor and ICU physicians' predictions, most predictions were close to the diagonal axis indicating good agreement between predicted and true CrCls (**Figure 8.A.5**).

8.A.2 Supplementary tables

Table 8.A.1 Physicians' characteristics for the one-day-ahead prediction task

	Physicians
Number of participants	54
Age, year, median (IQR)	29 (29 - 30)
Male gender, number (%)	33 (61.1)
Seniority level, number (%)	
Junior resident	37 (68.5)
Senior resident	8 (14.8)
Staff member	9 (16.7)

Table 8.A.2 Description of physicians' predictions per seniority and confidence levels for the one-day-ahead prediction task

	Junior resident (n=664 questionnaires, 566 days, 182 patients, 37 physicians)		Senior resident (n=524 questionnaires, 521 days, 152 patients, 8 physicians)		Staff member (n=110 questionnaires, 110 days, 63 patients, 9 physicians)	
	Mean absolute error (ml/min) (95% CI)	**Root-mean-square error (ml/min)**	**Mean absolute error (ml/min) (95% CI)**	**Root-mean-square error (ml/min)**	**Mean absolute error (ml/min) (95% CI)**	**Root-mean-square error (ml/min)**
Physicians' predictions	23.3 (20.9-25.7)	39.1	20.9 (18.5-23.3)	34.8	16.9 (13.5-20.4)	25.0
Core model	20.2 (18.4-22.0)	31.2	19.7 (17.6-21.7)	31.1	16.5 (13.5-19.5)	23.1
Core+BGA model	20.0 (18.2-21.9)	31.4	19.2 (17.1-21.3)	31.0	16.4 (13.3-19.6)	23.5
Core+BGA+Monitoring model	19.5 (17.7-21.3)	30.5	19.0 (16.9-21.0)	30.5	16.7 (13.6-19.8)	23.5

Table 8.A.3 Performance of clinicians split by confidence level for the one-day-ahead prediction task

	Physicians
Number of questionnaires	1298
Low or medium confidence	
Number (%)	966 (74.4)
MAE (ml/min)	22.4 (20.6-24.2)
RMSE (ml/min)	36.2
High confidence	
Number (%)	332 (25.6)
MAE (ml/min)	20.0 (16.6-23.3)
RMSE (ml/min)	36.9

MAE, mean absolute error; RMSE, root mean square error

Table 8.A.4 Patient characteristics and clinical outcomes for the average of two days ahead prediction task

	All patients (n=120)
Age, years, median (IQR)	65 (56 – 74)
Gender male, number (%)	77 (64.2)
Day from ICU admission, median (IQR)	5.8 (4.5 – 8.5)
BMI, median (IQR)	24.7 (22.3 – 27.1)
Mean creatinine clearance over the ICU stay, ml/min, median (IQR)	89.1 (47.4 – 120.1)
Baseline serum creatinine, mg/dl, median (IQR)	0.9 (0.7 – 1.1)
APACHE II score, median (IQR)	16.0 (13.0 – 20.5)
Diabetes, number (%)*	
Emergency admission, number (%)*	
Sepsis on ICU admission, number (%)*	
Pharmacological hemodynamic support on ICU admission, number (%)*	
Reason for admission	
Cardiac surgery, number (%)	20 (16.7)
Medical disease, number (%)	26 (21.7)
Neurology, number (%)	13 (10.8)
Trauma and other surgery, number (%)	45 (37.5)
Transplantation, number (%)	16 (13.3)
Deceased during the study, number (%)	15 (12.5)
Length of stay until the end of study, median (IQR)	10 (8 – 16)

*The following variables are currently being retrieved: diabetes, emergency admission, sepsis on ICU admission, pharmacological hemodynamic support on ICU admission.

Table 8.A.5 Summary of mean absolute error between the predicted and actual creatinine clearances, and root-mean-square error for the developed models and the physicians' predictions for the average CrCl of two days ahead prediction task

	Study cohort for the average CrCl of two days ahead prediction task	
	Mean absolute error (ml/min) (95% CI)	**Root-mean-square error (ml/min)**
Physicians' predictions	21.1 (19.3-23.0)	34.0
Core model	18.1 (16.5-19.6)	28.4
Core+BGA model	17.7 (16.2-19.2)	28.0
Core+BGA+Monitoring model	17.8 (16.3-19.3)	28.0

Table 8.A.6 Physicians' characteristics for the average of two days ahead prediction task

	Physicians
Number of participants	51
Age, year, median (IQR)	29 (29 - 30)
Male gender, number (%)	32 (62.8)
Seniority level, number (%)	
Junior resident	35 (68.6)
Senior resident	8 (15.7)
Staff member	8 (15.7)

Table 8.A.7 Description of physicians' predictions per seniority and confidence levels for the average of two days ahead prediction task

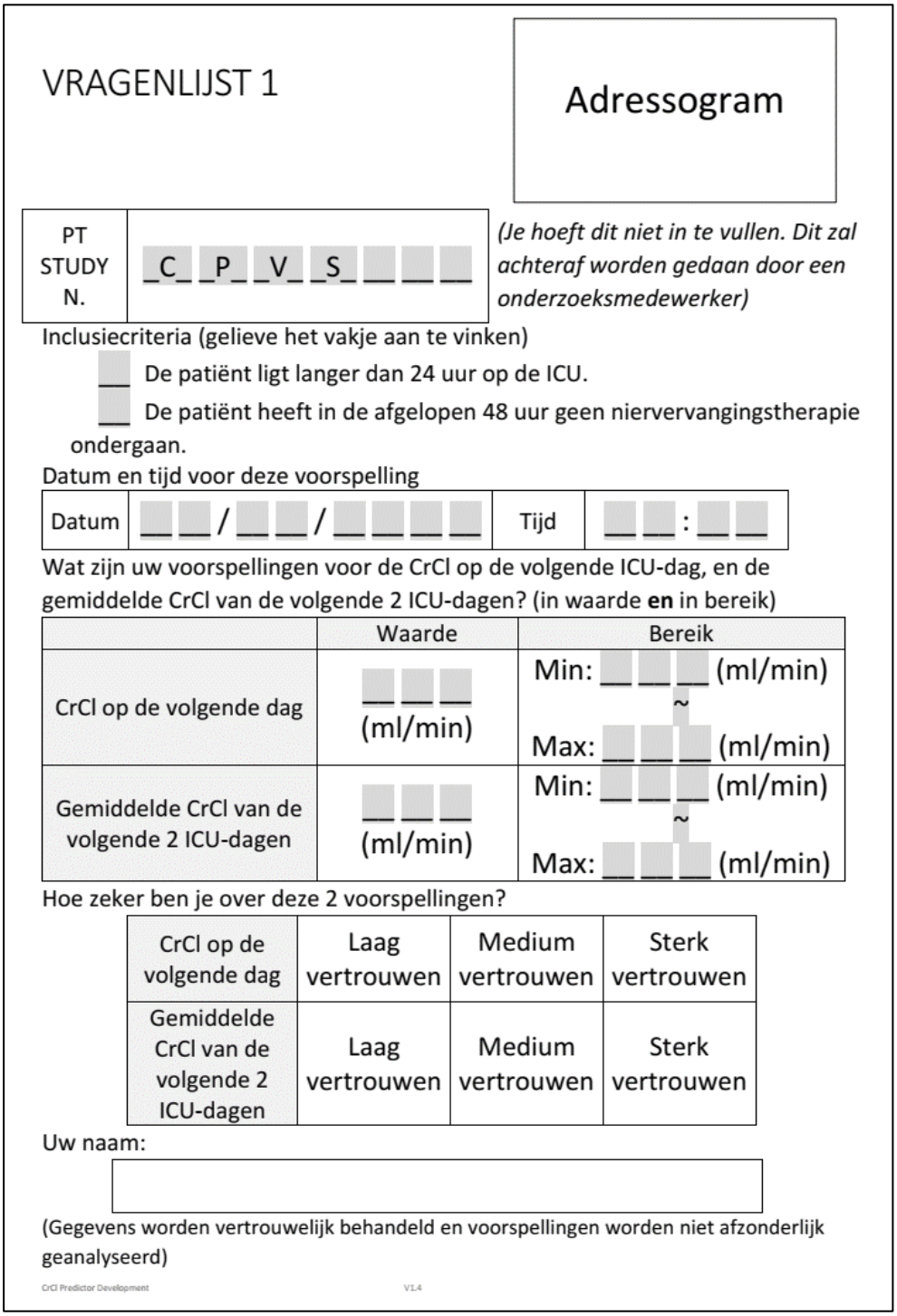
	Junior resident (n=402 questionnaires, 344 days, 112 patients, 35 physicians)		Senior resident (n=334 questionnaires, 331 days, 100 patients, 8 physicians)		Staff member (n=60 questionnaires, 60 days, 34 patients, 8 physicians)	
	Mean absolute error (ml/min) (95% CI)	**Root-mean-square error (ml/min)**	**Mean absolute error (ml/min) (95% CI)**	**Root-mean-square error (ml/min)**	**Mean absolute error (ml/min) (95% CI)**	**Root-mean-square error (ml/min)**
Physicians' predictions	22.6 (19.8-25.3)	36.1	19.5 (16.8-22.2)	31.9	20.6 (14.8-26.5)	30.9
Core model	18.6 (16.4-20.8)	29.3	17.9 (15.6-20.3)	28.3	15.0 (10.8-19.2)	22.3
Core+BGA model	18.1 (15.9-20.3)	28.6	17.8 (15.5-20.1)	27.9	15.1 (10.4-19.8)	23.7
Core+BGA+Monitoring model	18.3 (16.2-20.5)	28.8	17.7 (15.4-20.0)	27.8	14.7 (10.2-19.1)	22.8

Table 8.A.8 Performance of clinicians split by confidence level for the average of two days ahead prediction task

	Physicians
Number of questionnaires	796
Low or medium confidence	
Number (%)	676 (84.9)
MAE (ml/min)	21.2 (19.3-23.2)
RMSE (ml/min)	33.2
High confidence	
Number (%)	120 (15.1)
MAE (ml/min)	20.6 (14.8-26.4)
RMSE (ml/min)	38.2

MAE, mean absolute error; RMSE, root mean square error

8.A.3 Supplementary figures



VRAGENLIJST 1

Adressogram

PT STUDY N.	_C_ _P_ _V_ _S_ __ __ __

(Je hoeft dit niet in te vullen. Dit zal achteraf worden gedaan door een onderzoeksmedewerker)

Inclusiecriteria (gelieve het vakje aan te vinken)

__ De patiënt ligt langer dan 24 uur op de ICU.

__ De patiënt heeft in de afgelopen 48 uur geen niervervangingstherapie ondergaan.

Datum en tijd voor deze voorspelling

Datum	__ __ / __ __ / __ __ __ __	Tijd	__ __ : __ __

Wat zijn uw voorspellingen voor de CrCl op de volgende ICU-dag, en de gemiddelde CrCl van de volgende 2 ICU-dagen? (in waarde **en** in bereik)

	Waarde	Bereik
CrCl op de volgende dag	__ __ __ (ml/min)	Min: __ __ __ (ml/min) ~ Max: __ __ __ (ml/min)
Gemiddelde CrCl van de volgende 2 ICU-dagen	__ __ __ (ml/min)	Min: __ __ __ (ml/min) ~ Max: __ __ __ (ml/min)

Hoe zeker ben je over deze 2 voorspellingen?

CrCl op de volgende dag	Laag vertrouwen	Medium vertrouwen	Sterk vertrouwen
Gemiddelde CrCl van de volgende 2 ICU-dagen	Laag vertrouwen	Medium vertrouwen	Sterk vertrouwen

Uw naam:

(Gegevens worden vertrouwelijk behandeld en voorspellingen worden niet afzonderlijk geanalyseerd)

CrCl Predictor Development V1.4

Figure 8.A.1 Prediction questionnaire

CrClpredictor Kwaliteitsbewaking
VRAGENLIJST 2 (Artsen)

Om de voorspellingen van de artsen op leeftijd, geslacht, anciënniteit en aantal jaren ervaring op intensieve zorgen te analyseren, willen we deze gegevens ook van de artsen verzamelen. We willen graag benadrukken dat we de voorspellingen op een niveau van individuele arts niet gaan analyseren. Uw hoedanigheid als voorspeller is daarom niet getest en het is zeker geen evaluatie van uw vaardigheden als arts. De enige reden om uw naam op elke voorspelling te plaatsen, is om ze te koppelen aan deze vragenlijst 2.

Uw naam:

Wat is uw positie als een arts in de intensieve afdeling Geneeskunde?

ASSISTENT	SENIOR ASSISTENT	VAST STAFLID

Hoe oud ben je?

__ __ jaar

Wat is je geslacht?

MAN	VROUW

Hoeveel jaar ervaring heeft u in intensieve geneeskunde?

Uw cumulatieve klinische ervaring telt, inclusief de tijd die u hebt gewerkt als assistent, senior assistent of medewerker. Stagiairs, student of doctoraatsstudent-ervaring telt niet, en heeft geen onderzoekservaring.

- Als u in ICU voor minder dan 1 maand hebt gewerkt, voert u de tijd in weken in.
- Als u tussen 1 maand en 48 maanden (2 jaar) in ICU hebt gewerkt, voert u de tijd in maanden in.
- Als u langer dan 2 jaar in ICU hebt gewerkt, voert u de tijd in jaren in.

__ __ week __ __ maand __ __ jaar

Bedankt voor het samenwerken!

Figure 8.A.2 Physician questionnaire

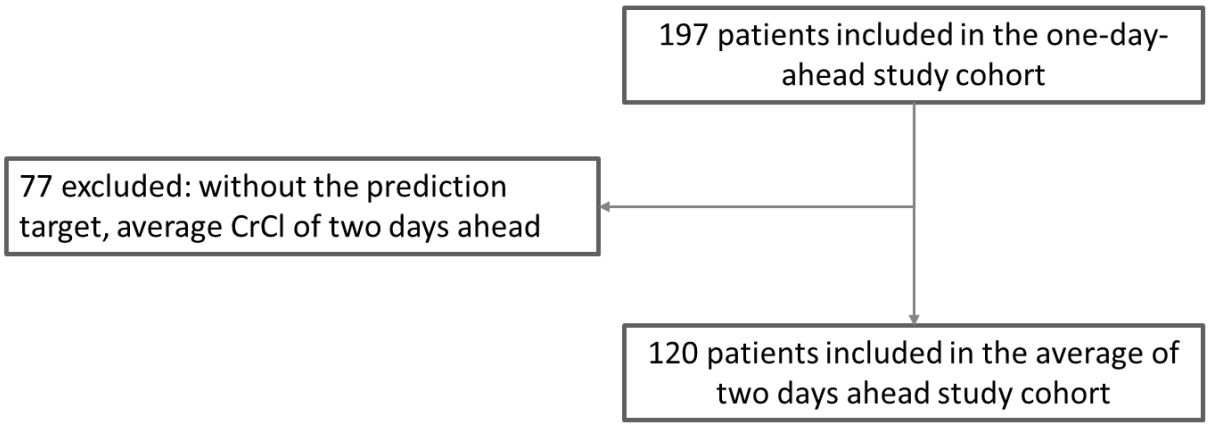


Figure 8.A.3 Consort diagram for prediction of average CrCl of two days ahead. CrCl, creatinine clearance; ICU, intensive care unit

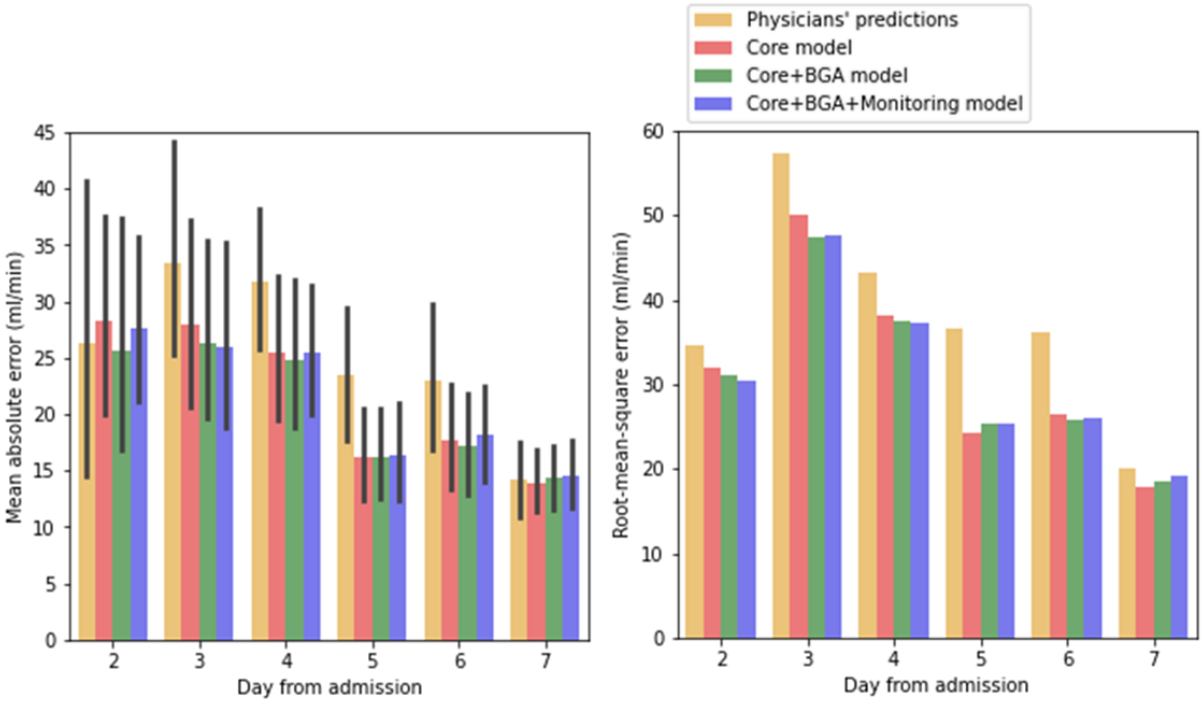


Figure 8.A.4 Temporal mean absolute error (left) and root-mean-square error (right) of different models within the first week of ICU admission for the average of two days ahead prediction task. Error bars represent 95% confidence intervals. There was no significant difference in mean absolute error between the predictions made by each of the CrCl predictor models and the physicians on each day of the first week of ICU admission.

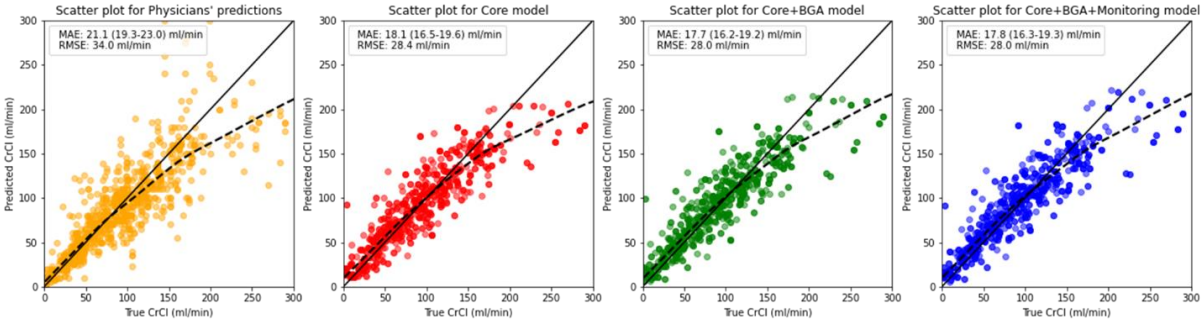


Figure 8.A.5 Relationships between predicted and actual CrCls for different models for the average of two days ahead prediction task. The orange, red, green, and blue scatter plots show the results for the ICU physicians' predictions, Core, Core+BGA, Core+BGA+Monitoring models, respectively. The dashed and solid lines represent the lowess-based regression lines for the predictions and the diagonal axis. MAE, mean absolute error; RMSE, root mean square error; CrCl, creatinine clearance

GRANT SUPPORT AND CONFLICT OF INTEREST

C-Y.Huang receives funding from the Taiwan-KU Leuven scholarship. G.Meyfroidt is funded by the research foundation, Flanders (FWO) as senior clinical investigator (1843123N).

All the authors declare no competing interests that are relevant to the content of this article.

ACKNOWLEDGMENT AND PERSONAL CONTRIBUTION

Study concept and design	**Huang**, Güiza, De Vlieger, Meyfroidt
Data acquisition	**Huang**, Wouters, Mebis, Gidts, Klewais
Statistical analysis	**Huang**
Interpretation of results	**Huang**, Güiza, De Vlieger, Meyfroidt
Drafting of the manuscript	**Huang**, Güiza, De Vlieger, Meyfroidt
Manuscript revision	**All authors**
Principal investigator	Meyfroidt

The authors would like to thank the junior residents, senior residents, and staff members from surgical ICUs of the University Hospitals Leuven for participating the study. The authors are also grateful for the administration work, data entry, and logistic support from the clinical research assistants (Pieter Wouters, Liese Mebis, Sandra Hendrickx, Katleen Witpas, Heidi Utens, and Hanna Van Cleemput) of laboratory of intensive care medicine, KU Leuven, Belgium. The authors would like to mention that some authors (C-Y.Huang, P.Wouters, L.Mebis, F.Güiza, G. De Vlieger, G.Meyfroidt) participated in the development of the CrCl predictor.

BIBLIOGRAPHY

1. Case J, Khan S, Khalid R, Khan A (2013) Epidemiology of acute kidney injury in the intensive care unit. Crit Care Res Pract 2013:479730. https://doi.org/10.1155/2013/479730
2. Jiang L, Zhu Y, Luo X, et al (2019) Epidemiology of acute kidney injury in intensive care units in Beijing: The multi-center BAKIT study. BMC Nephrol 20:468. https://doi.org/10.1186/s12882-019-1660-z
3. Hoste EAJJ, Bagshaw SM, Bellomo R, et al (2015) Epidemiology of acute kidney injury in critically ill patients: the multinational AKI-EPI study. Intensive Care Med 41:1411–1423. https://doi.org/10.1007/s00134-015-3934-7
4. Susantitaphong P, Cruz DN, Cerda J, et al (2013) World Incidence of AKI: A Meta-Analysis. Clin J Am Soc Nephrol 8:1482–1493. https://doi.org/10.2215/CJN.00710113
5. Wang HE, Muntner P, Chertow GM, Warnock DG (2012) Acute kidney injury and mortality in hospitalized patients. Am J Nephrol 35:349–355. https://doi.org/10.1159/000337487
6. Chertow GM, Burdick E, Honour M, et al (2005) Acute kidney injury, mortality, length of stay, and costs in hospitalized patients. J Am Soc Nephrol 16:3365–3370. https://doi.org/10.1681/ASN.2004090740
7. Federspiel CK, Itenov TS, Mehta K, et al (2018) Duration of acute kidney injury in critically ill patients. Ann Intensive Care 8:30. https://doi.org/10.1186/s13613-018-0374-x
8. Bilbao-Meseguer I, Rodríguez-Gascón A, Barrasa H, et al (2018) Augmented Renal Clearance in Critically Ill Patients: A Systematic Review. Clin Pharmacokinet 57:1107–1121. https://doi.org/10.1007/s40262-018-0636-7
9. Abdel El Naeem HEM, Abdelhamid MHE, Atteya DAM (2017) Impact of augmented renal clearance on enoxaparin therapy in critically ill patients. Egypt J Anaesth 33:113–117. https://doi.org/https://doi.org/10.1016/j.egja.2016.11.001
10. He J, Yang Z-TT, Qian X, et al (2020) A higher dose of vancomycin is needed in critically ill patients with augmented renal clearance. Transl Androl Urol 9:2166–2171. https://doi.org/10.21037/tau-20-1048
11. Huttner A, Von Dach E, Renzoni A, et al (2015) Augmented renal clearance, low β-lactam concentrations and clinical outcomes in the critically ill: An observational prospective cohort study. Int J Antimicrob Agents 45:385–392. https://doi.org/10.1016/j.ijantimicag.2014.12.017
12. Udy AA, Varghese JM, Altukroni M, et al (2012) Subtherapeutic initial β-lactam concentrations in select critically ill patients: association between augmented renal clearance and low trough drug concentrations. Chest 142:30–39. https://doi.org/10.1378/chest.11-1671
13. Claus BOMM, Hoste EA, Colpaert K, et al (2013) Augmented renal clearance is a common finding with worse clinical outcome in critically ill patients receiving antimicrobial therapy. J Crit Care 28:695–700. https://doi.org/10.1016/j.jcrc.2013.03.003
14. Carlier M, Dumoulin A, Janssen A, et al (2015) Comparison of different equations to assess glomerular filtration in critically ill patients. Intensive Care Med 41:427–435. https://doi.org/10.1007/s00134-014-3641-9

15. Le S, Allen A, Calvert J, et al (2021) Convolutional Neural Network Model for Intensive Care Unit Acute Kidney Injury Prediction. Kidney Int Reports 6:1289–1298. https://doi.org/https://doi.org/10.1016/j.ekir.2021.02.031

16. Malhotra R, Kashani KB, Macedo E, et al (2017) A risk prediction score for acute kidney injury in the intensive care unit. Nephrol Dial Transplant 32:814–822. https://doi.org/10.1093/ndt/gfx026

17. Wang Q, Tang Y, Zhou J, Qin W (2019) A prospective study of acute kidney injury in the intensive care unit: development and validation of a risk prediction model. J Transl Med 17:359. https://doi.org/10.1186/s12967-019-2118-6

18. Chiofolo C, Chbat N, Ghosh E, et al (2019) Automated Continuous Acute Kidney Injury Prediction and Surveillance: A Random Forest Model. Mayo Clin Proc 94:783–792. https://doi.org/10.1016/j.mayocp.2019.02.009

19. Sato N, Uchino E, Kojima R, et al (2021) Prediction and visualization of acute kidney injury in intensive care unit using one-dimensional convolutional neural networks based on routinely collected data. Comput Methods Programs Biomed 206:106129. https://doi.org/https://doi.org/10.1016/j.cmpb.2021.106129

20. Li Y, Yao L, Mao C, et al (2018) Early Prediction of Acute Kidney Injury in Critical Care Setting Using Clinical Notes. In: 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp 683–686

21. Gao W, Wang J, Zhou L, et al (2021) Prediction of acute kidney injury in ICU with gradient boosting decision tree algorithms. Comput Biol Med 140:105097. https://doi.org/10.1016/j.compbiomed.2021.105097

22. Chen Y, Feng F, Li M, et al (2019) Development of a risk stratification-based model for prediction of acute kidney injury in critically ill patients. Med (United States) 98:e16867. https://doi.org/10.1097/MD.0000000000016867

23. Flechet M, Güiza F, Schetz M, et al (2017) AKIpredictor, an online prognostic calculator for acute kidney injury in adult critically ill patients: development, validation and comparison to serum neutrophil gelatinase-associated lipocalin. Intensive Care Med 43:764–773. https://doi.org/10.1007/s00134-017-4678-3

24. Gijsen M, Huang CY, Flechet M, et al (2020) Development and External Validation of an Online Clinical Prediction Model for Augmented Renal Clearance in Adult Mixed Critically Ill Patients: The Augmented Renal Clearance Predictor. Crit Care Med 48:e1260–e1268. https://doi.org/10.1097/CCM.0000000000004667

25. Tang L, Ding XY, Duan LF, et al (2021) A Regression Model to Predict Augmented Renal Clearance in Critically Ill Obstetric Patients and Effects on Vancomycin Treatment. Front Pharmacol 12:1–12. https://doi.org/10.3389/fphar.2021.622948

26. Nei AM, Kashani KB, Dierkhising R, Barreto EF (2020) Predictors of Augmented Renal Clearance in a Heterogeneous ICU Population as Defined by Creatinine and Cystatin C. Nephron 55905:313–320.

https://doi.org/10.1159/000507255

27. Barletta JF, Mangram AJ, Byrne M, et al (2017) Identifying augmented renal clearance in trauma patients: Validation of the Augmented Renal Clearance in Trauma Intensive Care scoring system. In: The journal of trauma and acute care surgery. United States, pp 665–671

28. Udy AA, Roberts JA, Shorr AF, et al (2013) Augmented renal clearance in septic and traumatized patients with normal plasma creatinine concentrations: Identifying at-risk patients. Crit Care 17:. https://doi.org/10.1186/cc12544

29. Weiner SJ, Wang S, Kelly B, et al (2020) How accurate is the medical record? A comparison of the physician's note with a concealed audio recording in unannounced standardized patient encounters. J Am Med Inform Assoc 27:770–775. https://doi.org/10.1093/jamia/ocaa027

30. Flechet M, Falini S, Bonetti C, et al (2019) Machine learning versus physicians' prediction of acute kidney injury in critically ill adults: A prospective evaluation of the AKIpredictor. Crit Care 23:. https://doi.org/10.1186/s13054-019-2563-x

31. AKI prediction. https://www.akipredictor.com/en/. Accessed 10 Feb 2022

32. Halford GS, Baker R, McCredden JE, Bain JD (2005) How many variables can humans process? Psychol Sci 16:70–76

33. Huang C-Y, Güiza F, De Vlieger G, Meyfroidt G (2022) Daily fluctuations in kidney function in critically ill adults. Crit Care 26:347. https://doi.org/10.1186/s13054-022-04226-3

34. Dawes RM, Faust D, Meehl PE (1989) Clinical versus actuarial judgment. Science (80-) 243:1668–1674. https://doi.org/10.1126/science.2648573

9

Development of a prototype software to visualize the temporal evolution and prediction of kidney function

Adapted from: **Chao-Yuan Huang**, Fabian Güiza, Greet De Vlieger, Geert Meyfroidt. “Development of a prototype software to visualize the temporal evolution and prediction of kidney function”. Ready for submission.

Abstract

Purpose: Continuous kidney function prediction is important and necessary for drug dosage adjustment of renally cleared drugs and for risk stratification of patients with high kidney function instability. Therefore, we have developed the creatinine clearance (CrCl) predictor to predict short-term kidney function using machine-learning algorithm and regularly collected clinical data. However, technical knowledges are still required to use the CrCl predictor. Consequently, we aim to develop a prototype software that automatically performs the prediction algorithm in the background and visualizes the prediction results along with explanations in a user-friendly manner.

Methods: The study cohort described in **Chapter 7** was utilized to evaluate the functionality of the prototype program. We integrated the CrCl predictor into a software that can automatically make predictions for CrCl of the next ICU day. We visualized the temporal kidney function evolution and CrCl predictions along with the confidence intervals. Additionally, individual prediction explanation plot can be generated to increase model interpretability. The software was developed in Python 3.7.4 with Kivy version 1.11.1 library.

Results: We constructed a prototype software that can automatically collect necessary data from the patient data management system databases, pre-process the collected data, transform the pre-processed data into informative features, make predictions for the CrCl of the next ICU day, and present results with user-friendly graphical user interface in a real-time manner.

Conclusion: The proposed software successfully integrates the previously developed CrCl predictor and visualizes the prediction results along with explanations in a continuous way. Future studies are still warranted to investigate whether this software can help improving the kidney function management, patient care and outcome in clinical practice.

9.1 Introduction

In critically ill patients, clinically significant changes in kidney function were found to happen to 35-40% of days during the ICU stay (**Chapter 6**) [1]. In particular, acute kidney injury (AKI) and augmented renal clearance (ARC) happen frequently with reported prevalence of 40-60% and 20-65% in critically ill patients [2–4]. They both have adverse effects such as treatment failure and drug toxicity [3, 5], especially for renally cleared drugs such as vancomycin and β-lactam anti-microbials [6–8], as well as anticoagulants [9, 10]. The currently used consensus to diagnose AKI is based on serum creatinine (SCr) and urine output (UO) [11], both of which have their limitations. Specifically, muscle mass and hydration status have an impact on SCr [12], while diuretics influence UO. For ARC, the, commonly used definition is based on the creatinine clearance (CrCl) measured from a 24-hour urine collection, which slows down decision-making and renally excreted drug administration. Hence, there is a need for kidney function prediction to allow for quick decision making and timely drug dosage adjustment. However, most prediction models for kidney function were built to predict the onset of two extremes of kidney function spectrum: AKI [13–21] and ARC [22–26]. Since the kidney function is continuous, predicting the full range of kidney function is more in line with clinical and physiological reality.

Therefore, in a previous study (**Chapter 7**), we developed the CrCl predictor models for daily prediction of measured CrCl on the next day. The measured CrCl based on urinary volume and creatinine concentrations in serum and urine is currently the best surrogate for glomerular filtration rate (GFR) during critical illness [27]. In an external validation cohort, the developed models had smaller prediction errors than the reference reflecting the current clinical practice that assumed CrCl remained, mainly on the days when the kidney function was unstable. Subsequently, a prospective validation demonstrated that the CrCl predictor had a comparable accuracy to that observed in the original model development study and had smaller prediction errors than the predictions made by the treating physicians (**Chapter 8**).

Despite the many successes of the developed CrCl predictor and its potential as a tool to optimize drug dosing regimen and to stratify patients at risk, the implementation in clinical settings may be hampered by the time-consuming data retrieval and pre-processing procedures. As a result, the goal of this study is to create a prototype software that incorporates the previously developed CrCl prediction model and visualizes the prediction results and corresponding explanations in a user-friendly way, with an ultimate goal to facilitate the kidney function evaluation of critically ill patients.

9.2 Methods

To examine the functionality of the prototype software, the study cohort included in the **Chapter 7** was used.

9.2.1 CrCl definition

The following formula was used to determine CrCl, without correction for body surface area: CrCl (ml/min) = Urine creatinine (UCr) (mg/dL) × 24h UO (ml/day) / SCr (mg/dL) / 1440 (min/day). When multiple values were available on the same ICU day, summation was used for UO, and mean was applied to UCr and SCr. The UOs and CrCls measured on the first patient-days were corrected by hours in ICU due to the incomplete urine output collection.

9.2.2 CrCl predictor

The CrCl predictor included a series of models that predict CrCl on the next ICU day utilizing more features with increasing data resolution (Core model, Core+BGA model, and Core+BGA+Monitoring model) (**Chapter 7**). The Core model used only admission data and daily routine laboratory results. The Core+BGA model added to the above, blood gas analysis data. The Core+BGA+Monitoring model added to the above, monitoring data (heart rate, mean arterial blood pressure, and respiratory rate). The gradient-boosting trees method was used for model development, with features chosen after careful deliberation with two knowledgeable ICU physicians, consideration of the data availability, and the backward elimination method. In this study, as a proof of concept, only the most complete Core+BGA+Monitoring model was integrated in the software.

9.2.3 Prediction confidence interval

To indicate uncertainty levels of the prediction, prediction intervals were created by employing the quantile regression [28]. In contrast to the least squares method that calculates the conditional mean, quantile regression seeks to estimate the conditional quantiles. In this software, an upper bound of 0.975 quantile and a lower bound of 0.025 quantile were set by default, so that 95% confidence intervals can be generated, but the threshold can also be adapted, with different hyperparameters used in the CrCl predictor model development.

9.2.4 Individual prediction explanation

Machine-learning models have been notorious for the black-box characteristics. However, clinicians prefer strong evidence-based scientific backing to provide meaningful intervention and to modify their treatment approaches to vulnerable ICU patients. Therefore, many explanation techniques have been proposed for individual predictions, such as permutation importance [29], SHapley Additive exPlanations (SHAP) [30], and local interpretable model-agnostic explanations (LIME) [31]. Their consistency within literature and clinical interpretation have been showed [32–34]. LIME was used in this software for model explanation purpose due to its model-agnostic characteristics that treats the model as a black box and making no assumptions about the model, and self-explanatory visualization.

9.2.5 Software development

The software was developed in Python 3.7.4 (Python Software Foundation, http://www.python.org) with Kivy library 1.11.1 (kivy.org). Data were retrieved from the patient data management system (PDMS) database (Microsoft SQL Server®; Microsoft®, Redmond, Washington, USA), which allowed for a safe and efficient transmission of the necessary patient data and enabled the real-time applications.

9.2.6 User interface

To ensure the software comfortability for users to interact with, and to reduce the potential cognitive load, the software user interface was iteratively improved based on the comments given by the two ICU physicians of the University Hospitals Leuven (GDV and GM).

9.3 Results

9.3.1 Overall functionality

The patients assessed by the prototype software had their data queried and predictions computed successfully. **Figure 9.1** provides a graphical description of all activities included in the proposed software, which consists of three screens: the password-protected user login screen (**Figure 9.2**), the hospital number entry screen (**Figure 9.3**), and the main screen where the CrCl prediction and kidney function evolution are visualized (**Figure 9.4**). Furthermore, a popup window containing a prediction explanation plot can be generated on demand (**Figure 9.5**).

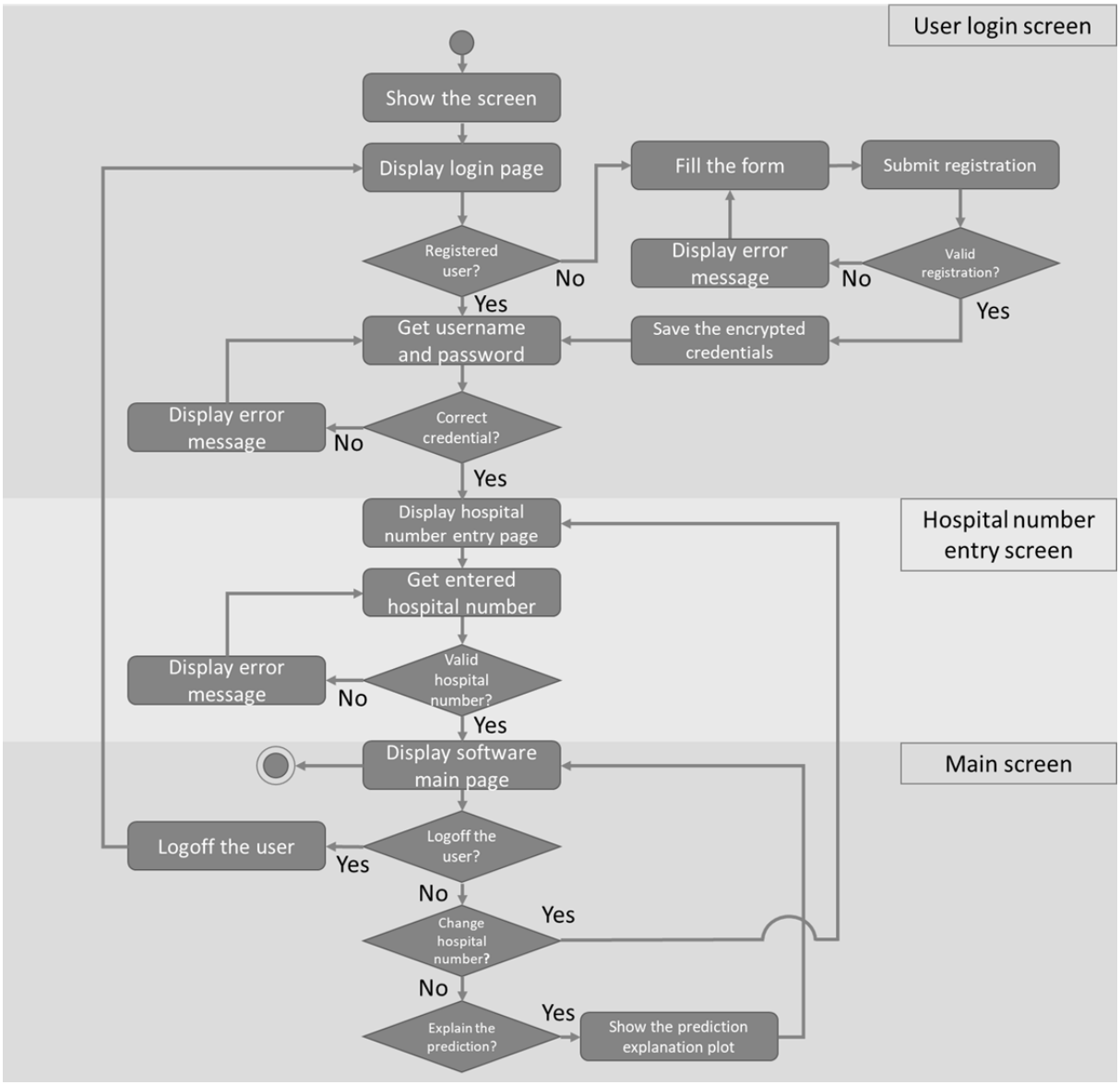


Figure 9.1. Activity diagram of the developed software. The upper, middle, and lower regions indicate the activities that are performed in the user login screen, hospital number entry screen, and the main screen respectively.

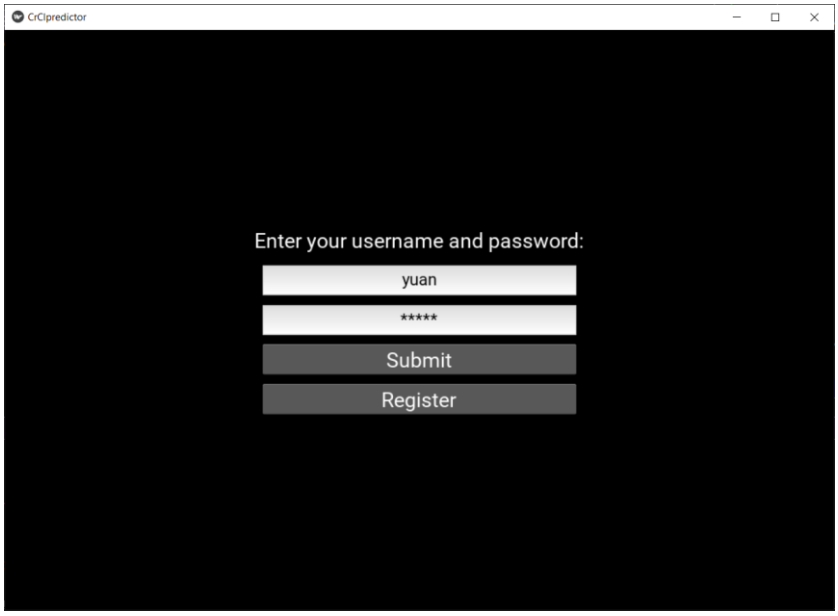


Figure 9.2. Screenshot of the password-protected user login screen

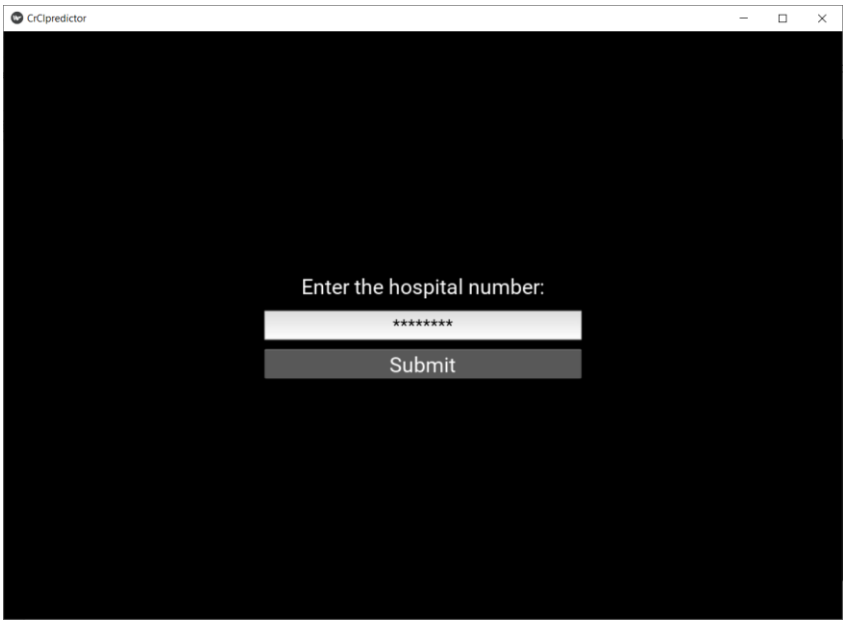


Figure 9.3. Screenshot of the screen for hospital number entry

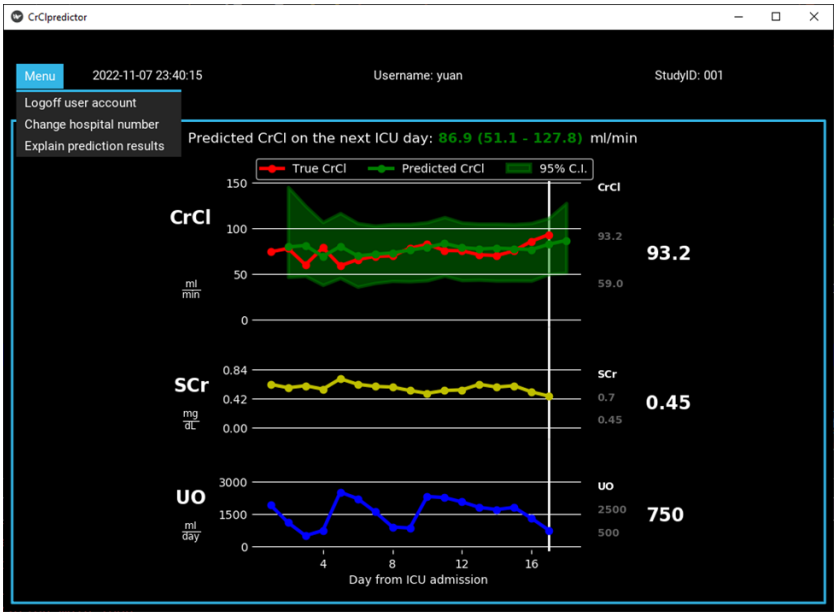


Figure 9.4. Screenshot of the main screen

Figure 9.5. Screenshot of the popup window containing a prediction explanation plot

9.3.2 User login screen

When conducting a clinical trial, it is necessary to have a password-protected login screen, in order to defend against any unauthorized access to crucial patient information. As a result, to meet the high clinical security level, the stored registered user credentials are encrypted with Secure Hash Algorithm 256 (SHA-256), a powerful cryptographic hash function. Contrary to several other well-known hashing algorithms such as MD5 message-digest algorithm and Secure Hash Algorithm 1 (SHA-1), SHA-256 has no known security flaws.

9.3.3 Hospital number entry screen

The hospital number entry screen allows to choose the specific patient of interest. Since one patient may have multiple ICU admissions, to identify the correct admission, the entered patient identification number is further matched with the closest ICU admission datetime, based on which the corresponding patient data are retrieved directly from the PDMS database. Subsequently, for the deidentification purpose, the entered patient identification number is replaced by the study number.

9.3.4 Main screen

The main screen contains two parts. In the upper header panel, the corresponding username and study number for the patient under investigation are indicated in the header for clarification purpose. Furthermore, there is a menu bar providing the following functions: 1) logoff user account, 2) change hospital number, and 3) explain prediction results. The first and the second functions change the current screen to the password-protected user login screen and the hospital number entry screen respectively, and the third function shows a popup window including a prediction explanation plot for the current prediction result.

In the middle main panel, there are four traces for true CrCl, predicted CrCl, SCr, and UO on each ICU day, where the vertical white line represents the current ICU day. The SCr and UO traces are represented in yellow and blue. For the CrCl traces, the true CrCl and predicted CrCl are indicated in red and green individually, with the green shaded area indicating the confidence intervals. The parameters' names and corresponding units are displayed on the left-hand side, and their minimum, maximum, and last available values are shown on the right-hand side. Every day at 7 a.m., based on the past information since ICU admission, the prediction for the following day is made and added to the graph.

9.3.5 Prediction explanation plot

Features that are positively correlated with the one-day-head CrCl are shown in red, otherwise blue. The length of bars indicates the magnitude of the corresponding linear regression coefficient of the model fitted locally to the predictions from the original model. The longer the bar is, the more important the corresponding feature is to the prediction of one-day-ahead CrCl.

9.3.6 Visualization examples of individual prediction

To demonstrate the prototype software's performance, we selected three cases where the CrCls were at medium, high, and low ranges respectively. First, **Figure 9.4** and **Figure 9.5** show screenshots of the main screen and prediction explanation plot for a 83-year-old woman. Her last available [minimum-maximum] SCr was 0.45 [0.45-0.7] mg/dL, last available [minimum-maximum] UO was 750 [500-2500] ml/day, and last available [minimum-maximum] CrCl was 93.2 [59.0-93.2] ml/min. It was predicted that the woman would have 86.9 [95% CI 51.1-127.8] ml/min CrCl on the next ICU day. Regarding the prediction explanation, the most important three features were "CrCl of the previous day", "Mean of CrCl of all past days during ICU stay", and "Age". Specifically, a "CrCl of the previous day" in a range between 81.7 and 124.0 ml/min positively correlated to the one-day-ahead CrCl, contributing the most to the prediction result. Furthermore, a "Mean of CrCl of all past days during ICU stay" in a range between 47.8 and 81.2 ml/min and an age larger than 74 years negatively correlated to the one-day-ahead CrCl.

Second, **Figure 9.6** and **Figure 9.7** show screenshots of the main screen and prediction explanation plot for a 35-year-old man. His last available [minimum-maximum] SCr was 0.54 [0.38-0.93] mg/dL, last available [minimum-maximum] UO was 4350 [900-4600] ml/day, and last available [minimum-maximum] CrCl was 198.03 [28.7-294.5] ml/min. It was predicted that the man would have a 206.5 [95% CI 103.4-253.5] ml/min CrCl on the next ICU day. Regarding the prediction explanation, the most important three features were "CrCl of the previous day", "Mean of CrCl of all past days during ICU stay", and "Age". Specifically, a "CrCl of the previous day" larger than 124.0 ml/min, a "Mean of CrCl of all past days during ICU stay" larger than 122.3 ml/min, and an age smaller than 54 years positively correlated to the one-day-ahead CrCl.

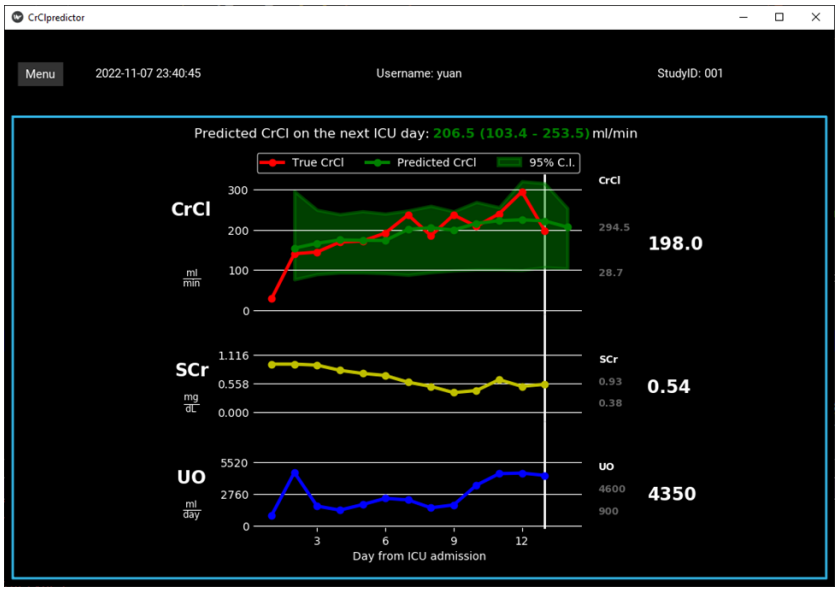


Figure 9.6. The main screen for an example of a 35-year-old man.

Figure 9.7. The prediction explanation plot for an example of a 35-year-old man.

Third, **Figure 9.8** and **Figure 9.9** show screenshots of the main screen and prediction explanation plot for a 68-year-old man. His last available [minimum-maximum] SCr was 2.36 [2.36-3.05] mg/dL, last available [minimum-maximum] UO was 1100 [300-2200] ml/day, and last available [minimum-maximum] CrCl was 23.9 [7.6-23.9] ml/min. It was predicted that the man would have a 26.5 [95% CI 15.7-38.2] ml/min CrCl on the next ICU day. Regarding the prediction explanation, the most important three features were "CrCl of the previous day", "Mean of CrCl of all past days during ICU stay", and "Mean of SCr of all past days during ICU stay". Specifically, a "CrCl of the previous day" lower than 47.5 m/min, a "Mean of CrCl of all past days during ICU stay" lower than 47.8 ml/min, and a "Mean of SCr of all past days during ICU stay" larger than 1.27 mg/dL negatively correlated to the one-day-ahead CrCl.

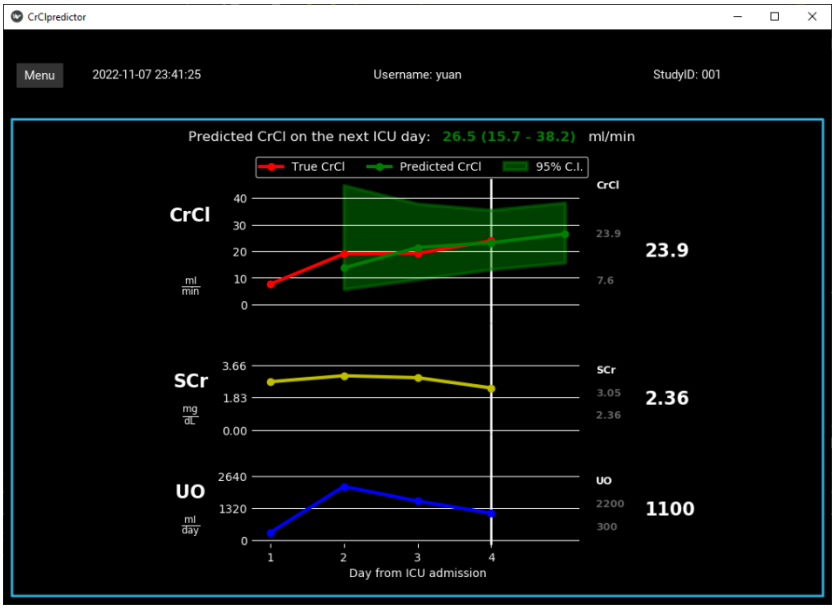


Figure 9.8. The main screen for an example of a 68-year-old man.

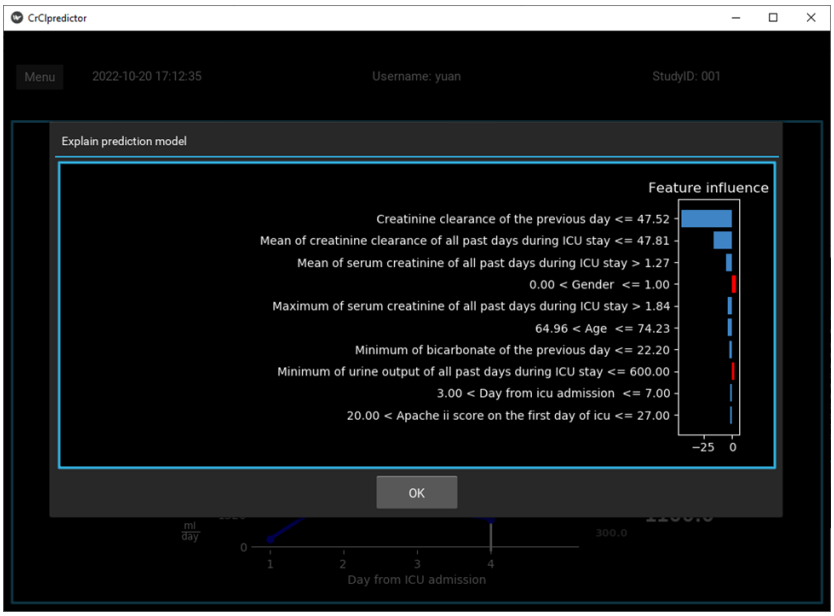


Figure 9.9. The prediction explanation plot for an example of a 68-year-old man.

Interestingly, from the above three examples, it is noteworthy that the predicted 95% confidence intervals are wider when the CrCls are larger, ranging from 23 ml/min in third example with low CrCl values, to 76 ml/min in the first example with medium CrCl values, and to 150 ml/min in the second example with high CrCl values. This finding reflects the true physiology that the kidney function instability is more pronounced at higher CrCl values (**Chapter 6**).

The median (IQR) ICU length of stay was 4 (2-9) days in the study cohort included in **Chapter 7**. For patients staying 2, 4, and 9 days, retrieving and processing the data and generating the prediction results took 2, 7, and 25 seconds on a computer with 12-core central processing unit (CPU), while the software processing time was 43 minutes for a patient that stayed in the ICU for 133 days.

9.4 Discussion

In this study, we have showed that daily predictions of short-term CrCl made by the previously developed CrCl predictor can be integrated into a graphical user interface software for individual patients on each day of their ICU stay, without any manual entry of patient data. Additionally, the prediction results were presented in a user-friendly manner, and explanations for model predictions can be generated if necessary, increasing its clinical utility.

This study has many strengths. First, all modeling methods were integrated including data collection, data preprocessing, feature engineering, model prediction, and result presentation. Consequently, this software completely avoids any manual patient data entry and requires no prior technical knowledges, additional tool, or support from the IT professionals. Second, a self-explanatory prediction explanation plot can be generated on demand for each patient on each ICU day. Individual prediction explanation is necessary for clinical applications, so that clinicians can trust the provided prediction results and use them to tailor treatment strategies for individual vulnerable patients. Furthermore, the explanation plot reflecting the real-time response to the given treatment may improve the kidney function knowledges for the treating physicians. Third, the time needed for a prediction was short, and thus it was expected that the software can run smoothly in a real-time setting. Specially, less than 30 seconds were needed to display the results for 75% of the included patients. Fourth, once the entered user credentials are validated, the software has direct access to the PDMS system. The use of data queried from the PDMS system made this software simple to be transferred to other centers with similar electronic health records settings. Additionally, despite not implemented yet, the prediction results and/or the corresponding prediction explanations can be saved to the PDMS system so that they can be directly accessible at the patient bedside or on other devices and linked to other applications (**Figure 9.10**).

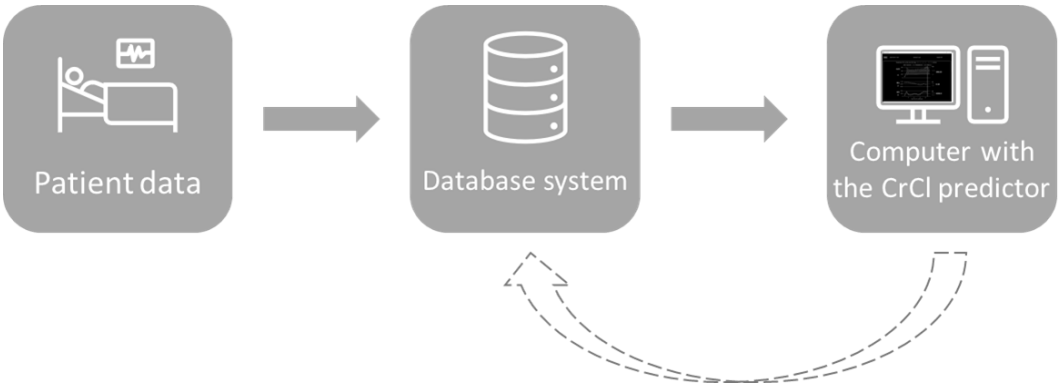


Figure 9.10. An illustration of the CrCl predictor's potential to save results back to the database.

This study has some limitations. First, by virtue of the software development and non-interventional nature of the study, it is currently unknown whether the software could aid in achieving pharmacokinetic aims, optimizing renally cleared medication dosage, and/or improving patient outcomes. Second, this is a single-center study in Belgium, but this is a pilot study to examining the feasibility of model-based software. Third, missing values were automatically imputed with the mean for continuous data and the mode for categorical data from the original study without further consideration, which may lead to an erroneous patient profile. However, this imputation method was to ensure the least efforts made by the clinicians, and future software versions may take this into account and allow users to impute individually. Fourth, while the CrCl predictor has three prediction models, only the most complete model was included in the software. Nevertheless, there was only a slight performance difference between the three models in the previous studies (**Chapter 7** and **Chapter 8**), and it was simple to remove features in the future after the integration of the most complete model. Finally, given the day-by-day prediction nature of the original CrCl predictor, the current software can only make predictions on a daily basis. Future versions of the software may consider integrating prediction models with higher temporal resolution such as predicting kidney function every 8 hours [35].

9.5 Conclusion

In this study, we developed a prototype software for kidney function management of critically ill patients. The prototype software can automatically predict CrCl for the next ICU day that is not available in the clinical environment yet and display the results and correspond explanations in a user-friendly manner. The developed software is ready to be prospectively validated in a real-time setting. Once the technical validation is finished, future interventional studies can be conducted to examine whether the use of this software in the clinical practice can help improving the kidney function management, patient care and outcome.

GRANT SUPPORT AND CONFLICT OF INTEREST

C-Y.Huang receives funding from the Taiwan-KU Leuven scholarship. G.Meyfroidt is funded by the research foundation, Flanders (FWO) as senior clinical investigator (1843123N).

All the authors declare no competing interests that are relevant to the content of this article.

ACKNOWLEDGMENT AND PERSONAL CONTRIBUTION

Study concept and design	Güiza, De Vlieger, Meyfroidt
Data acquisition	**Huang**
Statistical analysis	**Huang**
Interpretation of results	**Huang**, Güiza, De Vlieger, Meyfroidt
Drafting of the manuscript	**Huang**, Güiza, De Vlieger, Meyfroidt
Manuscript revision	**All authors**
Principal investigator	Meyfroidt

The authors are also grateful for the administration work and logistic support from the clinical research assistants (Pieter Wouters, Liese Mebis) of laboratory of intensive care medicine, KU Leuven, Belgium. The authors would like to mention that some authors (C-Y.Huang, F.Güiza, G. De Vlieger, G.Meyfroidt) participated in the development of the CrCl predictor.

BIBLIOGRAPHY

1. Huang C-Y, Güiza F, De Vlieger G, Meyfroidt G (2022) Daily fluctuations in kidney function in critically ill adults. Crit Care 26:347. https://doi.org/10.1186/s13054-022-04226-3
2. Nisula S, Kaukonen KM, Vaara ST, et al (2013) Incidence, risk factors and 90-day mortality of patients with acute kidney injury in Finnish intensive care units: The FINNAKI study. Intensive Care Med 39:420–428. https://doi.org/10.1007/s00134-012-2796-5
3. Hoste EAJJ, Bagshaw SM, Bellomo R, et al (2015) Epidemiology of acute kidney injury in critically ill patients: the multinational AKI-EPI study. Intensive Care Med 41:1411–1423. https://doi.org/10.1007/s00134-015-3934-7
4. Bilbao-Meseguer I, Rodríguez-Gascón A, Barrasa H, et al (2018) Augmented Renal Clearance in Critically Ill Patients: A Systematic Review. Clin Pharmacokinet 57:1107–1121. https://doi.org/10.1007/s40262-018-0636-7
5. Claus BOMM, Hoste EA, Colpaert K, et al (2013) Augmented renal clearance is a common finding with worse clinical outcome in critically ill patients receiving antimicrobial therapy. J Crit Care 28:695–700. https://doi.org/10.1016/j.jcrc.2013.03.003
6. He J, Yang Z-TT, Qian X, et al (2020) A higher dose of vancomycin is needed in critically ill patients with augmented renal clearance. Transl Androl Urol 9:2166–2171. https://doi.org/10.21037/tau-20-1048
7. Huttner A, Von Dach E, Renzoni A, et al (2015) Augmented renal clearance, low β-lactam concentrations and clinical outcomes in the critically ill: An observational prospective cohort study. Int J Antimicrob Agents 45:385–392. https://doi.org/10.1016/j.ijantimicag.2014.12.017
8. Udy AA, Varghese JM, Altukroni M, et al (2012) Subtherapeutic initial β-lactam concentrations in select critically ill patients: association between augmented renal clearance and low trough drug concentrations. Chest 142:30–39. https://doi.org/10.1378/chest.11-1671
9. Abdel El Naeem HEM, Abdelhamid MHE, Atteya DAM (2017) Impact of augmented renal clearance on enoxaparin therapy in critically ill patients. Egypt J Anaesth 33:113–117. https://doi.org/https://doi.org/10.1016/j.egja.2016.11.001
10. Shaikh SA, Regal RE (2017) Dosing of Enoxaparin in Renal Impairment. P T 42:245–249
11. Kellum JA, Lameire N, Aspelin P, et al (2012) Kidney disease: Improving global outcomes (KDIGO) acute kidney injury work group. KDIGO clinical practice guideline for acute kidney injury. Kidney Int Suppl 2:1–138. https://doi.org/10.1038/kisup.2012.1
12. Baxmann AC, Ahmed MS, Marques NC, et al (2008) Influence of muscle mass and physical activity on serum and urinary creatinine and serum cystatin C. Clin J Am Soc Nephrol 3:348–354. https://doi.org/10.2215/CJN.02870707
13. Le S, Allen A, Calvert J, et al (2021) Convolutional Neural Network Model for Intensive Care Unit Acute Kidney Injury Prediction. Kidney Int Reports 6:1289–1298. https://doi.org/https://doi.org/10.1016/j.ekir.2021.02.031

14. Malhotra R, Kashani KB, Macedo E, et al (2017) A risk prediction score for acute kidney injury in the intensive care unit. Nephrol Dial Transplant 32:814–822. https://doi.org/10.1093/ndt/gfx026

15. Wang Q, Tang Y, Zhou J, Qin W (2019) A prospective study of acute kidney injury in the intensive care unit: development and validation of a risk prediction model. J Transl Med 17:359. https://doi.org/10.1186/s12967-019-2118-6

16. Chiofolo C, Chbat N, Ghosh E, et al (2019) Automated Continuous Acute Kidney Injury Prediction and Surveillance: A Random Forest Model. Mayo Clin Proc 94:783–792. https://doi.org/10.1016/j.mayocp.2019.02.009

17. Sato N, Uchino E, Kojima R, et al (2021) Prediction and visualization of acute kidney injury in intensive care unit using one-dimensional convolutional neural networks based on routinely collected data. Comput Methods Programs Biomed 206:106129. https://doi.org/https://doi.org/10.1016/j.cmpb.2021.106129

18. Li Y, Yao L, Mao C, et al (2018) Early Prediction of Acute Kidney Injury in Critical Care Setting Using Clinical Notes. In: 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp 683–686

19. Gao W, Wang J, Zhou L, et al (2021) Prediction of acute kidney injury in ICU with gradient boosting decision tree algorithms. Comput Biol Med 140:105097. https://doi.org/10.1016/j.compbiomed.2021.105097

20. Chen Y, Feng F, Li M, et al (2019) Development of a risk stratification-based model for prediction of acute kidney injury in critically ill patients. Med (United States) 98:e16867. https://doi.org/10.1097/MD.0000000000016867

21. Flechet M, Güiza F, Schetz M, et al (2017) AKIpredictor, an online prognostic calculator for acute kidney injury in adult critically ill patients: development, validation and comparison to serum neutrophil gelatinase-associated lipocalin. Intensive Care Med 43:764–773. https://doi.org/10.1007/s00134-017-4678-3

22. Tang L, Ding XY, Duan LF, et al (2021) A Regression Model to Predict Augmented Renal Clearance in Critically Ill Obstetric Patients and Effects on Vancomycin Treatment. Front Pharmacol 12:1–12. https://doi.org/10.3389/fphar.2021.622948

23. Nei AM, Kashani KB, Dierkhising R, Barreto EF (2020) Predictors of Augmented Renal Clearance in a Heterogeneous ICU Population as Defined by Creatinine and Cystatin C. Nephron 55905:313–320. https://doi.org/10.1159/000507255

24. Barletta JF, Mangram AJ, Byrne M, et al (2017) Identifying augmented renal clearance in trauma patients: Validation of the Augmented Renal Clearance in Trauma Intensive Care scoring system. In: The journal of trauma and acute care surgery. United States, pp 665–671

25. Udy AA, Roberts JA, Shorr AF, et al (2013) Augmented renal clearance in septic and traumatized patients with normal plasma creatinine concentrations: Identifying at-risk patients. Crit Care 17:. https://doi.org/10.1186/cc12544

26. Gijsen M, Huang CY, Flechet M, et al (2020) Development and External Validation of an Online Clinical Prediction Model for Augmented Renal Clearance in Adult Mixed Critically Ill Patients: The Augmented Renal Clearance Predictor. Crit Care Med 48:e1260–e1268. https://doi.org/10.1097/CCM.0000000000004667

27. Carlier M, Dumoulin A, Janssen A, et al (2015) Comparison of different equations to assess glomerular filtration in critically ill patients. Intensive Care Med 41:427–435. https://doi.org/10.1007/s00134-014-3641-9

28. Lê Cook B, Manning WG (2013) Thinking beyond the mean: a practical guide for using quantile regression methods for health services research. Shanghai Arch psychiatry 25:55–59. https://doi.org/10.3969/j.issn.1002-0829.2013.01.011

29. Random Forests. https://www.stat.berkeley.edu/~breiman/RandomForests/. Accessed 19 May 2022

30. Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems. pp 4766–4775

31. Ribeiro MT, Singh S, Guestrin C (2016) "Why Should {I} Trust You?": Explaining the Predictions of Any Classifier. In: Proceedings of the 22nd {ACM} {SIGKDD} International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016. pp 1135–1144

32. Cava W La, Bauer C, Moore JH, Pendergrass SA (2019) Interpretation of machine learning predictions for patient outcomes in electronic health records. AMIA . Annu Symp proceedings AMIA Symp 2019:572–581

33. Lundberg SM, Nair B, Vavilala MS, et al (2018) Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat Biomed Eng 2:749–760. https://doi.org/10.1038/s41551-018-0304-0

34. Janssen A, Hoogendoorn M, Cnossen MH, et al (2022) Application of SHAP values for inferring the optimal functional form of covariates in pharmacokinetic modeling. CPT Pharmacometrics \& Syst Pharmacol 11:1100–1110. https://doi.org/https://doi.org/10.1002/psp4.12828

35. Cherry RA, Eachempati SR, Hydo L, Barie PS (2002) Accuracy of short-duration creatinine clearance determinations in predicting 24-hour creatinine clearance in critically ill and injured patients. J Trauma 53:267–271. https://doi.org/10.1097/00005373-200208000-00013

10
Discussion

In critically ill patients, both increased and decreased renal clearance occur frequently, and kidney function may vary during the course of critical illness. Acute kidney injury (AKI) is a highly prevalent syndrome occurring in many conditions that cause severe critical illness, and has an association with worse short- and long-term clinical outcomes. Fluctuations in renal clearance influence the plasma levels and clearance of renally excreted drugs, leading to unfavorable consequences such as drug toxicity and treatment failure. Predicting these fluctuations in kidney function could allow for more personalized drug dosage, and less treatment failure, potentially leading to better clinical outcomes. In addition, even though the prevention and management of AKI are mainly supportive, early predictions of a reducing or fluctuating kidney function can stratify patients according to their risk, and may aid in developing targeted therapies to prevent or mitigate the course of kidney injury.

The focus of current machine-learning models in this field has been the prediction of consensus-based classifications of changes in renal function, such as AKI or augmented renal clearance (ARC). Some of these models have demonstrated good predictive performance and outperformed the physicians in external validation, but still could benefit from additional external validations in independent datasets before they can be applied to different clinical settings. Furthermore, since the kidney function is continuous, predicting the complete spectrum of kidney function is more in accordance with clinical and physiological reality.

In this thesis, we applied advanced data analytical techniques and machine-learning algorithms to gain insights into kidney function in critically ill patients. The first objective focuses on AKI, and we validated a machine-learning model, the AKIpredictor, for early prediction of AKI within the first week of ICU admission. Additionally, we developed and validated prediction models for AKI recovery at hospital discharge in patients with AKI stage 3 during their ICU stay. In the second objective, we validated a prediction model for the onset of ARC on the next ICU day for coronavirus disease 19 (COVID-19) pneumonia patients admitted to critical care in the University Hospitals Leuven from February 2020 to January 2021. In the third objective, we first investigated the daily kidney function instability in critically ill patients. In addition, we developed and validated models for daily prediction of short-term measured creatinine clearance (CrCl). Furthermore, we externally validated the CrCl prediction models and compared their performance against ICU physicians in a prospective observational study. Finally, we developed a prototype software to incorporate the developed CrCl prediction models and to visualize the prediction results along with the prediction explanations in a user-friendly interface.

10.1 Prediction of acute kidney injury development and recovery

10.1.1 Main findings

According to epidemiologic research, 40–60% of critically ill patients have AKI depending on the use of AKI definition criteria and study cohort [1, 2]. AKI has been shown associated with ICU mortality greater than 50% [3], longer length of stay [4, 5], and higher financial cost [4, 6]. The current consensus to define AKI is based on the Kidney Disease: Improving Global Outcomes (KDIGO) AKI criteria proposed in 2012, which categorize AKI by serum creatinine (SCr) and urine output (UO).

Given the high prevalence and strong association of AKI with adverse events, early prediction of AKI may be beneficial. As a result, many prediction models have been developed for critically ill patients [7–15]. Specifically, Le et al. developed a prediction model to predict severe AKI 48 hours in advance using convolutional neural networks, which showed superior performance than the XGBoost-based model and sequential organ failure assessment (SOFA) scoring system [7]. Additionally, Gao et al. developed multiple prediction models using different machine-learning algorithms and time before the onset of AKI (24, 48, and 72 hours), and they explained the feature impact by the SHAP analysis [13]. The ensemble model was demonstrated with the best AUROC of 0.92 to predict AKI 24 hours in advance. Furthermore, Sato et al. proposed a prediction model using one-dimensional convolutional neural networks, which achieved higher performance compared to the other models, with AUROCs of 0.74 and 0.84 for stage-1 and stage-2 AKI respectively [11]. Finally, based on a large multi-center EPaNIC randomized controlled trial [16], Flechet et al. developed a series of prediction models named "AKIpredictor" to predict the presence of AKI during the first week of ICU stay [15] and made it accessible as a web-based application [17]. Four models were developed, each utilizing a unique set of features as they became available at different stages along the clinical course: prior to admission, upon admission, the first morning following admission, and after 24 hours. The models were able to predict any stage (1–2–3) and severe stages of AKI (2–3). The AKIpredictor's accuracy was tested in a validation cohort, where it excelled serum neutrophil gelatinase-associated lipocalin (NGAL), the most studied AKI biomarker at that time. Following that, Flechet et al. conducted another prospective observational study to compare the AKIpredictor with ICU physicians' predictions, where AKIpredictor demonstrated comparable discrimination and a higher net benefit for severe AKI (2–3) predictions [18]. Although the AKIpredictor represented a significant advancement in the early prediction of AKI, additional external validations were needed to evaluate its generalizability in various patient populations before it could be applied in clinical practice.

In **Chapter 3**, we externally validated the AKIpredictor of Flechet et al. [15] on a large heterogeneous cohort from the University Hospitals Leuven included in the M@tric database [19], with high-quality and intricately linked data from every adult patient admitted to the ICU from 2013 to 2018. In this external validation study, the AKIpredictor showed comparable predictive performance to that observed in the original model development research. This finding confirmed the AKIpredictor models' robustness despite the fact that the original database from which the models were developed was ten years old, a period during which the clinical setting and medical procedures underwent significant change.

AKI recovery prediction may also be helpful to guide clinical practice as persisting AKI is associated with higher short- and long-term mortality [20], chronic kidney disease [21], end-stage kidney disease [22], and need for kidney replacement therapy (KRT). However, the development of AKI recovery prediction models remains challenging, partially attributed to the differences in baseline SCr definitions, patient populations being studied, recovery evaluation time points, and definitions for recovery [23]. Particularly, the prediction models developed by Itenov et al. demonstrated fair AKI recovery predictive performance using the Cox regression [24], which may result from the pre-specified recovery timepoint of 28 days and predictions made on ICU admission. Additionally, Lee et al. developed classification and regression tree (CART) and logistic regression models for predicting AKI recovery [25], but only poor discrimination was demonstrated without external validation. Hence, in **Chapter 4**, with an aim to identify patients with high chance of AKI recovery, we developed and validated machine-learning prediction models for AKI recovery at hospital discharge in critically ill patients with AKI

stage 3 during their ICU stay using routinely collected clinical data. Models were developed and validated on a development cohort and a matched validation cohort from the critically ill adults admitted to ICUs of the University Hospitals Leuven, included in the EPaNIC randomized controlled trial (RCT) database. Two models were developed: one for complete recovery prediction and the other one for complete or partial recovery prediction. Complete recovery was defined as the absence of any stage of AKI and being alive without KRT at hospital discharge, and partial recovery was defined as AKI stage 1 or 2 and being alive without KRT at hospital discharge. Model performance was compared with plasma NGAL measured on first AKI-3 day (NGAL_AKI3) and a reference model that only based on age.

Five features were selected for each task based on the identified associations with kidney function recovery from the literatures, machine-learning techniques (Lasso, correlation-based feature selection method, and permutation feature importance), and discussion with two experienced ICU physicians. In the validation cohort, the developed AKI recovery prediction models only showed similar poor discrimination as the NGAL_AKI3 for the general ICU population, which were worse than the reference models that only based on age. For cardiac surgery patients, the developed models had better performance over NGAL_AKI3 and the reference models that only based on age.

10.1.2 Current impact of the research and future perspectives

The study presented in **Chapter 3** shows the necessity of the AKIpredictor, given the high prevalence of AKI in this mixed population of critically ill adults (10.97%), and the significantly higher mortality and ICU length of stay in patients with AKI (18% vs. 4%, $p<0.01$, and 6 days vs.3 days, $p<0.01$). Besides, **Chapter 3** confirms the generalizability of the AKIpredictor, implying that its potential as a promising tool to identify AKI patients and to prevent potential kidney damage at an early stage. Further interventional studies are needed to evaluate whether the use of AKIpredictor in clinical practice can improve outcome.

It is noteworthy that this external validation study performed in **Chapter 3** is retrospective and single-center, while a prospective international multicenter study is required to fully reflect the clinical usefulness of the AKIpredictor. Importantly, since the AKIpredictor was developed based on a multicenter database from Belgium, it is expected that the AKIpredictor may not achieve the same performance in patients from other countries due to the difference in genetic and environmental factors, both of which have an association with the kidney function. After a thorough understanding of which patient populations and under what conditions the AKIpredictor may perform worse, the addition of a correction term or recalibration of the AKIpredictor may be followed to make the AKIpredictor more applicable to the studied patient population.

In **Chapter 3**, we exclusively focused on critically ill adults with any stage of AKI, since care should be taken to prevent any stage of AKI, given the strong association of all stages of AKI with unfavorable long-term outcomes [2]. Besides, since the AKIpredictor models on the website application predict the risk of all stages of AKI, the investigation of all AKI stages enabled a better benchmarking with other studies. However, the AKIpredictor was also developed to predict severe stages of AKI, and patients with severe stages of AKI may benefit more from a focused therapy. Future studies may examine the generalizability of the AKIpredictor on predicting the risk of severe AKI stages.

Only KDIGO SCr criteria were used for AKI definition in **Chapter 3**, as the AKIpredictor was originally developed with SCr criteria only, and adhering to the original study's methodology enabled better interpretation and comparison. Furthermore, in a prospective study where UO criteria were added to SCr criteria, the AKIpredictor showed worse discrimination (median area under the receiver operating characteristic curve (AUROC): 0.78 vs. 0.76, 0.94 vs. 0.87, and 0.93 vs. 0.85 upon admission, on day 1, and after 24 hours respectively) [18]. Finally, hourly registration of the UO is often difficult in clinical setting. However, the UO criteria were demonstrated to be equally essential to the SCr criteria. In particular, Koeze et al. found that the addition of UO criteria to the SCr criteria may double the prevalence of AKI in critically ill patients and identify individuals with AKI eleven hours earlier [26]. Hence the generalizability of the AKIpredictor with inclusion of UO criteria remains to be investigated.

The relatively poor performance of the models in **Chapter 4** suggested that it was difficult to predict AKI recovery upon hospital discharge at an early stage, which may be partially explained by the following factors. First, we exclusively focused on critically ill patients with AKI-3 during their ICU stay. Despite the strong associations with adverse outcomes and the high clinical relevance, only a small number of AKI-3 patients were available, and only a small proportion of these patients actually recovered within the timeframe of the follow-up. Additionally, with the small dataset and low recovery prevalence, to reduce the risk of overfitting, we were limited to using a small number of features and relatively simple machine-learning algorithm (i.e., Lasso) for each prediction task, which may not suffice to capture the complex kidney function recovery signals. To address the imbalanced and limited dataset issue, one technique that synthesizes new examples from the existing ones may be considered in future studies. In particular, as a type of data augmentation, synthetic minority oversampling method (SMOTE) has demonstrated its value in enhancing the prediction performance [27, 28]. However, care should be taken when using this technique, since it has the potential drawbacks of introducing false information due to the generation of fake data and the potential for overfitting due to the strong similarity between produced data. Once more patient data are available, we suggest future researchers to consider employing more sophisticated machine-learning algorithms, such as tree-based algorithms or deep learning techniques, which are expected to give better performance for the underlying complicated research question.

Second, in a systematic review where Ohnuma et al. investigated the mortality prediction models for patients with AKI [29], they found out that the majority of the AKI mortality prediction models exhibited poor discrimination with an AUROC below 0.7 in the external validation. This absence of reliable and generalizable prediction models for AKI outcomes may suggest that the information available until the time of AKI diagnosis may not be adequate to anticipate patients' outcomes at a distant time point in the future. Additionally, it is anticipated that events and patient' conditions that occur after an AKI diagnosis may be more important for AKI recovery than those that occur before the diagnosis due to the effectiveness of treatments given after the presence of AKI, which may weaken the initial strong associations of AKI-3 with adverse outcomes. Therefore, future studies may consider combining data after the onset of AKI-3 such as the treatment and corresponding response, to enhance prediction accuracy. However, this potential increase in performance comes at the cost of a smaller window for clinical intervention, which lowers its clinical value.

10.2 Prediction of augmented renal clearance development

10.2.1 Main findings

Augmented renal clearance (ARC) occurs in 20-65% of the critically ill patients [30] and has been linked to reduced exposure to commonly administered medications such as beta-lactams, vancomycin, and anticoagulants [31, 32] and more treatment failure [33]. Although there is no universal agreement on how to define ARC, it is typically defined as a measured urinary creatinine clearance (CrCl) larger than 130 ml/min/1.73m^2 based on a 24-hour urine collection. However, kidney function is known to vary rapidly in critically ill patients, and by the time the 24-hour urine collection is finished, the kidney function may have already changed.

Given the significance and necessity of ARC prediction, many prediction models have been developed to predict the onset of ARC [34–38], the majority of which was based on a small and specific subgroup of critically ill patients and was not externally validated. Specifically, the prediction model developed by Udy et al. is a point-based scoring system to predict ARC based on the adjusted odds ratios [38]. The model demonstrated decent predictive performance with an AUROC of 0.89, but it may not be generalizable as only 71 septic and traumatized critically ill patients were included. Another scoring system for ARC prediction proposed by Barletta et al. was also point-based [37]. Despite the decent discrimination with an AUROC of 0.81, model development was based on 133 critically ill trauma patients, whose external validity might be questionable in other patient populations. Therefore, based on a large multi-center M@tric database [19], Gijsen et al. developed and externally validated a prediction model named "ARC predictor" to predict the onset of ARC on the next ICU day for critically ill patients [34]. Nevertheless, despite the excellent performance in the original study, the generalizability of the ARC predictor still has to be investigated in more external validations before it can be applied to clinical practice.

Hence, in **Chapter 5**, the ARC predictor was externally validated using the adult COVID-19 pneumonia patients admitted to critical care in the University Hospitals Leuven from February 2020 to January 2021. The similar ARC predictor predictive performance demonstrated in this study as compared to the original study served as confirmation of the ARC predictor's robustness. Results from **Chapter 5** demonstrated the ARC predictor's generalizability; performance was unaffected by the different patient characteristics between the critically ill COVID-19 dataset and the original ARC predictor development cohort.

10.2.2 Current impact of the research and future perspectives

The findings in **Chapter 5** first confirmed the high prevalence of ARC, showing that ARC was found on at least one ICU day in 57 (47.5%) patients, corresponding to 246 (23.1%) patient-days. As a result, ARC should be considered carefully when adjusting the dosage of renally excreted drugs. Importantly, although only slightly worse discrimination was observed compared to the original study, the calibration plot indicated that the ARC predictor underestimated the risk of ARC. This may be partly attributed to the unidentified increased risk of ARC resulting from the systemic inflammatory response syndrome (SIRS) that COVID-19 patients might experience, which were originally associated with the two ARC predictor features: trauma-related and cardiac-related diagnosis at admission. The results suggested that the ARC predictor requires an adjustment to the considered features to more appropriately reflect the influence of SIRS.

The ARC predictor's robustness was confirmed through the excellent predictive performance in the external validation. However, before using the ARC predictor for drug dosage optimization in clinical practice,

a prospective multicenter validation is still warranted to ensure the ARC predictor's generalizability in cohorts with different patient characteristics. In addition, whether the ARC predictor can help optimizing the drug dosing and improving the patient outcome remains to be investigated in large RCT, where several steps are necessary, including real-time testing in different clinical settings, implementation into clinical workflow with examination of model performance, safety, and effect on patients' outcome.

Currently, the ARC predictor is available online to facilitate external evaluation across different centers (**Figure 10.1**) [39]. The website has been assessed more than 910 times by more than 720 users from 60 countries since October 2021. The majority of the users were from United States, China, Belgium, Netherlands, and United Kingdom. Despite the convenient access as a website, patient data still need to be manually entered. To maximize the utility of the ARC predictor, it is ideal to automate every step of the process, including data extraction from the electronic health record, data processing of raw data, feature engineering, data entry into the prediction model, and results loading into a database or pharmacokinetic model. Knowledge in intensive care medicine, clinical information technology, pharmacometrics, and data science are necessary to complete this challenging task. As a consequence, close collaboration between intensivists, specialists in clinical databases, pharmacometricians, and data scientists would be required.

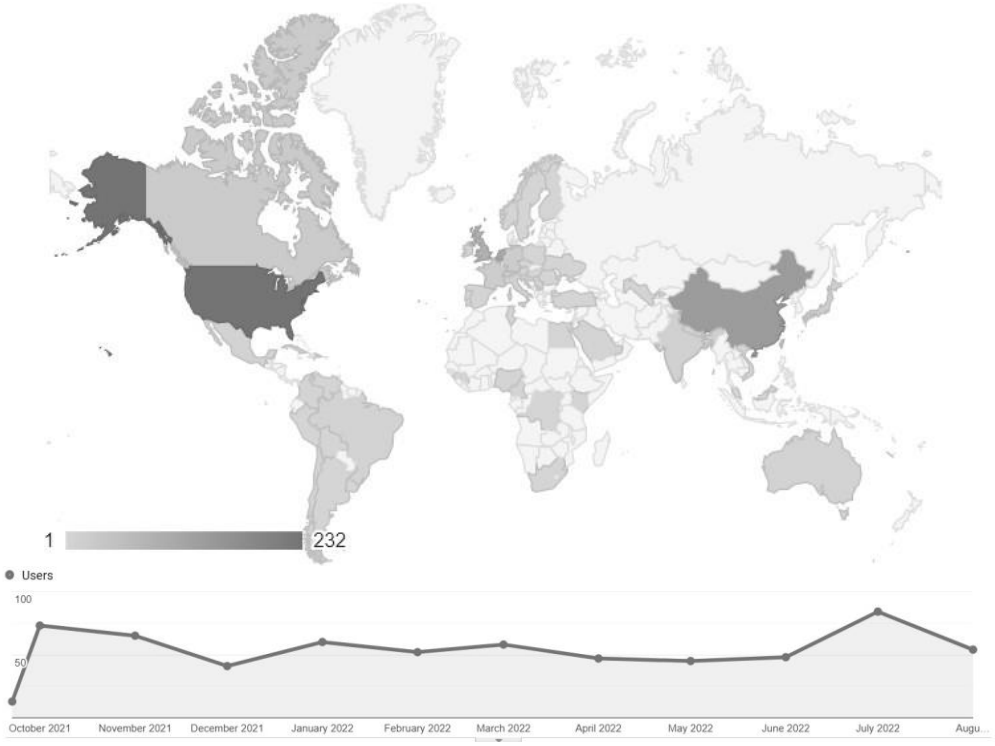


Figure 10.1 User statistics of the ARC predictor website from September 2021 to August 2022 Geographic reports of the users (upper panel) and line chart of the monthly usage (lower panel)

10.3 Prediction of creatinine clearance

10.3.1 Main findings

The kidney function changes rapidly in critically ill patients, with either sudden increases or decreases in renal clearance. Nevertheless, the actual incidence of kidney function instability has never been described systematically in critically ill adults.

Hence, in **Chapter 6**, we examined the daily fluctuations of kidney function, as defined by the daily differences in CrCl, in critically ill adults admitted to ICUs of the University Hospitals Leuven, included in the EPaNIC RCT database. Our results showed that in 35-40% of days, critically ill adults may experience potentially clinically relevant alterations in kidney function on a daily basis. This instability was more noticeable when the CrCl of the current day was higher and generally happened in the first week after ICU admission. The findings from **Chapter 6** suggest that current clinical practice assuming the CrCl remains stable may put patients at risk for receiving inappropriate dosages for renally cleared drugs such as commonly used antibiotics such as beta-lactams and vancomycin, and may increase the risk of drug toxicity and/or treatment failure.

Because of the high prevalence of clinically relevant kidney function changes and strong associations of adverse outcomes of AKI and ARC in critically ill patients, kidney function prediction is important. However, the majority of research about kidney function prediction focused on predicting AKI and ARC. Despite the success of these models (e.g., AKIpredictor [15, 18, 40] and ARC predictor [34]), the kidney function is in fact continuous instead of categorized. Predicting the complete kidney function spectrum is more in accordance with clinical and physiological reality. However, to the best of our knowledge, there is currently no prediction model for continuous kidney function.

Given the necessity and importance of continuous kidney function prediction, in **Chapter 7**, we developed and validated a set of machine-learning models for daily prediction of CrCl using routinely collected clinical data, named “CrCl predictor”. Three models were developed on critically ill adults admitted to ICUs of the University Hospitals Leuven, included in the EPaNIC RCT database, progressively including features with increasing data resolution: the Core model using only admission data and daily routine laboratory results, the Core+BGA model that adds to the above, blood gas analysis data, and the Core+BGA+Monitoring model that adds to the above, monitoring data (heart rate, mean arterial blood pressure, and respiratory rate). The CrCl predictor was able to predict CrCl of the next day and average CrCl of the next two days. Model performance was compared against the reference reflecting the current clinical practice: assuming the same CrCl value on the day of prediction. The results showed that the CrCl predictor demonstrated good performance when tested on a large external validation dataset of 20930 patients of the University Hospitals Leuven who were included in the large multicenter M@tric database between 2013 and 2018.

Before implementation of the CrCl predictor in the clinical practice, there are still some necessary steps, one of which is to verify the model’s generalizability by external validation in more studies with different clinical settings [41]. Additionally, the clinical usefulness of the model needs to be examined. It is unclear whether the model can perform at least as well as ICU physicians and whether it can provide complementary information to physicians with different seniority levels. The comparison with healthcare professionals is essential in order to investigate the potential added values as a clinical decision support tool. For instance, the AKIpredictor

developed by Flechet et al. was prospectively validated and compared with ICU physicians [18], where the AKIpredictor was demonstrated to have comparable discrimination as physicians and thus had the potential added value of identifying patients at risk of AKI, especially when the physicians were less confident.

Therefore, in **Chapter 8**, we conducted a prospective observational study to compare the predictive performance of the CrClpredictor and ICU physicians for daily prediction of CrCl in 197 critically ill adults admitted to surgical ICUs of the University Hospitals Leuven between January 2022 and April 2022. The observed comparable prediction errors to the original development study confirmed the external validity of the CrCl predictor. Additionally, it was shown that the ICU physicians had larger prediction errors than the CrCl predictor. Specifically, the CrCl predictor had slightly smaller prediction errors than the staff members and senior residents, and much lower prediction errors than the junior residents. The absence of statistically significant differences indicated that the CrCl predictor could perform at least as well as the ICU physicians. These results suggested the potential added value of the CrCl predictor to physicians' predictions, especially for junior residents. Future studies are still needed to examine its potential as an efficient and scalable screening tool to identify patients with high risk of changing kidney function.

Despite the good performance in the external validations of the previous two studies, it remains to be investigated whether the CrCl predictor can help improve patient care and outcome. To allow for evaluation of the CrCl predictor in a real clinical setting in future studies, the models have to be directly available at the patient bedside. Therefore, in **Chapter 9**, we developed a prototype software that integrated the developed CrCl predictor model and visualized the prediction results along with prediction explanations in a user-friendly manner. The prototype software successfully queried the data and calculated the predictions for the patients included in **Chapter 7**. Specifically, the software displayed the predicted one-day-ahead CrCl made by the CrCl predictor along with a 95% confidence interval (CI). To provide a complete understanding of the patient's kidney function trajectory, the temporal evolution of SCr, UO, and actual CrCl were also integrated. Finally, an individual prediction explanation plot corresponding to the prediction result can be generated on demand for each patient on each day of their ICU stay, for the promotion of prediction model transparency.

10.3.2 Current impact of the research and future perspectives

The findings presented in **Chapter 6** highlighted the risk of assuming that kidney function remains the same on the following ICU day. However, as a first step to future studies, the study did not examine the clinical implications of high CrCl fluctuations. Therefore, whether high CrCl fluctuations correspond to a less adequate drug dosing regimen for renally cleared drugs such as vancomycin and β-lactam anti-microbials, and thus a poorer patient outcome remains to be investigated in future studies. If the association between high CrCl fluctuations and adverse consequences is confirmed, an additional covariate reflecting kidney function stableness may be incorporated in pharmacokinetic (PK) models for a potentially more tailored drug dosing regimen [42].

Personalized medicine is an ambitious aim, but the work in **Chapter 7** represents a huge step toward realizing it. The prediction of the entire kidney function spectrum shifted the focus from the prevention of the most extreme kidney function conditions (e.g., AKI and ARC) to the daily optimization of kidney function management in all critically ill patients. From the clinical viewpoint, the predictions made by the CrCl predictor can be added as a covariate in the PK model to help optimizing the drug exposure. From the research perspective, the CrCl predictor could help identifying patients with high kidney function instability who might benefit from

innovative interventions, providing rigorous study population definition and reducing the cost and administration work [43]. After the identification, the factors associated with large fluctuations in renal clearance can also be determined and further studied.

Although the CrCl predictor has demonstrated good performance not only in a retrospective comparison with a reference reflecting the current clinical practice in the patients from a large dataset [19] (**Chapter 7**) but also in a prospective comparison against ICU physicians (**Chapter 8**), more large multicenter prospective studies are still warranted to examine its validity. More external validations of the CrCl predictor are important, regardless of the setting. In the same setting, clinical environment and measurement procedures may evolve considerably over time, which may inevitably worsen the accuracy of the CrCl predictor [44], so repeated evaluation in the same setting is suggested before application to new settings [45]. When the CrCl predictor is tested in different settings, more decreases in generalizability are expected, and model updating may be considered to ensure a better fit in the external validation population [46]. Although some researchers advocated model updating, others regarded updated models as developing new models, and thus the validity of updated models has to be confirmed in internal and external validation [47].

In addition to more external validations, as abovementioned, future studies involving the CrCl predictor may use the predicted CrCl as a covariate in PK models to facilitate dosage adjustment of renally cleared drugs. Since the data used to develop the PK model are mostly small and sparse, in order to have sufficient measurements to build a reliable PK model, the chosen drug should be commonly administrated renally cleared medication such as beta-lactams and vancomycin. For instance, Kim et al. integrated time-varying kidney function estimation as a covariate in development of population PK model for vancomycin prediction [42]. They found out that the inclusion of covariates that explained changes in kidney function improved the PK model structure and provided a better explanation for vancomycin pharmacokinetics. Since the kidney function estimation methods have been demonstrated with limited capacity to accurately reflect kidney function in critically ill patients [48–52], it is expected that the use of predicted CrCl made by the CrCl predictor as a covariate in PK model will offer better vancomycin prediction results, but it remains to be investigated in future studies. Once the added value of CrCl predictor in PK model target attainment is confirmed, a large multicenter RCT can be set up to see whether the help of the CrCl predictor in target attainment can result in a better clinical outcome.

The performed prospective observational study in **Chapter 8** to compare model performance to standard of care is an indispensable step before workflow implementation [53]. Without the comparison, we would not have discovered the potentially added clinical value of the CrCl predictor to the junior physicians' predictions, so it is strongly recommended to test the promising models in a prospective study. However, according to a recent system review [54], only 1.8% of the clinical prediction model studies have prospectively tested their models. It may be partly attributed to the complexity of massive amounts of high-dimensional data, the lack of external validations and high-quality clinical data, and the unclear understanding of clinical context, all of which hinder the further extension of potential models. Therefore, the author would advocate more close collaborations between the data scientists and ICU physicians in the future. Only by combining the knowledges from both sides will we be able to overcome these barriers and bring prediction models into clinical practice to examine whether they aid physicians in their decision making.

In **Chapter 9**, as a proof of concept, the most complete model (i.e., the Core+BGA+Monitoring model) of the CrCl predictor was integrated into the prototype software, since it was considered simpler to first integrate the most complete model and change to simpler models with less features in a later phase if necessary. Despite the integration of the most complete model, the software only took less than 30 seconds to calculate the predictions for the patients staying less than 9 days in the ICU, consisting of 75% of the study cohort in **Chapter 7**. Nonetheless, only small differences in predictive performance with other models were found in the previous studies (**Chapter 7** and **Chapter 8**), so future versions of the software may consider adding the simpler models (i.e., Core and Core+BGA models), which may further improve the already efficient real-time execution, reduce the need of powerful computers, and increase the clinical applicability in other ICUs.

In **chapter 9**, predicted CIs were generated for each prediction result, and it was found that the predicted CIs were wider when the CrCls were larger, reflecting the true physiology that kidney function instability is more pronounced at higher CrCl values (**Chapter 6**) [55]. The indication of predicted CIs is crucial, since it provides physicians extra information about the uncertainty level of the CrCl predictor. If the CIs are wide which represents that the CrCl predictor is less certain, the physicians need to be more careful when using the predictions, and it is suggested to examine the corresponding prediction explanation plots to ensure that the predictions are reasonable and trustworthy. Given that the model uncertainty level and model interpretability might influence the trust of physicians on the predictions, the design of the CIs and prediction explanation plots are considered necessary, and it is expected that this extra information would become routine and potentially mandatory in clinical prediction models of the future.

In the studies presented in **Chapter 6-9**, we used measured CrCl to indicate the kidney function, since it has been described that although CrCl measurements overestimate the glomerular filtration rate (GFR), they are still reliable (AUROC: 0.93-0.98 depending on the range of GFR as compared with the golden standard inulin clearance) [56]. In addition, measured CrCl is a fast and cost-effective assay, which are important characteristics as the examination of kidney function in these studies was performed on a daily basis. Finally, it was stated in the KDIGO AKI guidelines that CrCl is still the best clinical surrogate marker of kidney function, and patients with AKI should have it checked whenever possible [57]. However, the kidney function fluctuations may be partially explained by the inadequacy of SCr and its derivatives, as it is known to be inadequate to assess the GFR during critical illness [58]. Therefore, future studies may be conducted to assess to what extent the measured CrCl reflects the actual kidney function. Afterwards, a correction term may be applied to the CrCl (such as kinetic estimated GFR that may better describe kidney function than the MDRD equation in acute setting [59, 60]), and the CrCl predictor may be relearned.

In **Chapter 6-9**, kidney function was evaluated on a daily basis, since the use of 24-hour urine collection for measured CrCl provides the most reliable understanding of patient's kidney function. A higher temporal resolution may be considered in future studies, since a minimum urine collection period of 8 hours was suggested [61–63]. The adoption of short-duration CrCls may promote quicker decision-making processes and more timely drug dosing regimens. Furthermore, it gives rise to larger sample sizes within the same study period, attesting to more robust findings. However, care should be taken because the use of 8-hour collection may result in less accurate indication of kidney function, given the variability in urine output. And it might be infeasible to accurately collect urine output every 8 hours, which causes the risk of inducing more missing values and lowering the data quality and reliability.

In this thesis, we specifically focused on the development, validation, and application of prediction models, without inclusion of the biomarker tests, except for the comparison with the most studied AKI biomarker NGAL in **Chapter 4**. However, despite the efficiency and low cost of prediction models, we do not contend that prediction models are superior to the biomarker tests. Instead, a combination of different assays, including but certainly not limited to the prediction results and biomarker assessment, could be the next step toward personalized medicine.

Importantly, in this thesis, we focused on model performance using quantitative objective evaluation metrics such as AUROC for discrimination of classifiers and RMSEs for accuracy of regressors. Further steps are needed to evaluate whether the use of these prediction models in clinical practice is associated with patient outcome.

10.4 General conclusions

In this thesis, we applied machine learning algorithms and data analytic techniques to the routinely collected clinical data, to gain insights into the kidney function assessment in critically ill patients. First, we validated a machine-learning model for the early identification of AKI, the AKIpredictor, in critically ill adults. The identified robustness of the AKIpredictor highlighted its potential as a promising tool to identify AKI patients at an early stage. Additionally, we developed and validated prediction models for AKI recovery at hospital discharge in critically ill patients with AKI stage 3 during their ICU stay. The poor predictive performance suggested that it was difficult to predict non-reversible AKI at an early stage. Second, we demonstrated that a previously developed prediction model for ARC, the ARC predictor, had robust predictive performance when externally validated on COVID-19 pneumonia patients admitted to critical care in the University Hospitals Leuven from February 2020 to January 2021. Third, we investigated the incidence and degree of fluctuations of daily kidney function in critically ill adults, which has been known to change rapidly but without systematic description. Moreover, we developed and validated models for daily prediction of measured CrCl, the CrCl predictor, in critically ill patients. The CrCl predictor demonstrated good performance when evaluated on a large external validation dataset and performed as least equally well as the treating physicians. Finally, to increase the usability in the clinical practice, we developed a prototype software that integrated the CrCl predictor and visualized the prediction results along with prediction explanations in a user-friendly interface.

BIBLIOGRAPHY

1. Nisula S, Kaukonen KM, Vaara ST, et al (2013) Incidence, risk factors and 90-day mortality of patients with acute kidney injury in Finnish intensive care units: The FINNAKI study. Intensive Care Med 39:420–428. https://doi.org/10.1007/s00134-012-2796-5
2. Hoste EAJJ, Bagshaw SM, Bellomo R, et al (2015) Epidemiology of acute kidney injury in critically ill patients: the multinational AKI-EPI study. Intensive Care Med 41:1411–1423. https://doi.org/10.1007/s00134-015-3934-7
3. Case J, Khan S, Khalid R, Khan A (2013) Epidemiology of acute kidney injury in the intensive care unit. Crit Care Res Pract 2013:479730. https://doi.org/10.1155/2013/479730
4. Chertow GM, Burdick E, Honour M, et al (2005) Acute kidney injury, mortality, length of stay, and costs in hospitalized patients. J Am Soc Nephrol 16:3365–3370. https://doi.org/10.1681/ASN.2004090740
5. Federspiel CK, Itenov TS, Mehta K, et al (2018) Duration of acute kidney injury in critically ill patients. Ann Intensive Care 8:30. https://doi.org/10.1186/s13613-018-0374-x
6. Silver SA, Chertow GM (2017) The Economic Consequences of Acute Kidney Injury. Nephron 137:297–301
7. Le S, Allen A, Calvert J, et al (2021) Convolutional Neural Network Model for Intensive Care Unit Acute Kidney Injury Prediction. Kidney Int Reports 6:1289–1298. https://doi.org/https://doi.org/10.1016/j.ekir.2021.02.031
8. Malhotra R, Kashani KB, Macedo E, et al (2017) A risk prediction score for acute kidney injury in the intensive care unit. Nephrol Dial Transplant 32:814–822. https://doi.org/10.1093/ndt/gfx026
9. Wang Q, Tang Y, Zhou J, Qin W (2019) A prospective study of acute kidney injury in the intensive care unit: development and validation of a risk prediction model. J Transl Med 17:359. https://doi.org/10.1186/s12967-019-2118-6
10. Chiofolo C, Chbat N, Ghosh E, et al (2019) Automated Continuous Acute Kidney Injury Prediction and Surveillance: A Random Forest Model. Mayo Clin Proc 94:783–792. https://doi.org/10.1016/j.mayocp.2019.02.009
11. Sato N, Uchino E, Kojima R, et al (2021) Prediction and visualization of acute kidney injury in intensive care unit using one-dimensional convolutional neural networks based on routinely collected data. Comput Methods Programs Biomed 206:106129. https://doi.org/https://doi.org/10.1016/j.cmpb.2021.106129
12. Li Y, Yao L, Mao C, et al (2018) Early Prediction of Acute Kidney Injury in Critical Care Setting Using Clinical Notes. In: 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp 683–686
13. Gao W, Wang J, Zhou L, et al (2021) Prediction of acute kidney injury in ICU with gradient boosting decision tree algorithms. Comput Biol Med 140:105097. https://doi.org/10.1016/j.compbiomed.2021.105097

14. Chen Y, Feng F, Li M, et al (2019) Development of a risk stratification-based model for prediction of acute kidney injury in critically ill patients. Med (United States) 98:e16867. https://doi.org/10.1097/MD.0000000000016867

15. Flechet M, Güiza F, Schetz M, et al (2017) AKIpredictor, an online prognostic calculator for acute kidney injury in adult critically ill patients: development, validation and comparison to serum neutrophil gelatinase-associated lipocalin. Intensive Care Med 43:764–773. https://doi.org/10.1007/s00134-017-4678-3

16. Casaer MP, Mesotten D, Hermans G, et al (2011) Early versus Late Parenteral Nutrition in Critically Ill Adults. N Engl J Med 365:506–517. https://doi.org/10.1056/NEJMoa1102662

17. AKI prediction. https://www.akipredictor.com/en/. Accessed 10 Feb 2022

18. Flechet M, Falini S, Bonetti C, et al (2019) Machine learning versus physicians' prediction of acute kidney injury in critically ill adults: A prospective evaluation of the AKIpredictor. Crit Care 23:. https://doi.org/10.1186/s13054-019-2563-x

19. M@tric project. https://www.matric.be/. Accessed 5 Oct 2021

20. Lopez-Delgado JC, Esteve F, Torrado H, et al (2013) Influence of acute kidney injury on short- and long-term outcomes in patients undergoing cardiac surgery: risk factors and prognostic value of a modified RIFLE classification. Crit Care 17:R293. https://doi.org/10.1186/cc13159

21. Belayev LY, Palevsky PM (2014) The link between acute kidney injury and chronic kidney disease. Curr Opin Nephrol Hypertens 23:149–154. https://doi.org/10.1097/01.mnh.0000441051.36783.f3

22. Ishani A, Xue JL, Himmelfarb J, et al (2009) Acute kidney injury increases risk of ESRD among elderly. J Am Soc Nephrol 20:223–228. https://doi.org/10.1681/ASN.2007080837

23. Forni LG, Darmon M, Ostermann M, et al (2017) Renal recovery after acute kidney injury. Intensive Care Med 43:855–866. https://doi.org/10.1007/s00134-017-4809-x

24. Itenov TS, Berthelsen RE, Jensen JU, et al (2018) Predicting recovery from acute kidney injury in critically ill patients: Development and validation of a prediction model. Crit Care Resusc 20:54–60

25. Lee BJ, Hsu C yuan, Parikh R, et al (2019) Predicting Renal Recovery After Dialysis-Requiring Acute Kidney Injury. Kidney Int Reports 4:571–581. https://doi.org/10.1016/j.ekir.2019.01.015

26. Koeze J, Keus F, Dieperink W, et al (2017) Incidence, timing and outcome of AKI in critically ill patients varies with the definition used and the addition of urine output criteria. BMC Nephrol 18:70. https://doi.org/10.1186/s12882-017-0487-8

27. Zhou R, Yin W, Li W, et al (2022) Prediction Model for Infectious Disease Health Literacy Based on Synthetic Minority Oversampling Technique Algorithm. Comput Math Methods Med 2022:8498159. https://doi.org/10.1155/2022/8498159

28. Zhang X, Fei N, Zhang X, et al (2022) Machine Learning Prediction Models for Postoperative Stroke in Elderly Patients: Analyses of the MIMIC Database. Front Aging Neurosci 14:897611. https://doi.org/10.3389/fnagi.2022.897611

29. Ohnuma T, Uchino S (2017) Prediction models and their external validation studies for mortality of patients with acute kidney injury: A systematic review. PLoS One 12:e0169341. https://doi.org/10.1371/journal.pone.0169341

30. Bilbao-Meseguer I, Rodríguez-Gascón A, Barrasa H, et al (2018) Augmented Renal Clearance in Critically Ill Patients: A Systematic Review. Clin Pharmacokinet 57:1107–1121. https://doi.org/10.1007/s40262-018-0636-7

31. He J, Yang Z-TT, Qian X, et al (2020) A higher dose of vancomycin is needed in critically ill patients with augmented renal clearance. Transl Androl Urol 9:2166–2171. https://doi.org/10.21037/tau-20-1048

32. Abdel El Naeem HEM, Abdelhamid MHE, Atteya DAM (2017) Impact of augmented renal clearance on enoxaparin therapy in critically ill patients. Egypt J Anaesth 33:113–117. https://doi.org/https://doi.org/10.1016/j.egja.2016.11.001

33. Claus BOMM, Hoste EA, Colpaert K, et al (2013) Augmented renal clearance is a common finding with worse clinical outcome in critically ill patients receiving antimicrobial therapy. J Crit Care 28:695–700. https://doi.org/10.1016/j.jcrc.2013.03.003

34. Gijsen M, Huang CY, Flechet M, et al (2020) Development and External Validation of an Online Clinical Prediction Model for Augmented Renal Clearance in Adult Mixed Critically Ill Patients: The Augmented Renal Clearance Predictor. Crit Care Med 48:e1260–e1268. https://doi.org/10.1097/CCM.0000000000004667

35. Tang L, Ding XY, Duan LF, et al (2021) A Regression Model to Predict Augmented Renal Clearance in Critically Ill Obstetric Patients and Effects on Vancomycin Treatment. Front Pharmacol 12:1–12. https://doi.org/10.3389/fphar.2021.622948

36. Nei AM, Kashani KB, Dierkhising R, Barreto EF (2020) Predictors of Augmented Renal Clearance in a Heterogeneous ICU Population as Defined by Creatinine and Cystatin C. Nephron 55905:313–320. https://doi.org/10.1159/000507255

37. Barletta JF, Mangram AJ, Byrne M, et al (2017) Identifying augmented renal clearance in trauma patients: Validation of the Augmented Renal Clearance in Trauma Intensive Care scoring system. In: The journal of trauma and acute care surgery. United States, pp 665–671

38. Udy AA, Roberts JA, Shorr AF, et al (2013) Augmented renal clearance in septic and traumatized patients with normal plasma creatinine concentrations: Identifying at-risk patients. Crit Care 17:. https://doi.org/10.1186/cc12544

39. ARC prediction. https://www.arcpredictor.com/. Accessed 13 Apr 2022

40. Huang C-Y, Güiza F, De Vlieger G, Meyfroidt G (2022) External validation of the AKIpredictor in critically ill adults. Intensive Care Med 48:952–953. https://doi.org/10.1007/s00134-022-06746-6

41. Bleeker SE, Moll HA, Steyerberg EW, et al (2003) External validation is necessary in prediction research:: A clinical example. J Clin Epidemiol 56:826–832. https://doi.org/https://doi.org/10.1016/S0895-4356(03)00207-5

42. Kim D-J, Lee D-H, Ahn S, et al (2019) A new population pharmacokinetic model for vancomycin in

patients with variable renal function: Therapeutic drug monitoring based on extended covariate model using CKD-EPI estimation. J Clin Pharm Ther 44:750–759. https://doi.org/https://doi.org/10.1111/jcpt.12995

43. Rajkomar A, Dean J, Kohane I (2019) Machine Learning in Medicine. N Engl J Med 380:1347–1358. https://doi.org/10.1056/NEJMra1814259

44. Luijken K, Wynants L, van Smeden M, et al (2020) Changing predictor measurement procedures affected the performance of prediction models in clinical examples. J Clin Epidemiol 119:7–18. https://doi.org/https://doi.org/10.1016/j.jclinepi.2019.11.001

45. Dawes RM, Faust D, Meehl PE (1989) Clinical versus actuarial judgment. Science (80-) 243:1668–1674. https://doi.org/10.1126/science.2648573

46. Janssen KJM, Vergouwe Y, Kalkman CJ, et al (2009) A simple method to adjust clinical prediction models to local circumstances. Can J Anaesth 56:194–201. https://doi.org/10.1007/s12630-009-9041-x

47. Ramspek CL, Jager KJ, Dekker FW, et al (2020) External validation of prognostic models: what, why, how, when and where? Clin Kidney J 14:49–58. https://doi.org/10.1093/ckj/sfaa188

48. Al-Dorzi HM, Alsadhan AA, Almozaini AS, et al (2021) The Performance of Equations That Estimate Glomerular Filtration Rate against Measured Urinary Creatinine Clearance in Critically Ill Patients. Crit Care Res Pract 2021:. https://doi.org/10.1155/2021/5520653

49. Mulder MB, Eidelson SA, Sussman MS, et al (2019) Risk Factors and Clinical Outcomes Associated With Augmented Renal Clearance in Trauma Patients. J Surg Res 244:477–483. https://doi.org/10.1016/j.jss.2019.06.087

50. Adnan S, Ratnam S, Kumar S, et al (2014) Select critically ill patients at risk of augmented renal clearance: experience in a Malaysian intensive care unit. Anaesth Intensive Care 42:715–722. https://doi.org/10.1177/0310057x1404200606

51. Grootaert V, Willems L, Debaveye Y, et al (2012) Augmented Renal Clearance in the Critically Ill: How to Assess Kidney Function. Ann Pharmacother 46:925–959. https://doi.org/10.1345/aph.1Q708

52. Wongpraphairot S, Thongrueang A, Bhurayanontachai R (2022) Glomerular filtration rate correlation and agreement between common predictive equations and standard 24-hour urinary creatinine clearance in medical critically ill patients. PeerJ 10:e13556. https://doi.org/10.7717/peerj.13556

53. Fleuren LM, Thoral P, Shillan D, et al (2020) Machine learning in intensive care medicine: ready for take-off? Intensive Care Med. 46:1486–1488

54. van de Sande D, van Genderen ME, Huiskens J, et al (2021) Moving from bytes to bedside: a systematic review on the use of artificial intelligence in the intensive care unit. Intensive Care Med. 47:750–760

55. Huang C-Y, Güiza F, De Vlieger G, Meyfroidt G (2022) Daily fluctuations in kidney function in critically ill adults. Crit Care 26:347. https://doi.org/10.1186/s13054-022-04226-3

56. Carlier M, Dumoulin A, Janssen A, et al (2015) Comparison of different equations to assess glomerular filtration in critically ill patients. Intensive Care Med 41:427–435. https://doi.org/10.1007/s00134-014-

3641-9

57. Kellum JA, Lameire N, Aspelin P, et al (2012) Kidney disease: Improving global outcomes (KDIGO) acute kidney injury work group. KDIGO clinical practice guideline for acute kidney injury. Kidney Int Suppl 2:1–138. https://doi.org/10.1038/kisup.2012.1

58. Bragadottir G, Redfors B, Ricksten S-E (2013) Assessing glomerular filtration rate (GFR) in critically ill patients with acute kidney injury - true GFR versus urinary creatinine clearance and estimating equations. Crit Care 17:R108. https://doi.org/10.1186/cc12777

59. Chen S (2013) Retooling the creatinine clearance equation to estimate kinetic GFR when the plasma creatinine is changing acutely. J Am Soc Nephrol 24:877–888. https://doi.org/10.1681/ASN.2012070653

60. O'Sullivan ED, Doyle A (2017) The clinical utility of kinetic glomerular filtration rate. Clin Kidney J 10:202–208. https://doi.org/10.1093/ckj/sfw108

61. Cherry RA, Eachempati SR, Hydo L, Barie PS (2002) Accuracy of short-duration creatinine clearance determinations in predicting 24-hour creatinine clearance in critically ill and injured patients. J Trauma 53:267–271. https://doi.org/10.1097/00005373-200208000-00013

62. Baumann TJ, Staddon JE, Horst HM, Bivins BA (1987) Minimum urine collection periods for accurate determination of creatinine clearance in critically ill patients. Clin Pharm 6:393–398

63. O'Connell MB, Wong MO, Bannick-Mohrland SD, Dwinell AM (1993) Accuracy of 2- and 8-Hour Urine Collections for Measuring Creatinine Clearance in the Hospitalized Elderly. Pharmacother J Hum Pharmacol Drug Ther 13:135–142. https://doi.org/https://doi.org/10.1002/j.1875-9114.1993.tb02698.x

Summary

In critically ill patients, both increased and decreased renal clearance are highly prevalent and may vary during the course of critical illness. Acute kidney injury is a type of organ failure that is highly prevalent in many conditions that cause severe critical illness, and has an association with worse short- and long-term clinical outcomes. Fluctuations in renal clearance will influence the plasma levels and clearance of renally excreted drugs, and may result in unfavorable consequences such as drug toxicity and treatment failure. Predicting these fluctuations in kidney function could allow for more personalized drug dosage, and less treatment failure, potentially leading to better clinical outcomes. In addition, even though the prevention and management of AKI are mainly supportive, early predictions of a reducing or fluctuating kidney function could act as in silico biomarkers, to stratify patients according to their risk, and may aid in developing targeted therapies to prevent or mitigate the course of kidney injury. Despite the importance of kidney function prediction models, the focus of current predictions for kidney function made by machine-learning models has been on predicting the development of AKI or augmented renal clearance. Although some models have been demonstrated with good predictive performance and outperformed the physicians in external validation, they still need more external validations in independent datasets before they can be applied to clinical practice. Furthermore, since kidney function is continuous, predicting the complete spectrum of kidney function is more in accordance with clinical and physiological reality.

The general objective of this thesis is to apply advanced data analysis techniques and machine-learning algorithms to routinely collected clinical data from critically ill patients to develop and validate prediction models for kidney function. This thesis consisted of three primary objectives, where we focused on common medical phenomena and measurements with important clinical implications in the ICU.

In the **first part**, we performed an external validation of a prediction model for acute kidney injury (AKI), the AKIpredictor, in critically ill adults of the University Hospitals Leuven who were included in the large multicenter M@tric database between 2013 and 2018. This M@tric database contains high-quality and intricately interconnected data from all adult patients annually admitted to the ICU from 2013 to 2018. Even though this external validation dataset was collected 10 years after the original development cohort, the AKIpredictor still demonstrated its robustness. These results verified the AKIpredictor's potential to be a useful tool for the early detection of AKI patients.

AKI's progression and recovery are crucial since they are closely related to end-stage kidney disease and progressive renal dysfunction. Nevertheless, large databases frequently lack a good AKI recovery evaluation because the recovery definition is not uniform and baseline serum creatinine is frequently unknown, which presents difficulties for the development of AKI recovery prediction models. Using a large multicenter EPaNIC randomized controlled trial database, where two parenteral nutrition strategies were compared in 4640 critically ill adults between August 2007 and November 2010, we developed and externally validated prediction models for AKI recovery at hospital discharge in critically ill patients with AKI stage 3 during their ICU stay. For the overall ICU population, the developed AKI recovery models only exhibited similarly unsatisfactory discrimination to the plasma neutrophil gelatinase-associated lipocalin (NGAL) measured on the first day of AKI stage 3, which was poorer than the reference that only based on age. For patients with cardiac surgery, the

developed models had better performance over NGAL_AKI3 and the reference. The model is of limited clinical utility due to its poor predictive performance, which may be caused by the multiple pathophysiological processes but also by the definition of AKI recovery, which is still debated.

In the **second part**, we carried out an external validation of a prediction model for augmented renal clearance (ARC), the ARC predictor, in adult COVID-19 pneumonia patients admitted to critical care at the University Hospitals Leuven from February 2020 to January 2021. ARC affects 20-65% of critically ill patients and is associated with decreased exposure to commonly used antibiotics and anticoagulants. Therefore, in this study, we externally validated the ARC predictor in a recent critically ill COVID-19 cohort. Despite the slightly worse calibration, the ARC predictor showed robust performance with good discrimination and a wide clinical usefulness range. The robust performance is noteworthy, given the large differences in patient characteristics between this critically ill COVID-19 cohort and the original ARC predictor development cohort (the cohort in the presented study showed an ICU length of stay almost twice as long, 14 vs. 8 days). The promising performance identified in this study confirmed the ARC predictor's potential to be a useful tool for the identification of patients with high risk of ARC.

In the **third part**, we focused on the evaluation of daily kidney function instability and prediction of daily kidney function, based on daily measured creatinine clearance (CrCl). It is well known that kidney function can change rapidly during critical illness, and that this change may have important implications for modifying the dosage of drugs that are excreted through the kidneys. Nevertheless, the actual incidence and degree of fluctuations have never been described systematically. Therefore, in this project, we investigated daily changes in kidney function as defined by the daily differences in CrCl in critically ill adults admitted to the ICUs of the University Hospitals Leuven, included in the EPaNIC RCT database. Using a large amount of daily creatinine clearance measurements, we discovered that on approximately 35–40% of days, critically ill patients may experience potentially clinically significant changes in kidney function on a daily basis. Furthermore, this instability was more pronounced in the first week of ICU admission and at higher CrCl values. Future studies in independent cohorts of critically ill patients are needed to confirm these findings and to examine the factors associated with fluctuations in renal clearance.

In patients with critical illness, potentially clinically significant changes in kidney function were found to happen to 35-40% of days during the ICU stay in the previous study. However, the current methods for kidney function estimation had limited capacity to accurately reflect kidney function in critically ill patients, since they were developed based on healthy people. Additionally, these methods were based on measurements from the past, which may lead to an estimated kidney function lagging behind the actual kidney function. Therefore, we developed and validated models for daily prediction of measured CrCl, named "CrCl predictor", in critically ill adults admitted to ICUs of the University Hospitals Leuven, included in the EPaNIC RCT database. Three models with progressively more features with increasing data resolution were developed. All models demonstrated good performance when tested on a large external validation dataset of 20930 patients of the University Hospitals Leuven who were included in the large multicenter M@tric database between 2013 and 2018. The same good performance was observed when the model was compared with the reference reflecting current clinical practice, which assumed that CrCl remained unchanged.

To understand the external validity in independent datasets and the added value of these models to physicians' predictions, we conducted a prospective observational study in 197 critically ill adults admitted to surgical ICUs of the University Hospitals Leuven between January 2022 and April 2022. Treating physicians of the surgical ICU team were asked to predict CrCl and to report their corresponding levels of confidence via a well-designed questionnaire survey. The predictions made by ICU physicians were compared with the ones made by the developed CrCl predictor models, where the CrCl predictor showed robust performance with comparable accuracy to that observed in the original model development study. Furthermore, the CrCl predictor demonstrated slightly smaller prediction errors than the staff members and senior residents, and a much smaller prediction error than the junior residents. The lack of statistically significant differences implied that the CrCl predictor could perform at least as well as the ICU physicians. These findings suggested the potential added value of the CrCl predictor to physicians' predictions, especially for junior residents, and the potential of the CrCl predictor as a covariate to be integrated in the PK model to help optimize the renally cleared drug exposure.

Despite the good performance in the comparison with the reference reflecting current clinical practice and the ICU physicians, whether the CrCl predictor can help improve the patient care and outcome remains to be investigated in future interventional studies. For the purpose of evaluating the model performance in a real clinical setting, we developed a prototype software that integrated the developed CrCl predictor models and visualized the prediction results along with prediction explanations in a user-friendly manner. Prediction explanation is crucial since physicians prefer sufficient evidence-based scientific support, in order to provide reasonable intervention and to modify their treatment strategies for vulnerable ICU patients. The developed software was designed to function normally in a real-time setting, which was able to calculate the predictions within less than 30 seconds for the patients staying less than 9 days in the ICU, consisting of 75% of the study cohort in **Chapter 7**. Nevertheless, a prospective study is still needed to technically validate its functionality. Once a technical validation is finished, the developed software will be ready for an interventional study to investigate whether patient kidney function management and/or patient outcome can be improved with the additional patient kidney function information provided by the developed software.

In conclusion, this thesis focused on the application of advanced data analysis techniques and machine-learning algorithms to routinely collected clinical information from critically ill patients, in order to gain new insights into kidney function of critically ill patients. Many research questions have been thoroughly studied, encompassing the development and validation of prediction models for kidney function, examination of daily kidney function fluctuations, prospective comparison of models with ICU physicians, and integration of the developed kidney function prediction models into a user-friendly software. The developed prototype software provides the possibility for interventional study and assessment of the research findings' efficacy and safety, as well as their impact on kidney function management, patient care, and outcome.

CURRICULUM VITAE
Chao-Yuan Huang
Electrical Engineer

Personal information

Linkedin	https://www.linkedin.com/in/chao-yuan-huang/
Mobile	+32 483228340
Email	andy8307.eo01@gmail.com

Summary

I am a goal-oriented electrical engineer, with a focus on Machine Learning. During my PhD, I had diverse responsibilities including developing and validating prediction models, deploying prediction models into applications, and conducting a clinical trial. I am passionate about challenges where I can apply and sharpen my skills in deriving insights from data, through data visualization, data analytical techniques, and machine learning algorithms.

Education

KU Leuven, Leuven, Belgium	Expected end date: March 2023
PhD in Biomedical Science	
National Chiao Tung University, Hsinchu, Taiwan	August 2018
Master of Electrical and Computer Engineering	
KU Leuven, Leuven, Belgium	August 2017
Master of Electrical Engineering (dual-master program)	
National Chiao Tung University, Hsinchu, Taiwan	June 2016
Bachelor of Engineering, Department of Photonics	

List of publications

1. **Chao-Yuan Huang**, Fabian Güiza, Pieter Wouters, Liese Mebis, Joachim Gidts, Lore Klewais, Greet Van den Berghe, Greet De Vlieger, Geert Meyfroidt. A prospective comparison of ICU physicians and machine learning for the daily prediction of creatinine clearance in critically ill adults. Ready for submission.
2. **Chao-Yuan Huang**, Fabian Güiza, Pieter Wouters, Liese Mebis, Giorgia Carra, Jan Gunst, Philippe Meersseman, Michael Casaer, Greet Van den Berghe, Greet De Vlieger, Geert Meyfroidt. Development and validation of the creatinine clearance predictor machine learning models in critically ill adults. Ready for submission.
3. **Chao-Yuan Huang**, Fabian Güiza, Matthias Gijsen, Isabel Spriet, Dieter Dauwe, Yves Debaveye, Marijke Peetermans, Joost Wauters, Greet Van den Berghe, Geert Meyfroidt, Greet De Vlieger. External validation of the augmented renal clearance predictor in critically ill COVID-19 patients. Submitted for publication.

4. **Chao-Yuan Huang**, Fabian Güiza, Greet De Vlieger, Geert Meyfroidt. Daily fluctuations in kidney function in critically ill adults. *Critical Care*. 2022;26(1):347. doi:10.1186/s13054-022-04226-3
5. **Chao-Yuan Huang**, Fabian Güiza, Greet De Vlieger, Geert Meyfroidt. External validation of the AKIpredictor in critically ill adults. *Intensive Care Medicine*. 2022;48(7):952-953. doi:10.1007/s00134-022-06746-6
6. **Chao-Yuan Huang**, Fabian Güiza, Greet De Vlieger, Pieter Wouters, Jan Gunst, Michael Casaer, Ilse Vanhorebeek, Inge Derese, Greet Van den Berghe, Geert Meyfroidt. Development and validation of clinical prediction models for acute kidney injury recovery at hospital discharge in critically ill adults. *Journal of Clinical Monitoring and Computing*. Published online 2022. doi:10.1007/s10877-022-00865-7
7. Chiara Robba, Erika Banzato, Paola Rebora, Carolina Iaquaniello, **Chao-Yuan Huang**, Eveline J A Wiegers, Geert Meyfroidt, Giuseppe Citerio, Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) ICU Participants and Investigators. Acute Kidney Injury in Traumatic Brain Injury Patients: Results from the Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury Study. *Critical Care Medicine*. 2021;49(1):112-126. doi:10.1097/CCM.0000000000004673
8. Matthias Gijsen, **Chao-Yuan Huang**, Marine Flechet, Ruth Van Daele, Peter Declercq, Yves Debaveye, Philippe Meersseman, Geert Meyfroidt, Joost Wauters, Isabel Spriet. Development and External Validation of an Online Clinical Prediction Model for Augmented Renal Clearance in Adult Mixed Critically Ill Patients: The Augmented Renal Clearance Predictor. *Critical Care Medicine*. 2020;48(12):e1260-e1268. doi:10.1097/CCM.0000000000004667
9. **Chao-Yuan Huang**, Fabian Güiza Grandas, Marine Flechet, Geert Meyfroidt. Clinical prediction models for acute kidney injury in the intensive care unit: A systematic review. *Brazilian Journal of Intensive Care*. 2020;32(1):123-132. doi:10.5935/0103-507X.20200018
10. Yih-Liang Shen, **Chao-Yuan Huang**, Syu-Siang Wang, Yu Tsao, Hsin-Min Wang, Tai-Shih Chi. Reinforcement Learning Based Speech Enhancement for Robust Speech Recognition. In: *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*. Vol 2019-May. IEEE; 2019:6750-6754. doi:10.1109/ICASSP.2019.8683648

List of abstracts for international conferences

1. Belgian Society of Intensive Care Medicine (SIZ) 2022, Brussels. Augmented renal clearance predictor in critically ill COVID-19 patients. Poster presentation.
2. Belgian Society of Intensive Care Medicine (SIZ) 2022, Brussels. Acute Kidney Injury predictor in critically ill COVID-19 patients. Poster presentation.
3. European Society of Intensive Care Medicine (ESICM) Lives 2021, Online. Gradient-boosting machine learning model to predict average creatinine clearance of the next two days. Poster presentation.
4. European Society of Intensive Care Medicine (ESICM) Lives 2020, Online. Gradient-boosting machine learning models to predict creatinine clearance one day ahead. Poster presentation.
5. E-International Symposium on Intensive Care and Emergency Medicine (ISICEM) 2020, Online. Development and external validation of an online clinical prediction model for augmented renal clearance in adult mixed critically ill patients: the ARC predictor. Poster presentation.

6. International Symposium on Intensive Care and Emergency Medicine (ISICEM) 2020, Online. Development and internal validation of a model to predict acute kidney injury recovery at hospital discharge. Poster presentation.
7. European Society of Intensive Care Medicine (ESICM) Lives 2019, Berlin. Clinical prediction model for acute kidney injury recovery in the intensive care unit. Poster presentation.
8. Belgian Society of Intensive Care Medicine (SIZ) 2019, Brussels. Epidemiology of pediatric AKI: Analysis of prospectively collected data. Poster presentation.

Awards

- Award for an outstanding pitch (top 6 pitch winners competing against other 38 PhD researchers), VLIR/ULiège Interactive Career Day Data Sciences 2021 at Janssen Pharmaceutica, Online
- Second prize for the Best PhD team at KU Leuven datathon 2020 (our team developed and validated a prediction model for air quality on the next day, with an aim to provide guidance to patients with respiratory disease)

Technical skills

- Machine learning (Scikit-learn, Pandas)
- Structured query language (MSSQL, MySQL)
- Deep learning (Tensorflow, Keras)
- Visualization (Seaborn, Plotly)
- Website development (Django, Flask)
- Graphical user interface (Tkinter, Kivy)
- Statistics, Linux, Raspberry Pi, Git

Soft skills

- Problem solving
- Teamwork
- Analytical skills
- Project management
- Communication skills
- Scientific writing
- Qualitative and quantitative analysis

Languages

- Chinese: Native proficiency
- English: Professional working proficiency
- Dutch: Elementary proficiency
- Japanese: Elementary proficiency

Courses

2022	Discover your career profile ONLINE
2021	Brand your profile ONLINE
2021	Workshop research integrity part II
2020	Essentials of image editing
2020	Good clinical practice training meeting “Minimum criteria for ICH E6 GCP Investigator Site Personnel Training”
2019	Python software engineering
2019	Linux scripting
2019	Linux for HPC

2019 Creating effective research posters
2019 HPC introduction
2019 Python for data science
2019 Managing your PhD
2018 Central lecture research integrity
2018 Online statistics
2019 English intonation and pronunciation
2019 Presentation Skills for Biomedical Researchers
2019 Version control
2018 Writing Skills for Biomedical Researchers

Acknowledgment

Time flies; in a blink of an eye, years of research will be coming to an end, and I am about to graduate with a bag full of priceless lessons and precious blessings. When I look back, my heart is full of mixed feelings and gratitude.

My deepest gratitude goes to my (co-)supervisors, Prof. Dr. Geert Meyfroidt, Dr. Fabian Güiza Grandas, and Prof. Dr. Greet De Vlieger, for the successful completion of this thesis and PhD project. Geert, it is such an honor to have you as a promotor and follow your meticulous way of conducting research. I'm appreciative of your professional guidance throughout my PhD journey. There is never a shortcut or easy path on the way to success, and you are crystal clear about it and willing to bring this up unhesitatingly from the beginning. You taught me that pushing the boundaries of what we can do may not be comforting, but this is how we have accomplished so many milestones and improved ourselves so much. I'm grateful for you being such a straightforward and decisive mentor so that we are always aware of where the project is heading without getting lost. At the same time, you are open-minded, and knowing that your office door is always there open for me allows me to feel at ease to discuss both the academic and daily lives with you.

Fabian, thank you for being the strongest technical support of our machine-learning squad. Having you on board always calms me down, as you are so creative that you always think outside the box and point out the blind spots that nobody notices. On top of that, you are also a motivated person. Your insistence on details without compromising any chance to achieve the highest quality of research taught me the right way of conducting research. Also, your sense of beauty explicitly expressed when you gave suggestions on how the visualization could be improved. Finally, you are such a warm mentor, seamlessly and tightly catching me at the darkest moment of my PhD. I'm appreciative!

Greet, meeting you so late in the third year of my PhD is such a pity, but meeting late is absolutely better than never. Thank you for your efficient and in-depth replies all the time, making it such a pleasant working experience with you. Your broad and in-depth knowledge always significantly improves the projects in which you are involved. Your eyes are so sharp, pinpointing the key points right away. Your consistent passion for research also encourages me, convincing me that we can always do better. Apart from the professional aspect that I admire, your kindness and genuine care for all the people around you lit up the end of my PhD, for which I would like to express my sincere gratitude.

Professor Van den Berghe, thank you for setting up such an open laboratory for scientific discussion. It is a great honor for me to join this world-renowned laboratory, to perform studies with so many intelligent and skilled colleagues, and to get the chance to learn the right research methodology and attitude.

I'm appreciative of the reviewing work of my jury members. Professor Björn Meijers and Professor Celine Vens, thank you for following my PhD project and sharing your critical insights in the past 4 years. The provision of your innovative perspective undoubtedly helps improve the quality of this PhD project. I also want to express my gratitude to Professor Eric Hoste and Professor Marlies Ostermann for reviewing my thesis and providing insightful scientific feedback. Finally, I would like to thank Professor Koen Milisen for chairing the reading committee for my thesis and Professor Frank Claessens for chairing my public defense.

I gratefully acknowledge the support from the Taiwan-KU Leuven scholarship program. Particularly, thank you Taiwanese Ministry of Education (Taipei, Taiwan) and KU Leuven (Leuven, Belgium) for generously funding this project and providing an opportunity to researchers in need without expecting anything in return.

I want to express my gratitude to all the critically ill patients and their beloved families for participating in any of the retrospective and prospective studies, especially at the most difficult moment in their lives. The selfless contribution to science is noble and should not be emphasized too much.

For the AKI-related research, I am grateful to Marine and all the contributors to the original AKIpredictor study, from which the models were used in the external validation study. Thank you, Pieter, Professor Schetz, Prof. Gunst, Prof. Casaer, Prof. Vanhorebeek, and Inge for the contribution to the EPaNIC database, which was used to develop and validate the AKI recovery prediction models.

For the ARC-related research, thank you Matthias and Professor Spriet for involving me in the original ARC predictor development study, and thank you for the precious chance to externally validate such an impressive model (thank you Greet for the connection) and the invited talk at the symposium (thank you Omar and Anna for the interactive presentation) a few months ago. This collaboration could not have been possible without the connection help of Geert, which I am appreciative of. I am also thankful to Marine for setting up the ARC predictor website with me, where I learned about the systematic organization of projects. Additionally, I would like to thank the members of the University Hospitals Leuven COVID-19 units for providing the necessary data used to validate the ARC predictor.

For the CrCl-related research, thank you, Fabian, for the suggestion to submit Chapter 6 as a concise scientific report. Thank you, Greet, for the very helpful and convincing references. Thank you, Geert, for insisting on conducting the highest-quality research. Thank you all for the endless model optimization and patient brainstorming with me. Now I can confidently say that here are the things that I tried and did work! For the prospective study, the biggest thank goes to Pieter, Liese, Hanna, Sandra, Katleen, and Heidi. Without your professional help in study protocol submission, database setup, data entry, data examination, data cleaning, data export, and administrative work, I as an engineer would not have had the chance to perform clinical research, which will definitely be one of the most precious experiences of my entire life. Furthermore, I would like to thank Lore and Joachim for their help in encouraging other physicians to participate in the prospective study, patiently answering their questions and guiding them, and examining the baseline characteristics from piles of medical records.

I'm so grateful for meeting the most talented LICM master students, PhD students, or postdocs (in random order): Giorgia, Grégoire, Marine, Chloë, Sarah D, Sofie, Bavo, Lisa, An, Nathalie, Ines, Astrid, Arno, Ruben, Wouter, Alexander, Brenda, Caroline, and Lauren. On top of the food, drinks, and conferences, there were so many memorable good times that will last forever in my mind: lab excursions, datathon, fitness sessions, the PhD on tour 1.0/2.0 (I did not see it! xD), pizza Fridays, the cheese-and-wine evening, and Christmas market exploration. Thank you all for creating such a pleasant atmosphere in the lab. The unconditional help to each other and the company at all times helped maintain my mental health throughout the PhD journey. It is of my greatest honor to be the colleague of you guys. You are the best! Finally, I would like to express my gratitude to the principal investigators and administrative and technical colleagues (in random order): Professor Lies Langouche (thank you for the encouraging and warm words when I wrote the thesis), Professor Ilse Vanhorebeek

(thank you for organizing the research meetings), Sarah VP (thank you for supporting the entire lab), Lies P (let's forget the boat that luckily did not sink xD), and Inge (thank you for the small talks). A particular thank goes to Sarah VP, for the help with my thesis printing and the organization of my public defense and the reception.

I am thankful to my friends, who indirectly contributed to this PhD project (in random order): 翊展、泰翔、苡萱、炘哲、雲安、令幫、翊瑄、菲菲、查理、珊姐、小佘、尹蔓、小樂、怡鈞、小茜、柏涵、孟娟、立馨、承學、佳玲、秋月、哲睿、Zoey、 基永、家宥、志浩、奶茶哥、士翔、崔西、暐婷、振銘、Patty、曼青、信甫、粲琁、小麥。

Last but not least, I would like to thank my indispensable and beloved family. Dad, you are a self-demanding person, setting a role model for me to pursue advancement. Mom, your unconditional love for me with or without degrees stands as my strongest support. My elder sister, your helpful sharing of life experiences always teaches me precious lessons. My younger sister, your optimism and wisdoms in dealing with people always admire me. I would not stand here without any of you!

Yuan / 肇元