

ACTA BIOMEDICA LOVANIENSIA 760
KU Leuven
Biomedical Sciences Group
Faculty of Medicine
Department of Cellular and Molecular Medicine
Laboratory of Intensive Care Medicine

Marine FLECHET

CLINICAL PREDICTION MODELS IN CRITICAL ILLNESS:
FROM COMPUTER TO BEDSIDE



LEUVEN UNIVERSITY PRESS

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

Jury:
Promoter: Prof. dr. Geert Meyfroidt
Co-promoter: Prof. dr. Greet Van den Berghe
Dr. Fabian Güiza
Chair Examining Committee: Prof. dr. Ghislain Opdenakker
Chair Public Defense: Prof. dr. Joost Schymkowitz
Jury members: Prof. dr. Marlies Ostermann
Prof. dr. Fabio Taccone
Prof. dr. Johan van Loon
Prof. dr. Celine Vens

©2018 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 276 8
D/2018/1869/66

The best way to predict the future is to create it.

— Abraham Lincoln

ACKNOWLEDGMENTS

The completion of this thesis would not have been possible without the help and guidance of many people. Thus, I would like to express my gratitude to everyone who contributed to this work.

I am thankful to have been given the opportunity to pursue my PhD at the Katholiek Universiteit Leuven (KULEUVEN). Therefore, I would like to thank the rector, Professor Luc Sels and the former rector, Professor Rik Torfs. I am also thankful to the Fonds Wetenschappelijk Onderzoek (FWO) for the funding provided to perform this research.

Geert, it was an honor and a privilege to have you as a promoter, and to be your first PhD student. I am extremely grateful for your excellent guidance throughout my PhD. Since our first meeting, until the end of my thesis, you have always challenged me, and made me push my boundaries far away. Thanks to you, I never experienced the 2nd-year-PhD-disappointment-frustration-feeling. You are a great teacher, who loves sharing his research and clinical expertise, and made me understand the clinical impact of our work. I wish your future students learn as much from you as I had the chance to do. I am also grateful for challenging me to present our research at medical conferences, where you shared your network with me. Thank you for being this outstanding promoter and mentor; thanks to you, today I will become a doctor ;)

Greet, it has been a true honor to conduct my PhD in your laboratory, and to have such a talented and visionary researcher as a co-promoter. Your enthusiasm and confidence in our research always made me feel that we could always do better and that nothing was impossible. You helped me get fresh perspective on our work and look at the bigger picture. Although being an engineer in a medical environment could be challenging, you enthusiastically shared your clinical knowledge and expertise, which helped me feel at ease. You were and will remain a great source of inspiration. Thank you for everything!

Fabian, I could not have wished to have a better engineer co-promoter than you. By sharing your research and statistical expertise, you helped me gain

the required knowledge to pursue this PhD and develop a critical mindset. You always made time for me and taught me to see the good side in each failure. Your guidance was not only research-oriented but also for personal development and growth. Through various experiments, you helped me gain confidence and realize what are my strengths and ambitions. Your empathy and creativity were a great source of inspiration. You were a teacher, a confident, a mentor. I will greatly miss our close collaboration. Thanks a million.

I am also grateful to my Jury members. Professor Johan van Loon and Professor Celine Vens, thank you for your enthusiasm in our research and for following my work during the past 4 years as members of my Thesis Advisory Committee. Professor Marlies Ostermann, I had the pleasure to interact with you at several conferences. Being a jury member at my poster (ISICEM15) or chair at my presentation (ESICM15), you already followed my work before being officially introduced thanks to Miet at AKI&CRRT16. Thank you for agreeing to be part of my jury committee and for reviewing my thesis. Professor Fabio Taccone, I had the pleasure to meet you at ISICEM earlier this year. Thank you for your interest in our research and for your valuable feedback. Finally, I would like to thank Professor Ghislain Opdenakker for chairing the thesis committee and Professor Joost Schymkowitz for chairing the public defense.

I am very grateful to all my colleagues who have been involved in the different thesis projects.

For the kidney-related research, thank you Miet for your thorough manuscript reviews which taught me a lot, your valuable insights and ideas, and for the nice time together and sharing your network at AKI&CRRT. Stefano, it was a true pleasure collaborating with you on this project. Your help in the planning and realization of the AKIpredictor validation study was invaluable. Claudia, thank you so much for your enthusiastic participation in this study. Isabelle, thank you for your thorough data checks. Jan, Michael and Ilse, thank you for the in-depth manuscript reviews. Inge, thank you for the 2000+ NGAL measurements; after spending some time in the lab, I realized how much work this required. A particular thanks to all ICU assistants, residents and staff members who took part in the AKIpredictor validation study: Alexander Beaumont, An Tournicourt, An Schrijvers, Anneleen Vandebroek, Annemijn Witdouck, Astrid Barbe, Bart Lauwereins, Cathelle Cornelissen, Catherine Ingels, Dieter Dauwe, Yves Debaveye, Shaun de Meiroman, Erwin de Troy, Eva Boonen, Fabian Dhondt, Frederic de Rydt, Geert Meyfroidt, Greet de Vlieger, Hendrik Stragier, Ine Meeus, Jan Muller, Jan Gunst, Jolien Schildermans, Katelijne Buyck, Katrien Haeck, Lander Vanhulle, Laurenz de Donder, Lien Torisaen, Lore Winters, Maarten Brusseleers, Maarten Hendrickx, Michael Casaer, Michiel Brands, Michiel Boon, Neil Rayyan, Peter Huybrechts, Philippe

Moyaery, Philippe Leire, Renata Haghedaanen, Sofie Maebe, Tine Honinx, Wout van Oosterwyck, Dirk Vlasselaers.

For the brain-related research, thank you Bart for your valuable insights and our inspiring discussions in Cambridge, at ICP2016, at meetings and dinners. Your feedback is and will be of critical importance for the development of the brain software. I am also grateful to all contributors of the Brain-IT database. For their help in the planning and (upcoming) involvement in the BrainSoftware study, many thanks to Bram Vercammen, Dominiek Cotte, Fredrik Hermans, Sergey Balyaev, Mario Opsomer, Pieter, and the research nurses, Liese, Sylvia, Sandra, and Heidi.

For the NIRS-related research, thank you Dirk and Lars for your help in the planning of the study and for the feedback on the manuscripts. An additional thank you Dirk for your help to take some NIRS pictures. I am also grateful to Heidi, Marc and Stoffel for the data collection. Michaël, thank you for providing feedback on the manuscripts. An additional thanks to Sören, An, Sandra, Liese, Pieter and the psychologists, Sandra, Hanna, and Astrid, for helping with the 2 year follow-up evaluation of the additional non-PEPANIC children.

I am also thankful to everyone who contributed to the creation of the EPANIC and PEPANIC databases, which are invaluable high quality data sources. A particular thank to Pieter who is behind all data exports, data checks, and databases creation.

I am very grateful to my friends and (former) colleagues: Giorgia, An, Lisa, Chloë, Marc, Bram, Sam, Ruben, Ines, Thomas & Thomas, Sarah & Sarah, Lies, Inge, Nathalie, Wouter, Lies P, Steven, Sören, Arno, Chao-Yuan, Astrid and Marissa. Thank you for the nice chats during lunch and breaks, mental support, clinical explications, fun moments at drinks and dinners, dancing and drinking at Christmas parties, discovering the actor in everyone when filming the Christmas movies. A particular thank to Lisa and Inge for their explanations in the lab. After several years working close to each other, it was a pleasure to discover the lab word. Giorgia, it has been a true pleasure having another engineer on board. I enjoyed sharing my knowledge with you, and soon with Chao-Yuan. Sam, a particular thank for your explanations in the ICU. An, it has been a true pleasure working on additional research topics together. Thomas, thanks for sharing your thesis template. It definitely helped speed up the writing process. Chloë, my friendly desk-sharing colleague, thank you so much for your advice related to the administrative burden of the defense. Finally, a particular additional thank you to Sarah for her help with the organization of the defense and the reception. Thanks a million to all of you!

Finally, I would like to thank my parents and sister, as I would not be standing here today without their support over the years. I would also like to

thank my family, in particular my parents-in-law, and friends for their support and presence during these 4 years. Last but not least, I would like to thank my partner Romain for being the most nasty reviewer of this thesis, for challenging me with new knowledge, and for his unconditional help, support and love for more than 8 years. As my grandfather used to say: tu me tires toujours vers le haut. Merci mon amour.

CONTENTS

1	INTRODUCTION	1
1.1	Critical illness	2
1.1.1	Overview	2
1.1.2	Information overload	2
1.1.3	Electronic health records	3
1.2	Big data analytics in critical illness	4
1.2.1	Overview	4
1.2.2	Application to critical illness	5
1.2.3	Challenges	7
1.3	Research questions	8
1.3.1	Acute kidney injury	9
1.3.2	Post-operative care after corrective surgery for a congenital heart defect	11
1.3.3	Traumatic brain injury	13
1.4	Development of clinical prediction models	17
1.4.1	Feature selection	18
1.4.2	Data mining algorithms	20
1.4.3	Evaluation metrics	23
1.4.4	Validation techniques	28
1.5	Thesis overview	32
2	OBJECTIVES	43
2.1	General aim	43
2.2	Specific objectives	43
3	CLINICAL PREDICTION MODELS FOR ACUTE KIDNEY INJURY	45
3.1	Introduction	47
3.2	Methods	47
3.2.1	Study population: development and validation cohorts	47
3.2.2	AKI definition	48
3.2.3	Clinical prediction models	48
3.2.4	Variable selection	49
3.2.5	Model development and validation	49
3.2.6	NGAL quantification	49
3.2.7	NGAL as a predictor	49
3.2.8	Performance evaluation criteria	50

3.3	Results	50
3.3.1	Study population: development and validation cohorts . .	50
3.3.2	Variable selection and model development	51
3.3.3	Development cohort model performance	51
3.3.4	Validation cohort model performance	52
3.3.5	Validation cohort NGAL performance	56
3.3.6	Validation cohort: other clinical outcomes	56
3.3.7	Online AKI-123 prognostic calculator	57
3.4	Discussion	58
3.5	Conclusion	60
3.A	Appendix	62
3.A.1	Supplementary methods	62
3.A.2	Supplementary results	68
3.A.3	Supplementary tables and figures	70
4	MACHINE LEARNING VERSUS PHYSICIANS' PREDICTIONS OF ACUTE KIDNEY INJURY	97
4.1	Introduction	99
4.2	Methodology	100
4.2.1	Study population	100
4.2.2	Endpoint	100
4.2.3	Acute kidney injury	101
4.2.4	AKIpredictor predictions	101
4.2.5	Physicians' predictions	101
4.2.6	Statistical analysis	102
4.3	Results	104
4.3.1	Study population	104
4.3.2	AKIpredictor predictions	104
4.3.3	Physicians' predictions	106
4.3.4	Combining AKIpredictor with physicians' predictions . . .	108
4.4	Discussion	108
4.4.1	Strengths and limitations	111
4.5	Conclusion	111
4.A	Appendix	113
5	ASSOCIATION BETWEEN CEREBRAL OXIMETRY AND ACUTE OUTCOME	129
5.1	Introduction	131
5.2	Materials and Methods	132
5.2.1	Study population	132
5.2.2	Cerebral NIRS monitoring	132

5.2.3	Endpoints	133
5.2.4	SCTO ₂ predictors	133
5.2.5	Association between SCTO ₂ predictors and outcomes	135
5.2.6	Statistical analysis	136
5.3	Results	136
5.3.1	Study population	136
5.3.2	Association between SCTO ₂ predictors and PICU LOS, hospital LOS and duration of mechanical ventilation	137
5.3.3	Association between SCTO ₂ predictors and hospital and 90- day mortality	140
5.4	Discussion	140
5.5	Conclusions	144
5.A	Appendix	146
6	PREDICTION OF AKI WITH NEAR-INFRARED SPECTROSCOPY	155
6.1	Introduction	157
6.2	Methods	157
6.2.1	Study design	157
6.2.2	Study population	158
6.2.3	Near-infrared spectroscopy monitoring	158
6.2.4	Prediction of acute kidney injury	158
6.2.5	Statistical analysis	159
6.3	Results	160
6.3.1	Study population	160
6.3.2	Performance of near-infrared spectroscopy (NIRS) model . .	161
6.3.3	Performance of clinical model	163
6.3.4	Comparison of NIRS and clinical models performance . . .	165
6.4	Discussion	166
6.4.1	Strengths and limitations	167
6.5	Conclusion	168
6.A	Appendix	170
7	VISUALIZING CEREBROVASCULAR AUTOREGULATION INSULTS	179
7.1	Introduction	181
7.2	Methods	181
7.2.1	Study population	181
7.2.2	Patient management	182
7.2.3	Visualization method	183
7.3	Results	183
7.4	Discussion	186
7.5	Conclusion	187

8	PROTOTYPE OF A BRAIN INJURY MONITOR	191
8.1	Introduction	193
8.2	Methods	194
8.2.1	Patient population	194
8.2.2	Cerebral autoregulation	194
8.2.3	Visualization of patient-specific ICP and CAR dose	194
8.2.4	Software development	194
8.3	Results	197
8.3.1	Visualization of patient-specific ICP and CAR dose	197
8.3.2	Prototype software	199
8.4	Discussion	199
8.5	Conclusion	200
9	DISCUSSION	205
9.1	Early detection of acute kidney injury	205
9.1.1	Main findings	205
9.1.2	Current impact of research	208
9.1.3	Future perspectives	209
9.2	Assessment of the utility of NIRS-based cerebral oximetry	210
9.2.1	Main findings	210
9.2.2	Current impact of research	212
9.2.3	Future perspectives	212
9.3	Detection and refinement of secondary brain injuries	213
9.3.1	Main findings	213
9.3.2	Current impact of research	214
9.3.3	Future perspectives	215
9.4	Implementation details	216
9.5	Overall challenges	216
9.6	General conclusion	217
	SUMMARY	224
	CURRICULUM VITAE	229

ACRONYMS

APACHE	Acute physiology and chronic health evaluation
AKI	Acute kidney injury
AUC	Area under the curve
AUROC	Area under the receiver operating characteristic curve
BMI	Body mass index
BTF	Brain Trauma Foundation
CA	California
CAR	Cerebrovascular pressure autoregulation
CBF	Cerebral blood flow
CER	Comparative effectiveness research
CHD	Congenital heart defect
CI	Confidence interval
CKD	Chronic kidney disease
CPP	Cerebral perfusion pressure
CT	Computed tomography
DC	Decompressive craniectomy
DHCA	Deep hypothermic circulatory arrest
DS	Discrimination slope
ECMO	Extra corporal membrane oxygenation
EGFR	Estimated glomerular filtration rate
EHR	Electronic health record
ELISA	Enzyme-linked immunosorbent assay

EPANIC	Early versus Late Parenteral Nutrition in Critically Ill Adults
ERC	European Research Council
ESRD	End-stage renal disease
FN	False negative
FP	False positive
FWO	Fonds Wetenschappelijk Onderzoek
GCS	Glasgow Coma Scale
GDPR	General Data Protection Regulation
GFR	Glomerular filtration rate
GOS	Glasgow Outcome Score
HRV	Heart rate variability
ICP	Intracranial pressure
ICU	Intensive care unit
IDE	Integrated development environment
IDI	Integrated discrimination index
IGFBP-7	Insulin-like growth factor-binding protein 7
INTBIR	International Initiative for Traumatic Brain Injury Research
IQR	Interquartile range
IT	Information technology
IWT	Flemish Government Agency for Innovation by Science and Technology
KDIGO	Kidney Disease: Improving Global Outcome
KDIGO-SC	Serum creatinine criterion of the KDIGO guidelines
KDIGO-UO	Urine output criterion of the KDIGO guidelines
KOF	Clinical research Foundation of the University Hospitals Leuven
LAX	Low Frequency Autoregulation Index

LOS	Length of stay
MA	Massachusetts
MAP	Mean arterial blood pressure
MDRD	Modification of Diet in Renal Disease
MSSQL	Microsoft SQL
MX	Mean flow index
MYSQL	MySQL
NB	Net benefit
NEMO	Individualized Targeted Monitoring in Neurocritical Care
NGAL	Neutrophil gelatinase-associated lipocalin
NIR	Near-infrared
NIRS	Near-infrared spectroscopy
NPV	Negative predictive value
NRI	Net reclassification index
OOB	Out-of bag
OR	Odds ratio
PCTL	Percentile
PDMS	Patient Data Management System
PICU	Pediatric intensive care unit
PIM	Pediatric Index of Mortality
PPV	Positive predictive value
PRX	Pressure Reactivity Index
RACHS	Risk Adjustment in Congenital Heart Surgery
RCT	Randomized controlled trial
REMAP	Randomized, embedded, multifactorial, adaptive platform
RMSSD	Root mean of successive squared differences

ROC	Receiver operating characteristic
RRT	Renal replacement therapy
SC	Serum creatinine
SAO ₂	Arterial oxygen saturation
SCTO ₂	Cerebral tissue oxygen saturation
SD	Standard deviation
SD-S	Standard deviation of a smoothed signal
SOFA	Sequential Organ Failure Assessment
SQL	Structured query language
STROBE	Strengthening the reporting of observational studies in epidemiology
SVO ₂	Central venous oxygen saturation
TBI	Traumatic brain injury
TIMP-2	Tissue inhibitor of metalloproteinase-2
TN	True negative
TP	True positive
UK	United Kingdom
UO	Urine output
USA	United States
UZLEUVEN	University Hospitals Leuven

INTRODUCTION

Partially adapted from:

- **Flech M** et al. Informatics in neurocritical care: new ideas for Big Data. *Curr Opin Crit Care*. 2016;22(2):87-93. doi:10.1097/MCC.000000000000287
- **Flech M** et al. 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*. 2017;43(6):764-773. doi: 10.1007/s00134-017-4678-3

2 INTRODUCTION

1.1 CRITICAL ILLNESS

1.1.1 *Overview*

Critical Care Medicine is a relatively young and high-tech branch in modern medicine that combines clinical skills, powerful drugs and sophisticated mechanical devices to support the function of vital organs. This allows patients to survive a variety of previously lethal insults such as multiple trauma, major surgery, complication of chronic illness, extensive burn, or severe infection. Patients admitted to the intensive care unit (ICU) have acute life-threatening conditions and require continuous observation and monitoring.

Despite this dedicated care, mortality among critically ill patients who require intensive care for more than a few days remains around 20% worldwide. Critical illness affects millions of patients each year, and consumes a large fraction of health care resources [1]. The diagnostic-therapeutic cycle in these patients is short as their clinical situation may vary rapidly. It is therefore of great interest to detect those patients most vulnerable to specific organ deterioration as early as possible, in order to administer dedicated therapies earlier and hopefully prevent the chronic and lethal phases of critical illness. Prediction is at the heart of intensive care medicine.

1.1.2 *Information overload*

The typical ICU generates vast amounts of data from several devices for each patient, including patient monitors, mechanical ventilators, syringe and infusion pumps for delivery of drugs and fluids, or renal replacement therapy machines (Figure 1.1). These monitors and therapeutic devices generate data on an intermittent or continuous basis. In addition, laboratory tests of blood samples are analyzed several times a day, and microbiology samples several times a week. Chest x-ray are often performed daily in almost all patients. In some, computed tomography (CT) scans are necessary once to several times during their stay. The patient chart also contains drug prescription and notes from nurses and intensivists. The amount of data generated per patient is gigantic.

Nowadays, sophisticated monitors display the physiological data in various combinations, including numerical values but also in traces over the past minutes, hours or days. ICU physicians combine their medical knowledge and experience with the large amount of patient-related data (including artefacts and measurement failures that they have to identify and discard) to foresee changes in patients' state and adapt therapy accordingly. However,



Figure 1.1 Picture illustrating the vast number of monitors connected to a critically ill child in the pediatric intensive care unit.

integrating all data from various sources over time to manage the care of several patients simultaneously exceeds human capacity. This problem is referred to as *information overload*.

A possible amenable consequence is that important, but inaccessible, pieces of information are discarded when clinicians have to make decisions fast.

1.1.3 *Electronic health records*

In the past decades, computerization has revolutionized many sectors. Progressively, electronic devices have been integrated in the ICU. On the one hand, the computerization of the ICU has further contributed to the problem of information overload, by providing support for the implementation of electronic alarms to warn the clinical staff of patient deterioration. Although the development of such alarms is motivated by the improvement of patient care, in practice, clinicians and nurses get overwhelmed by false alarms. On the other hand, the computerization of the ICU has also provided an opportunity to simplify data visualization and improve usability.

The digitalization of patient charts has led to the creation of electronic health records (EHRs). EHRs are large databases that contain the complete electronic medical chart for each critically ill patient. The database integrates all patient-related data including demographics, medical history, interventional reports, and data collected upon admission and during ICU stay.

There are several advantages of the digitalization of medical charts. The EHR provides a better quality of medical charting, less administrative workload, hence providing more time for patient care, increased clinical staff satisfaction, and nursing retention [2, 3]. It also allows speeding up medical research with a faster and easier access to patient data. The adoption of the EHR allows to perform secondary use of data, at the level of each medical center, or multicenter, hence providing unique benchmarking opportunities. ICU computerization has already happened in numerous centers and is spreading worldwide.

At the University Hospitals Leuven (UZLEUVEN), the EHR is implemented through a Patient Data Management System (PDMS) (MetaVision, iMD-Soft, Needham, MA, USA) that stores continuous data on a minute-by-minute basis. The average amount of data generated per patient per day is 6.3 MB. With 124 ICU beds at an occupancy rate of 90% during the whole year, the ICUs from the UZLEUVEN generate approximately 250 GB per year. The current database contains more than 3400 GB. The EHR itself has provided support for the creation of large clinical trial databases.

1.2 BIG DATA ANALYTICS IN CRITICAL ILLNESS

1.2.1 *Overview*

Big Data refers to the current speed and volume at which computerized data are generated by web applications and by the digitization of information that was previously available on paper only. Additionally, it refers to the improvement in technology that enables to store, process, and analyze these data [4]. Informally, these characteristics are referred to as the “the three Vs” of big data: Volume, Velocity, and Variety. Surprisingly, nowadays it is more cost-effective to invest in data storage than in cleaning old databases [5]. However, the lack of standardization and the uncertain data quality has added a fourth V for “Veracity” to the definition. Finally, big data is useless without a clear vision and knowledge of the “Value” it will bring to business or healthcare [4]. Together, Volume, Velocity, Variety, Veracity and Value are the five Vs of *Big Data*.

These data are too voluminous and complex to be processed and analyzed by traditional methods. Big data analytics use techniques from computer science such as machine learning to create new knowledge by identifying patterns in very large amounts of data. Machine learning encompasses a various range of algorithms that are able to learn from multidimensional data to perform pattern recognition, predictions, or generate data-driven hypothesis.

Together with digitalization, big data analytics have become increasingly popular and have been used in various fields including genome sequencing, omics, administrative database, social networks, connected objects, and telemedicine among others. The main applications of big data analytics include knowledge discovery and predictive modeling.

Knowledge discovery is defined as an “exploratory analysis and modeling of data and the organized process of identifying valid, novel, useful and understandable patterns from these data sets ” [6]. The generation of new knowledge can be performed by findings hidden associations or visualizing the data differently, among others.

Predictive modeling refers to the development of an algorithm for classification or regression. In a classification problem, one wants to predict a binary or categorical outcome, such as whether a patient will develop a particular complication in the following days; while in a regression problem, the outcome of interest is a continuous value, for instance, the blood pressure of a particular patient in one hour from now. Such algorithms can be developed using supervised or unsupervised techniques. Supervised predictive modeling uses a labeled dataset (a dataset for which the outcome is known), while unsupervised modeling uses unlabeled data.

1.2.2 *Application to critical illness*

Big data analytics can give a new perspective on the many challenges that clinicians face in medicine and critical care [7]. The massive quantities of data from EHRs are currently underused and hold the promise of supporting a wide range of novel applications, to improve care, save lives and lower costs [8, 9]. These novel applications might be patient-specific but also center-specific. They include among others,

- Clinical decision support tools to assist physicians in clinical decision-making and in tailoring diagnostic and therapeutic strategies for each patient;
- Early warning systems to foresee patient worsening or for the counseling of patients or their relatives to provide an estimate of ICU or later outcomes;
- User-friendly visualization applications that extract insights from clinical data and display relevant patient information [8].

These applications will help the physician to focus his or her attention on what really matters. They will provide more time for clinician-to-patient

interaction, which might result in improved care and allow clinicians to record more phenotypes [10]. The ultimate goal of big data analytics in critical illness is to achieve improvements in both the process of care and patient outcomes [11].

1.2.2.1 *Complementary approach to randomized controlled trials*

A randomized controlled trial (RCT) is the gold standard to demonstrate efficacy and safety in medical research. The aim of a RCT is to investigate the impact of a medical intervention by randomizing patients to an arm receiving the specific intervention (intervention group) or to one receiving a placebo or an alternative intervention (control group). The randomization process is introduced to minimize selection bias and distribute evenly potential confounding factors in each arm.

It is very challenging to design RCTs that take into account population heterogeneity, and to delineate those subgroups that might benefit most from a certain intervention. On the one hand, randomizing patients is still the best way to establish a causal inference between an intervention and outcome. On the other hand, ethical, financial, timing or compliance constraints can limit the feasibility of translating a research question into a RCT [12]. Additionally, in some research fields, results of RCTs have been discouraging. For instance, most RCTs on neuro-protective strategies in the neuro-critical care unit have been negative [13]. The exact reasons for this include insufficient preclinical experimental work, incomplete (patho-)physiological understanding or methodological flaws. In addition, the primary endpoints might not have been sensitive enough, or might have missed a link with the pathophysiology. In such specific settings, the need for alternative methods to expand medical knowledge and help improve clinical decision-making is evident.

Big data analytics has been proposed as a complement to RCTs by performing comparative effectiveness research (CER) [14, 15]. CER aims at measuring differences in outcomes in large heterogeneous populations [15]. Therefore, CER can be used to analyze the results from large observational studies to identify specific subgroups that may benefit from a particular therapy and thus should be recruited for a targeted RCT [12]. Consequently, adaptive clinical trials could be a possible way forward, where a trial is designed in such a way that it can "learn", through big data analytics, to identify subgroups or conditions in which the intervention could be beneficial. An example of such framework is the randomized, embedded, multifactorial, adaptive platform (REMAP) trial [14], starting to be implemented in critical care [16].

1.2.3 Challenges

Researchers and clinicians need to be aware of the opportunities of big data analytics, but at the same time will need to look beyond the hype. Several challenges have to be addressed in order to use big data analytics optimally. The first challenge refers to the proper data organization and structuring and labeling of the data. The second challenge encompasses the translation of the analytics to clinical practice. Finally, the acceptance of such analytics by clinicians is the third challenge.

1.2.3.1 Databases

Proper data organization and structuring of the data is of key importance for the use of big data analytics. At the center level, the diversity of the data, their different modalities and the frequency of storage challenge their integration. Although the EHR implementation has tackled part of the difficulties, artefactual, spurious, and incomplete data are inherent to the use of continuous non-validated monitoring data and have to be dealt with. At the multi-center level, lack of standardization between centers leads to further complexity as data coming from different EHR systems have to be integrated. Therefore, development of multi-center databases amenable to big data research requires initiatives such as the M@TRIC , a collaboration between 3 university hospitals from Flanders (Belgium) that contains records of more than 9000 ICU patients [17]; the MIMIC, a database from more than 30 000 ICU patients admitted to Beth Israel Deaconess Medical Center (Boston, MA, USA) from 2001 to 2008 [18]; or center-TBI, a collaboration currently collecting highly detailed information from 1800 patients with severe traumatic brain injury [19].

Early 2018, the adoption of the new European General Data Protection Regulation (GDPR) has also affected the development of such databases. Secondary use of personal data is now better protected. As a consequence, the data have to be strictly anonymized and the politics on obtaining informed consent have been strengthened.

1.2.3.2 Translation to bedside

The translation of big data analytics to the patient bedside is a necessary step to assess the impact of the novel analytics in the clinical setting. For this purpose, several avenues can be considered: paper charts (for simple analytics), websites, smartphone or smartwatch apps, or software integrated into existing or new bedside monitors. Examples of successful translations are the clinical scores or rules derived from original research, such as the acute physiology and chronic health evaluation (APACHE) to stratify disease severity or the Pediatric Index of

Mortality (PIM) score for mortality prediction. However, overall, there are few translations from big data analytics research to the bedside [20].

A potential explanation for this lack of bedside translation is the complexity of the task. The knowledge required to develop a software, a website or an app is different from the data analyst skills required to develop the analytics or the clinical expertise to understand and pose the research question. Additionally, particular care should be given to improve end-user usability and reduce the burden resulting from false positive alerts. Therefore, successful bedside implementation requires a continuous collaboration between clinical experts, statisticians, data analysts, and engineers.

Ultimately, once the analytics have been translated to the bedside, its effectiveness has to be evaluated in a randomized controlled trial where patients are randomized to an arm receiving treatment based on the novel analytics or to an arm receiving normal care. Upon success, healthcare protocols can be updated to include the novel techniques in day-to-day care.

1.2.3.3 *Use in clinical settings*

Finally, acceptance of big data analytics in the intensive care should not be underestimated. Although optimization of usability will help facilitate the translation to the bedside, precautions need to be taken when presenting the tool to the clinical team. Indeed, a common concern is that doctors will not be needed anymore due to development of big data analytics. This is not true. Computer and physician must work together [10], both at the level of model development and afterwards, when applications are turned into bedside tools. Physicians are needed to interpret the displayed information to make decisions [21]. Physicians will always decide on the actions to be taken. They can choose optimally using all tools available at the bedside, including recommendations from big data analytics [10]. Hence, instead of a replacement, big data analytics will be a partner in patient care.

1.3 RESEARCH QUESTIONS

This thesis aims to develop big data decision support applications for conditions with important clinical consequences during critical illness. The analytics will be applied in three different fields where alternative methods have failed to demonstrate benefit to improve patient care, or for which there is a need for knowledge discovery: acute kidney injury, post-operative care after corrective surgery for a congenital heart defect, and traumatic brain injury.

Although the size of the databases used in some projects in this dissertation are not strictly *Big Data*, the terms big data analytics will be used, interchangeably with machine learning, to refer to the advanced data-driven techniques applied throughout the thesis.

1.3.1 Acute kidney injury

Acute kidney injury (AKI) is a rapid decline in estimated glomerular filtration rate, i.e. in renal excretory function [22]. In critically ill patients, the disease can result from several pathways, usually grouped in pre-renal, intrinsic and post-renal categories (Figure 1.2) [22–25]. In pre-renal disease, the glomerular filtration rate decreases following renal hypoperfusion caused by non-renal insults such as systemic hypotension, volume depletion or cardiac failure. Intrinsic AKI is more frequently caused by an acute tubular necrosis, usually following renal ischemia-reperfusion injury, sepsis, or nephrotoxicity [23]. Other causes includes acute glomerulopathies, vasculitis, or interstitial nephritis. Finally, in post-renal disease, the glomerular filtration rate decreases following obstruction of the urinary track.

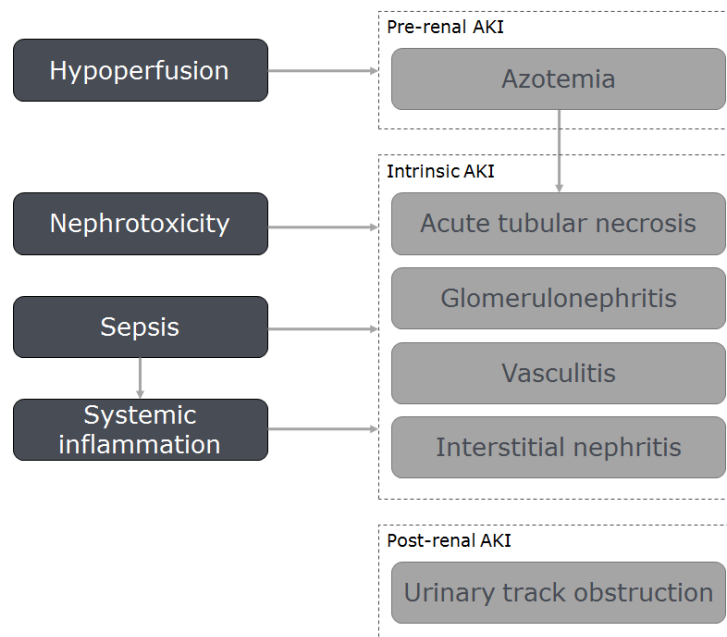


Figure 1.2 Main causes of AKI.

Although the causes of AKI are diverse, they are not reflected in current approaches to its diagnosis. Indeed, AKI is defined by an increase in serum creatinine or a decline in urine output (oliguria) according to the most recent consensus-based criteria by the international Kidney Disease: Improving Global

Outcome (KDIGO) working group [26]. KDIGO stratifies AKI in three stages of ascending severity (AKI 1, 2 and 3), according to the level of increase in serum creatinine as compared to its baseline or to the level of oliguria (Table 1.1).

AKI is highly prevalent in critically ill patients and is associated with increased risk of morbidity, including development of chronic kidney disease (CKD) or end-stage renal disease (ESRD), increased risk of mortality, and with high financial costs [22, 27–29]. Recognizing patients at risk, optimization of hemodynamics and prevention of nephrotoxicity remain the cornerstone in the prevention of AKI [30]. Indeed, treatment is largely supportive and consists of general measures [26]: early optimization of fluid status, maintenance of perfusion pressure, and avoidance or discontinuation of potentially nephrotoxic drugs. In case of severe loss of kidney function, intermittent or continuous renal replacement therapy (RRT) can be used to filter the blood.

Early diagnosis of AKI remains a major clinical challenge [31]. Its imprecise early identification could partially explain why the search for strategies and interventions to mitigate the course of AKI has been unsuccessful [32–37]. The AKI definition is based on late and non-specific markers of the underlying pathophysiological complication. Therefore, these markers have only a limited ability in predicting patients at risk for AKI.

AKI stage	Serum creatinine	Urine output
1	1.5-1.9 * baseline or ≥ 0.3 mg/dl in 48h	< 0.5 ml/kg/h for 6-12h
2	2.0-2.9 * baseline	< 0.5 ml/kg/h for ≥ 12 h
3	≥ 3 * baseline or ≥ 4.0 mg/dl or initiation of RRT or EGFR < 35 ml/min/ 1.73m^2 in patients aged < 18 y	< 0.3 ml/kg/h for ≥ 24 hr or anuria for ≥ 12 h

Table 1.1 AKI classification [26]

Biomarkers have been proposed for AKI stratification and prediction. The most prominently studied biomarker is Neutrophil gelatinase-associated lipocalin (NGAL) [38, 39]. NGAL is a protein expressed in low concentration by the kidney, lung and gastrointestinal tissue under normal physiological conditions. Following an ischemic or a nephrotoxic insult, NGAL is up-regulated in the kidney, thus increasing both urine and plasma NGAL concentration. Therefore, increased levels of NGAL upon ICU admission or during the first day

could be an early sign of future development of AKI. However, the performance of NGAL for that purpose varies from very good to unacceptable depending on the considered patient population [33, 40, 41]. Potential explanations lie in the nature of the protein, expressed in various tissues, hence being non-kidney specific, and in the pathophysiology of AKI, being broader than ischemia and nephrotoxicity.

Hence, up to now, the added clinical value of NGAL and of biomarkers in general appears to be limited, their measurement timing and the optimal population target are unclear, and their quantification costly [33, 42]. Prediction models based on patient information routinely collected at the bedside could be a cost-effective alternative to biomarkers for AKI prognostication [31, 42, 43].

1.3.2 *Post-operative care after corrective surgery for a congenital heart defect*

Congenital heart defect (CHD) is a structural abnormality of the heart that is present at birth [44]. The defect may affect different parts of the heart, including the valves, the myocardial wall and/or the great vessels. CHD can be diagnosed prenatally during the screening echography of the fetus. When CHD is not detected before birth, some children are symptomatic and pediatricians will be alerted to actively look for the diagnosis. When asymptomatic, CHD can be a coincidental finding on physical examination, for instance when a cardiac murmur is found on auscultation. The exact diagnosis is usually established through echocardiography, and sometimes cardiac catheterization is necessary [45]. In case of cyanotic defect, the blood oxygen saturation from the general circulation is low as part of the blood bypasses the lungs and fails to be oxygenated. In general, cyanotic diseases are associated with increased risk of morbidity and mortality.

The incidence of CHD varies across countries, likely because genetic and environmental factors play a role in its occurrence. The overall incidence of CHD is 8-9 per 1000 live births, affecting a total of 1.35 million of newborns each year [46]. Corrective cardiac surgery is often necessary. Before and after surgery, patients might receive appropriate care to prevent hemodynamic deterioration in the ICU.

Although the outcome of children with congenital heart disease undergoing cardiac surgery has improved over the past years, significant morbidity and mortality still exist. Inadequate tissue perfusion and oxygenation are often associated with their poor prognosis [47]. Commonly used techniques to monitor the balance between tissue oxygen delivery and consumption are either invasive or not-continuous [47]. In pediatric patients, because of their small body size, invasive cardiac output monitoring is not routinely used.

Additionally, a fall in blood pressure can be a late indicator of hemodynamic deterioration. It has been hypothesized that non-invasive monitoring of cerebral tissue oxygen saturation (SCTO₂) via a near-infrared (NIR) based cerebral oximeter could provide an early indication of critical changes in the patient's hemodynamic status that may adversely affect the brain [48].

Near-infrared spectroscopy (NIRS) is a technique used to monitor tissue oxygen saturation non-invasively and continuously [49]. Determination of tissue oxygen saturation is performed by measuring the difference in intensity between a transmitted and a received light delivered at specific wavelength [50]. Specifically, an electrode with a light emitter and one to several receivers are placed over the organ to be monitored (Figure 1.3). The emitter projects NIR light between 650 nm and 900 nm into the tissues, which are relatively transparent to these wavelengths, allowing the light to penetrate several centimeters. In the tissue, a proportion of the light will be absorbed by chromophores with an absorption spectrum in the NIR range [50].

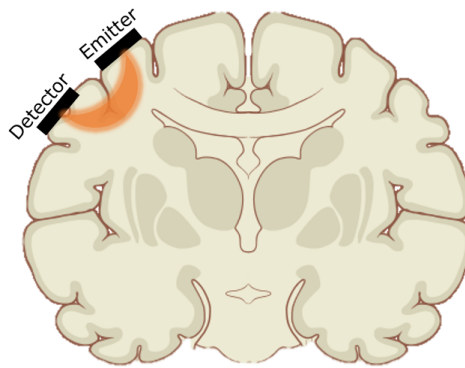


Figure 1.3 Schematic representation of the propagation of NIR light in the brain. The orange shadow indicates the most likely traveled path by the detected photons¹.

Deoxyhemoglobin, oxyhemoglobin and cytochrome c oxidase are light-absorbing molecules involved in the oxygenation process. The principle behind the cerebral NIRS technology is to take advantage of the fact that an attenuation in light from the NIR range is caused by a change in concentration between oxy-, deoxy-hemoglobin and cytochrome C oxidase. As each of these chromophores present a slightly different absorption range, using different wavelengths, the difference of concentration for each chromophore can be established.

Due to their appealing non-invasiveness, in recent years, several pediatric ICUs started to use such tissue oxygen monitors for hemodynamic monitoring [49] (Figure 1.4), as hemodynamic instabilities contribute significantly to further

¹Adapted from Wikimedia Commons, Patrick J. Lynch, medical illustrator [CC BY 2.5].

organ failure. However, there are no guidelines on how NIRS monitoring should be used in clinical practice and studies showing its benefits in post-operative care settings are lacking [51–54]. Additionally, different brands of NIRS oximeters are not comparable as they use different algorithms, thus they report different oxygenations levels [55]. Although a non-invasive monitor that provides early signs of hemodynamic deterioration would clearly be an asset, a study showing the performance of $\text{ScT}O_2$ for this purpose is currently lacking.

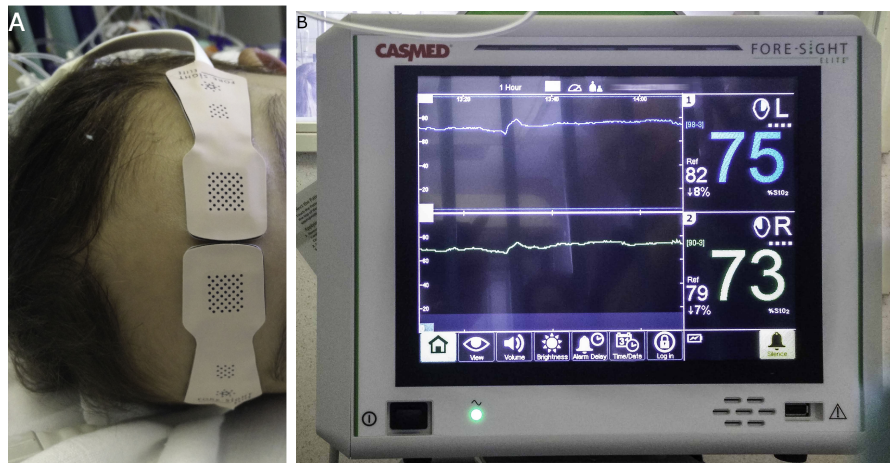


Figure 1.4 Example of NIRS monitoring in the pediatric intensive care unit (PICU).

Whether NIRS cerebral oximetry can detect or foresee potentially deleterious events of impaired brain perfusion and oxygenation during ICU stay, and whether these events are independently associated with worse clinical outcomes, remains unclear in pediatric postoperative care.

1.3.3 Traumatic brain injury

Traumatic brain injury (TBI) is defined as an alteration in brain function, or other evidence of brain pathology, caused by an external force [56]. This can occur in traffic, at home, at work, during sports activities, or on the battlefield. This results in TBI being one of the most important health care problems worldwide. The World Health Organization estimates that it will surpass many diseases as the major cause of death and disability by 2020 [57]. TBI affects approximately 2.5 million people each year and yields 75 000 deaths [58]. In the United States only, the annual burden of TBI is estimated more than \$75 billion [59].

TBI varies in severity from mild (including concussion) to moderate and severe TBI (Glasgow Coma Scale ≤ 8) [60]. In the hours and days following the initial traumatic insult (or *primary injury*), destructive and self-propagating

biological changes in the brain can lead to subsequent additional damage, referred to as *secondary injury* [61]. Prevention and treatment of secondary injury is the main goal of neuro-critical care in patients with TBI, and for that purpose, continuous and usually invasive monitoring is applied. Because the brain is enclosed and protected by the rigid skull, added intracranial volume, caused for example by hemorrhage or brain edema, will eventually lead to a rise in intracranial pressure (ICP). The cerebral perfusion pressure (CPP) is the driving pressure gradient for blood to enter the brain, calculated as the difference between mean arterial blood pressure (MAP) and ICP. When CPP is critically low, brain perfusion is at risk, leading to secondary ischemic injury.

International guidelines for the management of critically ill patients with TBI emphasize the prevention of secondary insults, and for patients with severe TBI, optimization of cardiorespiratory physiology, control of ICP, and maintenance of CPP [62, 63]. To this end, ICU management (Table 1.2) consists of sedation, hyperosmotic infusions, hyperventilation, drainage of cerebrospinal fluid, or temperature control. Additional more aggressive measures are referred to as third-tier therapies and involve aggressive cooling, deep sedation, intensive hyperventilation, and decompressive craniectomy [64]. All these treatments have potential side effects that have to be taken into account when deciding on patient therapy. In particular, third-tier therapies should be considered as last resort therapies in case all other options have been exhausted.

A number of neuromonitoring modalities are available to detect secondary brain injuries, the most used being ICP monitoring. ICP is usually monitored using an invasive intraparenchymal or intraventricular catheter. Some techniques have been introduced to monitor ICP non-invasively [66]. Albeit promising, they suffer from drawbacks, being intermittent monitoring or lack of clinical validation, which limit their clinical application [65]. Consequently, due to the risks associated with invasive monitors, ICP monitoring is only recommended in patients with a severe TBI.

International guidelines recommend ICP lowering therapies at a threshold of 20-25 mmHg, based on data from observational studies where increased ICP is clearly associated with worse outcomes [61, 63, 67–69]. However, it has been observed that too aggressive ICP management might lead to an increased use of sedatives, barbiturates, vasopressors, and fluids, possibly prolonging mechanical ventilation and ICU stay without necessarily improving outcome [70]. In addition, randomized controlled trials in patients with TBI have failed to demonstrate the outcome benefit of aggressive strategies aimed at keeping the ICP below 20 mmHg [71, 72]. Although most experts agree that ICP monitoring should still be standard after severe TBI [73], there is an urgent need for better

Therapeutic measure	Effect
Sedation	To decrease metabolic activity
Drainage of cerebrospinal fluid	To reduce intracranial volume
Hyperosmolar therapy	To reduce brain oedema
Hyperventilation	To reduce intracranial volume through hypocapnic cerebral vasoconstriction
Deep sedation	To achieve deep metabolic suppression
Intensive hyperventilation	To further reduce intracranial volume through hypocapnic cerebral vasoconstriction
Decompressive craniectomy	To accommodate brain swelling by removing a portion of the skull
Temperature control / Aggressive cooling	Various effects, including decrease of metabolic activity

Table 1.2 Therapeutic measures with their intended actions in TBI management, ordered from the least aggressive (top) to the most aggressive (bottom) [65].

definitions of secondary insults and of therapeutic thresholds for ICP [65, 70, 74].

To this end, several avenues can be considered. First, not merely an ICP rise above a certain threshold, but rather the cumulative *dose*, referring to the combination of ICP duration and intensity, is associated with poor neurological outcome [75–77]. The ICU research group from UZLEUVEN has investigated the association between the Glasgow Outcome Score (GOS) and the dose of elevated ICP in subpopulations of adult and pediatric TBI patients [78]. They found an exponential transition curve between the zones associated with good (higher GOS categories) or poor (lower GOS categories) neurological outcomes (Figure 1.5), suggesting that secondary brain insults are defined by both intensity and duration of episodes of increased ICP. The term *ICP-insult* refers to such an ICP dose associated with worse clinical outcomes.

Second, cerebrovascular pressure autoregulation (CAR), which refers to the physiological mechanism by which the brain is able to maintain a constant cerebral blood flow, likely plays a role in the capacity to tolerate insults of increased ICP or reduced CPP [79–82]. CAR is often impaired in patients with severe TBI [66, 83–86]. Computational methods can be used to assess the dynamic rate of CAR continuously, such as the Pressure Reactivity Index

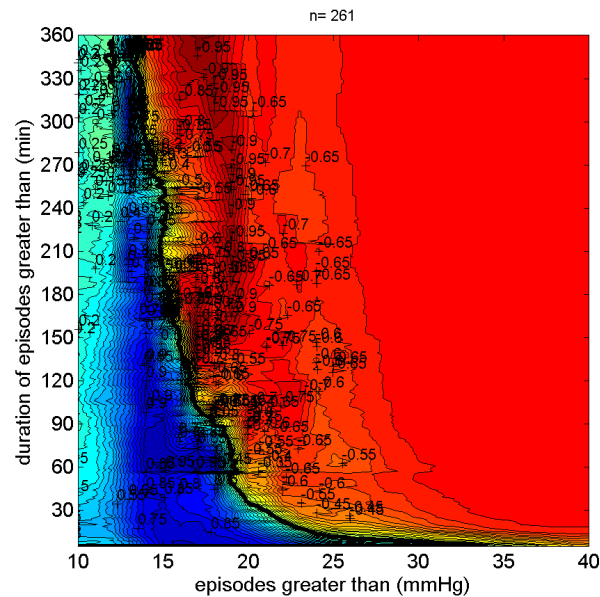


Figure 1.5 Visualization of correlation between GOS and average number of ICP doses per GOS category [78].

Dark red episodes mean that such ICP doses, on average, are associated with worse outcome; dark blue episodes mean that such ICP doses, on average, are associated with better outcome. The black line corresponds to the transition curve between regions of good or poor outcomes.

(PRX) [79], or the Low Frequency Autoregulation Index (LAX) [80]. These indices are calculated as a moving correlation coefficient between ICP and MAP, respectively sampled at a higher and a lower frequency. Other available indices of cerebral autoregulation exist such as the mean flow index (MX). This index is derived from the association between cerebral blood flow estimated using transcranial Doppler and either cerebral perfusion pressure (MX_c) or arterial blood pressure (MX_a) sampled at high frequency. Studies have demonstrated that PRX or MX greater than 0.3 indicates impaired autoregulation and lower than 0-0.05 functional autoregulation [87, 88]. The interval between 0-0.05 and 0.3 is referred to as a "grey zone" as no clear association with outcome emerged. The ability to tolerate episodes of elevated ICP is significantly reduced when CAR is impaired [78].

Third, ICP and CPP thresholds are likely not the same in all age groups. Indeed, lower thresholds have been identified for children in an observational study [76] and retrospectively [78].

Therefore, there is an urgent need for better understanding of secondary insults and of therapeutic thresholds in terms of ICP and CAR, in pediatric and adult populations.

Figure 1.7 schematizes the different steps of development of a prediction model. Each of the steps is described in the following sections.

1.4.1 *Feature selection*

During the modeling process, a right balance has to be found between a good fit of the model and good generalizability, i.e. achieving similar performance in unseen data. A possible problem when developing a model is that the model gets really good at predicting cases from which it has learned, i.e. model performance is good in the development cohort but not in unseen data. This problem is called *overfitting*.

The variables included in the prediction model are named features. The higher the number of features from the development cohort included in the model, the higher the chances the model will overfit (Figure 1.8).

To select the modeling features of interest, several avenues can be considered. First, prior clinical knowledge can be used to make a first selection of relevant features. For instance, in the case of the prediction of a medical complication, a thorough literature search can identify the features already known to be associated with the complication. Second, in case of secondary use of data from a retrospective study, the size of the database will limit the number of features available for modeling. Finally, statistical techniques such as stepwise feature selection can be used to determine a subset of the features most associated with the outcome of interest.

Stepwise feature selection are automated statistical methods to include only the most significant features in the model. Significance of the features are determined based on, among others, P-values for logistic regression or feature importance for decision trees (further explained in section 1.4.2). The stepwise process consists of learning a model with a different number of features at each step. Three different types of stepwise selection exist: forward, backwards and forward-backwards.

In forward feature selection, the model starts without any feature and the most significant feature is included at each step. In backward feature elimination, the model starts including all available features and the least significant feature is removed at each step. Finally, in forward-backwards feature selection, forward selection and backward elimination are combined in an iterative procedure.

It is necessary to define a stopping criterion for further inclusion/removal of features. Such stopping criterion can be based on the feature significance, for

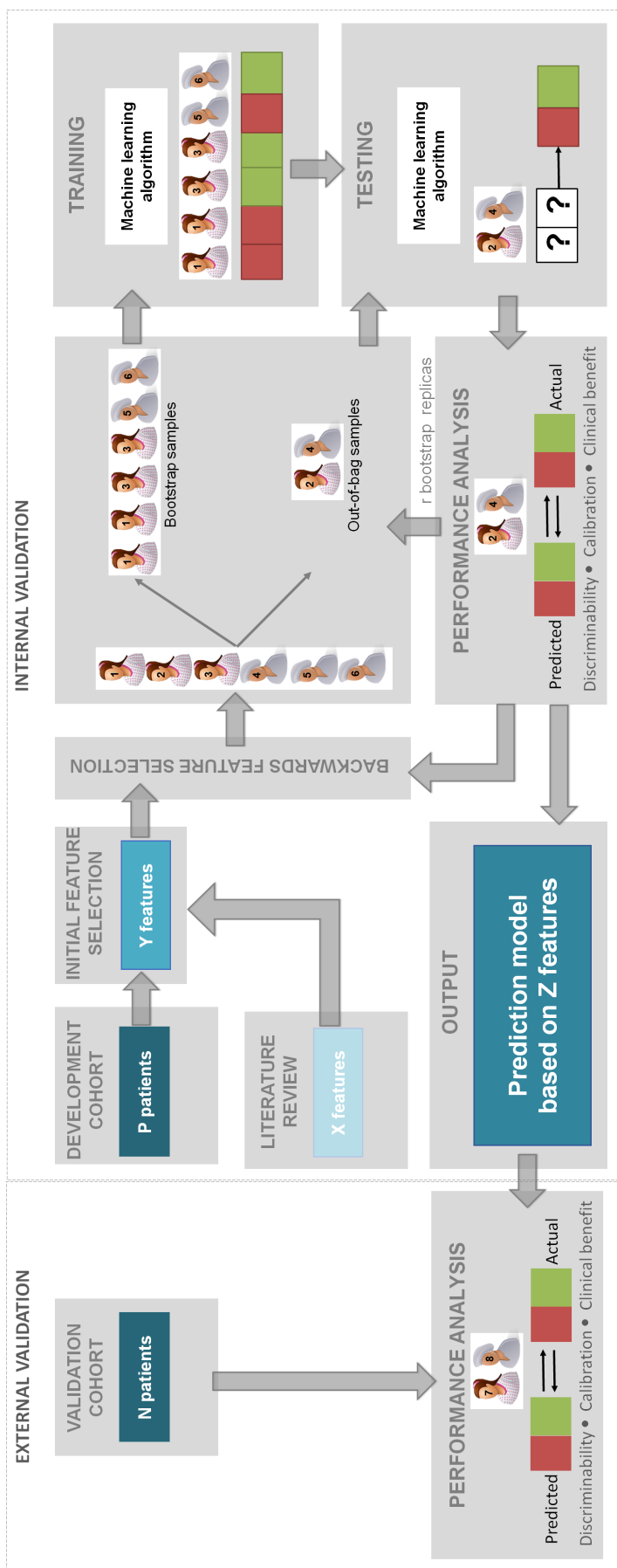


Figure 1.7 Overview of methodology to develop a prediction model.

using a logistic function (Figure 1.9). Mathematically, the probability that a data instance \mathbf{x} belong to class $y = 1$ is given by

$$P(y = 1|\mathbf{x}) = F\left(\sum_{k=0}^l w_k x_k\right)$$

where $F(x)$ is the logistic function (Figure 1.9) and w_k are the weights given to the features x_k .

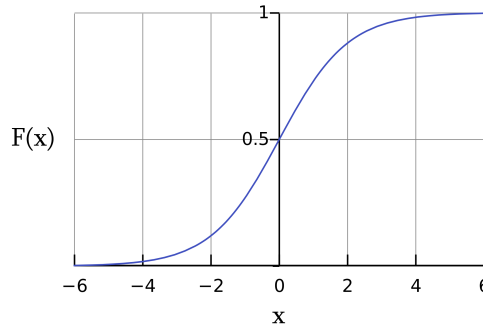


Figure 1.9 Plot of the logistic function $f(x) = \frac{1}{1+e^{-x}}$.

The probability that an event will occur divided by the probability that it will not occur is called the odds. The change of odds of an outcome is measured using odds ratio. The coefficient of each feature in the logistic regression determines the odds ratio for the outcome associated with a 1-unit change of the considered feature variable. Therefore, the odds ratio provides a measure of the magnitude of the association of each feature with the outcome. As such, logistic regression can be used to adjust for confounding factors to assess whether the association between the feature of interest and the outcome remains or is distorted by the confounding factor(s) [92]. A logistic regression with multiple features is referred to as multivariable logistic regression.

Logistic regression has several limitations [92]. First, the method assumes the features are independent from each other. In case several features are collinear (i.e. they carry overlapping information), a minor change in one of these features can have a considerable and unpredictable impact on the odds ratio of the other feature(s). Second, logistic regression assumes that the log of the odds ratio of each feature is linearly associated with the outcome. Therefore, each feature is expected to have the same magnitude of association with the outcome across its entire range. When this is not the case, transforming the continuous feature in categorical features is necessary. For instance, age does not have a consistent association with mortality and age may need to be split in categories $< 50y$, $50-60y$, $60-70y$, $>80y$. Logistic regression is therefore also sensitive to

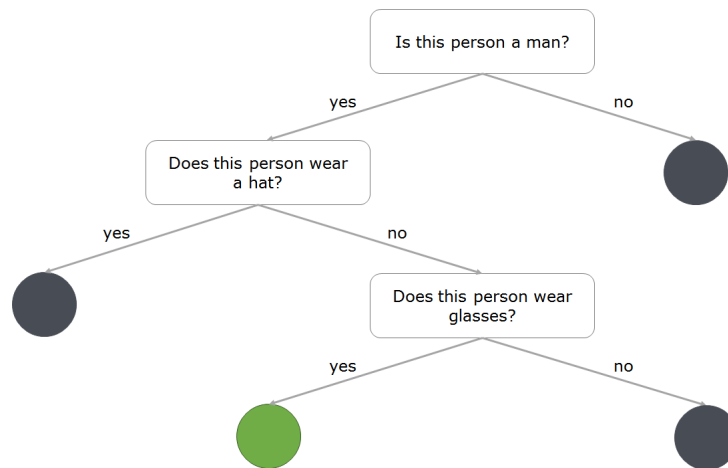


Figure 1.10 Example of a decision tree from the game “guess who”.

outliers. Third, logistic regression is unable to model interaction terms, i.e. when the modification of a value of a feature affects another feature. Therefore, interaction terms have to be identified and explicitly encoded as features. These limitations have to be taken into consideration when interpreting the odds ratio. Additionally, when its assumptions are not met, more complex predictive modeling algorithm should be considered.

DECISION TREES A decision tree is a tree-based classifier [93]. Its structure is similar to how humans perform classification. For instance, in the game “Guess Who?”, each player tries to deduce a selected character by asking questions once at a time. The organization of these questions follows a decision tree (Figure 1.10). The sequence of questions leads to classifying each character between being the person of interest or not. In this particular scenario, only one character is selected.

A decision tree starts at a root node where a question is answered to split the data to be classified, in two distinct branches. The answer will lead to the next node where the data are further split by an additional question (decision rule). Data will be split until reaching a terminal node (leaf node) where the data are assigned to a particular class.

Several algorithms exist to construct decision trees automatically, based on the training data [91, 93]. Generally, trees are built in a recursive fashion. At each node, a decision rule is selected to split the data, randomly or by maximizing a particular split criterion such as the information gain or the Gini impurity. The two datasets created from the split are recursively split, creating different levels of nodes. When reaching a stopping criterion, a leaf node is created with the remaining data and assigned to a class. The stopping criterion

can be specified as the maximum depth of the tree, the minimum number of samples to split a node, the minimum number of samples per leaf, or it can be based on the split criterion among others. When no stopping criterion is specified, the tree is created until each leaf contains a single data point or a group of data points belonging to a similar class.

Decision trees can handle non-linear and multi-output problems. One of their main advantages is that they are easily interpretable and they can be visualized. Their drawback is that they tend to overfit by creating over-complex trees if it is allowed to become too deep. Pruning can be used to reduce the complexity of the tree, by removing sections of the tree that carry little power to classify data. To reduce further the chance of overfitting, one can specify to use only a limited number of features to build the decision rule at each node. An additional disadvantage is that decision trees are unstable as a small variation in the training data can lead to the modeling of a completely different tree.

RANDOM FORESTS Random forests belong to the category of ensemble algorithms, which combines multiple machine-learning algorithms to improve their performance. As its name suggests, a random forest is a combination of decision trees [94]. Performance and robustness are improved by building each decision tree on a bootstrap replica (i.e. random selection of samples with replacement, explained in section 1.4.4) from the development set and by averaging the predicted probabilities of the different trees. Therefore, an additional optimization parameter to reduce overfitting is the number of trees contained in the forest, in addition to the tree-specific parameters. Random forests are very popular in particular thanks to their robustness to noise. An example of random forest is illustrated in Figure 1.11.

1.4.3 *Evaluation metrics*

Once the prediction model has been developed, its performance needs to be evaluated. Several statistical metrics can be used for that purpose. Each of them reports information on the model's capabilities and weaknesses, and they can be grouped in different categories: discrimination, calibration and net benefit.

As illustrated in Table 1.3, a correct classification by the model is called true positive (TP) or true negative (TN), respectively if the actual outcome is positive or negative. An incorrect classification by the model is called FP or FN, when the actual outcome is negative or positive, respectively.

Discrimination refers to how well the predictions allow to discriminate between the patients with and without the outcome. Traditional measures of

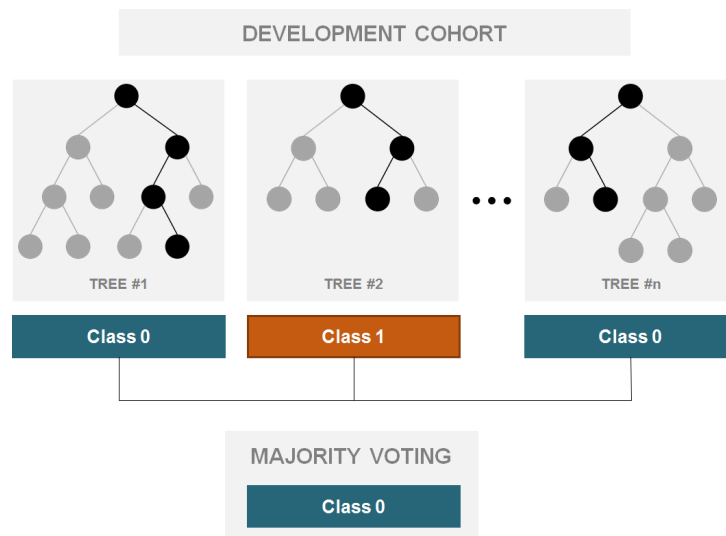


Figure 1.11 Example of a random forest with majority voting.

		Actual condition	
		True	False
Predicted condition	True	TP	FP
	False	FN	TN

Table 1.3 Type of classification.

discrimination include sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). These measures require choosing a threshold to classify patients as having a low or high risk of developing the outcome of interest. Hence, above this classification threshold, the score is positive. Sensitivity is the ability of the model to identify correctly the patients with the outcome of interest. Specificity is the ability of the model to identify correctly the patients without the outcome of interest. The PPV reflects how likely it is that a patient has the outcome of interest given that the predicted probability is higher than the classification threshold. Finally, the NPV reflects how likely it is that a patient does not have the outcome of interest given that the predicted probability is lower than the classification threshold.

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

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

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

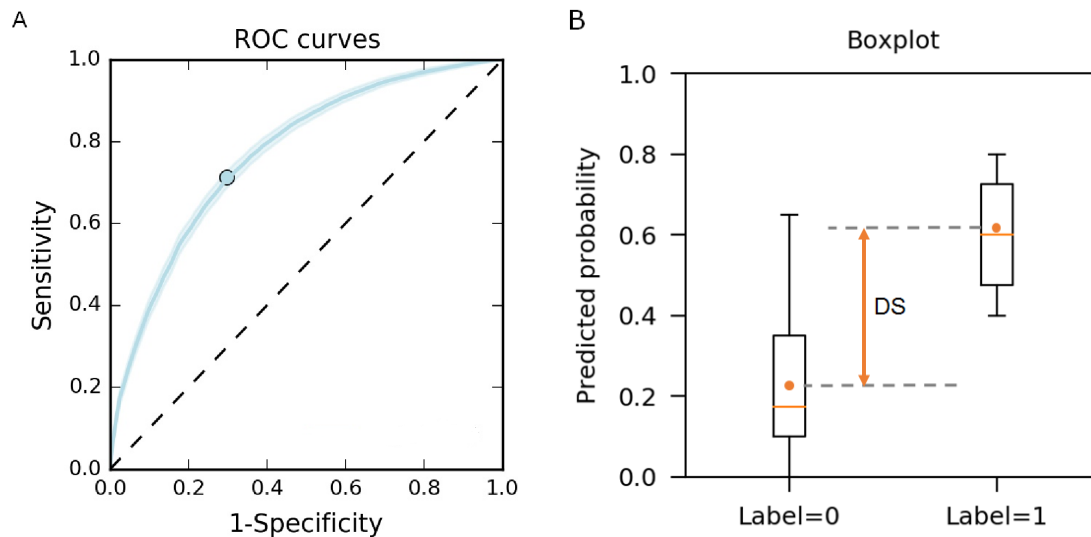


Figure 1.12 Illustration of ROC curve and DS.

(a) ROC curve of a prediction model with its 95% confidence interval. The dot on the curve shows the classification threshold maximizing sensitivity and specificity. Finally, the dotted line represents the ROC curve of a randomly guessing model. (b) Boxplot separating the predicted probabilities by the model according to their class label. The DS of the model corresponds to the differences between the mean predictions from the two classes.

Discrimination can also be evaluated visually with the receiver operating characteristic (ROC) curve, and quantified with the area under the receiver operating characteristic curve (AUROC) and the discrimination slope (DS). The ROC curve plots the sensitivity versus 1-specificity for all possible classification thresholds (Figure 1.12a). The classification threshold maximizing sensitivity and specificity can be calculated from the ROC curve as the point closest to the upper left corner. The DS (Figure 1.12b) measures how well the patients with (label=1) and without (label=0) the outcome are separated. It is calculated as the absolute difference in average predictions for those with and without the outcome. DS ranges from 0 (no discrimination) to 1 (complete discrimination). Finally, the integrated discrimination index (IDI) [95] can be used to compare discrimination between two models in the same patient subset. The IDI is the difference between the models' discrimination slopes and is preferred to the continuous net reclassification index (NRI) when no clear classification cut-off exists [96].

A perfectly discriminating model would maximize sensitivity and specificity and have an AUROC close to 1. In practice, AUROCs above 0.7 are considered adequate [97]. A well discriminating model therefore has higher predicted probabilities for patients with the outcome of interest and lower probabilities for patients without, which translates to a large DS. Alternatively, a trivial

model that consists in randomly guessing the outcome is represented with a curve close to the diagonal on the ROC curve and an AUROC close to 0.5. This trivial model would have similar predictions for patients with and without the outcome of interest, resulting in a DS close to zero.

Calibration refers to the agreement between the population outcomes and the model predictions. For example, if the observed frequency of the outcome of interest is 30 out of 100 patients in the population, a well-calibrated model should predict a 30% risk of developing this outcome for these 100 patients. Calibration can be evaluated visually with calibration curves, which plot the observed proportions in the population versus the model predicted probabilities, or with calibration belts that additionally shows the confidence interval around the curve (Figure 1.13) [98]. The observed proportions in the population are calculated in bins, using a loess algorithm that regresses the binary outcome of each patient into probabilities [99]. Calibration can additionally be quantified by the intercept (calibration-in-the-large) and the slope (calibration-slope) of a linear regression fitted to this curve. When calibration belts are reported, a statistically significant difference from perfect calibration is calculated by a calibration test ($P\text{-value} < 0.05$). Finally, a histogram of the predicted probabilities corresponding to the positive outcome of interest (top) and of predicted probabilities corresponding to negative outcome of interest (bottom) can be represented together with the calibration curve (Figure 1.13)

A well-calibrated model shows strong agreement between the population observed proportions and the probabilities given by the model. Therefore, it has a calibration curve close to the diagonal (calibration test $P\text{-value} \geq 0.05$), a calibration-slope close to 1, and a calibration-in-the-large close to 0.

The **clinical usefulness** of a prediction model refers to the potential clinical benefit achieved by using this model. In particular, it is important to report the clinical usefulness if the harm of misclassifying a patient without the outcome is different from the one of misclassifying a patient with the outcome, or if the benefit of classifying a patient with the outcome is different from the one of classifying a patient without the outcome.

The clinical usefulness can be quantified using the net benefit (NB). NB is the difference between the expected benefit and the expected harm associated with the classification, as shown in the equation below.

$$\text{NB} = \frac{\text{TP}}{n} - \frac{\text{FP}}{n} * \frac{p_t}{1 - p_t}$$

where n is the size of the dataset, the expected benefit ($\frac{\text{TP}}{n}$) is the percentage of patients who have the outcome and were classified as such (true positives) and

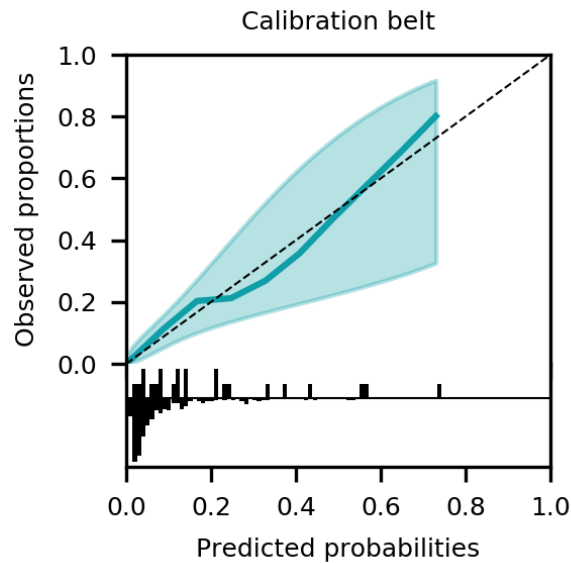


Figure 1.13 Illustration of calibration belt.

Example of calibration belt with the distribution of predictions, from the analysis presented later in Chapter 4. The blue curve shows the calibration curve with its 95% confidence interval (belt). The dotted line represents perfect calibration. Finally, the distribution of predicted probabilities is shown at the bottom of the graph.

who will therefore benefit from the treatment or measures taken to improve their care, and the expected harm ($\frac{FP}{n} * \frac{p_t}{1-p_t}$) is the percentage of patients who will be wrongly treated because they were classified as having the outcome (false positives) multiplied by a weighting factor depending on the classification threshold (p_t). The classification threshold (also called risk threshold) is chosen depending on the relative weight of harm and benefit associated with the classification. For instance, if the treatment associated with having the outcome can be considered harmful (e.g. has undesirable side effects, is very costly, etc.) it is important to have few false positive classifications; this corresponds to using a high classification threshold. Alternatively, if the treatment associated with having the outcome is benign, the number of false positive classifications is less important; this corresponds to using a low classification threshold. The weight given to any potential harm or benefit associated with the actions resulting from a classification is center, patient and task dependent, and therefore so is the choice of classification threshold.

The clinical usefulness can be visualized using decision curves (Figure 1.14). These decision curves plot the net benefit at each possible classification threshold [100]. For a model to be clinically useful at a given threshold it must have a positive NB, which means that it should improved as compared to trivially classifying all patients as not having the outcome (referred to as treat-

none strategy). In Figure 1.14, the treat-none curve corresponds to the line of 0 net benefit. The difference between the models' NB and the NB of classifying all patients as not having the outcome (treat-none strategy) is denoted $\Delta \text{NB}_{\text{None}}$. Second, it must have a higher net benefit than trivially classifying all patients as having the outcome (referred to as treat-all strategy). The difference between the model NB and the NB of classifying all patients as having the outcome (treat-all strategy) is denoted $\Delta \text{NB}_{\text{All}}$.

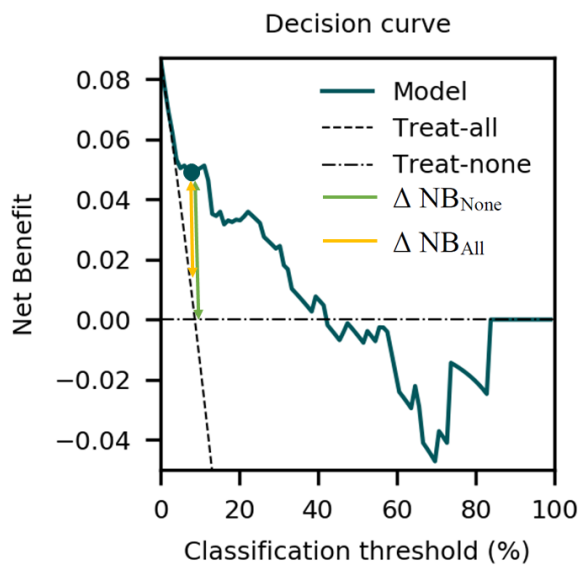


Figure 1.14 Example of a decision curve.

The dot on the curve shows the net benefit of the classification threshold identified in the ROC curve and maximizing sensitivity and specificity. The *treat-all* and *treat-none* lines corresponds to the net benefit achieved when considering that, respectively, all or no patients will develop the predicted outcome. For a particular classification threshold, the NB represents the fraction of patients who will get correctly detected without increasing the number of false positives.

Clinically useful models have larger net benefits than the two trivial strategies for all relevant classification thresholds. The wider the range of clinical usefulness, the more versatility the model offers, as it can be used with a larger number of different classification thresholds. Decision curves therefore allow for a straightforward visual comparison between the clinical usefulness of different models designed for the same task.

1.4.4 Validation techniques

When developing a model, one needs to ensure that this model is valid outside the specific data set on which it has been developed. In other words, one needs to assess whether the model is generalizable.

Internal validation techniques, such as cross-validation or bootstrapping, can be used to reduce bias in estimating model performance and thus help prevent overfitting. In cross-validation, the development set is split into k equally sized partitions (folds). Model training is performed k times using all folds but one, which is kept for testing (Figure 1.15). In leave-one-out cross validation, the number of folds k is equal to the number of samples in the development set. Therefore, testing is performed on each single sample individually. One of the main drawback of cross-validation is that by splitting the development set in folds, power is lost by decreasing the sample size. This technique is not recommended on small datasets.

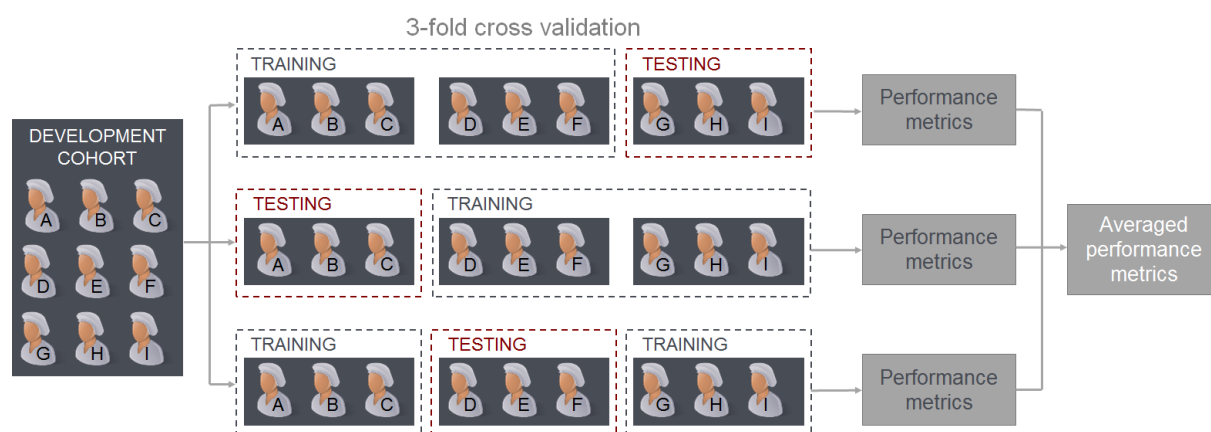


Figure 1.15 Example of 3-fold cross-validation.

The development cohort is split in 3 folds. Iteratively, two of them are used for training and one for testing. The overall performance is calculated by averaging the performance obtained from the 3 models.

In bootstrapping, bootstrap replicas of the same size as the development set are drawn by selecting samples randomly with replacement from the development set (Figure 1.16). A model is trained on each bootstrap replica and tested on the initial development set. The average performance of the trained models on the development set is an estimate of the performance in an unseen population. The confidence interval of the performance provides an estimate of the stability of the model: tight confidence intervals are synonymous with good stability.

A more accurate estimation of the model performance can be obtained by correcting for optimism [101]. Several variants exist, including the 0.632 method [102, 103]. In 0.632 bootstrapping (Figure 1.17), bootstrap replicas are created using the aforementioned methodology: a random selection of samples with replacement from the development set. Using this random process, approximately 63.2% of the samples will be included in each bootstrap replica.

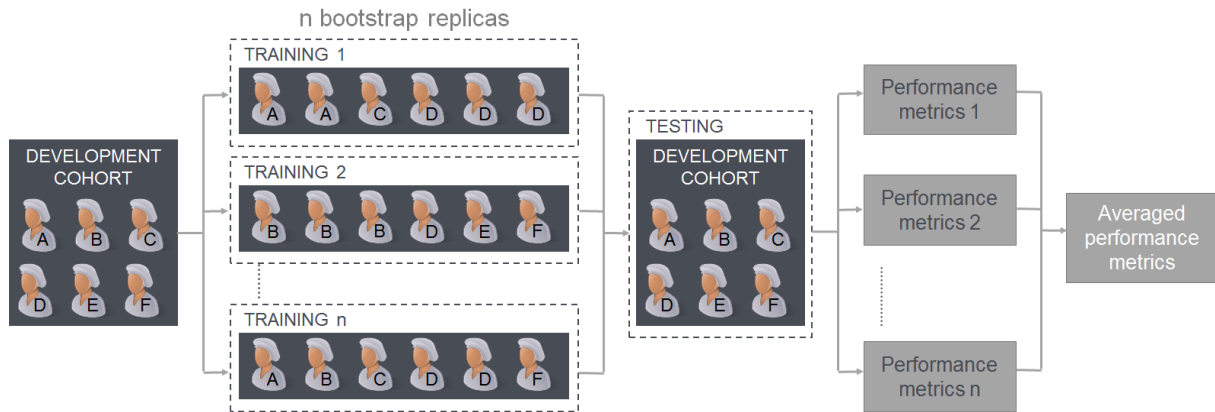


Figure 1.16 Example of bootstrapping.

A large number (n) of bootstrap replicas are created from the initial population. Each bootstrap replica is used to train a model, whose performance is tested in the development cohort. The overall performance is calculated by averaging the performance obtained for the n models.

The remaining samples are called the out-of-bag samples. A model is trained on each bootstrap replica and tested on the out-of-bag samples. An estimate of the performance (Performance_{est}) corrected for optimism can be obtained using the following formula:

$$\text{Performance}_{est} = \text{Performance}_{app} - 0.632(\text{Performance}_{app} - \text{Performance}_{oob})$$

Where Performance_{app} represents the performance of a model trained on the development set and tested on the same dataset and Performance_{oob} represents the performance in the out-of bag sample of a model trained on the bootstrap replica.

By calculating the average of the estimated performance on a large number of bootstraps replicas, a stable estimation of model performance in an unseen population can be obtained. This technique was applied in [104]. The bootstrap approach can be applied to any performance metric. In this thesis, it will be used to report AUROC [103], net benefits and calibration slopes, among others.

To improve the internal validation process, the bootstrap technique can be implemented at the level of feature selection, in addition to the level of model training. Ideally, feature selection will take place in each bootstrap replica to strengthen the stability of the findings and further decrease the risk of overfitting.

Internal validation techniques are necessary to obtain an honest impression of the model performance in a similar population. However, one can only be certain of the model's generalizability by assessing the model's performance in an unseen population. This process is referred to as **external validation**, a crucial but unfortunately uncommon step to validate prediction models [105].

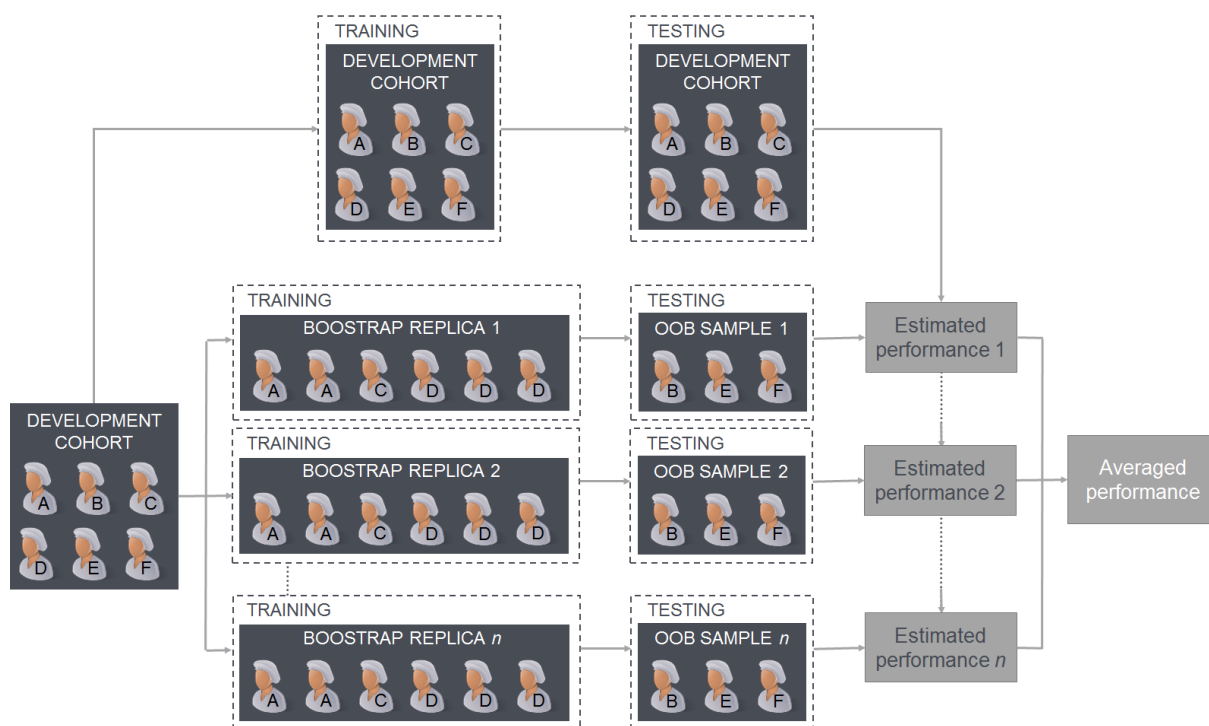


Figure 1.17 Example of 0.632 bootstrapping.

A large number (n) of bootstrap replicas are created from the initial population. Each bootstrap replica is used to train a model, whose performance is tested in the out-of bag (OOB) samples. The estimated performance for each bootstrap replica is calculated by subtracting the optimism from the performance obtained on the development cohort. The overall performance is calculated by averaging the estimated performance obtained for the n models.

The interaction between internal and external validation is schematized in Figure 1.7.

Ideally, the unseen population (validation cohort) is similar to the one from the development set but from a different geographical area, and/or from a different time. In case the unseen population corresponds to a different population, assessing model performance will allow identifying whether adjustments, such as model recalibration, have to be applied to use the model in this new population, or whether the specific model should not be used in such population. For instance, in chapter 3, we developed a model for acute kidney injury in an heterogeneous ICU population and found that the model achieved poorer performance in a population of septic patients.

Sample size is important for both internal and external validation. For internal validation, having sufficient sample size and using a large number of bootstrap replicas is necessary to ensure avoiding overfitting. For external validation, if the sample size is too low, it may result in concluding wrongly that a model performs satisfactorily [106].

Comparison of the performance of different models using external validation can be assessed using bootstrap-generated confidence intervals. First, the performance of each model is calculated for each bootstrap. Second, the distribution of the difference in performance between the models is drawn. Finally, a significant difference at an alpha level of 5% is found if 0 is not included in the 95% confidence interval of the distribution. In this thesis, bootstrap-generated confidence intervals will be used to assess the difference in model discrimination during external validation.

1.5 THESIS OVERVIEW

This thesis is divided into three main parts, each related to a different condition of critical illness. A flowchart of the thesis chapters and the translation of their findings into bedside clinical application can be found in Figure 1.18. The first part is related to the early detection of acute kidney injury and includes Chapters 3 and 4. The second part reports the results of a prospective blinded observational study on the use of near-infrared-based cerebral oximetry in critically ill children, and it includes Chapters 5 and 6. The last part is related to the visualization of secondary brain injuries in patients with a severe traumatic brain injury and it includes Chapters 7 and 8. In detail,

CHAPTER 2 lists the specific objectives of the thesis.

CHAPTER 3 describes the development and validation of AKIpredictor, a machine-learning-based prognostic calculator for acute kidney injury in critically ill adults. It compares its performance against NGAL, one of the most studied biomarkers for AKI prognostication.

CHAPTER 4 studies the clinical application of AKIpredictor by validating the model in an observational prospective study and comparing its performance with the estimated AKI risk by clinicians.

CHAPTER 5 introduces new statistical metrics of near-infrared-based cerebral oximetry and reports their association with acute outcomes in critically ill children after cardiac surgery.

CHAPTER 6 presents a prediction model for acute kidney injury based on routinely collected data from critically ill children after cardiac surgery and analyzes the added value of the near-infrared-based cerebral oximeter.

CHAPTER 7 investigates secondary brain injuries in terms of impaired cerebral autoregulation and present a visualization of their association with 6-month neurological outcomes.

CHAPTER 8 presents the development of a prototype software for visualization of secondary brain injuries from intracranial hypertension and impaired cerebral autoregulation, to be used at the bedside of patients with a severe traumatic brain injury.

CHAPTER 9 summarizes the main findings of this thesis and presents directions for future research or valorization avenues.

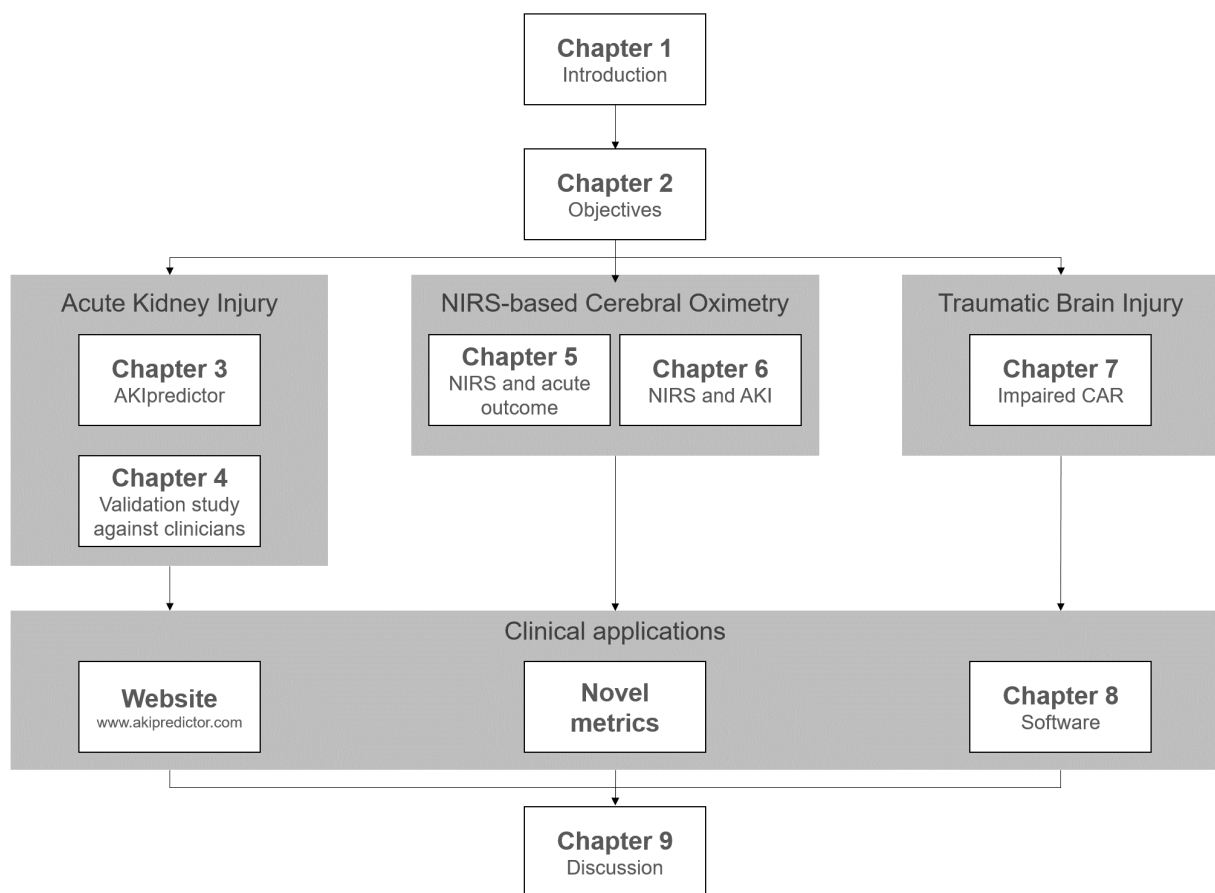


Figure 1.18 Thesis overview.

BIBLIOGRAPHY

- [1] N. K. Adhikari, R. A. Fowler, S. Bhagwanjee, and G. D. Rubenfeld. Critical care and the global burden of critical illness in adults. *Lancet* 376,9749 (Oct. 2010), pp. 1339–1346.
- [2] M. Apkon and P. Singhaviranon. Impact of an electronic information system on physician workflow and data collection in the intensive care unit. *Intensive Care Med.* 27,1 (2001), pp. 122–130.
- [3] J. Fraenkel David, J. Cowie Melleesa, and J. Daley Peter. Quality benefits of an intensive care clinical information system. *Crit. Care Med.* 31,1 (2003), pp. 120–125.
- [4] B. Marr. Why only one of the 5 Vs of big data really matters (Blog). 2015.
- [5] E. Slack. Lowering the Cost of Deleting Data. 2014.
- [6] A. Holzinger and D. Hutchison. *Interactive Knowledge Discovery and Data Mining in Biomedical Informatics*. Vol. 8401. 2014.
- [7] A. L. Beam and I. S. Kohane. Big Data and Machine Learning in Health Care. *JAMA* 319,13 (Apr. 2018), p. 1317.
- [8] W. Raghupathi and V. Raghupathi. Big data analytics in healthcare: promise and potential. *Heal. Inf. Sci. Syst.* 2,1 (2014), p. 3.
- [9] R. Platt and T. Lieu. Data enclaves for sharing information derived from clinical and administrative data. *JAMA* (Aug. 2018).
- [10] A. Vergheese, N. H. Shah, and R. A. Harrington. What This Computer Needs Is a Physician. *JAMA* 319,1 (Jan. 2018), p. 19.
- [11] N. A. Halpern. Early Warning Systems for Hospitalized Pediatric Patients. *Jama* (2018), pp. 27–28.
- [12] S. Bailly, G. Meyfroidt, and J.-F. Timsit. What's new in ICU in 2050: big data and machine learning. *Intensive Care Med.* (Dec. 2017), pp. 1–4.
- [13] N. Stocchetti et al. Neuroprotection in acute brain injury: an up-to-date review. *Crit. Care* 19,1 (Dec. 2015), p. 186.
- [14] D. C. Angus. Fusing Randomized Trials With Big Data: The Key to Self-learning Health Care Systems? *JAMA* 314,8 (2015), pp. 767–768.
- [15] C. Lazaridis et al. Alternative Clinical Trial Design in Neurocritical Care. *Neurocrit. Care* 22,3 (2015), pp. 378–384.

- [16] M. W. Semler et al. Balanced Crystalloids versus Saline in the Intensive Care Unit. The SALT Randomized Trial. *Am. J. Respir. Crit. Care Med.* 195,10 (May 2017), pp. 1362–1372.
- [17] M@TRIC. <http://matric.be/>.
- [18] M. Saeed, C. Lieu, G. Raber, and R. G. Mark. MIMIC II: a massive temporal ICU patient database to support research in intelligent patient monitoring. *Comput. Cardiol.* 29, (2002), pp. 641–644.
- [19] CENTER-TBI. <https://www.center-tbi.eu/>.
- [20] F. Güiza, J. Van Eyck, and G. Meyfroidt. Predictive data mining on monitoring data from the intensive care unit. *J. Clin. Monit. Comput.* 27,4 (Aug. 2013), pp. 449–53.
- [21] J. Labarère, B. Renaud, R. Bertrand, and M. J. Fine. How to derive and validate clinical prediction models for use in intensive care medicine. *Intensive Care Med.* 40,4 (Apr. 2014), pp. 513–27.
- [22] R. Bellomo, J. A. Kellum, and C. Ronco. Acute kidney injury. *Lancet* 380,9843 (Aug. 2012), pp. 756–766.
- [23] N. H. Lameire et al. Acute kidney injury: An increasing global concern. *Lancet* 382,9887 (2013), pp. 170–179.
- [24] J. A. Kellum and J. R. Prowle. Paradigms of acute kidney injury in the intensive care setting. *Nat. Rev. Nephrol.* 14,4 (Jan. 2018), pp. 217–230.
- [25] E. Macedo and R. L. Mehta. Prerenal failure: from old concepts to new paradigms. *Curr. Opin. Crit. Care* 15,6 (2009), pp. 467–473.
- [26] Kidney Disease: Improving Global Outcomes (KDIGO) Acute Kidney Injury Work Group. KDIGO Clinical Practice Guideline for Acute Kidney Injury. *Kidney Int. Suppl.* 2,2 (June 2012), pp. 1–138.
- [27] J. Gunst et al. Impact of Early Parenteral Nutrition on Metabolism and Kidney Injury. *J. Am. Soc. Nephrol.* 24,6 (June 2013), pp. 995–1005.
- [28] M. Joannidis et al. Acute kidney injury in critically ill patients classified by AKIN versus RIFLE using the SAPS 3 database. *Intensive Care Med.* 35,10 (2009), pp. 1692–1702.
- [29] S. Nisula et al. 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, (2013), pp. 420–8.
- [30] M. Joannidis et al. Prevention of acute kidney injury and protection of renal function in the intensive care unit. *Intensive Care Med.* 36,3 (Mar. 2010), pp. 392–411.

- [31] S. M. Sutherland et al. Utilizing electronic health records to predict acute kidney injury risk and outcomes: workgroup statements from the 15th ADQI Consensus Conference. *Can. J. Kidney Heal. Dis.* 3,1 (Dec. 2016), p. 11.
- [32] L. S. Chawla, P. W. Eggers, R. A. Star, and P. L. Kimmel. Acute Kidney Injury and Chronic Kidney Disease as Interconnected Syndromes. *N. Engl. J. Med.* 371,1 (July 2014), pp. 58–66.
- [33] M. Ostermann and M. Joannidis. Biomarkers for AKI improve clinical practice: no. *Intensive Care Med.* 41,4 (Apr. 2015), pp. 618–622.
- [34] F. T. Billings et al. High-Dose Perioperative Atorvastatin and Acute Kidney Injury Following Cardiac Surgery. *JAMA* 315,9 (Mar. 2016), pp. 877–888.
- [35] A. X. Garg et al. Perioperative Aspirin and Clonidine and Risk of Acute Kidney Injury: A Randomized Clinical Trial. *JAMA* 312,21 (Dec. 2014), pp. 2254–2264.
- [36] P. Young et al. Effect of a Buffered Crystalloid Solution vs Saline on Acute Kidney Injury Among Patients in the Intensive Care Unit. *JAMA* 314,16 (Oct. 2015), pp. 1701–1710.
- [37] F. P. Wilson et al. Automated, electronic alerts for acute kidney injury: a single-blind, parallel-group, randomised controlled trial. *Lancet* 385,9981 (May 2015), pp. 1966–1974.
- [38] D. Shemin and L. D. Dworkin. Neutrophil gelatinase-associated lipocalin (NGAL) as a Biomarker for Early Acute Kidney Injury. *Crit. Care Clin.* 27,2 (Apr. 2011), pp. 379–389.
- [39] K. Kashani et al. Discovery and validation of cell cycle arrest biomarkers in human acute kidney injury. *Crit. Care* 17,1 (Jan. 2013), R25.
- [40] J. Mishra et al. Neutrophil gelatinase-associated lipocalin (NGAL) as a biomarker for acute renal injury after cardiac surgery. *Lancet* 365,9466 (Apr. 2005), pp. 1231–1238.
- [41] Z. Zhang. Biomarkers, diagnosis and management of sepsis-induced acute kidney injury : a narrative review. *Hear. Lung Vessel.* 7,1 (2015), pp. 64–73.
- [42] R. Bellomo et al. Acute kidney injury in the ICU: from injury to recovery: reports from the 5th Paris International Conference. *Ann. Intensive Care* 7,1 (Dec. 2017), p. 49.
- [43] S. Sutherland, S. Goldstein, and S. Bagshaw. Acute Kidney Injury and Big Data. *Contributions to nephrology* 193, (2018), p. 55.

- [44] J. I. Hoffman and S. Kaplan. The incidence of congenital heart disease. *J. Am. Coll. Cardiol.* 39,12 (2002), pp. 1890–1900.
- [45] R. Sun, M. Liu, L. Lu, Y. Zheng, and P. Zhang. Congenital Heart Disease: Causes, Diagnosis, Symptoms, and Treatments. *Cell Biochem. Biophys.* 72,3 (July 2015), pp. 857–860.
- [46] A. C. Fahed, B. D. Gelb, J. G. Seidman, and C. E. Seidman. Genetics of congenital heart disease: The glass half empty. *Circ. Res.* 112,4 (2013), pp. 707–720.
- [47] G. V. Parr, E. H. Blackstone, and J. W. Kirklin. Cardiac performance and mortality early after intracardiac surgery in infants and young children. *Circulation* 51,5 (1975), pp. 867–874.
- [48] H. M. Phelps et al. Postoperative cerebral oxygenation in hypoplastic left heart syndrome after the Norwood procedure. *Ann. Thorac. Surg.* 87,5 (May 2009), pp. 1490–4.
- [49] N. S. Ghanayem and G. M. Hoffman. Near Infrared Spectroscopy as a Hemodynamic Monitor in Critical Illness. *Pediatr. Crit. Care Med.* 17, (2016), S201–S206.
- [50] J. M. Murkin and M. Arango. Near-infrared spectroscopy as an index of brain and tissue oxygenation. *Br. J. Anaesth.* 103,Supplement 1 (2009), pp. i3–i13.
- [51] J. S. Tweddell, N. S. Ghanayem, and G. M. Hoffman. Pro: NIRS is “Standard of Care” for postoperative management. *Semin. Thorac. Cardiovasc. Surg. Pediatr. Card. Surg. Annu.* 13,1 (2010), pp. 44–50.
- [52] J. C. Hirsch, J. R. Charpie, R. G. Ohye, and J. G. Gurney. Near infrared spectroscopy (NIRS) should not be standard of care for postoperative management. *Semin. Thorac. Cardiovasc. Surg. Pediatr. Card. Surg. Annu.* 13,1 (2010), pp. 51–54.
- [53] J. M. Kane and D. M. Steinhorn. Lack of irrefutable validation does not negate clinical utility of near-infrared spectroscopy monitoring: Learning to trust new technology. *J. Crit. Care* 24,3 (2009), 472.e1–472.e7.
- [54] J. Steppan and C. W. Hogue. Cerebral and tissue oximetry. *Best Pract. Res. Clin. Anaesthesiol.* 28,4 (2014), pp. 429–439.
- [55] C. Schmidt et al. The effects of systemic oxygenation on cerebral oxygen saturation and its relationship to mixed venous oxygen saturation: A prospective observational study comparison of the INVOS and ForeSight Elite cerebral oximeters. *Can. J. Anesth. Can. d’anesthésie* 83, (Feb. 2018).

- [56] G. T. Manley and A. I. R. Maas. Traumatic Brain Injury: An International Knowledge-Based Approach. *Jama* 310,5 (2013), pp. 473–4.
- [57] A. A. Hyder, C. A. Wunderlich, P. Puvanachandra, G. Gururaj, and O. C. Kobusingye. The impact of traumatic brain injuries: a global perspective. *NeuroRehabilitation* 22,5 (2007), pp. 341–353.
- [58] A. I. Maas et al. Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI). *Neurosurgery* 76,1 (2015), pp. 67–80.
- [59] V. G. Coronado et al. Trends in traumatic brain injury in the US and the public health response: 1995–2009. *J. Safety Res.* 43,4 (2012), pp. 299–307.
- [60] A. I. R. Maas et al. Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. *Lancet Neurol.* 4422,17 (2017).
- [61] K. N. Corps, T. L. Roth, and D. B. McGavern. Inflammation and Neuroprotection in Traumatic Brain Injury. *JAMA Neurol* (Jan. 2015).
- [62] Brain Trauma Foundation. Guidelines for the Management of Severe Traumatic Brain Injury 3rd Edition. *J. Neurotrauma* 24,212 (2007), pp. 1–106.
- [63] N. Carney et al. Guidelines for the Management of Severe Traumatic Brain Injury, Fourth Edition. *Neurosurgery* 80,1 (2017), pp. 6–15.
- [64] A. I. R. Maas, A. Marmarou, G. D. Murray, S. G. M. Teasdale, and E. W. Steyerberg. Prognosis and clinical trial design in traumatic brain injury: the IMPACT study. *J. Neurotrauma* 24,2 (2007), pp. 232–238.
- [65] N. Stocchetti and A. I. Maas. Traumatic intracranial hypertension. *N. Engl. J. Med.* 370,22 (May 2014), pp. 2121–2130.
- [66] C. Hawthorne and I. Piper. Monitoring of intracranial pressure in patients with traumatic brain injury. *Front. Neurol.* 5,July (Jan. 2014), p. 121.
- [67] L. F. Marshall, R. W. Smith, and H. M. Shapiro. The outcome with aggressive treatment in severe head injuries. Part II: acute and chronic barbiturate administration in the management of head injury. *J. Neurosurg.* 50,1 (Jan. 1979), pp. 26–30.
- [68] T. G. Saul and T. B. Ducker. Effect of intracranial pressure monitoring and aggressive treatment on mortality in severe head injury. *J. Neurosurg.* 56,4 (Apr. 1982), pp. 498–503.
- [69] R. K. Narayan et al. Intracranial pressure: to monitor or not to monitor? A review of our experience with severe head injury. *J. Neurosurg.* 56,5 (May 1982), pp. 650–659.

- [70] P. Le Roux. Intracranial pressure after the BEST TRIP trial: a call for more monitoring. *Curr. Opin. Crit. Care* 20,2 (Apr. 2014), pp. 141–7.
- [71] P. J. Andrews et al. Hypothermia for Intracranial Hypertension after Traumatic Brain Injury. *N. Engl. J. Med.* 373,25 (2015), pp. 2403–2412.
- [72] D. Cooper and J. Rosenfeld. Decompressive craniectomy in diffuse traumatic brain injury. *N. Engl. J. Med.* 364, (2011), pp. 1493–1502.
- [73] P. Le Roux et al. Consensus summary statement of the International Multidisciplinary Consensus Conference on Multimodality Monitoring in Neurocritical Care. *Intensive Care Med.* 40,9 (2014), pp. 1189–1209.
- [74] C. Lazaridis, C. G. Rusin, and C. S. Robertson. Secondary brain injury: Predicting and preventing insults. *Neuropharmacology* (June 2018), pp. 1–8.
- [75] A. Vik et al. Relationship of “dose” of intracranial hypertension to outcome in severe traumatic brain injury. *J. Neurosurg.* 109,4 (Oct. 2008), pp. 678–84.
- [76] I. R. Chambers et al. Critical thresholds of intracranial pressure and cerebral perfusion pressure related to age in paediatric head injury. *J. Neurol. Neurosurg. Psychiatry* 77,2 (Feb. 2006), pp. 234–40.
- [77] S. Kahraman et al. Automated measurement of pressure times time dose of intracranial hypertension best predicts outcome after severe traumatic brain injury. *J Trauma* 69,1 (2010), pp. 110–118.
- [78] F. Güiza et al. Visualizing the pressure and time burden of intracranial hypertension in adult and paediatric traumatic brain injury. *Intensive Care Med.* 41,6 (Apr. 2015), pp. 1067–1076.
- [79] M. Czosnyka, P. Smielewski, P. Kirkpatrick, R. J. Laing, D. Menon, and J. D. Pickard. Continuous Assessment of the Cerebral Vasomotor Reactivity in head injury. *Neurosurgery* 41,July (1997), pp. 11–19.
- [80] B. Depreitere et al. Pressure autoregulation monitoring and cerebral perfusion pressure target recommendation in patients with severe traumatic brain injury based on minute-by-minute monitoring data. *J. Neurosurg.* 120,6 (June 2014), pp. 1451–1457.
- [81] C. Lazaridis et al. Patient-specific thresholds of intracranial pressure in severe traumatic brain injury. *J. Neurosurg.* 120,4 (2014), pp. 893–900.
- [82] M. Jaeger, M. U. Schuhmann, M. Soehle, and J. Meixensberger. Continuous assessment of cerebrovascular autoregulation after traumatic brain injury using brain tissue oxygen pressure reactivity. *Crit. Care Med.* 34,6 (June 2006), pp. 1783–1788.

- [83] J. C. Hemphill, P. Andrews, and M. De Georgia. Multimodal monitoring and neurocritical care bioinformatics. *Nat. Rev. Neurol.* 7,8 (Aug. 2011), pp. 451–60.
- [84] F. Güiza, G. Meyfroidt, T. Y. M. Lo, P. A. Jones, G. Van Den Berghe, and B. Depreitere. Continuous optimal CPP based on minute-by-minute monitoring data: A study of a pediatric population. *Acta Neurochir. Suppl.* 122, (2016), pp. 187–191.
- [85] C. Zweifel et al. Continuous monitoring of cerebrovascular pressure reactivity in patients with head injury. *Neurosurg. Focus* 25,4 (Oct. 2008), E2.
- [86] L. a. Steiner et al. Continuous monitoring of cerebrovascular pressure reactivity allows determination of optimal cerebral perfusion pressure in patients with traumatic brain injury. *Crit. Care Med.* 30,4 (Apr. 2002), pp. 733–738.
- [87] E. Sorrentino et al. Critical Thresholds for Cerebrovascular Reactivity After Traumatic Brain Injury. *Neurocrit. Care* 16,2 (Apr. 2012), pp. 258–266.
- [88] E. Sorrentino et al. Critical thresholds for transcranial doppler indices of cerebral autoregulation in traumatic brain injury. *Neurocrit. Care* 14,2 (2011), pp. 188–193.
- [89] G. Meyfroidt et al. Computerized prediction of intensive care unit discharge after cardiac surgery: development and validation of a Gaussian processes model. *BMC Med. Inform. Decis. Mak.* 11,1 (Jan. 2011), p. 64.
- [90] N. Veronese, Y. Li, J. E. Manson, W. C. Willett, L. Fontana, and F. B. Hu. Combined associations of body weight and lifestyle factors with all cause and cause specific mortality in men and women: prospective cohort study. *BMJ* (Nov. 2016), p. i5855.
- [91] T. Hastie, R. Tibshirani, and J. Friedman. The Elements of Statistical Learning. *Bayesian Forecast. Dyn. Model.* 1, (2009), pp. 1–694.
- [92] J. Tolles and W. J. Meurer. Logistic Regression. *JAMA* 316,5 (Aug. 2016), p. 533.
- [93] L. Breiman, J. H. Friedman, and R. A. Olshen. *Classification and regression trees*. London : Chapman and Hall, 1993.
- [94] L. Breiman. Random forests. *Mach. Learn.* 45,1 (2001), pp. 5–32.

- [95] M. J. Pencina, R. B. D'Agostino, R. B. D'Agostino, and R. S. Vasan. Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. *Stat. Med.* 27,2 (Jan. 2008), pp. 157–172.
- [96] M. J. Leening, E. W. Steyerberg, B. Van Calster, R. B. D'Agostino, and M. J. Pencina. Net reclassification improvement and integrated discrimination improvement require calibrated models: relevance from a marker and model perspective. *Stat. Med.* 33,19 (Aug. 2014), pp. 3415–3418.
- [97] J. Swets. Measuring the accuracy of diagnostic systems. *Science.* 240,4857 (June 1988), pp. 1285–1293.
- [98] G. Nattino, S. Finazzi, and G. Bertolini. A new calibration test and a reappraisal of the calibration belt for the assessment of prediction models based on dichotomous outcomes. *Stat. Med.* 33,14 (2014), pp. 2390–2407.
- [99] P. C. Austin and E. W. Steyerberg. Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. *Stat. Med.* 33,3 (2014), pp. 517–535.
- [100] A. J. Vickers and E. B. Elkin. Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. *Med. Decis. Mak.* 26,6 (Nov. 2006), pp. 565–574.
- [101] E. W. Steyerberg, S. E. Bleeker, H. A. Moll, D. E. Grobbee, and K. G. Moons. Internal and external validation of predictive models: A simulation study of bias and precision in small samples. *J. Clin. Epidemiol.* 56,5 (May 2003), pp. 441–447.
- [102] B. Efron and R. Tibshirani. Improvements on Cross-Validation: The 632+ Bootstrap Method. *J. Am. Stat. Assoc.* 92,438 (June 1997), pp. 548–560.
- [103] B. Sahiner, H.-P. Chan, and L. Hadjiiski. Classifier performance prediction for computer-aided diagnosis using a limited dataset. *Med. Phys.* 35,4 (Mar. 2008), pp. 1559–1570.
- [104] H. Chahal et al. Heart failure risk prediction in the Multi-Ethnic Study of Atherosclerosis. *Heart* 101,1 (Jan. 2015), pp. 58–64.
- [105] G. C. Siontis, I. Tzoulaki, P. J. Castaldi, and J. P. Ioannidis. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *J. Clin. Epidemiol.* 68,1 (Jan. 2015), pp. 25–34.
- [106] E. W. Steyerberg. Validation in prediction research: the waste by data-splitting. *J. Clin. Epidemiol.* (2018).

OBJECTIVES

2.1 GENERAL AIM

The general aim of this thesis is to investigate the application of big data analytics to develop decision support applications for critically illness, that are validated, and can be used in clinical practice. The analytics will be applied in fields where alternative methods have failed to demonstrate benefit to improve patient care, or for which there is a need for knowledge discovery. When alternative methods exist, we hypothesized that data-driven analytics could solve the existing need better than these alternative methods. The developed analytics should achieve sufficient performance and robustness to be translated to the patient bedside where they can potentially show clinical usefulness.

2.2 SPECIFIC OBJECTIVES

More specifically, this project has focused on 3 highly prevalent conditions in the intensive care unit, with important clinical consequences. A specific objective is related to each condition.

OBJECTIVE 1 The first objective is to **develop machine-learning-based prediction models** for the development of **acute kidney injury** during the first week of intensive care unit stay, and to **compare their performance** with existing acute kidney injury **biomarkers**. In addition, we will investigate whether these models can be made available through an **open-access web-application**. Finally, we will design and conduct an **observational prospective blinded study** to compare the performance of the developed models with clinicians and assess their potential **clinical benefit**.

OBJECTIVE 2 The second objective is to **investigate the added and independent clinical usefulness of near-infrared-based cerebral oximetry** for the post-operative care of children after cardiac surgery. Using data from a prospective observational blinded study, we will investigate which properties

of the cerebral oximeter signal are predictive, first for **intensive care unit and hospital outcomes**, and second for **acute kidney injury**, a prevalent clinical condition in these children, with important prognostic implications.

OBJECTIVE 3 The third objective is to investigate whether episodes of **deficient cerebral autoregulation** are associated with worse clinical outcomes, and can be defined as a subtype of acute **secondary brain injury**, in patients with severe traumatic brain injury. We will investigate the association between 6-month neurological outcome and the dose of cerebral autoregulation deficiency in a large database of patients with severe traumatic brain injury. Second, we will develop a **prototype of a bedside monitor** that is able to calculate and display an index of **cerebral autoregulation** and to visualize **patient-specific secondary brain injury** in terms of intracranial pressure and of cerebral autoregulation, in a continuous way. The prototype will be ready to be tested in an interventional study.

AKIPREDICTOR, AN ONLINE PROGNOSTIC CALCULATOR
FOR ACUTE KIDNEY INJURY IN ADULT CRITICALLY ILL
PATIENTS

Adapted from: **Flechet M** et al. 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.* 2017;43(6):764-773. doi: [10.1007/s00134-017-4678-3](https://doi.org/10.1007/s00134-017-4678-3)

Presented as:

- Oral presentation at the 28th Annual Congress of the European Society of Intensive Care Medicine (*ESICM*). Berlin, Germany, October 2015.
- Abstract at the 14th International Symposium on Intelligent Data Analysis (*IDA*). Saint Etienne, France, October 2015.
- Abstract at the 21th International Conference on Advances in Critical Care Nephrology (*AKI-CRRT*). San Diego, USA, February 2016.

Published as science outreach: **Flechet M**, Meyfroidt G (2017). The AKIpredictor an online calculator to predict acute kidney injury. *ICU Management & Practice*.

ABSTRACT

PURPOSE: Early diagnosis of acute kidney injury (AKI) remains a major challenge. We developed and validated AKI prediction models in adult intensive care unit (ICU) patients and made these models available via an online prognostic calculator. We compared predictive performance against serum neutrophil gelatinase-associated lipocalin (NGAL) levels at ICU admission.

METHODS: Analysis of the large multicenter EPANIC database. Model development (n = 2123) and validation (n = 2367) were based on clinical information available (1) before and (2) upon ICU admission, (3) after 1 day in ICU and (4) including additional monitoring data from the first 24h. The primary outcome was a comparison of the predictive performance between models and NGAL for the development of any AKI (AKI-123) and AKI stages 2 or 3 (AKI-23) during the first week of ICU stay.

RESULTS: Validation cohort prevalence was 29% for AKI-123 and 15% for AKI-23. The AKI-123 model before ICU admission included age, baseline serum creatinine, diabetes and type of admission (medical/surgical, emergency/planned) and had an area under the receiver operating characteristic curve (AUROC) of 0.75 (95% CI 0.75-0.75). The AKI-23 model additionally included height and weight (AUROC 0.77 (95% CI 0.77-0.77)). Performance consistently improved with progressive data availability to AUROCs of 0.82 (95% CI 0.82-0.82) for AKI-123 and 0.84 (95% CI 0.83-0.84) for AKI-23 after 24 h. NGAL was less discriminant with AUROCs of 0.74 (95% CI 0.74-0.74) for AKI-123 and 0.79 (95% CI 0.79-0.79) for AKI-23.

CONCLUSIONS: AKI can be predicted early with models that only use routinely collected clinical information and outperform NGAL measured at ICU admission. The AKI-123 models are available at <http://akipredictor.com/>.

3.1 INTRODUCTION

Acute kidney injury (AKI), a rapid decline in renal excretory function, is highly prevalent in critically ill patients and is associated with increased risk of morbidity and mortality and with high financial costs [1–4]. AKI early diagnosis remains a major clinical challenge [5], because AKI is defined and classified by an increase in serum creatinine (sc) or a decline in urine output (uo), both late non-specific markers of the underlying phenomenon. Imprecise early identification of AKI could partially explain why the search for strategies or interventions to mitigate the course of AKI has been unsuccessful [6–13]. Recognizing patients at risk, optimization of hemodynamics and prevention of nephrotoxicity remain the mainstay in the prevention of AKI, whereas treatment is largely supportive [14, 15].

Biomarkers, most prominently neutrophil gelatinase-associated lipocalin (NGAL), have been studied for AKI stratification and prediction [16, 17] in mixed [18], cardiac [19] and septic populations of critically ill patients [20–22]. However, quantifying biomarkers remains expensive and their added clinical value appears limited [7, 23, 24].

The increase of computerization in the intensive care unit (ICU) has given rise to large electronic databases, thus making them amenable for ‘big data’ analytics. This term pertains to machine-learning algorithms to analyze large datasets and develop models for prognostication or decision-support in critical care [25, 26]. Such models could be a cost-effective alternative to biomarkers for AKI prognostication [5].

The aim of the present study was the development and validation of clinical prediction models for the development of AKI in the first week of ICU stay in a general ICU population. Their performance was compared to NGAL in the general ICU population and in subgroups of cardiac and septic patients. The validated models are available as an online prognostic calculator.

3.2 METHODS

3.2.1 *Study population: development and validation cohorts*

This is a retrospective analysis of the Early versus Late Parenteral Nutrition in Critically Ill Adults (EPANIC) multicenter randomized clinical trial [27] database. This trial compared two nutritional strategies in a heterogeneous population of 4640 adult patients included in seven ICUs between August 2007 and November 2010. Written informed consent was obtained from all patients or their designated representatives. The protocol was approved by

the institutional review board of the participating centers and by the Belgian authorities (Supplementary methods 3.A.1.1). Withholding parenteral nutrition in the first week of ICU stay did not impact AKI incidence or recovery, but reduced the median duration of renal replacement therapy (RRT) by 3 days [2].

We a priori decided to divide the database into development and validation cohorts, matched for baseline characteristics, nutritional strategies, AKI risk factors and outcome (Supplementary methods 3.A.1.1). Patients were excluded from the analysis if they had a history of end-stage renal disease, their baseline SC was ≥ 4 mg/dL or the SC measurements were unavailable to stage AKI.

3.2.2 AKI definition

AKI was staged each ICU stay day using the ‘Kidney Disease: Improving Global Outcomes’ serum creatinine criteria (KDIGO-SC) [14]. The UO criterion was not used as it was not prospectively collected and therefore not available hourly. Baseline SC determination is described in the Supplementary methods 3.A.1.2.

AKI prediction tasks performed:

- AKI-123: Prediction of development of any AKI stage during the first week of ICU stay.
- AKI-23: Prediction of development of AKI stages 2 or 3 during the first week of ICU stay.

3.2.3 Clinical prediction models

For each prediction task, four models were developed which are intended to be used serially, based on the clinical information available at the bedside at successive time points.

- A *Baseline model* using only demographic data and data known before ICU admission.
- An *Admission model* that adds to the above, data available upon ICU admission.
- A *Day1 model* that adds to the above, data available on the first day in ICU.
- A *Day1+ model* that adds to the above, data from the first 24h of monitoring and administered medication, as well as data on the use of radio-contrast agents during the week prior to ICU admission until the first ICU day.

Models were developed using a random forest machine-learning algorithm [28]. Data were obtained from the EPANIC research database (Filemaker Pro[®]; FileMaker Inc, FileMaker International). Data for the Day₁₊ model were retrieved from the clinical patient data management system database (MetaVision[®]; iMDSoft, Needham, MA, USA), and the radio-contrast data were retrieved from the pharmacy data warehouse.

3.2.4 *Variable selection*

Candidate variables were selected based on literature review, expert opinion and availability in the dataset. The final set of predictor variables was determined incrementally for each increasingly complex model via bootstrapped backwards elimination analysis [29] (Supplementary methods 3.A.1.3).

3.2.5 *Model development and validation*

Selection of predictor variables, and model development was performed in the development cohort only; performance and stability were internally validated via bootstrapping [29]. Model performance was subsequently evaluated in the validation cohort [30]. Models were also evaluated separately in cardiac and septic patients only and for prediction of other clinical outcomes (Supplementary methods 3.A.1.4).

3.2.6 *NGAL quantification*

Arterial blood samples were taken upon ICU admission. Blood was centrifuged, and the serum was snap-frozen and stored at -80 °C. In the validation cohort, serum NGAL levels were quantified with the Human Lipocalin-2/NGAL Quantikine enzyme-linked immunosorbent assay (ELISA) Kit from R&D Systems (Abingdon, UK) according to the manufacturer's instructions.

3.2.7 *NGAL as a predictor*

NGAL was evaluated as a continuous variable and at classification cut-offs of 150, 200 and 400 ng/mL. To compute calibration and decision curves, NGAL measurements were converted to probabilities via logistic recalibration [31]. The predictive value of NGAL when added to the Admission model was assessed via logistic regression.

3.2.8 *Performance evaluation criteria*

Discrimination was evaluated with the receiver operating characteristic (ROC) curve and quantified with the area under the receiver operating characteristic curve (AUROC) and the discrimination slope (DS). DS is the difference in average predictions for those with and without the outcome. AUROCs above 0.7 were considered adequate [32]. The integrated discrimination index (IDI) [33] was used to compare discrimination between two models in the same patient subset. Calibration was evaluated with calibration curves and quantified with calibration slopes and calibration-in-the-large. Adequate calibration results in calibration slopes close to 1, calibration-in-the-large close to 0, and calibration curves close to the diagonal. Decision curves are plots of net benefit at each possible classification threshold and were used to assess clinical usefulness [34]. Unless stated otherwise, to further evaluate model robustness, sensitivity, specificity, positive predictive value (PPV), and net benefit were evaluated at the classification thresholds that maximized sensitivity and specificity in the development cohort. These are, however, not intended as optimal, as any threshold within the model's range of clinical usefulness can be used. The choice of classification threshold is center-, patient- and task-specific and should reflect the intended use of the prediction models.

A brief and comprehensive description of all the criteria can be found in [35] (Supplementary methods 3.A.1.5).

Student's t tests were used for normally distributed and Mann-Whitney U tests for non-normally distributed continuous variables, and Fisher's exact tests for Boolean variables. To compare performance criteria of models in the same patient subset, the difference and corresponding bootstrap confidence interval (CI)s were computed. If the 95% CI excluded 0, the difference was declared statistically significant at the 0.05 level. Analyses were performed using Python 2.7.8, Scipy 0.15, scikit-learn library [36] and MATLAB 2015a (The MathWorks, Natick, MA, USA).

3.3 RESULTS

3.3.1 *Study population: development and validation cohorts*

A total of 4490 patients were included (Figure 3.1). The development cohort (n = 2123) did not differ from the validation cohort (n = 2367) regarding demographic and medical characteristics, risk factors for AKI or clinical outcomes (Table 3.A.1), except for younger median age (65.9 vs. 66.9, p = 0.02) and proportionally more transplant surgery patients (7.5 vs. 5.9%, p = 0.03).

Within the first week of ICU stay, 27.7 and 29.2% patients developed AKI-123 in the development and validation cohorts respectively, while 14.0 and 14.8% developed AKI-23.

Baseline SC was calculated in 22.8 and 22.9% patients, respectively, in each cohort. In the validation cohort, there were 1591 (65.0%) cardiac and 455 (18.6%) septic patients, and NGAL was measured upon ICU admission in 2081 patients.

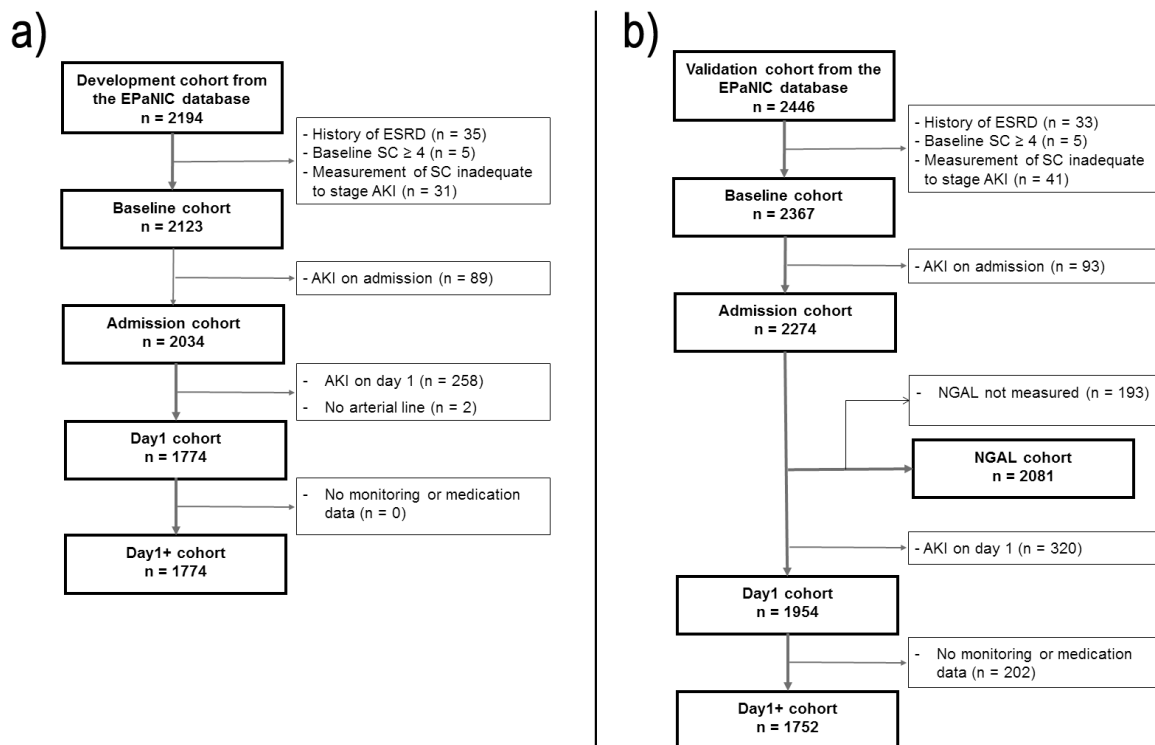


Figure 3.1 Consort diagrams. **(A)** Development cohort. **(B)** Validation cohort.

3.3.2 Variable selection and model development

Candidate predictors and univariable association with AKI-123 and AKI-23 are listed in Tables 3.A.2 and 3.A.4 (Supplementary results). Selected AKI-123 predictors are shown in Table 3.1 and for AKI-23 in Table 3.2. All models are random forests comprising 60 trees with a maximum depth of 15 and 10 patients minimum per leaf.

3.3.3 Development cohort model performance

All models performed well according to all evaluation criteria (Supplementary results, Figure 3.A.1; Table 3.A.5). Performance consistently improved with

Table 3.1 Selected predictors for the four AKI-123 clinical prediction models

Baseline	Age Baseline serum creatinine Surgical or medical category (Transplant surgery/ cardiovascular surgery (non-transplant)/ abdominal or pelvic surgery/ thorax surgery (non-cardiac)/ others (medical, trauma, other surgery) Planned admission (yes/no) Diabetes (yes/no)
Admission	Blood glucose upon ICU admission Suspected sepsis upon ICU admission (yes/no) Hemodynamic support upon ICU admission (none/mechanical/pharmacological/both)
Day 1	Serum creatinine APACHE II score Maximum lactate Bilirubin Hours of ICU stay
Day1+	Total amount of urine Urine slope ^a Time the mean arterial blood pressure is above its average value Time the mean arterial blood pressure is below 60 mmHg Pharmacologic hemodynamic support (cumulative dose of inotropes and vasopressors)

Each model also includes all predictors of the rows above it

^a The urine slope refers to the slope of a linear model fitted to the hourly urine flow

increasing data availability. The AKI-123 and AKI-23 Baseline models resulted in respective AUROCs of 0.77 and 0.81, increasing to 0.80 and 0.83 for Admission, and 0.86 and 0.88 for Day 1. Adding monitoring and medication data improved AUROC to 0.87 for the AKI-123 Day1+ model. Low prevalence precluded the development of an AKI-23 Day1+ model [37].

3.3.4 Validation cohort model performance

Table 3.3 and Figures 3.2a-c report the performance for AKI-123. They show good discrimination with AUROCs of 0.75, 0.77, 0.80 and 0.82 and DSS of 0.19, 0.20, 0.23 and 0.21 for Baseline, Admission, Day1 and Day1+, respectively. They are well calibrated with respective calibration slopes of

Table 3.2 Selected predictors for the three AKI-23 clinical prediction models

Baseline	Age Baseline serum creatinine Surgical or medical category (Transplant surgery/ cardiovascular surgery (non-transplant)/ abdominal or pelvic surgery/ thorax surgery (non-cardiac)/ others (medical, trauma, other surgery) Planned admission (yes/no) Diabetes (yes/no) Gender Height Weight
Admission	Blood glucose upon ICU admission Suspected sepsis upon ICU admission (yes/no) Hemodynamic support upon ICU admission (none/mechanical/pharmacological/both)
Day 1	Serum creatinine APACHE II score

Each model also includes all predictors of the rows above it

0.80, 0.85, 0.78 and 0.96, calibration-in-the-large close to 0, and calibration curves close to the diagonal. Decision curves show clinical usefulness for a wide range of risk thresholds, respectively 8.5-61.1, 2.5-65.7, 4.2-72.4 and 3.3-77.2%. The classification thresholds identified in the development cohort of 25.80, 20.71, 14.46 and 15.81% remained robust and resulted in similar sensitivities, specificities, PPV, net benefit and false positive reduction in the validation cohort. To improve sensitivity or specificity, other classification thresholds within the range of clinical usefulness can be chosen. Development and validation cohorts show similar performance for AKI-23 (Figure 3.A.2; Table 3.A.6). Further illustrating discrimination are the Admission models' predictions stratified by AKI stages (Supplementary results, Figures 3.3a and 3.A.3).

Model performance did not decrease when evaluated in the subset of cardiac patients, but AKI-23 prevalence was insufficient for the Day1 evaluation (Figure 3.A.4; Tables 3.A.7 and 3.A.8). Performance in septic patients could only be evaluated for AKI-123 where it decreased to AUROCs of 0.65, 0.76 and 0.77 and ranges of clinical usefulness decreased to 32.0-64.0, 6.8-61.9 and 12.5-73.5% for the Admission, Day1 and Day1+ models (Supplementary results, Figure 3.A.5; Table 3.A.9).

Table 3.3 Performance for AKI-123 prediction in the validation cohort

AKI-123	Validation cohort ^a				Validation NGAL cohort ^b					
	Baseline model	Admission model	Day1 model	Day1+ model	NGAL			Admission model	Combined model	
Number of patients	2367	2274	1954	1752	2081			2081	2081	
AKI-123 prevalence (%)	29.24	26.61	14.84	14.84	27.15			27.15	27.15	
AUROC	0.75 (0.75-0.75)	0.77 (0.77-0.77)	0.80 (0.80-0.80)	0.82 (0.82-0.82)	0.74 (0.74-0.74)				0.77 (0.77-0.77)	0.80 (0.80-0.80)
Discrimination slope	0.19 (0.19-0.19)	0.20 (0.20-0.20)	0.23 (0.23-0.23)	0.21 (0.21-0.21)	0.13 (0.13-0.13)				0.20 (0.20-0.20)	0.22 (0.22-0.22)
Calibration slope	0.80 (0.80-0.80)	0.85 (0.85-0.85)	0.78 (0.78-0.79)	0.96 (0.96-0.96)	0.83 (0.82-0.83)				0.84 (0.84-0.84)	0.87 (0.87-0.87)
Calibration in the large	-0.01 (-0.01 to -0.01)	-0.01 (-0.02 to -0.01)	-0.01 (-0.01 to -0.01)	-0.03 (-0.03 to -0.03)	0.00 (0.00 to 0.00)				-0.02 (-0.02 to -0.02)	0.00 (0.00-0.00)
Classification threshold	25.80%	20.71%	14.46%	15.81%	150 ng/mL	200 ng/mL	400 ng/mL	20.71%	20.98%	
Sensitivity	0.70 (0.70-0.70)	0.72 (0.72-0.72)	0.64 (0.64-0.64)	0.60 (0.60-0.61)	0.87 (0.86-0.87)	0.69 (0.69-0.69)	0.31 (0.31-0.31)	0.73 (0.73-0.73)	0.75 (0.75-0.75)	
Specificity	0.69 (0.69-0.69)	0.68 (0.68-0.69)	0.82 (0.82-0.82)	0.88 (0.88-0.88)	0.42 (0.42-0.42)	0.65 (0.65-0.65)	0.93 (0.93-0.93)	0.68 (0.68-0.68)	0.71 (0.71-0.71)	
Positive predictive value	0.48 (0.48-0.48)	0.45 (0.45-0.45)	0.38 (0.38-0.38)	0.46 (0.46-0.46)	0.36 (0.36-0.37)	0.42 (0.42-0.42)	0.62 (0.62-0.62)	0.46 (0.46-0.46)	0.49 (0.49-0.49)	
Δ Net benefit _{None} (wrt treat-none)	0.13 (0.13-0.13)	0.13 (0.13-0.14)	0.07 (0.07-0.07)	0.07 (0.07-0.07)	0.13 (0.13-0.13)	0.12 (0.12-0.12)	0.05 (0.05-0.05)	0.14 (0.14-0.14)	0.15 (0.15-0.15)	
Δ Net benefit _{All} (wrt treat-all)	0.08 (0.08-0.08)	0.05 (0.05-0.06)	0.05 (0.05-0.05)	0.08 (0.08-0.08)	0.03 (0.03-0.03)	0.05 (0.05-0.05)	0.20 (0.20-0.20)	0.05 (0.05-0.05)	0.06 (0.06-0.06)	

Estimated 95% confidence intervals obtained from 2000 bootstrap replicas are shown in parentheses. Values in the last five rows were computed at the respective classification thresholds for each column. Values highlighted in bold are significantly different from the value in the previous column at a statistical level of 0.05

^a Clinical prediction models

^b NGAL, admission model and combined model; combined model is NGAL plus the admission model

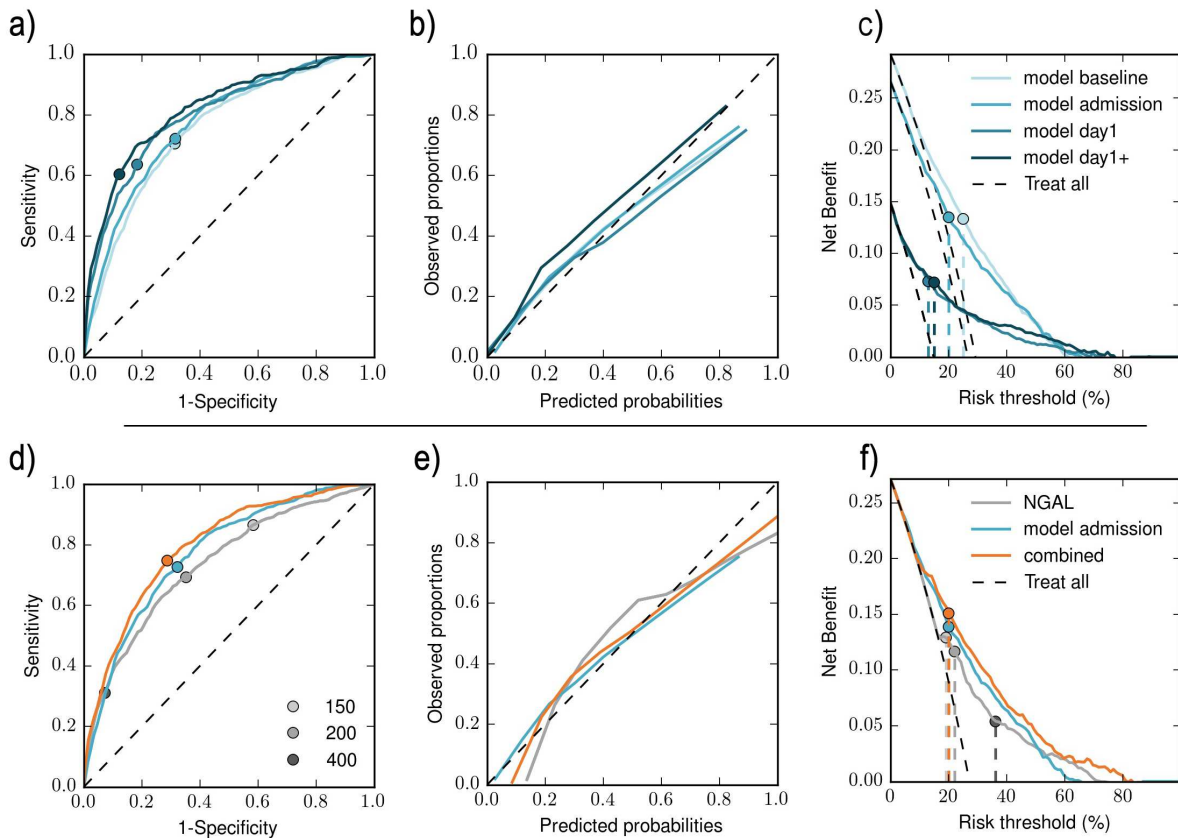


Figure 3.2 Performance for AKI-123 prediction in the validation cohort. **a, d** ROC curves, **b, e** calibration curves and **c, f** decision curves. *Top row* clinical prediction models. *Bottom row* admission model, NGAL and combined model; combined model is the admission model plus NGAL. Classification thresholds are overlaid as *dots* on the ROC and decision curves

3.3.5 *Validation cohort NGAL performance*

Table 3.3 and Figures 3.2d-f report the performance of admission NGAL, the Admission model and their combination in the subset of patients with an NGAL measurement. Respective AUROCs are 0.74, 0.77 and 0.80 and DSS are 0.13, 0.20, and 0.22. The statistically significant discrimination improvements correspond to a 0.073 IDI between the Admission model and NGAL and a smaller 0.016 IDI between the combined and the Admission models. The respective calibration-slopes of 0.83, 0.84 and 0.87 are significantly different and all calibrations-in-the-large are close to 0. Decision curves show clinical usefulness in the ranges of 14.3-72.4, 2.9-65.4 and 8.3-81.4%, respectively, with a noteworthy improvement in net benefit of the combined over the Admission model only for high risk thresholds above twice the population prevalence. A 212.54 ng/mL NGAL cut-off maximized sensitivity and specificity. Similar performance was observed for AKI-23 (Supplementary results, Table 3.A.10; Figure 3.A.6). NGAL measurements stratified by AKI stages and classification cut-offs of 150, 200 and 400 ng/mL are shown in Figure 3.3b.

Performance remained largely unchanged in the subset of cardiac patients (Figure 3.A.7; Table 3.A.11), but decreased in the subset of septic patients (Figure 3.A.8; Table 3.A.13) to AUROCs of 0.71, 0.65 and 0.74 and to narrower ranges of clinical usefulness of 35.7-78.2, 31.8-64.3 and 16.2-93.4%. At day 1 (n = 1779, AKI-123 prevalence = 15.06%), the Day1 model performed significantly better than admission NGAL, with respective AUROCs of 0.81 and 0.67 and DSS of 0.24 and 0.03, equivalent to a 0.21 IDI. Likewise, decision curves show a broader clinical usefulness range of 3.1-72.5 versus 10.3-35.3% (Supplementary results, Figure 3.A.9; Table 3.A.14a). A similar performance was observed for AKI-23 (Supplementary results, Figure 3.A.9; Table 3.A.14b).

3.3.6 *Validation cohort: other clinical outcomes*

All models performed well at predicting relevant related outcomes with AUROCs between 0.72 and 0.84 for the requirement of RRT and between 0.74 and 0.80 for mortality at hospital discharge and the combined outcome. NGAL performed similar to the Admission model for the requirement of RRT but had AUROCs below 0.7 for mortality and the combined outcome (Supplementary results, Table 3.A.16).

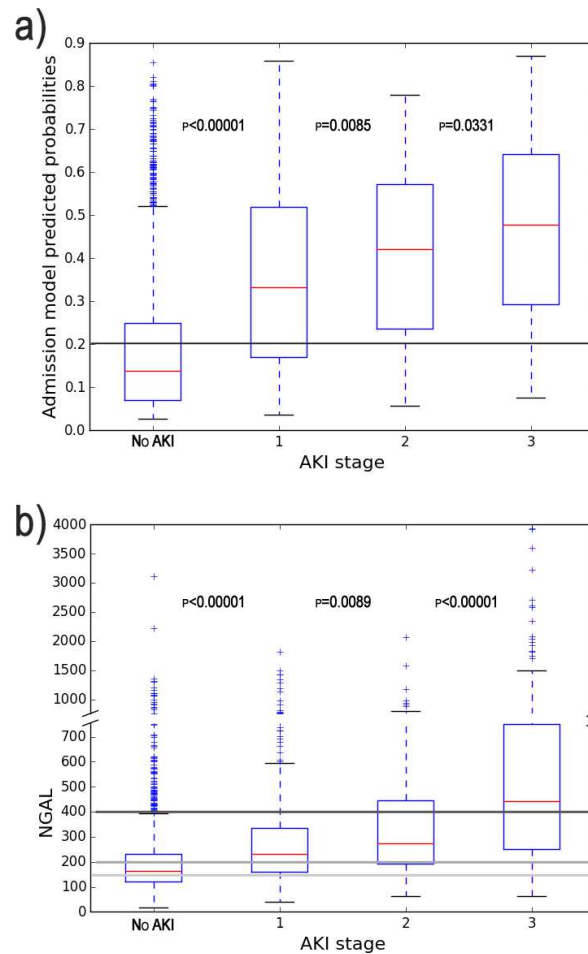


Figure 3.3 Predictions of the AKI-123 Admission model and admission NGAL measurements in the validation cohort stratified by Kidney Disease: Improving Global Outcome (KDIGO) AKI stages. **a)** Admission model ($n = 2274$). **b)** NGAL ($n = 2081$). Horizontal lines show the classification threshold of 20.71% for the admission model and of 150, 200 and 400 ng/mL for NGAL. Reported P-values are for Mann-Whitney U tests

3.3.7 Online AKI-123 prognostic calculator

The Baseline, Admission and Day1 models for AKI-123 are available at: <http://akipredictor.com/>. On this website, the user can enter patient characteristics and obtain a probability of developing AKI. The user can specify a classification threshold and see how this cut-off affects the model performance (including sensitivity, specificity, PPV, negative predictive value (NPV), Δ net benefit $(NB)_{None}$, ΔNB_{All}).

3.4 DISCUSSION

In this study, we evaluated the use of clinical prediction models and NGAL for AKI prognostication. Prediction of any AKI stage (AKI-123) is relevant for clinical practice and research, as it allows for risk stratification. Additionally, in clinical trials, it might aid in stratification or heterogeneity reduction. To emphasize the relevance of AKI-123, we only include these more sensitive models in the online calculator, although a recent consensus meeting [5] highlighted AKI-23 given its stronger association with outcome [4, 38]. In clinical practice, care should be aimed at the prevention of AKI in all patients, regardless of their risk. Nevertheless, higher-risk patients require special attention and might benefit from a more focused preventive management.

Our results clearly demonstrate that the development of AKI-123 and AKI-23 during the first week of ICU stay, as defined by the KDIGO-SC criteria, can be accurately predicted based only on clinical information routinely collected up to 24 hours after ICU admission. Already at baseline and with few predictor variables, the developed models performed well according to all evaluation criteria, and consistently improved with increasing data availability. The best performances were obtained when using all available data sources, including continuous monitoring.

In the subset of patients with biomarker measurements, the Admission model's predictions and admission NGAL concentrations related well to injury severity, thus highlighting their potential use for risk stratification. The AUROCs of NGAL are in accordance with previous large studies in general ICU patients [18, 39, 40]. The Admission model significantly outperformed the biomarker in all criteria regarding AKI-123 and performed equally for AKI-23. Combining NGAL and clinical information significantly increased AUROC, as reported in previous studies [18, 40]. However, decision curves revealed that this performance increase occurs only for the subset of patients at high risk of developing AKI, above twice the risk for the entire cohort. In these high-risk patients, it is disputable whether the additional cost of measuring NGAL with a small gain in discrimination is clinically relevant, especially since no available treatment exists [34, 35].

Overall, the same findings remained when evaluating the prediction models in cardiac patients. Additionally, the models had equal or better performance than risk scores designed specifically for cardiac patients when evaluated in large studies [41–43]. In septic patients, NGAL resulted in similar AUROCs as in previous studies [18, 42, 43], and likewise performed best at higher cutoffs. Although NGAL performed better than the Admission model, this was again only the case for the high-risk patients.

There is a general disillusionment in biomarkers for early AKI diagnosis [44] and their transferability to clinical practice [7]. Nevertheless, our decision curve analyses show that NGAL and possibly other fast-acting biomarkers [17, 23, 44, 45], could improve predictive performance (1) at or before ICU admission, (2) for the less prevalent and most severe AKI stages, (3) for the patients at highest risk, such as those with sepsis, (4) for early enrollment in interventions with potentially adverse side effects [8], and (5) possibly when only a calculated baseline creatinine is available (eResults). Furthermore, the online calculator and other risk scores could identify the high-risk patients who would benefit most from incurring the additional cost of a biomarker evaluation, which would reduce unnecessary and expensive testing.

This study has several strengths. First and foremost is its use of a large dataset of 4490 mixed critically ill patients from seven ICUs [27]. Second, admission NGAL was measured in 2081 patients, which to the best of our knowledge constitutes the largest study of any AKI biomarker and of NGAL in particular in a mixed critically ill population. Additionally, this also constituted the largest NGAL study in cardiac ($n = 1405$) or septic ($n = 438$) critically ill patients. Furthermore, the NGAL measurement timing was the same for all patients, namely at ICU admission. Third, AKI was staged according to the KDIGO criteria. Fourth, care was taken to obtain the most accurate values of baseline SC by searching back through hospital records. Fifth, a development and validation approach was followed, with robust internal validation techniques used for variable selection and model development. Sixth, a comprehensive battery of evaluation criteria, including decision curve analyses was used. Seventh, the models use a small number of predictors that are likely available in most ICUs, which improves their likelihood of worldwide use and generalizability. Lastly, and most importantly, the developed AKI-123 models are publicly available as an online prognostic calculator. This is the first of its kind and provides a benchmarking platform for future biomarker and AKI-prediction studies, and complements existing risk assessment scores [46] in use at the bedside and/or in low resource settings [47].

This study has the following limitations. First, we only evaluated NGAL while other biomarkers might be more discriminant [17, 45]. Second, NGAL was only measured in the validation cohort, such that its reported decision- and calibration-curve results are likely overestimations. Third, NGAL was measured only at ICU admission; however, earlier studies demonstrated that serial measurements do not improve predictive performance [18]. Fourth, as only the KDIGO-SC criteria were used, patients developing AKI according to just KDIGO-UO criteria could be missed. Nevertheless, the models predicted well the related and clinically relevant outcomes of RRT requirement and mortality at hospital

discharge. Fifth, despite the database size, the low AKI-23 prevalence precluded its complete study, and therefore these models were excluded from the online calculator. Sixth, models were developed for mixed critically ill patients and then evaluated in cardiac and septic settings; therefore, it is plausible that better-performing models could be developed specifically for each subpopulation. Seventh, since the models were developed and validated in the EPANIC database, which includes a high proportion of cardiac surgery patients, their performance might vary when evaluated in other datasets, particularly if they differ in patient characteristics and AKI prevalence. Therefore, the findings of this study require further external (and prospective) validation. Lastly, as with any tool, there is an inherent risk for potential misuse of the online calculator. Care should be taken to ensure the choice of a classification threshold that is suited for the intended use of the prediction models.

3.5 CONCLUSION

We have shown that AKI development within the first week of ICU stay, as defined by the KDIGO-SC criteria, can be predicted well with clinical prediction models that only use routinely collected clinical data. We have also shown that NGAL measured at admission, when used as a single predictor or in combination with clinical data, does not improve on the developed models. These findings require prospective validation. Future biomarker studies on AKI prediction must show improvement over what is possible with readily available clinical data or risk scores, not only regarding discrimination but also with decision curve analysis. To facilitate further (prospective) validation, the developed clinical prediction models are available online as a prognostic calculator.

GRANT SUPPORT AND CONFLICT OF INTEREST

M. Flechet receives funding from the Fonds Wetenschappelijk Onderzoek (FWO) as a PhD fellow (11Y1116 N). G. Meyfroidt receives funding from FWO as senior clinical investigator (1846113 N). M. Casaer receives funding from the UZLeuven KOF. G. Van den Berghe, through the KULeuven, receives long-term research financing via the Flemish government Methusalem-program.

The authors declare that they have no conflict of interest.

ACKNOWLEDGMENTS AND PERSONAL CONTRIBUTION

All analysis presented in this chapter were performed by Marine Flechet.

- Study concept and design: Meyfroidt, Schetz, Güiza, Van den Berghe
- Acquisition, analysis, or interpretation of data: **Flechet**, Güiza, Schetz, Derese, Gunst, Spriet, Casaer, Meyfroidt
- Drafting of the manuscript: Güiza, **Flechet**
- Critical revision of the manuscript: **All authors**
- Statistical analysis: **Flechet**, Güiza, Meyfroidt
- Creation of figures: **Flechet**
- Development of website: **Flechet**
- Administrative, technical, or material support: Wouters
- Study supervision: Meyfroidt

The authors would like to thank Romain Pirson and Eugène Stassen for their help during the development of the akipredictor.com website.

3.A APPENDIX

3.A.1 *Supplementary methods*

3.A.1.1 *Study population: Development and Validation cohorts*

This study is a secondary analysis of the database of the EPaNIC multicenter randomized clinical trial [27] which compared two nutritional strategies in a heterogeneous population of 4640 adult ICU patients included between August 2007 and November 2010. Withholding parenteral nutrition in the first week of ICU stay did not have an impact on AKI incidence or recovery, but it did reduce the median duration of RRT by 3 days [2]. Written informed consent was obtained from all patients or their designated representatives. The protocol was approved by the institutional review board of the participating centers and by the Belgian authorities.

The Informed consent and protocol, including exploratory metabolic studies and analyses allowing to quantify presence of AKI were approved by the "Institutional Review Board" decision ML4190 (Commissie Medische Ethiek Universitaire Ziekenhuizen KULeuven 10 mei 2007).

Initially, we only considered the subset of patients with complete electronic records including monitoring data. We a priori decided to divide this subset into development and validation cohorts matched for the following baseline characteristics, nutritional strategies, AKI risk factors and outcome:

- Age
- Gender
- APACHE-II
- Cardiac surgery
- Sepsis
- AKI stage
- AKI 3
- Renal replacement therapy
- ICU length of stay
- Hospital length of stay
- Hospital mortality

- Day-90 mortality
- Nutritional strategy (Randomization)

To perform the matching, the entire dataset was randomly permuted and split into two sets of equal size and the aforementioned characteristics were compared. This procedure was repeated until two sets 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 patients in the validation cohort for the models not requiring this information. NGAL was measured in all matched patients from the validation cohort. The inclusion of additional patients to the validation cohort happened after the NGAL measurements were already performed.

Patients were excluded from this study if they had a history of end-stage renal disease, if their baseline SC was ≥ 4 mg/dL or if the measurement of SC was inadequate to stage AKI (SC missing on day 1 or missing for more than 48 hours).

3.A.1.2 *AKI definition*

Baseline SC was determined for all patients as the lowest SC during the 3 months prior to ICU admission for elective admissions and the lowest SC from 3 months to 1 week before ICU admission for emergency admissions [48]. SC was searched for in the hospital database or manually retrieved by searching documents from referring hospitals/physicians. In case no baseline SC value was available, a normal glomerular filtration rate (GFR) of 75 mL/min per 1.73 m² was assumed, and the Modification of Diet in Renal Disease (MDRD) formula was used to estimate baseline SC [49].

The number (%) of patients with a true baseline creatinine in the development and validation cohorts were respectively, 1640 (77.2%) and 1825 (77.1%) (p-value= 0.91). Additionally, no patient with a calculated baseline creatinine was suspected to have chronic kidney disease (CKD).

AKI was staged for each day of ICU stay [48], using the KDIGO SC criteria [14].

3.A.1.3 *Variable selection*

First, we performed a review of the literature to select all variables that were associated with AKI and that were available in the EPaNIC database. The

identified variables were sorted in 4 categories depending on when they were measured: before ICU admission (Baseline), upon ICU admission (Admission), first morning of ICU stay (Day1) and first 24 hours of ICU stay (Day1+).

In the development cohort, we used a backwards feature selection technique, based on a correlation-based ranking algorithm (CfsSubsetEval) of the Weka machine-learning suite [50]. This algorithm determined via cross-validation the lowest-ranked attribute from the attributes available at baseline. One model was built with all the attributes available at baseline and another without the lowest-ranked attribute. The discrimination performance of the two models was compared with a bootstrapped approach [29]. If the difference in discrimination was not statistically significant, we removed the next lowest-ranked feature. This procedure was iterated until removal of the lowest-ranked feature resulted in a significant drop in performance. As such, we obtained the most discriminant model based on a small set of baseline attributes.

The process was then repeated in the development cohort for 1) the Admission model to identify the most discriminant attributes to be added to the already selected attributes for the simpler Baseline model; 2) the Day1 model to identify the most discriminant attributes to be added to the already selected attributes for the simpler Admission model; and 3) the Day1+ model to identify the most discriminant attributes to be added to the already selected attributes for the simpler Day1 model. With this incremental approach, as the complexity of the model increases and more data becomes available, no already selected attribute is removed but only new ones added so long as they provide an increase in performance.

3.A.1.4 *Model development and validation*

Selection of variables to be used as predictors, as well as model development was performed in the development cohort only. Model performance and stability were internally validated via bootstrapping [29] in the development cohort.

Models were developed using a random forest machine learning algorithm. Using a grid search approach, the following parameters of the random forest were optimized: numbers of trees in the forest, maximum depth of the decision tree, minimum number of samples to make a split, and minimum number of samples to be a leaf. Additionally, at each node of the decision trees, the square root of the number of available features for modeling was considered to make the split.

The performance of the different models was subsequently externally evaluated in the validation cohort. Models were also evaluated separately in

cardiac and septic patients only, but only subsets with at least 50 patients with and without the outcome were considered for validation [30, 51]. This meant that AKI-23 could not be studied in septic or cardiac patients with the Day1 or Day1+ models, or in general ICU patients with the Day1+ model. Models were further evaluated for prediction of other relevant clinical outcomes, namely requirement of RRT, mortality at hospital discharge and the combined outcome.

3.A.1.5 *Performance evaluation criteria*

To assess the performance of the developed models, we reported discrimination, calibration and clinical usefulness.

Discrimination refers to how well the predictions allow to discriminate between the patients with and without the outcome, i.e. in our study those who will develop AKI during the first week of ICU stay from those who will not. Discrimination was evaluated visually with the ROC curve, and quantified with the AUROC and the DS. The ROC curve plots the sensitivity versus 1-specificity for all possible classification thresholds. The classification threshold is the cut-off value above which the score is positive. In our study, a positive score corresponds to considering the patient to be at risk for developing AKI. The DS measures how well the patients with and without the outcome are separated. In our study, it is the difference between the average predictions of the patients with AKI and those without. DS ranges from 0 (no discrimination) to 1 (complete discrimination). Finally, the IDI [52] can be used to compare discrimination between two models in the same patient subset. The IDI is the difference between the models' discrimination slopes and is preferred to the continuous net reclassification index (NRI) when, as is the case for AKI, no clear classification cut-off exists [33].

A well discriminating model would maximize its sensitivity and specificity and would therefore have an AUROC close to 1. In practice, AUROCs above 0.7 are considered adequate [32]. A well discriminating model therefore has higher predicted probabilities for patients with AKI and lower probabilities for patients without AKI, which translates to a large DS. Alternatively, a trivial model that consists in randomly guessing the outcome is represented with a curve close to the diagonal on the ROC curve and an AUROC close to 0.5. This trivial model would have similar predictions for patients with and without AKI, resulting in a DS close to 0.

We also reported additional common measures of discrimination including sensitivity, specificity, positive predictive value and negative predictive value. These measures require choosing a threshold to classify patients as having a low or high risk of developing AKI. Commonly used thresholds are: 1)

the percentage of patients in a population with similar characteristics who developed AKI or 2) the cutoff that results in the highest sensitivity and specificity in a population with similar characteristics. We have chosen the latter (as determined for all patients in the development cohort) in the reported tables and as the default in the online calculator. This cutoff is equivalent to the point in the ROC curve closest to the upper left corner. Different classification thresholds are likely required to study populations at different risks for AKI. For instance, a higher cutoff would be considered when studying septic patients than for cardiac ones. The sensitivity shows the ability of the model to correctly identify the patients with AKI. The specificity shows the ability of the model to correctly identify the patients without AKI. The positive predictive value reflects how likely it is that a patient has AKI given that the predicted probability is higher than the classification threshold. Finally, the negative predictive value reflects how likely it is that a patient does not have AKI given that the predicted probability is lower than the classification threshold.

Calibration refers to the agreement between the population outcomes and the model predictions. For example, if the observed frequency of AKI is 30 out of 100 patients in the population, a well calibrated model should predict a 30% risk of developing AKI for these 100 patients. Calibration can be evaluated visually with calibration curves which plot the observed proportion in the population versus the model predicted probabilities. Calibration can additionally be quantified by the intercept (calibration-in-the-large) and the slope (calibration-slope) of a linear regression fitted to this curve.

A well calibrated model shows strong agreement between the population observed proportions and the probabilities given by the model. Therefore, it has a calibration curve close to the diagonal, a calibration-slope close to 1, and a calibration-in-the-large close to 0.

The clinical usefulness of a prediction model refers to the potential clinical benefit achieved by using this model. In particular, it is important to report the clinical usefulness if the harm of misclassifying a patient without the outcome is different than the one of misclassifying a patient with the outcome, or if the benefit of classifying a patient with the outcome is different than the one of classifying a patient without the outcome.

The clinical usefulness can be quantified using the NB. NB is the difference between the expected benefit and the expected harm associated with the classification. The expected benefit is the number of patients who have the outcome and were classified as such (true positives) and who will therefore benefit from the treatment or measures taken to improve their care. The expected harm is the number of patients who will be wrongly treated because they were classified as having the outcome (false positives) multiplied by a

weighting factor depending on the classification threshold. The classification threshold is chosen depending on the relative weight of harm and benefit associated with the classification. For instance, if the treatment associated with the classification of having the outcome can be considered harmful (e.g. has undesirable side-effects, is very costly, etc.) it is important to have few false positive classifications; this corresponds to using a high classification threshold. Alternatively, if the treatment associated with the classification of having the outcome is benign, the number of false positive classifications is less important; this corresponds to using a low classification threshold. The weight given to any potential harm or benefit associated with the actions resulting from a classification is center, patient and task dependent, and therefore so is the choice of classification threshold. In our study, early recognition of AKI will allow for earlier preventive measures. Because these preventive measures will arguably pose little harm to the patient, risk thresholds below 0.5 are recommended.

The clinical usefulness can be visualized using decision curves. These decision curves plot the net benefit at each possible classification threshold [53]. For a model to be clinically useful at a given threshold it must have a positive net benefit, which is equivalent to improving upon trivially classifying all patients as not having the outcome (referred to as treat-none strategy). The difference between the model's NB and the NB of classifying all patients as not having the outcome (treat-none strategy) is denoted $\Delta \text{NB}_{\text{None}}$. Second, it must have a higher net benefit than trivially classifying all patients as having the outcome (referred to as treat-all strategy). The difference between the model NB and the NB of classifying all patients as having the outcome (treat-all strategy) is denoted $\Delta \text{NB}_{\text{All}}$.

Clinically useful models have larger net benefits than the two trivial strategies for all relevant classification thresholds. The wider the range of clinical usefulness, the more versatility the model offers, as it can be used with a larger number of different classification thresholds. Decision curves therefore allow for a straightforward visual comparison between the clinical usefulness of different models designed for the same task.

The net benefit reported in the results tables was evaluated at the classification thresholds that maximized sensitivity and specificity in the development cohort, equivalent to the point in the ROC curve closest to the upper left corner. As an example, the $\Delta \text{NB}_{\text{None}}$ of the Baseline model for AKI-123 in the validation cohort at this classification threshold was 13%. Therefore, for every 100 patients classified by the Baseline model, 13 extra true positives will be identified without increasing the false positive rate, as compared to assuming no patient will develop AKI (treat-none strategy). The $\Delta \text{NB}_{\text{All}}$ was

8%. Therefore, for every 100 patients classified by the Baseline model, 8 extra true positives will be identified without increasing the false positive rate, as compared to assuming all patients will develop AKI (treat-all strategy).

A brief and comprehensive description of all the criteria can be found in [35].

3.A.2 *Supplementary results*

3.A.2.1 *Performance of clinical prediction models for AKI-23*

Table 3.A.6 and Figure 3.A.2 report the performance of the clinical prediction models for AKI-23 in the validation cohort. The low prevalence precluded the development of the more complex Day1+ model for this task [37]. Discrimination results are AUROCS of 0.77, 0.79 and 0.84 and DSS of 0.15, 0.15, and 0.12 for the Baseline, Admission and Day1 models, respectively. The models have respective calibration-slopes of 0.63, 0.67 and 0.44; calibrations-in-the-large close to 0; and calibration curves that deviate downward from the diagonal. Decision curves show clinical usefulness in the ranges of 3.1-35.8%, 1.0-39.1% and 0.1-28.7%, respectively. Admission model predictions stratified by AKI stages and the classification threshold are shown in Figure 3.A.4b. Baseline and Admission model performance, mainly calibration, decreased when compared to the development cohort (Figure 3.A.1, Table 3.A.5b), but remained unchanged when evaluated only in the subset of cardiac patients in the validation cohort (Figure 3.A.5, Table 3.A.9b). The number of septic patients in the validation cohort was insufficient for model evaluation.

3.A.2.2 *Performance of NGAL for AKI-23 prediction*

Table 3.A.8 and Figure 3.A.3 report the performance of admission NGAL, the Admission model and their combination for AKI-23 prediction in the subset of patients in the validation cohort with a biomarker measurement (n=2081, prevalence=12.59%). Discrimination respectively results in AUROCS of 0.79, 0.79 and 0.83 and DSS of 0.15, 0.15, and 0.19; equivalent to a 0.002 IDI between NGAL and the Admission model and a 0.042 IDI between the combined and the Admission models. Respective calibration-slopes are 0.70, 0.69 and 0.81 and calibrations-in-the-large are close to 0. Decision curves show clinical usefulness in the ranges of 5.2-72.4%, 2.2-39.8% and 4.0-86.4%, respectively. The NGAL classification threshold that maximized sensitivity and specificity in this cohort was 232.40 ng/mL. Performance remained largely unchanged when evaluated only in the subset of cardiac patients in the validation cohort (Figure 3.A.7, Table 3.A.12), but with a reduction in the ranges of clinical usefulness to 5.0-45.0%, 2.1-28.2% and 3.2-45.9%. The number of septic patients developing AKI-

23 was insufficient for model evaluation. At day 1 (n=1779, prevalence=4.5%) admission NGAL and the Day1 model respectively result in AUROCs of 0.70 and 0.83 and DSS of 0.02 and 0.12, equivalent to a 0.098 IDI. Decision curves show clinical usefulness in the ranges of 3.0-30.0% and 0.1-27.3%, respectively (Figure 3.A.9, Table 3.A.14b).

3.A.2.3 *Performance of NGAL and clinical prediction models for AKI-123 prediction with respect to true/calculated creatinine baseline*

Table 3.A.15 and Figure 3.A.10 report the performance of admission NGAL, the Admission model and their combination for AKI-123 prediction in the subset of patients in the validation cohort with a biomarker measurement (n=2081, prevalence=12.59%) separated into two subsets, one with true creatinine baseline and the other with a calculated baseline creatinine. Of the 2081 patient in the validation cohort with an NGAL measurement, 458 had a calculated baseline creatinine (141 developed AKI) and 1623 had a true baseline creatinine (424 developed AKI). ROC and decision curves are reported below for these two patient subsets. When the baseline is calculated, NGAL results in better AUROC than the Admission model, and the ROC curves show the increase in discriminability occurs for higher specificity. Likewise, the Decision curves show that NGAL improves performance over the Admission model for higher risk patients (risk thresholds approximately between 40 and 80%). However as only 141 patients developed AKI in this cohort, the performance improvement provided by NGAL (or potentially other biomarkers) when baseline creatinine is calculated should be validated in larger samples sizes.

When considering the much larger cohort of patients with true creatinine baseline, the Admission model outperforms NGAL.

3.A.2.4 *Performance of the prediction models and of admission NGAL for outcome prediction*

We report the discriminative performance of the different models for predicting other relevant outcomes, namely 1) Requirement of RRT, 2) Mortality or RRT at hospital discharge and 3) Mortality at hospital discharge. We performed the analysis for the 4 models for AKI-123 and for the 3 first models for AKI-23. We also compared NGAL and the Admission model in the subset of patients with a biomarker measurement.

The results shown in Table 3.A.16A for the AKI-123 case and in Table 3.A.16B for the AKI-23 case, indicate that the models retain good discriminatory performance for these tasks. It is noteworthy that unlike AKI, requirement of RRT is not necessarily captured by the change in serum creatinine. It is also

noteworthy that for RRT, the performance consistently improves with data availability (except for the Day1+ case).

All models performed well at predicting mortality related outcomes with AUROCs above 0.74. NGAL performed similar to Admission model for requirement of RRT but had AUROCs below 0.7 for mortality related tasks.

3.A.3 Supplementary tables and figures

Table 3.A.1 Patient characteristics and clinical outcomes

	Development cohort	Validation cohort	P-value
Number of patients	2123	2367	-
AKI, n(%)	588 (27.7)	692 (29.2)	.26
AKI stage, n(%)			
Stage 1	290 (13.7)	342 (14.4)	.46
Stage 2	94 (4.4)	117 (4.9)	.43
Stage 3	204 (9.6)	233 (9.8)	.80
CKD, n(%)	373 (17.6)	392 (16.6)	.38
Age, years, median (IQR)	65.9 (54.9-74.8)	66.9 (56.5-75.2)	.02
Diabetes, n(%)	359 (16.9)	408 (17.2)	.78
Baseline serum creatinine, mg/dl, median (IQR)	0.91 (0.76-1.05)	0.92 (0.76-1.05)	.37
True baseline serum creatinine, n(%)	1640 (77.2)	1825 (77.1)	0.91
Elective admission, n(%)	1237 (58.3)	1422 (60.1)	.22
Reason for ICU admission, n(%)			
Transplant surgery	159 (7.5)	139 (5.9)	.03
Cardiovascular surgery	1425 (67.1)	1591 (67.2)	.94
Abdominal /Thorax/ Pelvic surgery	277 (13.0)	356 (15.0)	.06
Other	262 (12.3)	281 (11.9)	.64
ICU LOS, days, median (IQR)	3.0 (2.0-8.0)	4.0 (2.0-8.0)	.26
Hospital LOS, days, median (IQR)	15.0 (10.0-27.0)	14.0 (9.0-28.0)	.17
Mortality at 90 days, n(%)	240 (11.5)	251 (10.8)	.44
Dialysis, n(%)	155 (7.3)	181 (7.6)	.69

AKI acute kidney injury, CKD chronic kidney disease, ICU intensive care unit, IQR interquartile range, LOS length of stay.

Table 3.A.2 Baseline model candidate predictor list and univariable association with AKI-123 and AKI-23

A)	Baseline development cohort	AKI-123	no AKI	AKI-123 P-value	AKI-23 P-value
Age, years	65.9 (54.9-74.8)	68.8 (58.4-77.1)	64.5 (53.8-73.9)	<.001	.05
Gender male, n(%)	1346 (63.4)	368 (62.6)	978 (63.7)	.65	.79
Weight, kg	75.0 (65.0-85.0)	75.0 (65.0-85.0)	75.0 (65.0-85.0)	.49	.14
Height, cm	170.0 (163.0-176.0)	170.0 (163.0-175.0)	170.0 (163.0-176.0)	.13	.48
Baseline serum creatinine, mg/dl	0.91 (0.76-1.05)	0.97 (0.75-1.13)	0.90 (0.77-1.03)	.001	.47
Diabetes, n(%)	359 (16.9)	134 (22.8)	225 (14.7)	<.001	.004
Malignancy, n(%)	282 (13.3)	105 (17.9)	177 (11.5)	<.001	.01
Elective admission, n(%)	1237 (58.3)	209 (35.5)	1028 (67.0)	<.001	<.001
Reason for ICU admission, n(%)					
Transplant surgery	159 (7.5)	65 (11.1)	94 (6.1)	<.001	.01
Cardiovascular surgery	1425 (67.1)	331 (56.3)	1094 (71.3)	<.001	<.001
Abdominal/Thorax/Pelvic surgery	277 (13.0)	116 (19.7)	161 (10.5)	<.001	<.001
Other	262 (12.3)	76 (12.9)	186 (12.1)	.54	.05
B)	Admission development cohort	AKI-123	no AKI	AKI-123 P-value	AKI-23 P-value
Age, years	65.9 (54.9-74.8)	69.7 (58.9-77.3)	64.8 (53.8-73.9)	<.001	.03
Gender male, n(%)	1289 (63.4)	315 (62.6)	974 (63.6)	.70	.82
Weight, kg	75.0 (65.0-85.0)	75.0 (65.0-84.0)	75.0 (65.0-85.0)	.32	.38
Height, cm	170.0 (163.0-176.0)	170.0 (162.0-175.0)	170.0 (163.0-176.0)	.04	.29
Baseline serum creatinine, mg/dl	0.92 (0.76-1.05)	0.97 (0.75-1.16)	0.90 (0.77-1.03)	<.001	.27
Diabetes, n(%)	335 (16.5)	111 (22.1)	224 (14.7)	<.001	.02
Malignancy, n(%)	266 (13.1)	91 (18.3)	175 (11.4)	<.001	.01
Elective admission, n(%)	1232 (60.6)	204 (40.6)	1028 (67.1)	<.001	<.001
Reason for ICU admission, n(%)					
Transplant surgery	148 (7.3)	55 (10.9)	93 (6.1)	<.001	.02
Cardiovascular surgery	1384 (65.9)	291 (57.9)	1093 (71.4)	<.001	<.001
Abdominal/Thorax/Pelvic surgery	261 (10.0)	100 (19.9)	161 (10.5)	<.001	<.001
Other	241 (11.1)	57 (11.3)	184 (12.0)	.75	.23

Sepsis upon ICU admission, n(%)	408 (20.1)	185 (36.8)	223 (14.6)	<.001	<.001
Blood glucose upon ICU admission, mg/dl	135.0 (112.0-162.0)	141.0 (113.0-174.5)	134.0 (112.0-159.0)	<.001	0.005
Hemodynamic support on ICU admission, n(%)					
Pharmacological	1687 (82.9)	450 (89.5)	1228 (80.2)	<.001	<.001
Mechanical	51 (2.5)	33 (6.6)	18 (1.2)	<.001	.005
C)	Day1 development cohort	AKI-123	no AKI	AKI-123 P-value	AKI-23 P-value
Age, years	65.6 (54.6-74.5)	69.7 (58.8-78.0)	64.8 (53.8-73.9)	<.001	.02
Gender male, n(%)	1123 (63.3)	150 (61.2)	973 (63.6)	.47	.62
Weight, kg	75.0 (65.0-84.0)	75.0 (65.0-82.0)	75.0 (65.0-85.0)	.26	.34
Height, cm	170.0 (163.0-176.0)	169.0 (162.0-175.0)	170.0 (163.0-176.0)	.05	.05
Baseline serum creatinine, mg/dl	0.92 (0.77-1.05)	1.03 (0.81-1.37)	0.90 (0.77-1.03)	<.001	<.001
Diabetes, n(%)	273 (15.4)	49 (20.0)	224 (14.7)	.03	.62
Malignancy, n(%)	205 (11.6)	30 (12.2)	175 (11.4)	.74	.07
Elective admission, n(%)	1154 (65.1)	126 (51.4)	1028 (67.2)	<.001	.001
Reason for ICU admission, n(%)					
Transplant surgery	130 (7.3)	37 (15.1)	93 (6.1)	<.001	<.001
Cardiovascular surgery	1242 (70.0)	150 (61.2)	1092 (71.4)	<.001	.12
Abdominal/Thorax/Pelvic surgery	197 (11.1)	36 (14.7)	161 (10.5)	.06	.57
Other	205 (11.6)	22 (9.0)	183 (12.0)	.19	.20
Sepsis upon ICU admission, n(%)	284 (16.0)	62 (25.3)	222 (14.5)	<.001	.07
Blood glucose upon ICU admission, mg/dl	135.0 (113.0-161.0)	143.0 (120.0-178.0)	134.0 (112.0-159.0)	<.001	<.001
Hemodynamic support on ICU admission, n(%)					
Pharmacological	1459 (82.2)	225 (91.8)	1226 (80.1)	<.001	<.001
Mechanical	34 (1.9)	16 (6.5)	18 (1.2)	<.001	<.001
Serum creatinine on day 1, mg/dl	0.87 (0.69-1.09)	1.13 (0.85-1.48)	0.85 (0.69-1.03)	<.001	<.001
Apache II on day 1	18.0 (13.0-29.0)	26.0 (18.0-35.0)	17.0 (13.0-27.0)	<.001	<.001
Maximum lactate on day 1, mg/dl	1.8 (1.4-2.5)	2.3 (1.6-3.2)	1.8 (1.3-2.4)	<.001	<.001
Bilirubin on day 1, mg/dl	0.75 (0.51-1.14)	0.86 (0.51-1.55)	0.74 (0.51-1.09)	0.001	.01
Sofa on day 1	7.0 (6.0-9.0)	9.0 (7.0-10.0)	7.0 (6.0-8.0)	<.001	<.001

Blood glucose on day 1, mg/dl	104.0 (90.0-120.0)	105.0 (89.0-132.0)	104.0 (91.0-119.0)	.16	.37
Urea on day 1, mg/dl	32.0 (25.0-41.0)	46.0 (32.0-62.3)	30.0 (23.8-39.0)	<.001	<.001
Urine over 24h					
Total amount, ml	1650.0 (1213.0-2210.3)	1200.0 (970.0-1666.5)	1710.0 (1311.0-2260.0)	<.001	n.a.
Slope ¹	-0.0003 (-0.0011-0.0004)	-0.0005 (-0.0011-0.0000)	-0.0003 (-0.0011-0.0005)	<.001	n.a.
Number of measurements ²	18.0 (15.0-20.0)	17.0 (15.0-19.5)	18.0 (15.0-20.0)	.15	n.a.
Protein, g/L	0.29 (0.26-0.60)	0.28 (0.23-0.43)	0.60 (0.43-1.11)	.21	n.a.
White blood cell, cells/ μ L	0.026 (0.012-0.048)	0.026 (0.013-0.055)	0.026 (0.012-0.045)	.23	n.a.
Red blood cell, cells/ μ L	0.026 (0.010-0.121)	0.048 (0.016-0.378)	0.022(0.009-0.100)	.04	n.a.
Heart frequency, beat/min					
Mean	81.9 (74.9-89.5)	84.0 (77.7-92.1)	81.4 (74.5-89.1)	<.001	n.a.
Median	81.0 (74.1-90.0)	84.0 (79.0-92.4)	81.0 (74.0-89.0)	<.001	n.a.
Standard deviation	7.5 (5.4-10.5)	7.6 (5.1-10.8)	7.5 (5.5-10.4)	.41	n.a.
Time ³ above 120	0.0 (0.0-0.0)	0.0 (0.0-3.0)	0.0 (0.0-0.0)	.06	n.a.
Time ³ above 110	0.0 (0.0-24.0)	0.0 (0.0-55.0)	0.0 (0.0-21.0)	<.001	n.a.
Time ³ above 100	8.0 (0.0-176.75)	18.0 (0.0-330.0)	6.0 (0.0-165.0)	.006	n.a.
Time ³ below 60	0.0 (0.0-28.0)	0.0(0.0-19.0)	0.0(0.0-30.0)	.48	n.a.
Time ³ below 50	0.0 (0.0-0.0)	0.0(0.0-0.0)	0.0(0.0-0.0)	.21	n.a.
Time ³ below 40	0.0 (0.0-0.0)	0.0(0.0-0.0)	0.0(0.0-0.0)	.39	n.a.
Time ³ inside the CI	2001.5 (1776.3-2077.0)	2019.0 (1774.0-2087.0)	1999.0 (1777.0-2076.0)	.15	n.a.
Time ³ outside the CI	608.0 (522.0-656.0)	601.0 (516.0-661.0)	609.0 (523.0-656.0)	.30	n.a.
Time ³ above the mean	657.5 (566.0-764.0)	661.0 (573.0-781.0)	657.0 (564.0-762.0)	.23	n.a.
Time ³ below the mean	712.5 (596.3-808.0)	717.0(616.0-843.0)	711.0(594.0-806.0)	.06	n.a.
Time ³ above the median	643.0 (567.0-684.0)	649.0 (574.0-686.0)	642.0 (563.0-684.0)	.21	n.a.
Time ³ below the median	644.0 (565.0-685.0)	649.0 (559.0-685.0)	643.0(565.0-685.0)	.36	n.a.
Mean arterial blood pressure, mmHg					
Mean	73.8 (69.3-79.2)	72.0 (67.4-77.7)	74.0 (69.6-79.3)	<.001	n.a.
Median	73.0 (68.0-78.0)	72.0 (66.0-77.0)	73.0 (68.0-79.0)	<.001	n.a.
Standard deviation	8.9 (7.3-10.9)	9.2 (7.4-11.2)	8.9 (7.3-10.8)	.15	n.a.
Time ³ above 120	0.0 (0.0-1.0)	0.0 (0.0-0.0)	0.0 (0.0-1.0)	.30	n.a.
Time ³ above 110	0.0 (0.0-7.0)	0.0 (0.0-7.0)	0.0(0.0-7.0)	.01	n.a.
Time ³ above 100	9.0 (0.03-41.0)	5.0 (0.0-36.0)	10.0 (0.0-42.0)	.01	n.a.
Time ³ below 60	52.0 (7.0-208.3)	114.0 (18.0-377.0)	47.0 (5.0-189.0)	<.001	n.a.

Time ³ below 50	0.0 (0.0-4.0)	1.0 (0.0-16.0)	0.0 (0.0-3.0)	<.001	n.a.
Time ³ below 40	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	.02	n.a.
Time ³ inside the CI	2062.0 (1956.0-2100.8)	2073.0 (2006.0-2103.0)	2060.0 (1947.0-2100.0)	.001	n.a.
Time ³ outside the CI	647.0 (607.0-676.0)	648.0 (325.0-676.0)	647.0 (604.0-676.0)	.03	n.a.
Time ³ above the mean	630.0 (565.3-692.0)	642.0 (589.0-701.0)	628.0 (561.0-691.0)	.006	n.a.
Time ³ below the mean	751.0 (676.0-815.0)	764.0 (691.0-816.0)	748.0 (674.0-814.0)	.09	n.a.
Time ³ above the median	670.0 (629.0-694.0)	672.0 (646.0-693.0)	669.0 (626.0-695.0)	.04	n.a.
Time ³ below the median	670.0 (627.0-694.0)	675.0 (644.0-697.0)	668.0 (623.0-693.0)	<.001	n.a.
Radio-contrasts					
n (%)	116 (6.5)	16 (6.5)	100 (6.5)	>.99	n.a.
Number of radio-contrast administrations	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0(0.0-0.0)	.49	n.a.
Maximum osmolality	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0(0.0-0.0)	.49	n.a.
Maximum viscosity	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	.49	n.a.
N ionic radio-contrast (%)	4 (0.2)	1 (0.4)	3 (0.2)	.44	n.a.
Medications					
Inotropes and vasopressors dose, mg	6.4e-06 (3.7e-07-4.7e-05)	3.4e-05 (5.5e-06-2.2e-04)	5.1e-06 (1.8e-07-3.0e-05)	<.001	n.a.
Inotropes and vasopressors, n (%)	1405 (79.2)	213 (86.9)	1192 (78.0)	<.001	n.a.
Vancomycin dose, mg	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	.02	n.a.
Vancomycin, n (%)	48 (2.7)	16 (6.5)	32 (2.1)	<.001	n.a.
Aminoglycoside, n (%)	83 (4.7)	26 (10.6)	57 (3.7)	<.001	n.a.
Antiviral, n (%)	91 (5.1)	26 (10.6)	65 (4.3)	<.001	n.a.
Beta-lactam antibiotics, n (%)	347 (19.6)	65 (26.5)	282 (18.4)	.004	n.a.
Diuretics, n (%)	604 (34.0)	64 (26.1)	540 (35.3)	.004	n.a.
Ciclosporin/tacrolimus, n (%)	121 (6.8)	28 (11.4)	93 (6.1)	.003	n.a.
Antifungals, n (%)	20 (1.1)	6 (2.4)	14 (0.9)	.04	n.a.
Anti-inflammatory drugs, n (%)	82 (4.6)	7 (2.9)	75 (4.9)	.18	n.a.
Ace inhibitors, n (%)	14 (0.7)	2 (0.8)	12 (0.8)	>.99	n.a.

Values are expressed as median (IQR) unless indicated otherwise. Antifungals are Diflucan, VFend and AmBisome. Antivirals are Cymevene, Zovirax, Valcyte and Docaciclo. Anti-inflammatory drugs are Taradyl, Ibuprofene and Voltaren.

AKI acute kidney injury, *ICU* intensive care unit, *IQR* interquartile range, *n.a.*, not assessed.

¹ The urine slope refers to the slope of a linear model fitted to the hourly urine values.

² The number of measurements refers to the number of time the urine output was measured at the patient's bedside.

³ Time in minutes.

Table 3.A.4 Radio-contrast predictor information

Radio-contrast	Ionicity (yes/no)	Osmolality (mOsm/kg water)	Viscosity (cp 37 °C)
DOTAREM	yes	1350	2.4
IOMERON 250	no	435	7.5
IOMERON 300	no	521	4.5
IOMERON 350	no	618	7.5
IOMERON 400	no	726	12.6
LIPIODOL	no	500	2.5
MULTIHANCE	yes	1950	5.3
OMNIPAQUE 240	no	520	3.4
OMNIPAQUE 300	no	672	6.3
OPTIRAY 350	no	651	9
TELEBRIX 12	yes	640	1.1
TELEBRIX 35	yes	2130	7.5
ULTRAVIST 240	no	483	2.8
ULTRAVIST 370	no	774	10
UROGRAFINE 30	no	710	1.4
VISIPAQUE 270	no	290	6.3
VISIPAQUE 320	no	290	11.8
XENETIX 350	no	915	10

Table 3.A.5 Performance for AKI-123 and AKI-23 prediction in the development cohort

A) AKI-123	Baseline model	Admission model	Day1 model	Day1+ model
Number of patients	2123	2034	1774	1774
AKI prevalence	27.70	24.72	13.81	13.81
AUROC	0.77 [0.77-0.77]	0.80 [0.80-0.80]	0.86 [0.86-0.86]	0.87 [0.86-0.87]
Sensitivity ¹	0.71 [0.71-0.72]	0.71 [0.71-0.72]	0.78 [0.78-0.78]	0.79 [0.78-0.79]
Specificity ¹	0.70 [0.70-0.71]	0.74 [0.74-0.74]	0.81 [0.80-0.81]	0.79 [0.80-0.80]
Calibration slope	0.83 [0.82-0.83]	0.92 [0.92-0.93]	0.87 [0.87-0.88]	0.96 [0.95-0.97]
Calibration in the large	-0.00 [-0.00 to 0.00]	-0.00 [-0.00 to -0.00]	-0.00 [-0.01 to -0.00]	-0.01 [-0.01 to -0.00]
Δ Net benefit _{None} (w.r.t. treat-none) ¹	0.13 [0.13-0.13]	0.12 [0.12-0.12]	0.08 [0.08-0.09]	0.09 [0.09-0.09]
Δ Net benefit _{All} (w.r.t. treat-all) ¹	0.09 [0.09-0.09]	0.08 [0.08-0.08]	0.07 [0.07-0.07]	0.08 [0.07-0.08]
Classification threshold (%)	25.80	20.71	14.46	15.81
B) AKI-23	Baseline model	Admission model	Day1 model	Day1+ model
Number of patients	2123	2034	1774	1774
AKI prevalence (%)	14.04	11.31	4.28	4.28
AUROC	0.81 [0.81-0.81]	0.83 [0.83-0.83]	0.88 [0.88-0.89]	n.a.
Sensitivity ¹	0.79 [0.78-0.79]	0.77 [0.76-0.77]	0.81 [0.81-0.81]	n.a.
Specificity ¹	0.71 [0.71-0.72]	0.75 [0.75-0.75]	0.82 [0.81-0.82]	n.a.
Discrimination slope	0.20 [0.20-0.20]	0.19 [0.19-0.20]	0.19 [0.18-0.19]	n.a.
Calibration slope	0.76 [0.76-0.77]	0.82 [0.81-0.83]	0.74 [0.73-0.77]	n.a.
Calibration in the large	-0.00 [-0.00 to -0.00]	-0.00 [-0.00 to -0.00]	-0.00 [-0.00 to -0.00]	n.a.
Δ Net benefit _{None} (w.r.t. treat-none) ¹	0.08 [0.08-0.08]	0.06 [0.06-0.06]	0.03 [0.03-0.03]	n.a.
Δ Net benefit _{All} (w.r.t. treat-all) ¹	0.06 [0.06-0.06]	0.06 [0.06-0.06]	0.03 [0.03-0.03]	n.a.
Classification threshold (%)	15.91	13.55	6.97	n.a.

Median and 95% CI were obtained from 400 bootstrap replicas.

AKI acute kidney injury, AUROC area under the ROC curve, n.a. not applicable, ROC receiver operator characteristic.

¹ Evaluated at the threshold that maximized sensitivity and specificity in the development cohort.

² The small number of patients developing AKI-23 precluded the development of the complex Day1+ model.

Table 3.A.6 Performance for AKI-23 prediction in the validation cohort

AKI-23	Baseline model	Admission model	Day1 model	Day1+ model
Number of patients	2367	2274	1954	n.a.
AKI prevalence (%)	14.78	12.00	4.19	n.a.
AUROC	0.77 [0.77-0.77]	0.79 [0.79-0.79]	0.84 [0.83-0.83]	n.a.
Sensitivity ¹	0.64 [0.64-0.64]	0.65 [0.65-0.65]	0.56 [0.56-0.56]	n.a.
Specificity ¹	0.76 [0.76-0.76]	0.78 [0.78-0.78]	0.90 [0.90-0.90]	n.a.
Positive predictive value ¹	0.31 [0.31-0.32]	0.29 [0.29-0.29]	0.20 [0.20-0.20]	n.a.
Discrimination slope	0.15 [0.15-0.15]	0.15 [0.15-0.15]	0.12 [0.12-0.13]	n.a.
Calibration-slope	0.63 [0.63-0.63]	0.67 [0.66-0.67]	0.44 [0.43-0.45]	n.a.
Calibration in the large	-0.01 [-0.01 to -0.01]	-0.01 [-0.01 to -0.01]	0.00 [-0.00 to 0.00]	n.a.
Δ Net benefit _{None} (w.r.t. treat-none) ¹	0.06 [0.06-0.06]	0.05 [0.05-0.05]	0.02 [0.02-0.02]	n.a.
Δ Net benefit _{All} (w.r.t. treat-all) ¹	0.06 [0.06-0.06]	0.06 [0.06-0.06]	0.04 [0.04-0.04]	n.a.
Classification threshold (%)	15.91	13.55	6.97	n.a.

Estimated 95% confidence intervals obtained from 2000 bootstrap replicas are shown between square brackets.

AKI acute kidney injury, AUROC area under the ROC curve, n.a. not applicable, ROC receiver operator characteristic.

¹ Evaluated at the threshold that maximized sensitivity and specificity in the development cohort, reported in the last row.

Table 3.A.7 Performance for AKI-123 prediction in the validation cohort in the subset of cardiac patients

AKI-123	Baseline model	Admission model	Day1 model	Day1+ model
Number of patients	1591	1559	1406	1245
AKI prevalence (%)	23.95	22.39	13.94	13.49
AUROC	0.77 [0.77-0.77]	0.79 [0.79-0.79]	0.82 [0.82-0.82]	0.84 [0.84-0.84]
Sensitivity ¹	0.60 [0.60-0.61]	0.64 [0.64-0.64]	0.59 [0.59-0.59]	0.62 [0.62-0.62]
Specificity ¹	0.80 [0.80-0.80]	0.79 [0.79-0.79]	0.86 [0.86-0.86]	0.90 [0.90-0.90]
Positive predictive value ¹	0.49 [0.49-0.49]	0.46 [0.46-0.46]	0.41 [0.41-0.41]	0.49 [0.49-0.49]
Discrimination slope	0.20 [0.20-0.21]	0.20 [0.20-0.20]	0.25 [0.25-0.25]	0.25 [0.25-0.25]
Calibration slope	0.88 [0.88-0.88]	0.91 [0.90-0.91]	0.86 [0.85-0.86]	1.03 [1.03-1.03]
Calibration in the large	-0.01 [-0.01 to -0.01]	-0.02 [-0.02 to -0.02]	-0.02 [-0.02 to -0.02]	-0.03 [-0.03 to -0.03]
Δ Net benefit _{None} (w.r.t. treat-none) ¹	0.09 [0.09-0.10]	0.10 [0.10-0.10]	0.07 [0.07-0.07]	0.07 [0.07-0.07]
Δ Net benefit _{All} (w.r.t. treat-all) ¹	0.11 [0.11-0.11]	0.07 [0.07-0.08]	0.06 [0.06-0.06]	0.09 [0.09-0.09]
Classification threshold (%)	25.80	20.71	14.46	15.81

Estimated 95% confidence intervals obtained from 2000 bootstrap replicas are shown between square brackets.

AKI acute kidney injury, AUROC area under the ROC curve, n.a. not applicable, ROC receiver operator characteristic.

¹ Evaluated at the threshold that maximized sensitivity and specificity in the development cohort, reported in the last row.

² External validation was not performed as there were less than 50 patients with AKI-23 in this cohort.

Table 3.A.8 Performance for AKI-23 prediction in the validation cohort in the subset of cardiac patients

AKI-23	Baseline model	Admission model	Day1 model ²	Day1+ model ²
Number of patients	1591	1559	1406	1406
AKI prevalence (%)	10.25	8.40	3.41	3.41
AUROC	0.77 [0.77-0.78]	0.78 [0.78-0.78]	n.a.	n.a.
Sensitivity ¹	0.60 [0.60-0.60]	0.49 [0.49-0.49]	n.a.	n.a.
Specificity ¹	0.85 [0.85-0.85]	0.86 [0.86-0.86]	n.a.	n.a.
Positive predictive value ¹	0.31 [0.31-0.31]	0.25 [0.25-0.25]	n.a.	n.a.
Discrimination slope	0.16 [0.16-0.16]	0.13 [0.13-0.13]	n.a.	n.a.
Calibration slope	0.54 [0.53-0.54]	0.40 [0.39-0.40]	n.a.	n.a.
Calibration in the large	-0.00 [-0.00 to -0.00]	0.00 [0.00-0.00]	n.a.	n.a.
Δ Net benefit _{None} (w.r.t. treat-none) ¹	0.04 [0.04-0.04]	0.02 [0.02-0.02]	n.a.	n.a.
Δ Net benefit _{All} (w.r.t. treat-all) ¹	0.09 [0.09-0.09]	0.08 [0.08-0.08]	n.a.	n.a.
Classification threshold (%)	15.91	13.55	6.97	n.a.

Estimated 95% confidence intervals obtained from 2000 bootstrap replicas are shown between square brackets.

AKI acute kidney injury, AUROC area under the ROC curve, n.a. not applicable, ROC receiver operator characteristic.

¹ Evaluated at the threshold that maximized sensitivity and specificity in the development cohort, reported in the last row.

² External validation was not performed as there were less than 50 patients with AKI-23 in this cohort.

Table 3.A.9 Performance for AKI-123 prediction in the validation cohort in the subset of septic patients

AKI-123	Baseline model ²	Admission model	Day1 model	Day1+ model
Number of patients	n.a.	455	302	290
AKI prevalence (%)	n.a.	48.35	22.85	23.45
AUROC	n.a.	0.65 [0.65-0.65]	0.76 [0.76-0.76]	0.77 [0.76-0.77]
Sensitivity ¹	n.a.	0.97 [0.97-0.97]	0.76 [0.75-0.76]	0.62 [0.62-0.63]
Specificity ¹	n.a.	0.11 [0.11-0.12]	0.64 [0.64-0.64]	0.78 [0.77-0.78]
Positive predictive value ¹	n.a.	0.50 [0.50-0.51]	0.39 [0.39-0.39]	0.46 [0.46-0.46]
Discrimination slope	n.a.	0.09 [0.09-0.10]	0.18 [0.18-0.18]	0.17 [0.17-0.17]
Calibration slope	n.a.	0.68 [0.67-0.68]	0.67 [0.67-0.69]	0.95 [0.96-0.98]
Calibration in the large	n.a.	0.01 [0.01-0.01]	-0.03 [-0.03 to -0.03]	-0.06 [-0.07 to -0.06]
Δ Net benefit _{None} (w.r.t. treat-none) ¹	n.a.	0.35 [0.35-0.35]	0.13 [0.13-0.13]	0.12 [0.12-0.12]
Δ Net benefit _{All} (w.r.t. treat-all) ¹	n.a.	0.00 [-0.00-0.00]	0.02 [0.02-0.02]	0.02 [0.02-0.02]
Classification Threshold (%)	n.a.	20.71	14.46	15.81

Estimated 95% confidence intervals obtained from 2000 bootstrap replicas are shown between square brackets.

AKI acute kidney injury, AUROC area under the ROC curve, n.a. not applicable, ROC receiver operator characteristic.

¹ Evaluated at the threshold that maximized sensitivity and specificity in the development cohort, reported in the last row.

² The baseline model cannot be evaluated in this cohort as sepsis is only known upon ICU admission.

Table 3.A.10 Performance for AKI-23 prediction in the validation NGAL cohort

AKI-23	NGAL			Admission model	Combined model
Classification threshold	150	200	400	13.55%	9.18%
	0.0798	0.0916	0.155		
Number of patients	2081			2081	2081
AKI prevalence (%)	12.59			12.59	12.59
AUROC	0.79 [0.79-0.79]			0.79 [0.79-0.79]	0.83 [0.83-0.83]
Sensitivity	0.92 [0.92-0.92]	0.80 [0.80-0.80]	0.45 [0.44-0.45]	0.65 [0.65-0.65] ¹	0.81 [0.81-0.81] ²
Specificity	0.38 [0.38-0.38]	0.61 [0.61-0.61]	0.91 [0.91-0.91]	0.78 [0.78-0.77] ¹	0.73 [0.73-0.73] ²
Positive predictive value	0.18 [0.17-0.18]	0.23 [0.23-0.23]	0.41 [0.41-0.41]	0.29 [0.30-0.30] ¹	0.31 [0.31-0.31] ²
Discrimination slope	0.15 [0.15-0.15]			0.15 [0.15-0.15]	0.19 [0.19-0.19]
Calibration-slope	0.70 [0.69-0.70]			0.69 [0.68-0.69]	0.81 [0.81-0.81]
Calibration in the large	0.00 [0.00-0.00]			-0.01 [-0.01-0.01]	0.00 [0.00-0.00]
Δ Net benefit _{None} (w.r.t. treat-none)	0.06 [0.06-0.06]	0.07 [0.07-0.07]	0.04 [0.04-0.04]	0.05 [0.05-0.05] ¹	0.08 [0.08-0.08] ²
Δ Net benefit _{All} (w.r.t. treat-all)	0.00 [0.00-0.00]	0.02 [0.02-0.02]	0.07 [0.07-0.07]	0.06 [0.06-0.06] ¹	0.03 [0.03-0.03] ²

Estimated 95% confidence intervals obtained from 2000 bootstrap replicas are shown between square brackets. Values highlighted in bold are significantly different from the value in the previous column at a statistical level of P below 0.05.

AKI acute kidney injury, AUROC area under the ROC curve, ROC receiver operator characteristic.

¹ Evaluated at the threshold that maximized sensitivity and specificity in the development cohort, reported in the last row.

² Evaluated at the threshold that maximized sensitivity and specificity in the validation cohort. Classification thresholds are reported in the first row..

Table 3.A.11 Performance for AKI-123 prediction in the validation NGAL cohort in the subset of cardiac patients

A) AKI-123	NGAL			Admission model	Combined model
Classification threshold	150	200	400	20.71%	20.98%
	0.19	0.22	0.36		
Number of patients	1405			1405	1405
AKI prevalence (%)	22.92			22.92	22.92
AUROC	0.74 [0.74-0.74]			0.80 [0.80-0.80]	0.82 [0.81-0.82]
Sensitivity	0.68 [0.68-0.68]	0.56 [0.56-0.56]	0.25 [0.25-0.25]	0.65 [0.65-0.65] ¹	0.68 [0.67-0.68] ²
Specificity	0.65 [0.65-0.65]	0.79 [0.79-0.79]	0.96 [0.96-0.96]	0.79 [0.79-0.79] ¹	0.82 [0.82-0.82] ²
Positive predictive value	0.37 [0.37-0.37]	0.45 [0.45-0.45]	0.65 [0.65-0.66]	0.48 [0.48-0.48] ¹	0.52 [0.52-0.53] ²
Discrimination slope	0.12 [0.12-0.12]			0.21 [0.21-0.21]	0.23 [0.23-0.23]
Calibration-slope	0.78 [0.77-0.78]			0.91 [0.91-0.92]	0.82 [0.82-0.82]
Calibration in the large	0.00 [0.00-0.00]	-0.02 [-0.02 to -0.02]	0.00 [0.00-0.00]		
Δ Net benefit _{None} (w.r.t. treat-none)	0.09 [0.09-0.09]	0.08 [0.08-0.08]	0.04 [0.04-0.04]	0.11 [0.11-0.11] ¹	0.12 [0.12-0.12] ²
Δ Net benefit _{All} (w.r.t. treat-all)	0.05 [0.05-0.05]	0.07 [0.07-0.07]	0.25 [0.25-0.25]	0.07 [0.07-0.07] ¹	0.07 [0.07-0.07] ²

Estimated 95% confidence intervals obtained from 2000 bootstrap replicas are shown between square brackets. Values highlighted in bold are significantly different from the value in the previous column at a statistical level of P below 0.05.

AKI acute kidney injury, AUROC area under the ROC curve, ROC receiver operator characteristic.

¹ Evaluated at the threshold that maximized sensitivity and specificity in the development cohort, reported in the last row.

² Evaluated at the threshold that maximized sensitivity and specificity in the entire validation cohort. Classification thresholds are reported in the first row.

Table 3.A.12 Performance for AKI-23 prediction in the validation NGAL cohort in the subset of cardiac patients

B) AKI-23	NGAL			Admission model	Combined model
Classification threshold	150	200	400	13.55%	9.18%
	0.08	0.09	0.16		
Number of patients	1405			1405	1405
AKI prevalence (%)	9.11			9.11	9.11
AUROC	0.80 [0.79-0.80]			0.78 [0.78-0.78]	0.84 [0.84-0.85]
Sensitivity	0.89 [0.89-0.89]	0.77 [0.77-0.78]	0.34 [0.34-0.34]	0.50 [0.50-0.50] ¹	0.70 [0.70-0.70] ²
Specificity	0.43 [0.42-0.43]	0.67 [0.67-0.68]	0.95 [0.95-0.95]	0.86 [0.86-0.86] ¹	0.84 [0.84-0.84] ²
Positive predictive value	0.13 [0.13-0.14]	0.19 [0.19-0.19]	0.39 [0.38-0.39]	0.26 [0.26-0.27] ¹	0.31 [0.31-0.31] ²
Discrimination slope	0.09 [0.09-0.09]			0.14 [0.13-0.14]	0.13 [0.12-0.13]
Calibration-slope	0.48 [0.46-0.48]			0.45 [0.44-0.46]	0.59 [0.58-0.60]
Calibration in the large	0.01 [0.01-0.01]			-0.01 [-0.01 to -0.00]	0.00 [0.00-0.00]
Δ Net benefit _{None} (w.r.t. treat-none)	0.03 [0.03-0.03]	0.04 [0.04-0.04]	0.02 [0.02-0.02]	0.03 [0.03-0.03] ¹	0.05 [0.05-0.05] ²
Δ Net benefit _{All} (w.r.t. treat-all)	0.01 [0.01-0.01]	0.03 [0.03-0.03]	0.09 [0.09-0.09]	0.07 [0.07-0.07] ¹	0.04 [0.04-0.04] ²

Estimated 95% confidence intervals obtained from 2000 bootstrap replicas are shown between square brackets. Values highlighted in bold are significantly different from the value in the previous column at a statistical level of P below 0.05.

AKI acute kidney injury, AUROC area under the ROC curve, ROC receiver operator characteristic.

¹ Evaluated at the threshold that maximized sensitivity and specificity in the development cohort, reported in the last row.

² Evaluated at the threshold that maximized sensitivity and specificity in the entire validation cohort. Classification thresholds are reported in the first row.

Table 3.A.13 Performance for AKI-123 prediction in the validation NGAL cohort in the subset of septic patients

AKI-123	NGAL			Admission model	Combined model
Classification threshold	150	200	400	20.71%	20.98%
	0.20	0.23	0.37		
Number of patients	438			438	438
AKI prevalence (%)	48.17			48.17	48.17
AUROC	0.71 [0.71-0.71]			0.65 [0.65-0.65]	0.74 [0.74-0.74]
Sensitivity	0.95 [0.95-0.95]	0.85 [0.85-0.85]	0.56 [0.56-0.57]	0.97 [0.97-0.97] ¹	0.98 [0.98-0.98] ²
Specificity	0.19 [0.18-0.19]	0.32 [0.32-0.32]	0.74 [0.73-0.74]	0.12 [0.12-0.12] ¹	0.10 [0.10-0.10] ²
Positive predictive value	0.52 [0.52-0.53]	0.54 [0.54-0.54]	0.67 [0.66-0.67]	0.50 [0.50-0.50] ¹	0.51 [0.50-0.51] ²
Discrimination slope	0.18 [0.18-0.18]			0.10 [0.09-0.10]	0.19 [0.19-0.19]
Calibration slope	0.74 [0.73-0.74]			0.69 [0.68-0.69]	0.97 [0.97-0.98]
Calibration in the large	-0.07 [-0.07 to -0.07]			0.02 [0.01-0.02]	0.05 [0.05-0.05]
Δ Net benefit _{None} (w.r.t. treat-none) ¹	0.36 [0.36-0.36]	0.32 [0.32-0.32]	0.19 [0.19-0.19]	0.35 [0.35-0.35] ¹	0.36 [0.36-0.36] ²
Δ Net benefit _{All} (w.r.t. treat-all) ¹	0.00 [0.00-0.00]	-0.01 [-0.01 to -0.01]	0.00 [0.00-0.01]	0.00 [-0.00 to -0.00] ¹	0.01 [0.01-0.01] ²

Estimated 95% confidence intervals obtained from 2000 bootstrap replicas are shown between square brackets. Values highlighted in bold are significantly different from the value in the previous column at a statistical level of P below 0.05.

AKI acute kidney injury, AUROC area under the ROC curve, ROC receiver operator characteristic.

¹ Evaluated at the threshold that maximized sensitivity and specificity in the development cohort, reported in the last row.

² Evaluated at the threshold that maximized sensitivity and specificity in the entire validation cohort. Classification thresholds are reported in the first row.

Table 3.A.14 Performance for AKI-123 and AKI-23 prediction using NGAL and Day1 models in the validation NGAL cohort

A) AKI-123		NGAL		Day1 model
Classification Threshold	150	200	400	14.46%
	0.13	0.14	0.19	
Number of patients	1779			1779
AKI prevalence (%)	15.06			15.06
AUROC	0.67 [0.67-0.67]			0.81 [0.81-0.81]
Sensitivity	0.81 [0.81-0.81]	0.60 [0.60-0.60]	0.17 [0.17-0.18]	0.67 [0.67-0.67] ¹
Specificity	0.42 [0.42-0.42]	0.65 [0.65-0.65]	0.93 [0.93-0.93]	0.81 [0.81-0.81] ¹
Positive predictive value	0.20 [0.20-0.20]	0.23 [0.23-0.23]	0.30 [0.30-0.30]	0.38 [0.38-0.38] ¹
Discrimination slope	0.03 [0.03-0.03]			0.24 [0.24-0.24]
Calibration slope	0.57 [0.55-0.58]			0.78 [0.78-0.79]
Calibration in the large	0.00 [0.0-0.00]			-0.01 [-0.01 to -0.01]
Δ Net benefit _{None} (w.r.t. treat-none) ¹	0.04 [0.04-0.04]	0.05 [0.05-0.05]	0.02 [0.02-0.02]	0.08 [0.08-0.08]
Δ Net benefit _{All} (w.r.t. treat-all) ¹	0.01 [0.01-0.01]	0.02 [0.02-0.03]	0.05 [0.05-0.06]	0.06 [0.06-0.06]
B) AKI-23		NGAL		Day1 model
Classification threshold	150	200	400	6.97%
	0.037	0.04	0.057	
Number of patients	1779			1779
AKI prevalence (%)	4.50			4.50
AUROC	0.70 [0.70-0.70]			0.83 [0.83-0.83]
Sensitivity	0.84 [0.84-0.84]	0.69 [0.69-0.69]	0.23 [0.23-0.23]	0.56 [0.56-0.56] ¹
Specificity	0.39 [0.39-0.40]	0.62 [0.62-0.63]	0.92 [0.92-0.92]	0.89 [0.89-0.89] ¹
Positive predictive value	0.06 [0.06-0.06]	0.08 [0.08-0.08]	0.12 [0.12-0.12]	0.20 [0.20-0.20] ¹
Discrimination slope	0.02 [0.02-0.02]			0.12 [0.12-0.12]
Calibration slope	0.37 [0.41-0.45]			0.42 [0.41-0.43]
Calibration in the large	0.00 [0.00-0.00]			-0.00 [-1.00-0.00]
Δ Net benefit _{None} (w.r.t. treat-none) ¹	0.02 [0.02-0.02]	0.02 [0.02-0.02]	0.01 [0.01-0.01]	0.02 [0.02-0.02]
Δ Net benefit _{All} (w.r.t. treat-all) ¹	0.00 [0.00-0.00]	0.00 [0.00-0.00]	0.02 [0.02-0.02]	0.04 [0.04-0.04]

A) Prediction of AKI-123. **B)** Prediction of AKI-23. Estimated 95% confidence intervals obtained from 2000 bootstrap replicas are shown between square brackets. Values highlighted in bold are significantly different from the value in the previous column at a statistical level of P below 0.05.

AKI acute kidney injury, AUROC area under the ROC curve, ROC receiver operator characteristic.

¹ Evaluated at the threshold that maximized sensitivity and specificity in the development cohort, reported in the last row.

Table 3.A.15 Performance for AKI-123 prediction in the subset of patients with true and with calculated baseline serum creatinine.

AKI-123	Validation NGAL cohort true baseline			Validation NGAL cohort calculated baseline		
	NGAL	Admission model	Combined model	NGAL	Admission model	Combined model
Number of patients	1623	1623	1623	458	458	458
AKI-123 prevalence (%)	26.12	26.12	26.12	30.79	30.79	30.79
AUROC	0.73 [0.73-0.73]	0.79 [0.79-0.79]	0.81 [0.81-0.81]	0.74 [0.74-0.74]	0.70 [0.70-0.70]	0.76 [0.76-0.76]

Estimated 95% confidence intervals obtained from 2000 bootstrap replicas are shown between square brackets. Values highlighted in bold are significantly different from the value in the previous column at a statistical level of P below 0.05.

AKI acute kidney injury, AUROC area under the ROC curve, ROC receiver operator characteristic

Table 3.A.16 Performance for AKI-123 and AKI-23 prediction models and of admission NGAL for outcome prediction

	Validation cohort				Validation NGAL cohort		
	Baseline AUROC	Admission AUROC	Day1 AUROC	Day1+ AUROC	Admission AUROC	NGAL AUROC	Combined AUROC
A) AKI-123 model							
Requirement of RRT	0.73 [0.73-0.73]	0.77 [0.77-0.78]	0.84 [0.84-0.84]	0.84 [0.84-0.84]	0.77 [0.77-0.77]	0.78 [0.78-0.78]	0.82 [0.82-0.82]
Mortality or RRT at hospital discharge	0.75 [0.75-0.75]	0.80 [0.79-0.80]	0.74 [0.74-0.74]	0.74 [0.74-0.75]	0.79 [0.79-0.79]	0.67 [0.67-0.67]	0.80 [0.80-0.80]
Mortality at hospital discharge	0.75 [0.75-0.75]	0.79 [0.79-0.79]	0.74 [0.74-0.74]	0.74 [0.74-0.74]	0.79 [0.79-0.79]	0.66 [0.66-0.66]	0.79 [0.79-0.79]
B) AKI-23 model							
Requirement of RRT	0.72 [0.72-0.72]	0.78 [0.78-0.78]	0.82 [0.82-0.82]	n.a.	0.78 [0.78-0.78]	0.78 [0.78-0.78]	0.83 [0.82-0.83]
Mortality or RRT at hospital discharge	0.76 [0.76-0.76]	0.80 [0.80-0.80]	0.76 [0.76-0.76]	n.a.	0.80 [0.80-0.80]	0.67 [0.67-0.67]	0.81 [0.81-0.81]
Mortality at hospital discharge	0.76 [0.76-0.76]	0.80 [0.80-0.80]	0.76 [0.76-0.76]	n.a.	0.80 [0.80-0.80]	0.66 [0.66-0.66]	0.80 [0.80-0.81]

A) AKI-123 models. **B)** AKI-23 models. Estimated 95% confidence intervals obtained from 2000 bootstrap replicas are shown between square brackets. Values highlighted in bold are significantly different from the value in the previous column at a statistical level of P below 0.05.

AKI acute kidney injury, AUROC area under the ROC curve, n.a. not applicable, ROC receiver operator characteristic, RRT renal replacement therapy.

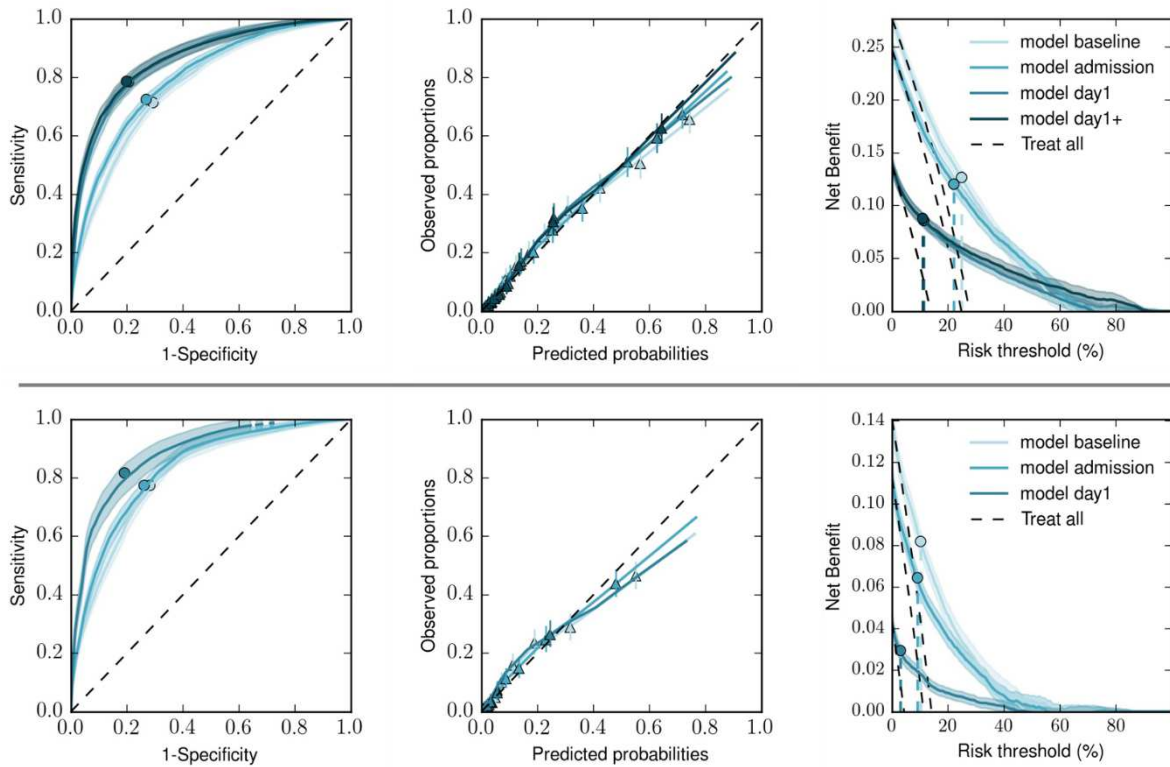


Figure 3.A.1 Performance of clinical prediction models in the development cohort.

a) Top row: AKI-123. b) Bottom row: AKI-23. First column: ROC curves. Second column: Calibration curves. Third column: Decision curves. Optimal thresholds were determined in development cohort and are overlaid to the ROC and Decision curves.

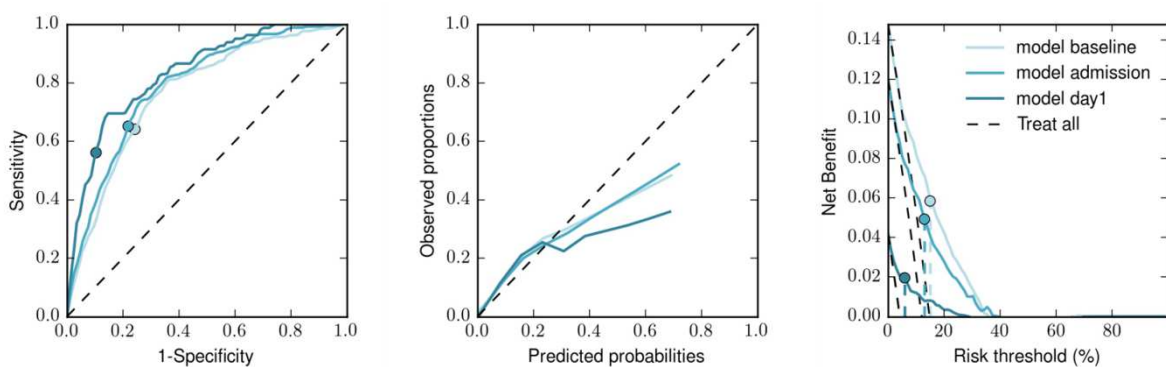


Figure 3.A.2 Performance of clinical prediction models in the validation cohort for prediction of AKI stage-23.

First column: ROC curves. Second column: Calibration curves. Third column: Decision curves. Optimal classification thresholds were determined in development cohort and are overlaid to the ROC and Decision curves.

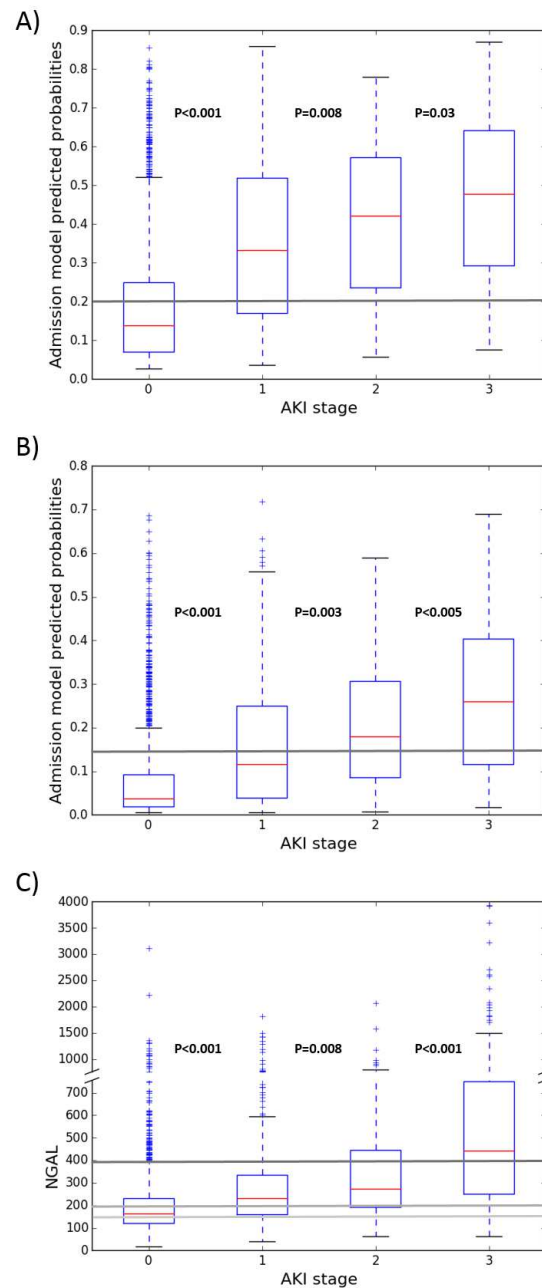


Figure 3.A.3 Predictions of admission models and admission NGAL concentrations in the validation cohort stratified by KDIGO AKI stages.

a) Admission Model for AKI-123 (n=2274). b) Admission Model for AKI-23 (n=2274). c) NGAL (n=2081). Horizontal lines show classification thresholds of 150 ng/mL, 200 ng/mL and 400 ng/mL for NGAL and determined in the development cohort for the clinical prediction models as 20.71% for AKI-123 and 13.55% for AKI-23. Reported p-values are for Mann-Whitney-U tests.

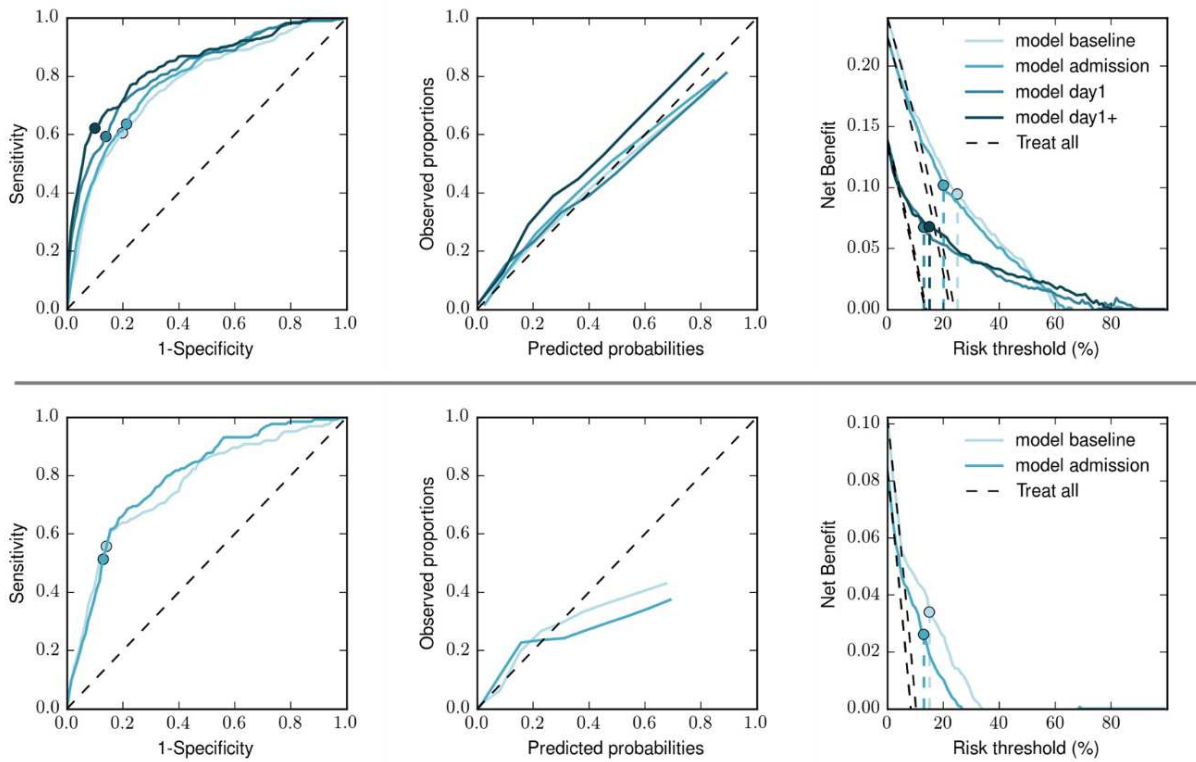


Figure 3.A.4 Performance of clinical prediction models in the validation cohort in cardiac patients.

a) Top row: AKI-123. b) Bottom row: AKI-23. First column: ROC curves. Second column: Calibration curves. Third column: Decision curves. Optimal thresholds were determined in development cohort and are overlaid to the ROC and Decision curves.

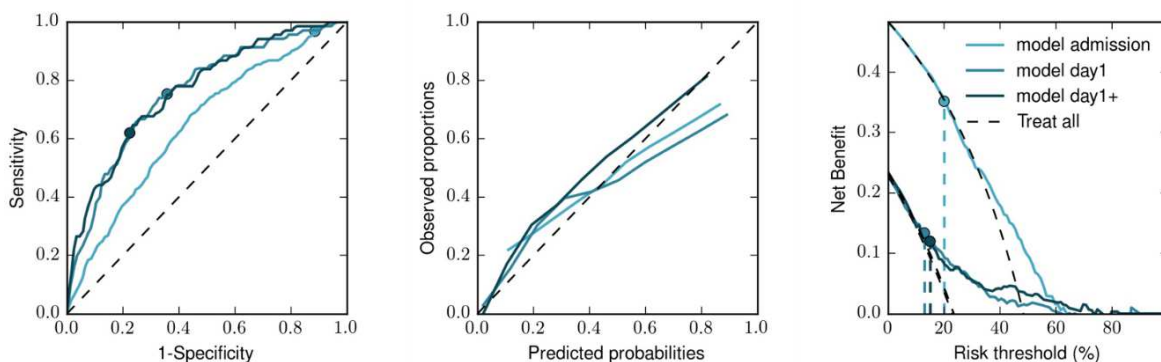


Figure 3.A.5 Performance of clinical prediction models in the validation cohort in septic patients for prediction of AKI-123.

The baseline model cannot be evaluated in this cohort as sepsis is only known upon ICU admission. First column: ROC curves. Second column: Calibration curves. Third column: Decision curves. Optimal thresholds were determined in development cohort and are overlaid to the ROC and Decision curves.

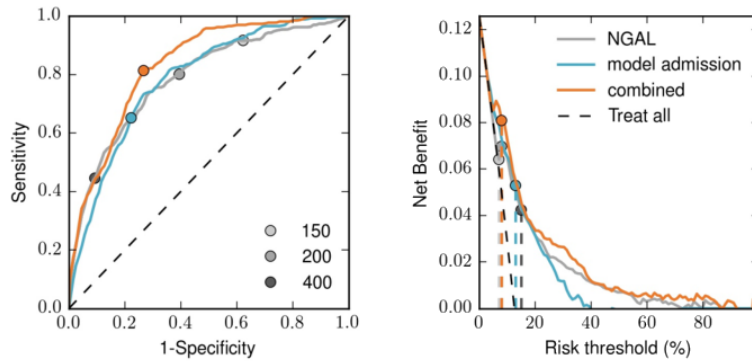


Figure 3.A.6 Performance of admission clinical prediction model, admission NGAL, and the combination in the validation cohort for prediction of AKI-23.

First column: ROC curves. Second column: Decision curves. Optimal classification thresholds were determined in the development cohort for the admission clinical prediction models; pre-defined as 150 ng/mL, 200 ng/mL and 400 ng/mL for NGAL; and for the combined models as the value that maximized sensitivity and specificity in the validation cohort. Thresholds are overlaid to the ROC and Decision curves.

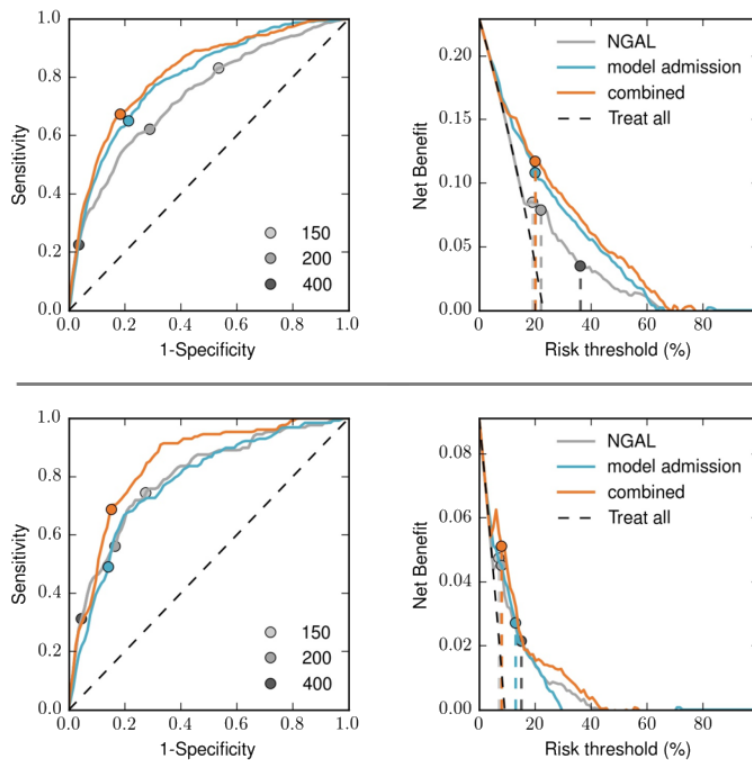


Figure 3.A.7 Performance of the admission clinical prediction model, admission NGAL, and the combination in the validation cohort for cardiac patients.

Top row: prediction of AKI-123. Bottom row: prediction of AKI-23. First column: ROC curves. Second column: Decision curves. Optimal thresholds were determined in development cohort for the clinical prediction models and pre-defined as 150, 200 and 400 ng/mL for NGAL and are overlaid to the ROC and Decision curves.

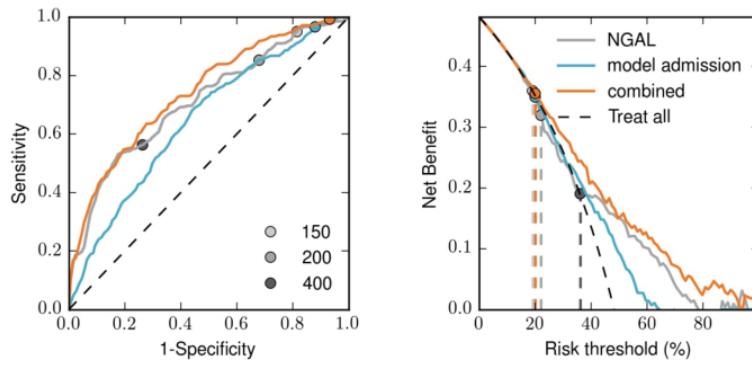


Figure 3.A.8 Performance of the admission clinical prediction model, admission NGAL, and the combination in the validation cohort for septic patients for prediction of AKI-123.

First column: ROC curves. Second column: Decision curves. Optimal thresholds were determined in development cohort for clinical prediction models and pre-defined as 150, 200 and 400 ng/mL for NGAL and are overlaid to the ROC and Decision curves.

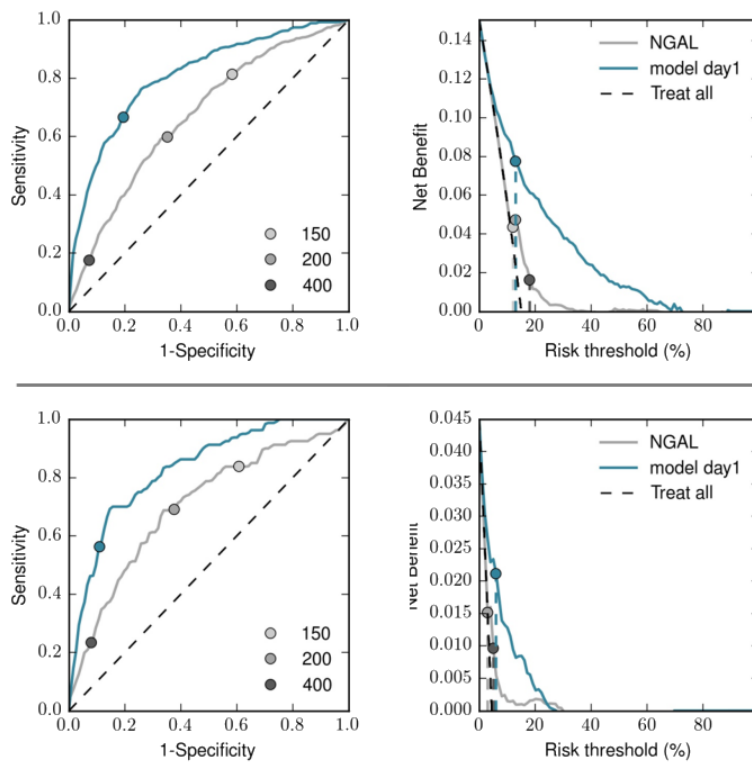


Figure 3.A.9 Performance of the admission NGAL and Day1 models in the validation cohort.

Top row: prediction of AKI-123. Bottom row: prediction of AKI-23. First column: ROC curves. Second column: Decision curves. Optimal thresholds were determined in development cohort for clinical prediction models and pre-defined as 150, 200 and 400 ng/mL for NGAL and are overlaid to the ROC and Decision curves.

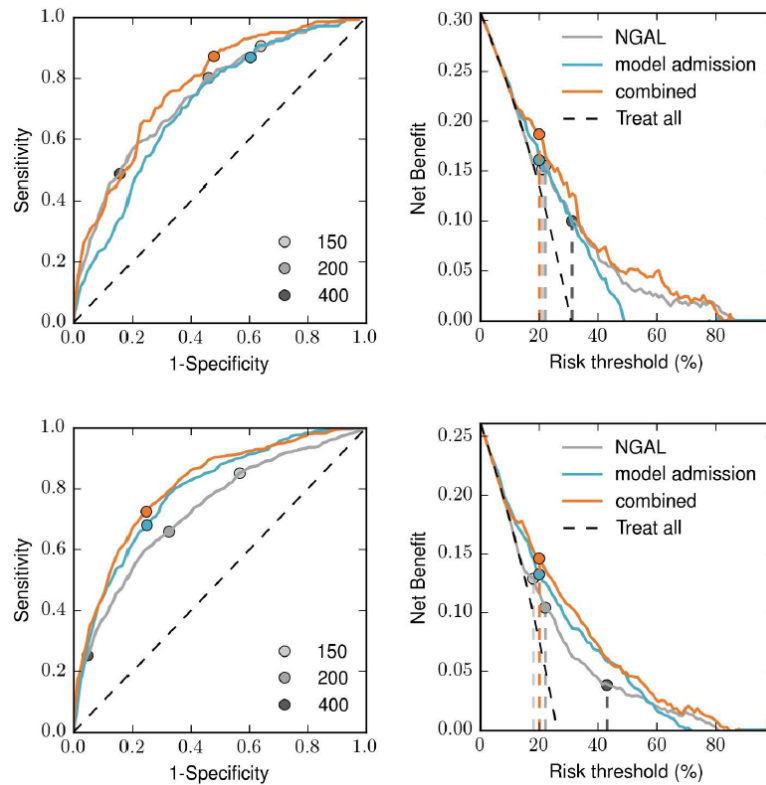


Figure 3.A.10 Performance of admission NGAL, the admission clinical prediction model and the combination in the validation cohort for the subset of patients with true and with calculated creatinine baseline for prediction of AKI-123.

Top row: prediction of AKI-123. Bottom row: prediction of AKI-23. First column: ROC curves. Second column: Decision curves. Optimal thresholds were determined in development cohort for clinical prediction models and pre-defined as 150, 200 and 400 ng/mL for NGAL and are overlaid to the ROC and Decision curves.

BIBLIOGRAPHY

- [1] R. Bellomo, J. A. Kellum, and C. Ronco. Acute kidney injury. *Lancet* 380,9843 (Aug. 2012), pp. 756–766.
- [2] J. Gunst et al. Impact of Early Parenteral Nutrition on Metabolism and Kidney Injury. *J. Am. Soc. Nephrol.* 24,6 (June 2013), pp. 995–1005.
- [3] M. Joannidis et al. Acute kidney injury in critically ill patients classified by AKIN versus RIFLE using the SAPS 3 database. *Intensive Care Med.* 35,10 (2009), pp. 1692–1702.
- [4] S. Nisula et al. 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, (2013), pp. 420–8.
- [5] S. M. Sutherland et al. Utilizing electronic health records to predict acute kidney injury risk and outcomes: workgroup statements from the 15th ADQI Consensus Conference. *Can. J. Kidney Heal. Dis.* 3,1 (Dec. 2016), p. 11.
- [6] L. S. Chawla, P. W. Eggers, R. A. Star, and P. L. Kimmel. Acute Kidney Injury and Chronic Kidney Disease as Interconnected Syndromes. *N. Engl. J. Med.* 371,1 (July 2014), pp. 58–66.
- [7] M. Ostermann and M. Joannidis. Biomarkers for AKI improve clinical practice: no. *Intensive Care Med.* 41,4 (Apr. 2015), pp. 618–622.
- [8] M. Park, S. G. Coca, S. U. Nigwekar, A. X. Garg, S. Garwood, and C. R. Parikh. Prevention and Treatment of Acute Kidney Injury in Patients Undergoing Cardiac Surgery: A Systematic Review. *Am. J. Nephrol.* 31,5 (2010), pp. 408–418.
- [9] K. E. A. Burns et al. Perioperative N-acetylcysteine to Prevent Renal Dysfunction in High-Risk Patients Undergoing CABG Surgery. *JAMA* 294,3 (July 2005), pp. 342–350.
- [10] F. T. Billings et al. High-Dose Perioperative Atorvastatin and Acute Kidney Injury Following Cardiac Surgery. *JAMA* 315,9 (Mar. 2016), pp. 877–888.
- [11] A. X. Garg et al. Perioperative Aspirin and Clonidine and Risk of Acute Kidney Injury: A Randomized Clinical Trial. *JAMA* 312,21 (Dec. 2014), pp. 2254–2264.

- [12] P. Young et al. Effect of a Buffered Crystalloid Solution vs Saline on Acute Kidney Injury Among Patients in the Intensive Care Unit. *JAMA* 314,16 (Oct. 2015), pp. 1701–1710.
- [13] F. P. Wilson et al. Automated, electronic alerts for acute kidney injury: a single-blind, parallel-group, randomised controlled trial. *Lancet* 385,9981 (May 2015), pp. 1966–1974.
- [14] Kidney Disease: Improving Global Outcomes (KDIGO) Acute Kidney Injury Work Group. KDIGO Clinical Practice Guideline for Acute Kidney Injury. *Kidney Int. Suppl.* 2,2 (June 2012), pp. 1–138.
- [15] M. Joannidis et al. Prevention of acute kidney injury and protection of renal function in the intensive care unit. *Intensive Care Med.* 36,3 (Mar. 2010), pp. 392–411.
- [16] D. Shemin and L. D. Dworkin. Neutrophil gelatinase-associated lipocalin (NGAL) as a Biomarker for Early Acute Kidney Injury. *Crit. Care Clin.* 27,2 (Apr. 2011), pp. 379–389.
- [17] K. Kashani et al. Discovery and validation of cell cycle arrest biomarkers in human acute kidney injury. *Crit. Care* 17,1 (Jan. 2013), R25.
- [18] H. R. H. de Geus, J. Bakker, E. M. E. H. Lesaffre, and J. L. M. L. le Noble. Neutrophil gelatinase-associated lipocalin at ICU admission predicts for acute kidney injury in adult patients. *Am. J. Respir. Crit. Care Med.* 183,7 (Apr. 2011), pp. 907–914.
- [19] J. Mishra et al. Neutrophil gelatinase-associated lipocalin (NGAL) as a biomarker for acute renal injury after cardiac surgery. *Lancet* 365,9466 (Apr. 2005), pp. 1231–1238.
- [20] Z. Zhang. Biomarkers, diagnosis and management of sepsis-induced acute kidney injury : a narrative review. *Hear. Lung Vessel.* 7,1 (2015), pp. 64–73.
- [21] S. M. Bagshaw, C. Langenberg, M. Haase, L. Wan, C. N. May, and R. Bellomo. Urinary biomarkers in septic acute kidney injury. *Intensive Care Med.* 33,7 (June 2007), pp. 1285–1296.
- [22] J. Mårtensson, M. Bell, A. Oldner, S. Xu, P. Venge, and C.-R. Martling. Neutrophil gelatinase-associated lipocalin in adult septic patients with and without acute kidney injury. *Intensive Care Med.* 36,8 (Aug. 2010), pp. 1333–1340.
- [23] M. Bell, A. Larsson, P. Venge, R. Bellomo, and J. Mårtensson. Assessment of Cell-Cycle Arrest Biomarkers to Predict Early and Delayed Acute Kidney Injury. *Dis. Markers* 2015, (2015), pp. 1–9.

- [24] J. R. Prowle. Measurement of AKI biomarkers in the ICU: still striving for appropriate clinical indications. *Intensive Care Med.* 41,3 (Jan. 2015), pp. 541–543.
- [25] R. Bellazi, B. Zupan, R. Bellazzi, and B. Zupan. Predictive data mining in clinical medicine: Current issues and guidelines. *Int. J. Med. Inform.* 77,2 (Feb. 2008), pp. 81–97.
- [26] K. A. Eagle et al. A Validated Prediction Model for All Forms of Acute Coronary Syndrome. *JAMA* 291,22 (2004), pp. 2727–2733.
- [27] M. P. Casaer et al. Early versus Late Parenteral Nutrition in Critically Ill Adults. *N. Engl. J. Med.* 365,6 (Aug. 2011), pp. 506–517.
- [28] L. Breiman. Random forests. *Mach. Learn.* 45,1 (2001), pp. 5–32.
- [29] B. Efron and R. Tibshirani. Improvements on Cross-Validation: The 632+ Bootstrap Method. *J. Am. Stat. Assoc.* 92,438 (June 1997), pp. 548–560.
- [30] E. W. Steyerberg, S. E. Bleeker, H. A. Moll, D. E. Grobbee, and K. G. Moons. Internal and external validation of predictive models: A simulation study of bias and precision in small samples. *J. Clin. Epidemiol.* 56,5 (May 2003), pp. 441–447.
- [31] K. Van Hoorde, Y. Vergouwe, D. Timmerman, S. Van Huffel, E. W. Steyerberg, and B. Van Calster. Assessing calibration of multinomial risk prediction models. *Stat. Med.* 33,15 (July 2014), pp. 2585–2596.
- [32] J. Swets. Measuring the accuracy of diagnostic systems. *Science.* 240,4857 (June 1988), pp. 1285–1293.
- [33] M. J. Leening, E. W. Steyerberg, B. Van Calster, R. B. D’Agostino, and M. J. Pencina. Net reclassification improvement and integrated discrimination improvement require calibrated models: relevance from a marker and model perspective. *Stat. Med.* 33,19 (Aug. 2014), pp. 3415–3418.
- [34] M. Fitzgerald, B. R. Savige, and R. J. Lewis. Decision Curve Analysis. *JAMA* 313,4 (Jan. 2015), pp. 409–410.
- [35] E. W. Steyerberg et al. Assessing the Performance of Prediction Models. *Epidemiology* 21,1 (Jan. 2010), pp. 128–138.
- [36] F. Pedregosa et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 12, (2011), pp. 2825–2830.
- [37] J. Concato. The Risk of Determining Risk with Multivariable Models. *Ann. Intern. Med.* 118,3 (Feb. 1993), pp. 201–210.

- [38] E. A. J. Hoste et al. Epidemiology of acute kidney injury in critically ill patients: the multinational AKI-EPI study. *Intensive Care Med.* 41,8 (Aug. 2015), pp. 1411–1423.
- [39] D. N. Cruz et al. Plasma neutrophil gelatinase-associated lipocalin is an early biomarker for acute kidney injury in an adult ICU population. *Intensive Care Med.* 36,3 (Mar. 2010), pp. 444–451.
- [40] M. Haase, R. Bellomo, P. Devarajan, P. Schlattmann, and A. Haase-Fielitz. Accuracy of Neutrophil Gelatinase-Associated Lipocalin (NGAL) in Diagnosis and Prognosis in Acute Kidney Injury: A Systematic Review and Meta-analysis. *Am. J. Kidney Dis.* 54,6 (Dec. 2009), pp. 1012–1024.
- [41] L. Englberger et al. Validation of Clinical Scores Predicting Severe Acute Kidney Injury After Cardiac Surgery. *Am. J. Kidney Dis.* 56,4 (Oct. 2010), pp. 623–631.
- [42] S. M. Bagshaw et al. Plasma and urine neutrophil gelatinase-associated lipocalin in septic versus non-septic acute kidney injury in critical illness. *Intensive Care Med.* 36,3 (Mar. 2010), pp. 452–461.
- [43] A. Zhang et al. Diagnosis and prognosis of neutrophil gelatinase-associated lipocalin for acute kidney injury with sepsis: a systematic review and meta-analysis. *Crit. Care* 20,1 (Dec. 2016), p. 41.
- [44] J. Vanmassenhove, R. Vanholder, E. Nagler, and W. Van Biesen. Urinary and serum biomarkers for the diagnosis of acute kidney injury: an in-depth review of the literature. *Nephrol. Dial. Transplant.* 28,2 (Feb. 2013), pp. 254–273.
- [45] G. Schley et al. Comparison of Plasma and Urine Biomarker Performance in Acute Kidney Injury. *PLoS One* 10,12 (Dec. 2015). Ed. by L. R. James, e0145042.
- [46] L. G. Forni et al. Identifying the patient at risk of acute kidney injury: A predictive scoring system for the development of acute kidney injury in acute medical patients. *Nephron - Clin. Pract.* 123,3-4 (2013), pp. 143–150.
- [47] R. L. Mehta et al. International Society of Nephrology's oby25 initiative for acute kidney injury (zero preventable deaths by 2025): a human rights case for nephrology. *Lancet* 385,9987 (June 2015), pp. 2616–2643.
- [48] M. Schetz, J. Gunst, and G. Van den Berghe. The impact of using estimated GFR versus creatinine clearance on the evaluation of recovery from acute kidney injury in the ICU. *Intensive Care Med.* 40,11 (Nov. 2014), pp. 1709–17.

- [49] J. A. Kellum. How can we define recovery after acute kidney injury? Considerations from epidemiology and clinical trial design. *Nephron - Clin. Pract.* 127,1-4 (2014), pp. 81–88.
- [50] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The WEKA Data Mining Software: An Update. *SIGKDD Explor. Newsl.* 11,1 (Nov. 2009), pp. 10–18.
- [51] N. Peek, D. G. T. Arts, R. J. Bosman, P. H. J. van der Voort, and N. F. de Keizer. External validation of prognostic models for critically ill patients required substantial sample sizes. *J. Clin. Epidemiol.* 60,5 (May 2007), pp. 491–501.
- [52] M. J. Pencina, R. B. D'Agostino, and R. S. Vasan. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Statistics in medicine* 27,2 (2008), pp. 157–172.
- [53] A. J. Vickers and E. B. Elkin. Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. *Med. Decis. Mak.* 26,6 (Nov. 2006), pp. 565–574.

MACHINE LEARNING VERSUS PHYSICIANS'
PREDICTIONS OF ACUTE KIDNEY INJURY
IN CRITICALLY ILL ADULTS

Flechet M et al. Machine learning versus physicians' predictions of acute kidney injury in critically ill adults: A prospective evaluation of the AKIpredictor (submitted for publication).

ABSTRACT

PURPOSE: Early diagnosis of acute kidney injury (AKI) is a major challenge in the intensive care unit (ICU). The AKIpredictor is a machine-learning-based prediction model using routinely collected patient information, and is available as an online calculator. In order to evaluate its clinical value, the AKIpredictor is compared to predictions by physicians.

METHODS: Prospective observational study in the surgical ICUs of a tertiary academic center. Critically ill adults without end-stage renal disease and without AKI upon admission were considered for enrollment. Using structured questionnaires, physicians were asked upon admission, on the first morning of ICU stay, and after 24 hours, to predict the development of AKI stages 2 or 3 (AKI-23) during the first week of ICU stay. Predictions were compared against those made by the machine-learning models.

RESULTS: 252 patients were included, 30 (12%) developed AKI-23. Performance of physicians and AKIpredictor were, respectively: upon ICU admission area under the receiver operating characteristic curve (AUROC) 0.80 versus 0.75 (n=120, P=0.25), with clinical benefit in ranges (0-26%) versus (0-74%); on the first morning, AUROC 0.94 versus 0.89 (n=187, P=0.27) with main clinical benefit in ranges (0-10%) versus (0-48%); after 24 hours, AUROC 0.94 versus 0.89 (n=89, P=0.09) with main clinical benefit in ranges (0-67%) versus (0-50%).

CONCLUSIONS: There was no significant differences in discrimination between predictions of AKI-23 by physicians and the AKIpredictor, although the latter showed better calibration and higher net benefit overall, because physicians tended to overestimate the risk of AKI. These findings suggest an added value of the AKIpredictor to physicians' predictions.

4.1 INTRODUCTION

Acute kidney injury (AKI) is an abrupt decline in kidney function that is highly prevalent in critically ill patients, with an incidence that appears to be increasing [1–3]. AKI has an unfavourable impact on both short- and long-term outcomes, and is associated with increased financial costs [4–6]. The 2012 guidelines by the international Kidney Disease: Improving Global Outcome (KDIGO) work group have classified AKI in 3 stages of ascending severity [7], according to a quantitative increase in serum creatinine (SC) and/or a decrease in urine output (UO). However, both are late and unspecific markers of the underlying pathological insult. It is hypothesized that the late recognition of AKI could be one of the factors to explain the current lack of evidence based therapeutic options that could prevent AKI or attenuate its course [8–10]. Early biomarkers have been proposed as accurate predictors of the disease, and have a good predictive performance in certain settings [11, 12]. However, as they are an additional lab test with a certain cost, it is necessary to identify which subgroups of patients would benefit most from biomarker testing.

Prediction models have been proposed for the early prediction of AKI. These models have the advantage that they do not require additional testing, but use the information already present in the (electronic) health records [13, 14]. In hospitalized patients, the AKI prediction score developed by Forni et al. [15] is a simple scoring system to detect hospital-acquired AKI. The score was recently externally validated [16] and showed moderate discrimination and acceptable calibration. For trauma patients admitted to critical care, Haines and colleagues developed a prediction model showing good discrimination for any stage of AKI and for its more severe stages [17]. Finally, in a general population of critically ill patients, the AKIpredictor models [18] were developed with advanced machine learning techniques, on a large patient database from a multi-center randomized controlled trial (EPANIC) [19]. The AKIpredictor predicted AKI at different time points in the clinical course of the patient (before admission, upon admission, on the first morning after admission, and after 24h), for any stage of AKI (AKI-123) or only the most severe forms (AKI-23). The models' performance was assessed in a separate validation cohort, and compared against serum neutrophil gelatinase-associated lipocalin (NGAL), a biomarker for AKI [20, 21]. The AKIpredictor not only proved a high degree of accuracy, but it also outperformed the biochemical test [18]. Some of the models have been made available as a free online calculator on the website <http://www.akipredictor.com>.

The potential usefulness of the AKIpredictor [16, 22–24], or similar prediction models [13, 25], has been recognized. However, it remains to be investigated

whether and how these models could be used in clinical practice. Using the available clinical data to estimate risk of critically ill patients to developing complication, such as AKI, is part of the daily practice of ICU physicians. It may be expected that they will perform well in this task, especially if they are experienced. Computerized AKI prediction models should be evaluated prospectively in new and previously unseen patient cohorts. In addition, comparing them against predictions by physicians could add a dimension to the evaluation. In the present study, performance of the AKIpredictor to predict the most severe stages of AKI (AKI-23) within the first week of ICU stay, will be evaluated prospectively, and compared against predictions by ICU physicians.

4.2 METHODOLOGY

This prospective observational study was performed between May and June 2018, in the surgical ICUs of the University Hospitals Leuven (UZLEUVEN), Leuven, Belgium. The Institutional Review Board approved the enrollment and the protocol for the collection of anonymized clinical data, providing waiver of consent for study participation. The study is registered at ClinicalTrials.gov (NCT03574896).

4.2.1 *Study population*

All adult critically ill patients were eligible for the study. Patients were excluded if they had pre-existing end-stage renal disease or had already developed any stage of AKI at ICU admission. Patients for whom the 3 prediction moments occurred during on-call time were excluded because of the unavailability of research staff to hand out the questionnaires. In case of multiple ICU admissions for a single patient, only the first admission was considered.

4.2.2 *Endpoint*

The primary objective of the study was the comparison of the performances of AKIpredictor and physicians in predicting AKI-23 in the first 7 days following ICU admission. Comparison was made upon ICU admission (admission cohort), on the first morning of ICU stay (day1 cohort) and after 24 hours of ICU stay (day1+ cohort). Secondary objective were a) to assess the influence of the level of seniority on the accuracy of physicians' predictions and the feasibility of making predictions within a 3-hour window for physicians; b) to compare the AKIpredictor performance using two definitions for AKI (SC versus SC and UO).

4.2.3 Acute kidney injury

AKI was staged each day of the week after ICU admission using the SC and UO criteria from the KDIGO guidelines [7]. To compare the performance of AKIpredictor with the development study [18], where AKI was classified only by SC, AKI was also staged each day based on the SC criterion only. Baseline SC values were defined as the lowest SC value that could be identified in the 3 months prior to ICU admission. In case no baseline SC value was available, a glomerular filtration rate of 75 ml/min/1.73m² [26] was assumed, and baseline SC was calculated using the Modification of Diet in Renal Disease (MDRD) formula.

4.2.4 AKIpredictor predictions

The AKIpredictor algorithm used routinely collected patient information for prediction of AKI-23. All predictions and confidence intervals from the AKIpredictor algorithm were retrospectively calculated.

4.2.5 Physicians' predictions

Questionnaires (Appendix, Figure 4.A.9) were handed to physicians at the same target time points than the AKIpredictor: upon ICU admission, on the first morning of ICU stay, and at 24 hours after admission. Prospective data collection included:

- Physicians' binary predictions: *Do you think this patient will develop AKI stage 2 or 3 over the next 7 days? (yes-no)*. Binary predictions were used to determine physicians' classification threshold and their derived sensitivity and specificity.
- Physicians' prediction as percentage: *What is your prediction that this patient will develop AKI stage 2 or 3 over the next 7 days? (scale 0-100%)*
- Physicians' level of confidence about their prediction: *How confident do you feel about this prediction? (low-medium-high)*

To accommodate for physicians' availability, questionnaires were considered valid if collected within 1 hour before up to 3 hours after the predefined target time point. When several predictions were available per patient, an average of the predictions was calculated and, in a subsequent analysis, compared to the average weighted by physician experience and level of confidence.

Two categories of physicians were interviewed: junior (junior residents) and senior (senior residents and staff members). Junior residents are in training for their basic specialty (anesthesiology, internal medicine, or pediatrics), and have at least 3 years of experience as a physician. Senior residents are physicians in training for intensive care medicine, and have already obtained their basic specialty (in anesthesiology, internal medicine, or pediatrics). All staff members are certified specialists in intensive care medicine. Their age, gender and seniority level (years of experience and category) were recorded (Appendix, Figure 4.A.10).

4.2.6 *Statistical analysis*

Data are presented as means and standard deviations (SD), medians and interquartile ranges (IQR), and numbers and proportions where appropriate. Statistical significance was set at $P < 0.05$. All analyses were performed using Python version 2.7.13 (Python Software Foundation, <http://www.python.org>), Scipy version 0.18.1 (SciPy.org) and R version 3.5.0.

Reporting of the study was performed using the STROBE guidelines [27].

4.2.6.1 *Diagnostic accuracy assessment*

Discrimination, calibration and clinical usefulness [28] were used to evaluate the performance of AKIpredictor and physicians. Discrimination refers to how well the predictions allow to distinguish patients with and without AKI. Discrimination was evaluated with the receiver operating characteristic (ROC) curve and the area under the receiver operating characteristic curve (AUROC) [29]. The DeLong test [30] from the PROC R package [31] was used for AUROC comparison. Calibration refers to the agreement between predicted probabilities and the observed frequency of AKI in the population. Calibration was assessed using calibration belts or curves where appropriate, together with the distribution of patient numbers [32]. Finally, the clinical usefulness of the model was assessed by the difference between the expected benefit and the expected harm associated with model classification of AKI. Clinical usefulness was visualized using decision curves and reported using ranges above treat-all and treat-none curves [33, 34]. An example of decision curve with its interpretation is given in Figure 4.1. To assess the added value of the AKIpredictor to the predictions by physicians, a multivariable logistic regression was used to combine the estimated AKI risk by the AKIpredictor with the one by physicians.

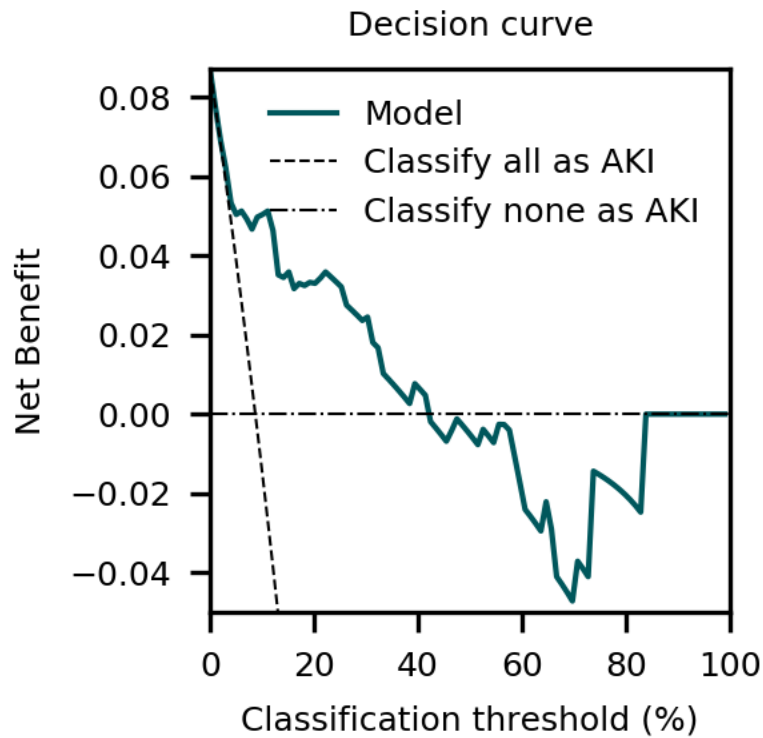


Figure 4.1 Illustration of decision curve analysis.

The example illustrates the decision curve of a model to predict whether patients will have AKI, from a population with an AKI prevalence of 9%. In decision curve analysis, the classification threshold corresponds to the cutoff above which a patient is classified as "will develop AKI". Knowing whether the patient will or will not have AKI will trigger different therapeutic interventions. Low classification thresholds are used when the associated therapy is not harmful, hence patients will not suffer from being classified as false positives. High classification thresholds are used when the associated therapy is toxic or has side effects and therefore, it is important to not classify patients as at risk for AKI when they are not (thereby limiting the number of false positive classifications). Currently, preventive measures for AKI are optimization of hemodynamics and prevention of nephrotoxicity, amenable to all patients and therefore corresponding to low classification thresholds.

The net benefit is a weighted measure between true and false positives depending on the classification threshold [33]. The maximum net benefit is obtained by detecting all patients who will later develop AKI, therefore this net benefit is the prevalence of AKI in the population (9%). The line corresponding to the trivial assumption that all patients will have AKI can be drawn (*classify all as aki*, traditionally called *treat-all*). Similarly, the minimum net benefit is obtained by considering that no patient will develop AKI and is 0 (*classify none as aki*, traditionally called *treat-none*). To be clinically useful, a model should have a higher net benefit than the two trivial classifications. Here, being slightly above the *classify all as aki* curve, the model shows usefulness in the range 0-43%. Above 43%, the models shows negative net benefit, which reflects harm and should be avoided in clinical practice. Here, the model is clinically relevant as it shows benefit for low risk thresholds corresponding to its associated preventive measures.

4.3 RESULTS

4.3.1 Study population

A total of 348 adults were considered for study inclusion, of which 58 were excluded because the 3 prediction moments occurred during on-call time, 23 due to AKI at ICU admission, 11 due to readmission, and 4 for end-stage renal disease (Figure 4.2). Two hundred and fifty-two patients remained for analysis.

Patients' characteristics are reported in Table 4.1, for all patients, and Table 4.A.4, for patients with predictions by physicians. Within the first week of ICU stay, 30 (12%) patients developed AKI-23 using both SC and UO criteria and 23 (9%) with SC criteria only. Baseline serum creatinine was available in 202 (80.2%) patients, and was calculated in the remaining 50 (19.8%).

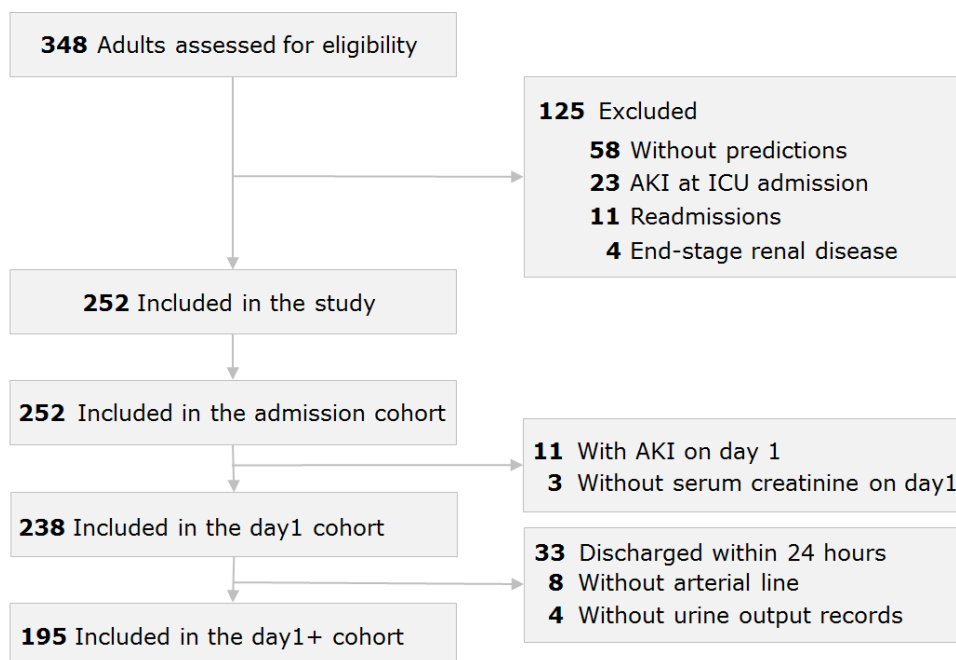


Figure 4.2 Flow chart

4.3.2 AKI predictor predictions

Predictions were calculated in 252 patients at ICU admission, in 238 patients on the first morning of ICU stay (Figure 4.2, 11 excluded for AKI on day 1, 3 excluded for missing AKI on day 1), and in 195 patients after 24 hours (Figure 4.2, 33 excluded for discharge within 24 hours, 8 without an arterial line, 4 without urine output records).

Table 4.1 Patient characteristics and clinical outcomes

	Admission cohort	Day1 cohort	Day1+ cohort
N	252	238	195
AKI-23 by SC and UO, n (%)	30 (12)	17 (7)	17 (7)
AKI-23 by SC, n (%)	23 (9)	13 (5)	13 (5)
Demographics			
Age, year	65.5 (52.8-74.0)	65 (52-74)	66 (54-74)
Male gender, n (%)	155 (61.5)	146 (61.3)	121 (62.1)
Height, cm	171 (165-178)	171 (165-178)	171 (165-178)
Weight, kg	75.2 (65.0-86.3)	75.0 (65.0-86.0)	75.0 (63.5-85.7)
Diabetic, n (%)	6 (2.4)	5 (2.1)	3 (1.5)
Baseline SC, mg/dL	0.88 (0.73-1.05)	0.88 (0.73-1.04)	0.88 (0.73-1.06)
Clinical parameters			
Elective admission, n (%)	154 (61.1)	149 (62.6)	124 (63.6)
Surgical category, n (%)			
Cardiac	103 (40.9)	98 (41.1)	85 (43.6)
Transplant	7 (2.8)	7 (2.9)	7 (3.6)
Others	92 (36.5)	86 (35.8)	71 (36.4)
Medical category, n (%)	50 (19.8)	47 (19.8)	32 (16.4)
Hemodynamic support at ICU admission, n (%)			
Pharmacological	165 (65.5)	158 (66.4)	140 (71.8)
Mechanical	4 (1.6)	3 (1.3)	3 (1.5)
Blood glucose at ICU admission, mg/dL	135 (113-155)	135 (114-155)	137 (116-160)
Sepsis upon ICU admission, n (%)	20 (7.9)	18 (7.6)	16 (8.2)
Maximum lactate on day 1, mg/dL	1.6 (1.1-2.4)	1.5 (1.1-2.3)	1.6 (1.1-2.4)
Bilirubin on day 1, mg/dL	0.54 (0.37-0.84)	0.54 (0.37-0.83)	0.53 (0.38-0.83)
APACHE II score on day 1	14 (10-17)	13.5 (10-17)	14 (11-17)
SOFA score on day 1	9 (5-11)	9 (5-11)	9 (6-12)
SC on day1, md/dL	0.87 (0.68-1.08)	0.87 (0.68-1.05)	0.88 (0.70-1.08)
Monitoring parameters^a			
Urine slope, ml/hour	-0.00014 (-0.00052 to 0.00034)	-0.00013 (-0.00051 to 0.00035)	-0.00014 (-0.00046 to 0.00032)
Total amount of urine, mL/hour	1025 (770-1437)	1047 (789-1451)	1048 (495-1473)
Blood pressure below 60 mmHg, min	10 (3-50)	10 (3-48)	11 (4-51)
Blood pressure above average, min	644 (569-696)	647 (569-698)	655 (593-716)
Dose of vasopressors, mg	2.7 (0-8.9)	2.7 (0-8.6)	4.3 (0-9.7)

Data are reported as median (IQR) unless otherwise indicated.

^a measured during first 24 hours of ICU stay.

When classifying AKI only by SC, the admission AKI predictor predicted AKI-23 with AUROC 0.78 and clinical benefit in ranges (0-74%) (Appendix, Figure 4.A.1). On day 1, AUROC was 0.94 with clinical benefit in ranges (0-48%). After 24 hours, AUROC was 0.93 with clinical benefit in ranges (3-43%).

When classifying AKI by SC and UO criteria, the admission AKI predictor predicted AKI-23 with AUROC 0.76 and clinical benefit in ranges (0-74%) (Figure 4.3). On day 1, AUROC was 0.87 with clinical benefit in ranges (0-48%). After 24 hours, AUROC was 0.85 with clinical benefit in ranges (0-43%).

4.3.3 Physicians' predictions

Forty-three physicians of which 24 (55.8%) junior residents, 8 (18.6%) senior residents, and 11 (25.6%) staff members filled questionnaires (Appendix, Table 4.A.1). They provided a total of 709 predictions (Appendix, Table 4.A.2): 183 predictions about 120 patients at ICU admission, 394 predictions about 187 patients on the morning of the first day, and 128 predictions about 89 patients after 24 hours of ICU stay. Although the protocol allowed gathering physicians' predictions 1 hour before the time point, the majority of the predictions were obtained later (Appendix, Table 4.A.2, 183 (100%) on admission, 383 (97.2%) on first morning, 77 (60.2%) at 24 hours). On average, predictions were obtained 68 minutes after ICU admission, 140 minutes after day 1 and 20 minutes after 24 hours of ICU stay. Table 4.A.3 (Appendix) presents the physicians' predictions by confidence level and seniority level. Overall, confidence obtained at later time points was higher, with 53 (29%) highly confident at admission, 147 (37.3%) highly confident at day 1, 55 (43%) highly confident after 24 hours.

Upon admission, physicians predicted AKI with AUROC 0.80 and clinical benefit in ranges (0-26%) (Figure 4.4). On day 1, AUROC was 0.94 with clinical benefit in ranges (0-10%+90-96%). After 24 hours, AUROC was 0.95 with clinical benefit in ranges (0-36%+40-48%+50-67%+80-100%). Figure 4.A.2 (Appendix) shows performance when using the binary predictions from the physicians, which allowed for the identification of the classification threshold they adopted: sensitivity and specificity were, respectively, 55% and 82% on admission, 85% and 86% on day 1, 75% and 90% after 24 hours. Averaging predictions by weighing for physicians' seniority and level of confidence did not improve predictive performance (Appendix, Figure 4.A.3). However, as compared to junior physicians, senior physicians showed higher discrimination and calibration at all time points (Appendix Figure 4.A.4, AUROC 0.81 vs 0.85 at admission; AUROC 0.87 vs 0.92 on day 1; AUROC 0.90 vs 0.96 at 24 hours for junior and senior, respectively). Finally, when physicians expressed low or medium confidence in their predictions, their performance was worse (Appendix Figure

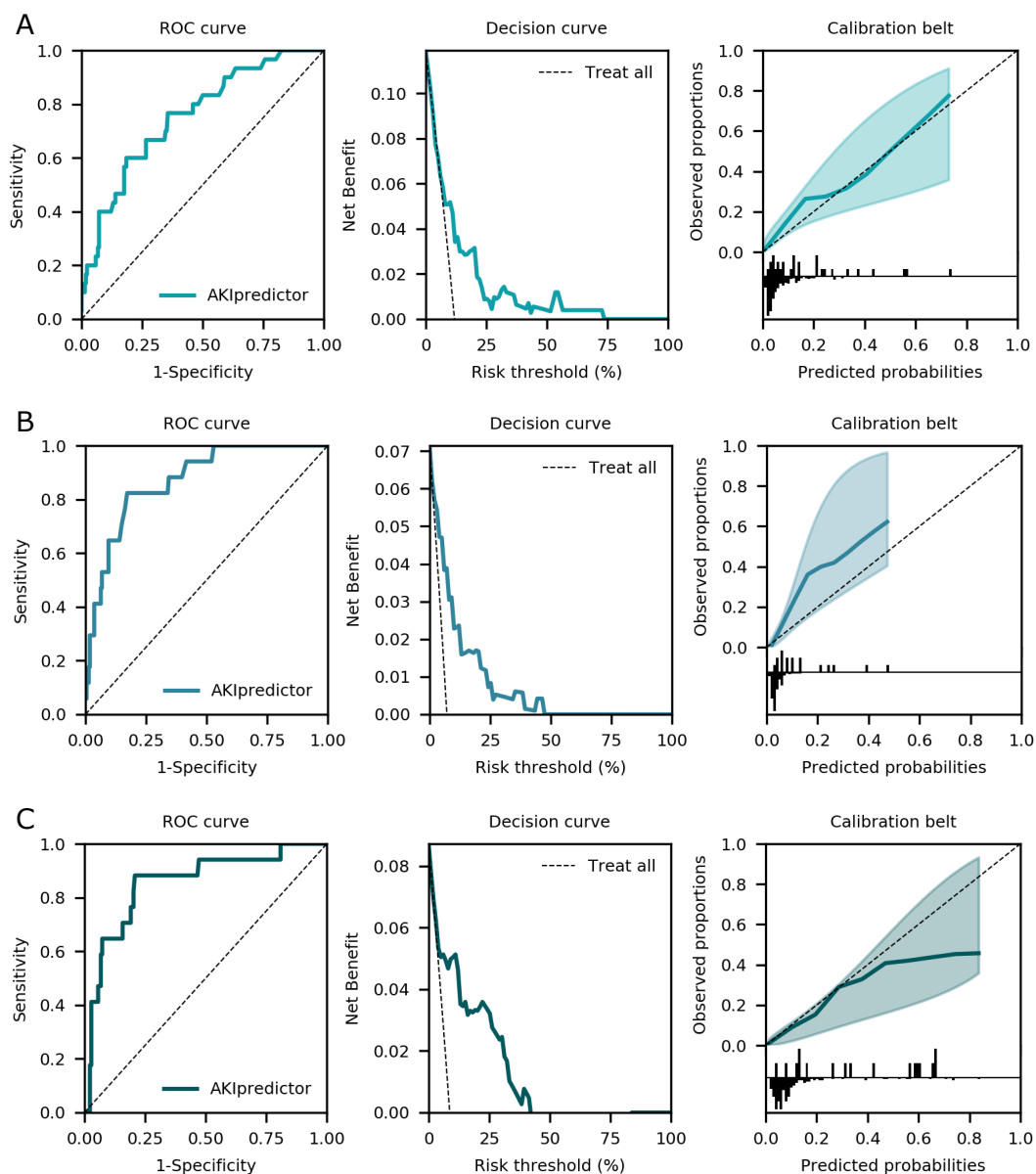


Figure 4.3 Performance of AKIpredictor for prediction of AKI-23 by SC and UO (a) At ICU admission (n=252), AUROC was 0.76, clinical benefit in ranges (0-74%); (b) On the first morning of ICU stay (n=238), AUROC was 0.87, clinical benefit in ranges (0-48%); (c) After 24 hours (n=195), AUROC was 0.85, clinical benefit in ranges (0-43%).

4.A.5, AUROC 0.74 versus 0.85 at admission, AUROC 0.93 versus 0.92 on day 1, and AUROC 0.89 versus 0.98 at 24 hours, for medium versus high confidence respectively).

In the subset of patients with physician predictions (Figure 4.4), AKIpredictor predicted AKI with AUROC 0.75 ($P=0.25$ as compared to clinicians) with clinical benefit in ranges (0-74%), higher than physicians in ranges (14-74%). On day 1, AUROC was 0.89 ($P=0.27$) with higher clinical benefit compared with physicians in ranges (0-48%). Finally, after 24 hours, AUROC was 0.89 ($P=0.09$) with higher clinical benefit compared with physicians in ranges (0-20%+23-50%).

4.3.4 Combining AKIpredictor with physicians' predictions

In the subset of patients where physicians' predictions were combined with the AKIpredictor, no improvement in discriminability was observed as compared to physicians ($P=0.96, 0.39, 0.41$ respectively for admission, day 1 and after 24 hours), but a better calibration resulted in wider and higher ranges of clinical benefit at all time points (Appendix, Figure 4.A.6). The same was observed for junior physicians only (Appendix, Figure 4.A.7), and for low-medium confidence predictions (Appendix, Figure 4.A.8).

4.4 DISCUSSION

In this study, we compared the performance of the AKI risk estimated by physicians versus the one provided by AKIpredictor, a machine learning based clinical prediction model [18]. The comparison was made at three different time points: upon ICU admission, at the first morning, and after 24 hours of ICU stay. There was no statistically significant difference in discrimination between physicians and AKIpredictor at any timepoint. However, on average, physicians required more time to provide predictions.

Decision curve analysis helps to identify the expected clinical benefit or harm when deciding for treatment at different risk levels. Compared to physicians, the AKIpredictor showed improved net benefit for AKI classification thresholds above 26% upon admission and for almost all ranges of AKI classification thresholds on day 1 and after 24 hours. This comparison provides meaningful insight on how the tool could be used in clinical practice.

As shown by the calibration curve, physicians tend to over-estimate the risk of AKI. In the decision curve, this behavior results in a net benefit similar to considering that all patients will develop AKI (*treat-all* curve) [35]. Currently there is no treatment for AKI and preventive measures are mainly supportive,

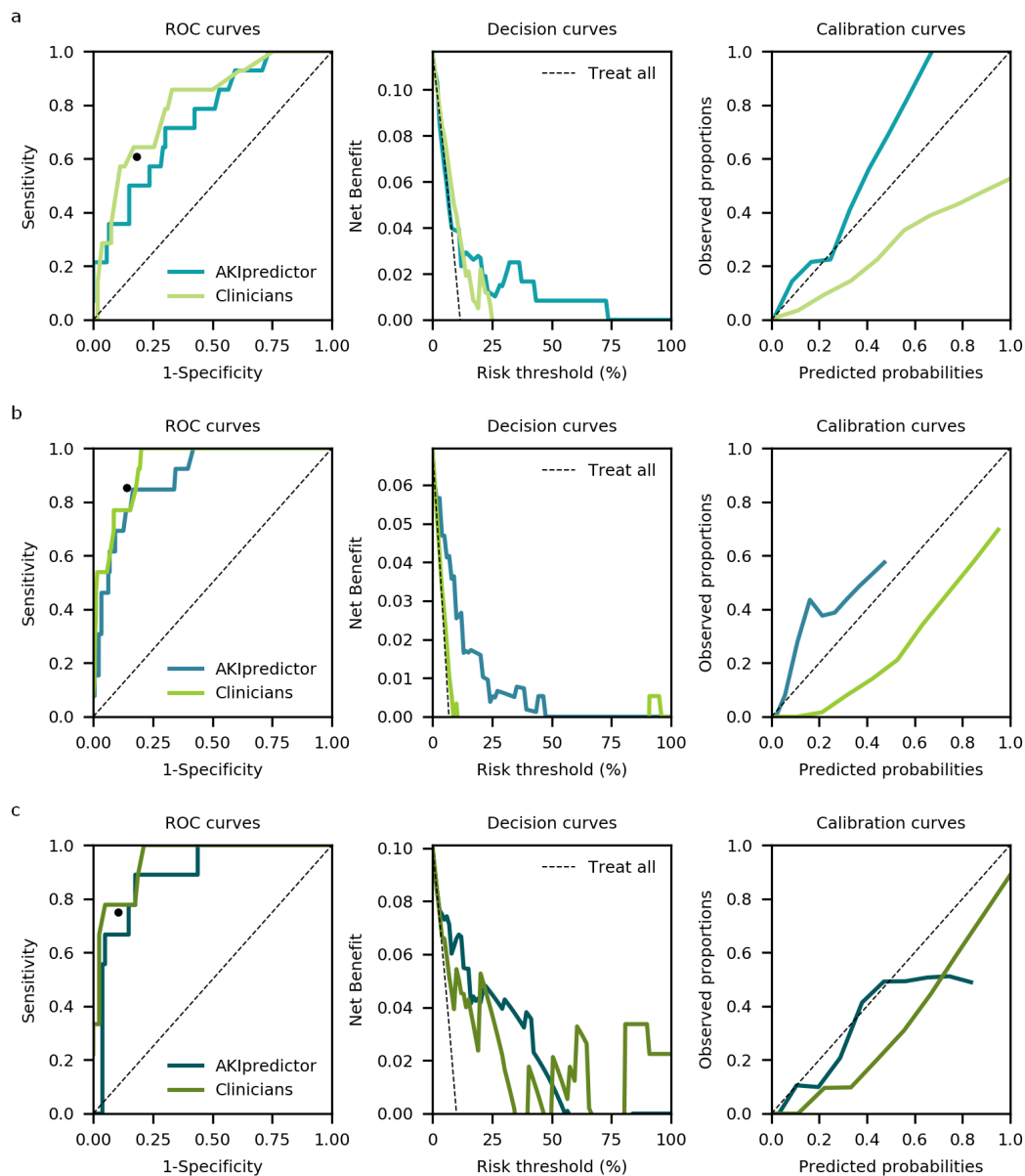


Figure 4.4 Comparison of performance of AKIpredictor and physicians for prediction of AKI-23 by SC and UO.

The black dot represents the classification threshold from the physicians. **(a)** At ICU admission ($n=120$), AUROCs were 0.80 and 0.75 ($P=0.25$), clinical benefit in ranges (0-26%), (0-74%) for physicians, and AKIpredictor respectively. Physicians' classification threshold achieved 55% sensitivity and 82% specificity. **(b)** On the first morning of ICU stay ($n=187$), AUROCs were 0.94 and 0.89 ($P=0.27$), clinical benefit in ranges (0-10%+90-96%), (0-48%) for physicians, and AKIpredictor respectively. Physicians' classification threshold achieved 85% sensitivity and 86% specificity. **(c)** After 24 hours ($n=89$), AUROCs were 0.95 and 0.89 ($P=0.09$), with clinical benefit in ranges (0-36%+40-48%+50-67%+80-100%), (0-58%) for physicians and AKIpredictor respectively. Physicians' classification threshold achieved 75% sensitivity and 90% specificity.

so there would be no harm from misclassifying a patient as high-risk. However, in the hypothetical context of enrolling high risk patients for a clinical trial or if a new potentially toxic or expensive therapy for AKI became available, this approach would induce selection bias, unnecessary exposure of patients to potential side-effects or higher costs. This situation corresponds to a high classification threshold for AKI, for which only the AKIpredictor showed clinical benefit.

Additional benefits of the AKIpredictor were highlighted by this study. First, it allows stratifying patients consistently, with similar performance to a well-trained clinical staff. Second, it provides predictions at fixed time points without delays: on average, physicians required instead more time to provide predictions. Third, although senior physicians are the best at predicting AKI, they have to supervise a higher number of patients and might benefit from an electronic warning system that draws their attention to patients at risk. Fourth, when doctors perceive that it is more difficult to predict and express a low or medium confidence in their predictions, these predictions are actually less performant. In such cases, they might find it useful to consult the AKIpredictor.

This study is the first prospective validation study of the AKIpredictor. Compared to the results of the original development study [18], the models showed similar performance upon ICU admission and an even higher performance on the first morning and after 24 hours of ICU stay. This observed improvement might be explained by the difference in patient population (higher prevalence of cardiac surgery and sicker patients with more comorbidities in the initial development population). Indeed, AKIpredictor has shown worse performance in septic patients, who are less prevalent in this study. This might be explained by not predicting during on-call time, which resulted in fewer unplanned admissions, such as sepsis, and more planned admissions, such as surgery. The difference in population might also explain the lower prevalence of severe AKI as compared to the initial development population (9% vs 12%) and other studies [36]. Furthermore, the design of the current study might have raised physicians' awareness towards the kidney, which in turn could have prevented AKI development and hence affected AKI incidence.

It is striking that, although the AKIpredictor was developed to predict AKI based on SC and not UO criteria, in this study, the model performed well, even when AKI was defined by both SC and UO.

4.4.1 *Strengths and limitations*

This study had several strengths. First, it is prospective in design, and hence detailed in data collection. Second, in order to reduce bias from lack of collaboration by physicians, interviewers made efforts to obtain questionnaires for all included patients: predictions for all but 12 patients were gathered for at least one time point. In addition, when feasible, predictions were obtained from both junior and senior physicians, allowing for secondary analysis based on physicians' experience. Finally, to the best of our knowledge, this study is the first of its kind to assess physicians' estimation of AKI risk, which provides benchmarking opportunities for comparison against other AKI prognosticators such as biomarkers.

This study had the following limitations: first, as a single center study, findings have to be used with caution, as they might not generalize to other centers. In particular, fewer predictions were available on admission and after 24 hours. Second, a bias in favor of physicians cannot be excluded as 1) the AKIpredictor is not optimized to predict AKI defined by UO as in the development study the definition of AKI was only base on the SC criterion [18]. Additionally, due to the low prevalence of AKI-23 after 24 hours of ICU stay, no model was developed [18] . Therefore, at 24 hours, the comparison is made using the AKIpredictor AKI-123 model. 2) Physicians received 3 more hours to provide their predictions. Therefore, they had access to later clinical information than the AKIpredictor. 3) Physicians did not provide predictions at all time points. However, we limited this bias by asking predictions for at least one time point in all but 12 patients. Third, predictions from junior and senior physicians were not available for all patients. Therefore, when averaging based on physicians' experience and level of confidence, performance was not improved although the separate analysis clearly showed a difference in performance on both levels. Finally, we cannot exclude a Hawthorne effect: the design of the current study might have raised physicians' awareness towards the kidney, which could have favored a more meticulous clinical management to prevent AKI.

4.5 CONCLUSION

Physicians can predict AKI with good discrimination, but tend to overestimate the risk of AKI, pointing to poor calibration in the low-risk patients. The AKIpredictor performed on par with physicians in terms on discrimination, but did better in terms of calibration and net benefit. This highlights the potential uses of AKIpredictor in clinical practice: selection of high risk

patients, or reducing false positives in studies evaluating new and potentially harmful therapies. Additionally, our findings suggest a potential for overall improvement of care with the concurrent use of physicians' expertise and the AKIpredictor. External validation and further studies of the AKIpredictor are warranted.

GRANT SUPPORT AND CONFLICT OF INTEREST

M.Flechet receives funding from the Fonds Wetenschappelijk Onderzoek (FWO) as a PhD. fellow (11Y1118N). G.Meyfroidt receives funding from FWO as senior clinical investigator (1843118N). G. Van den Berghe, via the University of Leuven, receives structural research financing via the Methusalem program funded by the Flemish Government (METH/14/06) and holds an European Research Council (ERC) Advanced Grant (AdvG-2017-785809) from the Horizon2020 Programme of the European Union.

The authors declare that they have no conflict of interest.

ACKNOWLEDGMENTS AND PERSONAL CONTRIBUTION

All analysis presented in this chapter were performed by Marine Flechet.

- Study concept and design: **Flechet**, Falini, Meyfroidt, Güiza, Schetz, Van den Berghe
- Acquisition, analysis, or interpretation of data: **Flechet**, Falini, Bonetti, Güiza, Meyfroidt, Schetz
- Drafting of the chapter: **Flechet**, Falini
- Critical revision of the chapter: **All authors**
- Creation of figures: **Flechet**
- Study supervision: Meyfroidt, **Flechet**

The authors are grateful to the ICU junior and senior residents and staff members for their involvement in the study. The authors are also grateful to Bram Vercammen, Dominiek Cotte, Lisa van Dyck and Wouter Vankrunkelsven for their logistic help.

4.A APPENDIX

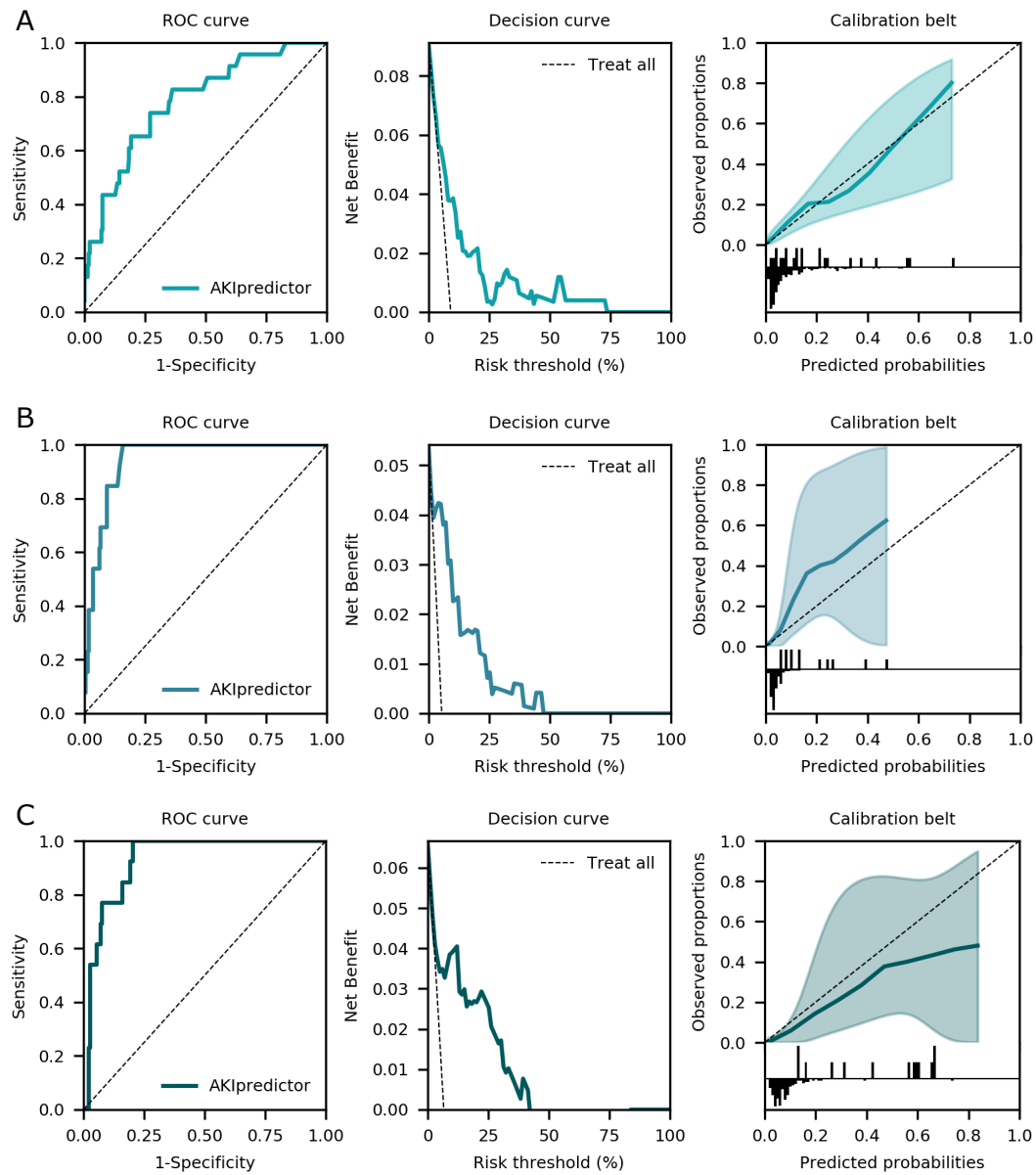


Figure 4.A.1 Performance of AKIpredictor for prediction of AKI-23 by SC
(a) At ICU admission (n=252), AUROC 0.78, clinical benefit in ranges (0-74%). Optimal cutoff was 8%; **(b)** On the first morning of ICU stay (n=240), AUROC 0.94, clinical benefit in ranges (0-48%). Optimal cutoff was 6%; **(c)** After 24 hours (n=195), AUROC 0.93, clinical benefit in ranges (3-43%). Optimal cutoff was 12%.

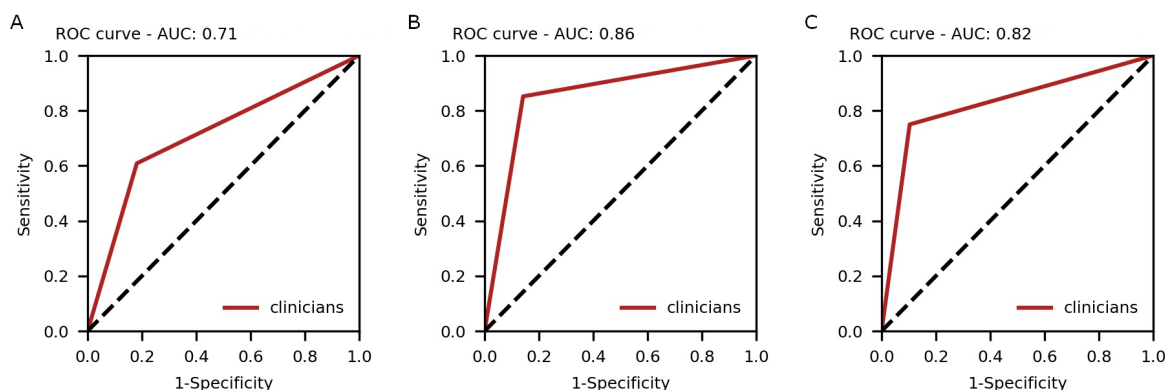


Figure 4.A.2 Performance of binary predictions by physicians.

(a) Upon ICU admission, AUROC 0.71 (n=120). Physicians' classification threshold achieved 55% sensitivity and 82% specificity. **(b)** On day 1, AUROC 0.86 (n=187). Physicians' classification threshold achieved 85% sensitivity and 86% specificity. **(c)** After 24 hours, AUROC 0.82 (n=89). Physicians' classification threshold achieved 75% sensitivity and 90% specificity.

Table 4.A.1 Physicians' generalities

	Physicians
N	43
Age, years, median (IQR)	30 (29-34)
Male gender, n (%)	25 (58.1)
Seniority level, n (%)	
Junior resident	24 (55.8)
Senior resident	8 (18.6)
Staff member	11 (25.6)

Table 4.A.2 Description of physicians' predictions

	Admission	Day1	Day1+
N patient	120	187	89
N total	183	394	128
Self-filled predictions, n (%)	18 (9.8)	4 (1.0)	16 (12.5)
Predictions collected by interviewer, n (%)	165 (90.2)	390 (99.0)	112 (87.5)
Predictions collected before time point, n (%)	0 (0)	11 (2.8)	51 (39.8)
Predictions collected after time point, n (%)	183 (100)	383 (97.2)	77 (60.2)
Time between prediction and time point, min, median (IQR)	68 (36-105)	140 (120-150)	20 (-26 to 83)

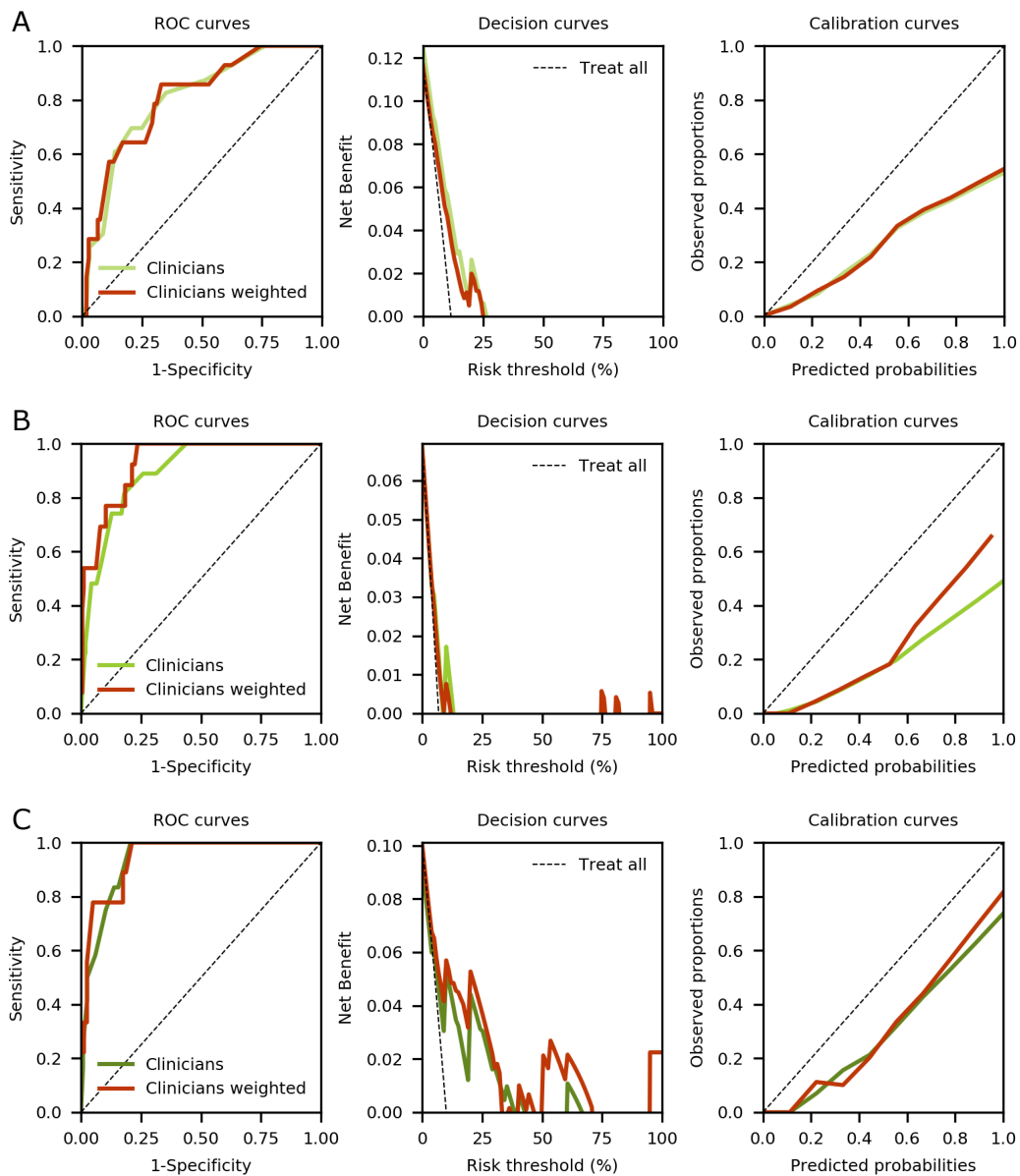


Figure 4.A.3 Performance of clinicians when predictions were averaged using weight for clinicians' expertise and level of confidence.

(a) Upon admission, AUROCs were 0.80, 0.80, for non-weighted and weighted average respectively; with similar clinical benefit ranges from (0-25%). **(b)** On day 1, AUROCs were 0.89, 0.93, for non-weighted and weighted average respectively; with similar clinical benefit ranges from (0-14%). **(c)** After 24 hours, AUROCs were 0.94, 0.94 for non-weighted and weighted average respectively, with higher clinical benefit ranges within (0-34% + 50-72%) for weighted average.

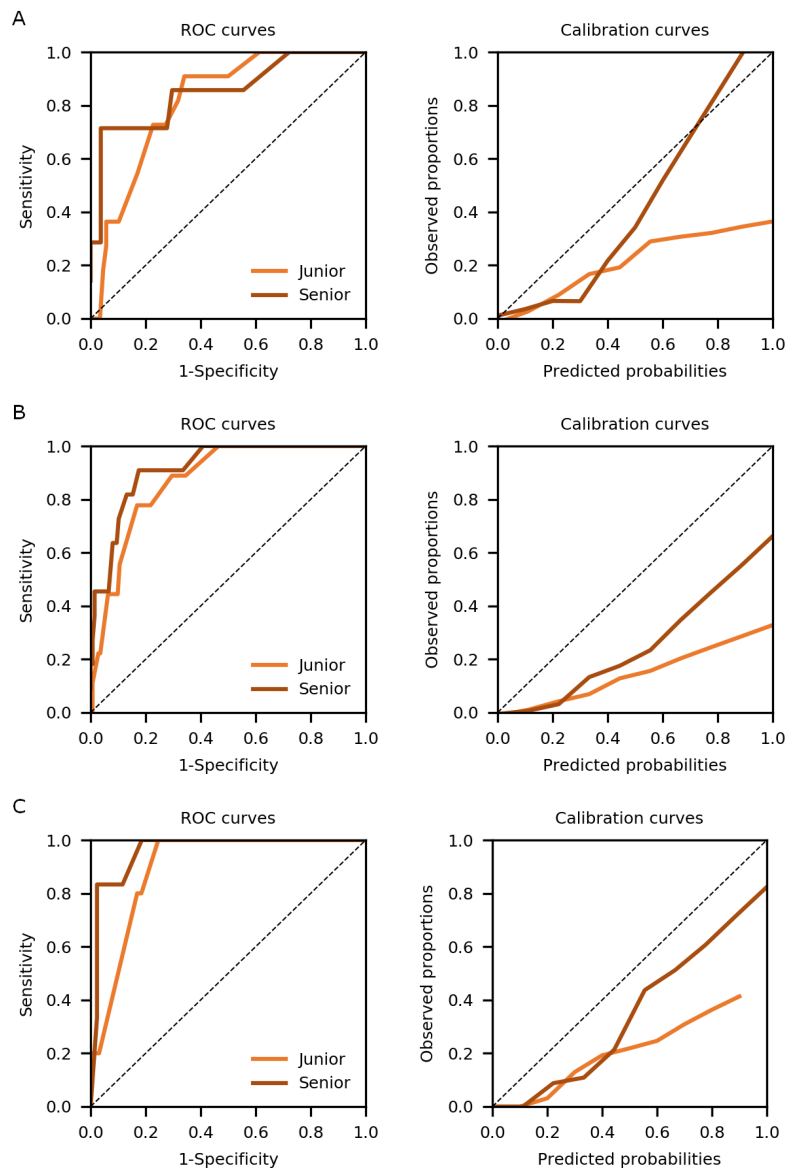


Figure 4.A.4 Performance of physicians split by seniority level.

(a) Upon admission, AUROCs were 0.81 and 0.85, for junior and senior respectively. Classification threshold had 55% sensitivity and 75% specificity for junior, 71% sensitivity and 94% specificity for senior. **(b)** On day 1, AUROCs were 0.87 and 0.92, for junior and senior respectively. Classification threshold had 78% sensitivity and 82% specificity for junior, 91% sensitivity and 84% specificity for senior. **(c)** After 24 hours, AUROCs were 0.90 and 0.96, for junior and senior respectively. Classification threshold had 60% sensitivity and 88% specificity for junior, 83% sensitivity and 91% specificity for senior. Decision curves are not represented, as they are not comparable (different patient population).

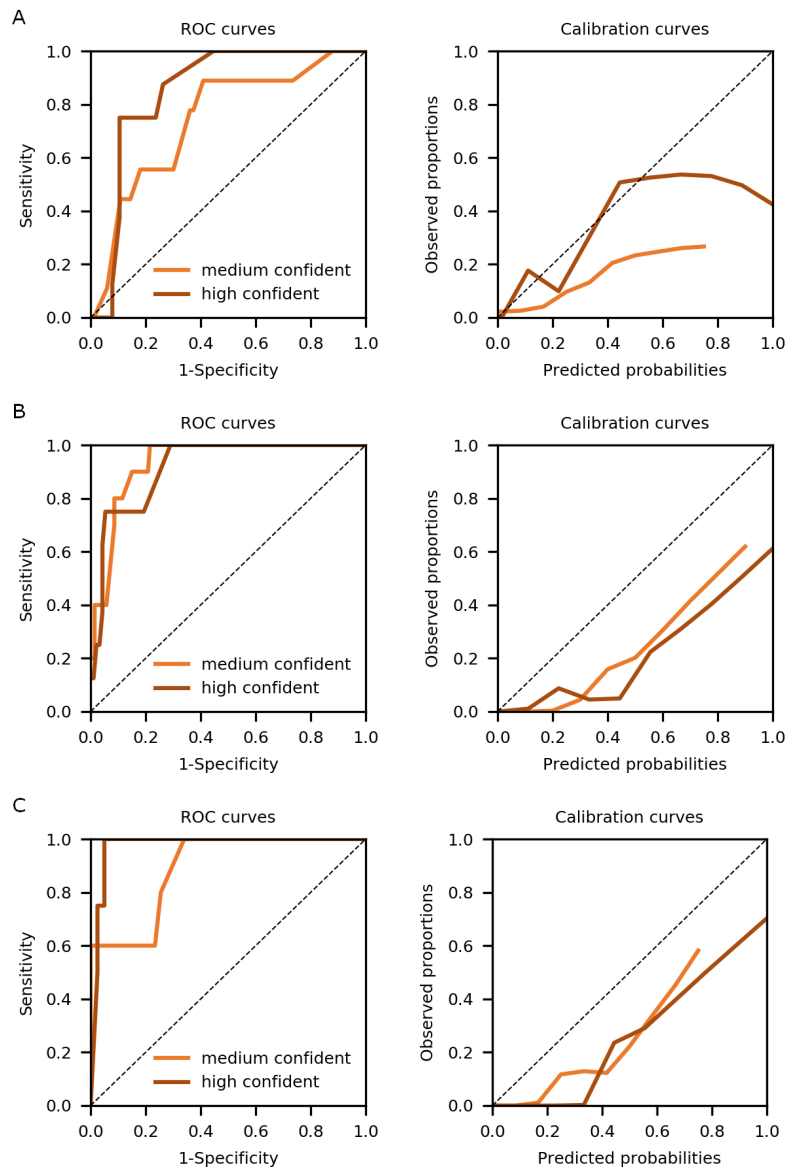


Figure 4.A.5 Performance of physicians split by confidence level.

Low confident predictions had too few occurrences and were combined to medium confident predictions. **(a)** Upon admission, AUROCs were 0.74 and 0.85, for medium and high confidence respectively. Classification threshold had 78% sensitivity and 64% specificity for medium confidence, 75% sensitivity and 89% specificity for high confidence. **(b)** On day 1, AUROCs were 0.93 and 0.92, for medium and high confidence respectively. Classification threshold had 90% sensitivity and 85% specificity for medium confidence, 75% sensitivity and 95% specificity for high confidence. **(c)** After 24 hours, AUROCs were 0.89 and 0.98, for medium and high confidence respectively. Classification threshold had 80% sensitivity and 74% specificity for medium confidence, 60% sensitivity and 95% specificity for high confidence. Decision curves are not represented, as they are not comparable (different patient population).

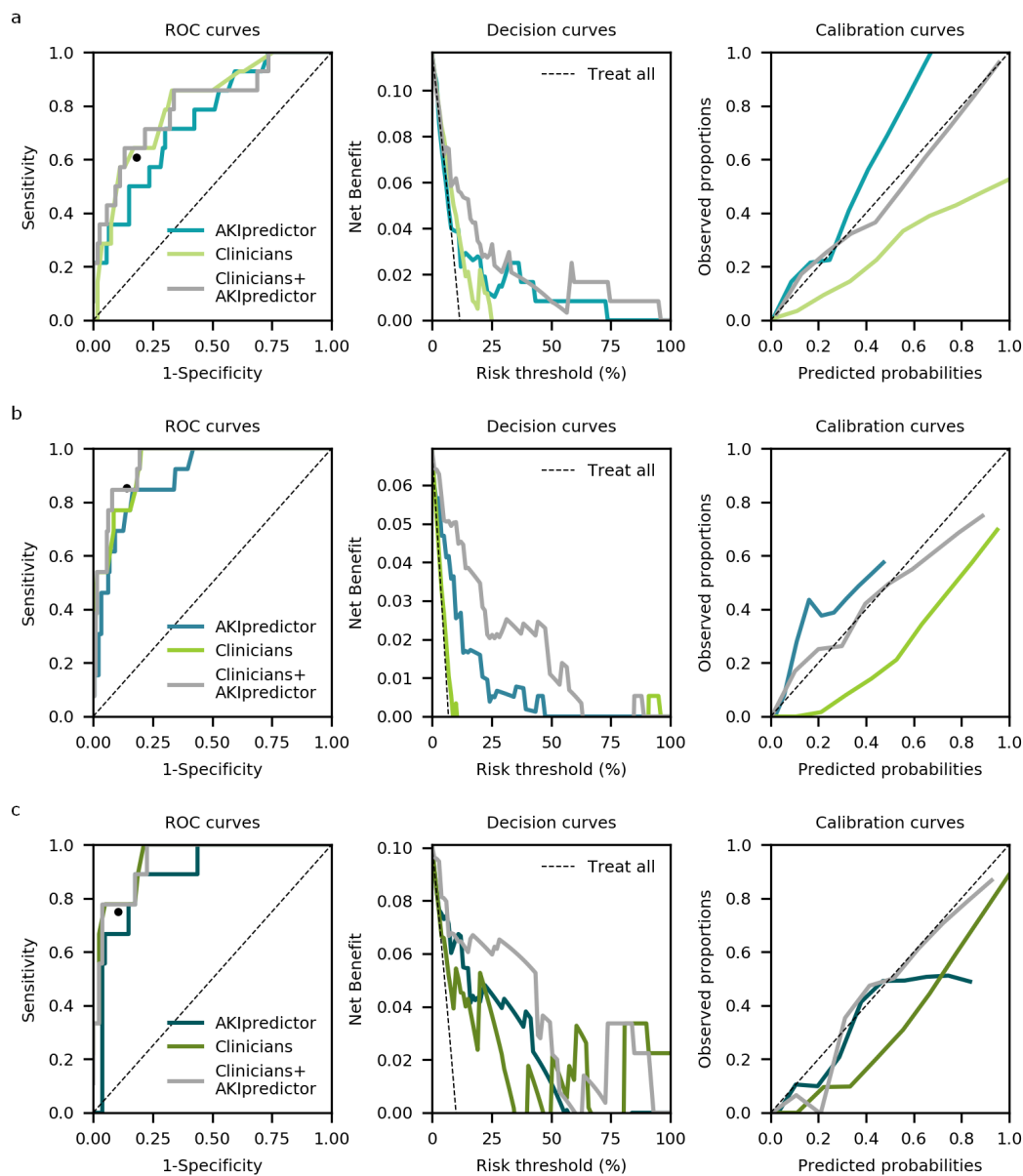


Figure 4.A.6 Comparison of performance of AKI predictor, physicians and their combination.

The black dot represents the classification threshold from the physicians. **(a)** At ICU admission ($n=120$), AUROCs were 0.80, 0.75 ($P=0.25$ as compared to physicians), 0.72 ($P=0.96$ as compared to physicians), clinical benefit in ranges (0-26%), (0-74%), (0-96%) for physicians, AKI predictor, and their combination respectively. **(b)** On the first morning of ICU stay ($n=187$), AUROCs were 0.94, 0.89 ($P=0.27$ as compared to physicians), 0.95 ($P=0.39$ as compared to physicians), clinical benefit in ranges (0-10%+90-96%), (0-48%), (0-64%+84-89%) for physicians, AKI predictor, and their combination respectively. **(c)** After 24 hours ($n=89$), AUROCs were 0.95, 0.89 ($P=0.09$ as compared to physicians), 0.94 ($P=0.41$ as compared to physicians), with clinical benefit in ranges (0-36%+40-48%+50-67%+80-100%), (0-58%), (0-61%+63-93%) for physicians, AKI predictor, and their combination respectively.

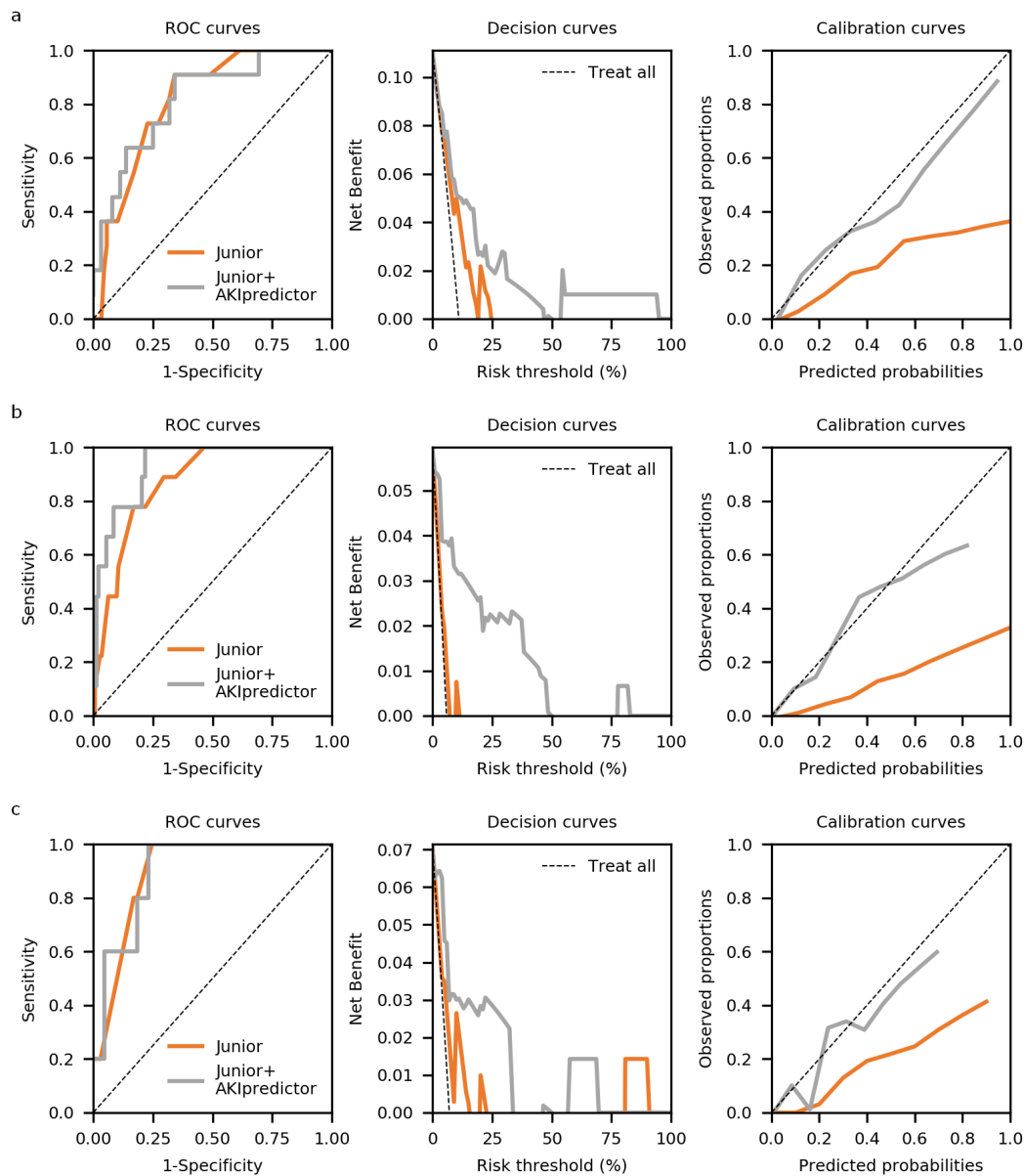


Figure 4.A.7 Comparison of performance of junior physicians and the combination of junior physicians with AKIpredictor.

A logistic regression is used to combine the predictions of junior physicians and of the AKIpredictor. Consequently, an optimal calibration is obtained and the associated net benefit represents the maximum net benefit that could be achieved by the junior physicians using the AKIpredictor. **(a)** At ICU admission ($n=99$), AUROCs were 81.8, 81.8, clinical benefit in ranges (0-26%), (4-48%+54-95%) for junior physicians and their combination with AKIpredictor respectively. **(b)** On the first morning of ICU stay ($n=151$), AUROCs were 91.4, 93.0, clinical benefit in ranges (10-13%), (0-51%+77-83%) for junior physicians, and their combination with AKIpredictor respectively. **(c)** After 24 hours ($n=70$), AUROCs were 90.5, 89.8, with clinical benefit in ranges (0-17%+20-24%+80-91%), (0-35%+46-51%+57-70%), for junior physicians, and their combination with AKIpredictor respectively.

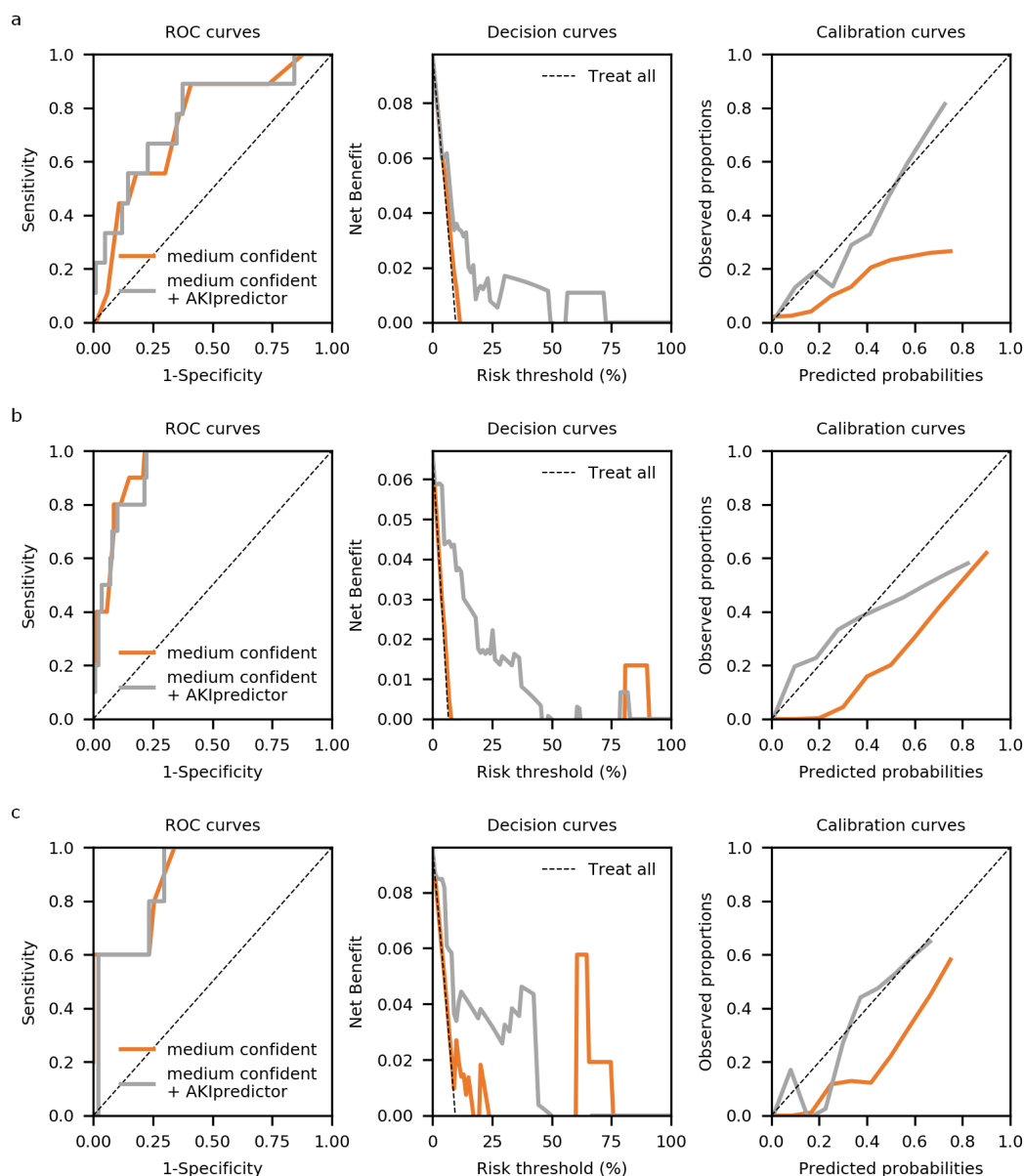


Figure 4.A.8 Comparison of performance of physicians with low-medium confidence in their predictions and the combination of their predictions with AKIpredictor.

A logistic regression is used to combine the low-medium confidence predictions with the ones from the AKIpredictor. Consequently, an optimal calibration is obtained and the associated net benefit represents the maximum net benefit that could be achieved using the AKIpredictor. **(a)** At ICU admission (n=99), AUROCs were 74.9, 76.4, clinical benefit in ranges (0-13%), (4-51%+56-73%) for low-medium confidence predictions and their combination with AKIpredictor respectively. **(b)** On the first morning of ICU stay (n=151), AUROCs were 90.9, 92.2, clinical benefit in ranges (0-9%+80-91%), (0-51%+60-63%+78-83%) for low-medium confidence predictions, and their combination with AKIpredictor respectively. **(c)** After 24 hours (n=70), AUROCs were 88.3, 88.1, with clinical benefit in ranges (0-19%+20-25%+60-76%), (0-51%), for junior physicians, and their combination with AKIpredictor respectively.

AKIpred Validation Study

The AKIpred Validation Study aims to compare the predictions of AKI made by a patient’s attending physician with those made by a computer algorithm. Please help us gather your predictions about AKI.

But remember, we are interested only in predictions made in one (or more!) of these moments:

- upon admission (+ 3 hours maximum)
- on the 1st morning after admission (between 7 at 10 am)
- and after 24 hours (up to a maximum of 27).

According to the 2012 KDIGO guidelines, AKI is defined (and classified) in the following way:

	Stage 1	Stage 2	Stage 3
Serum creatinine	1.5-1.9 × baseline OR ≥ 0.3 increase	2.0 – 2.9 × baseline	≥ 3.0 × baseline OR ≥ 4.0 increase OR RRT*
Urine output	< 0.5 ml/kg/h for ≥ 6 hours	< 0.5 ml/kg/h for ≥ 12 h	< 0.3 ml/kg/h for ≥ 24 h OR anuria for ≥ 12 h

*RRT = Renal Replacement Therapy

QUESTIONNAIRE

PATIENT EAD	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
-------------	---

✂

PT STUDY N.	(Don't worry about this, it will be assigned later!) A P V S	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
-------------	--	---

What is your prediction that this patient will develop AKI stage 2 or 3 over the next 7 days?
(Express your estimation on a scale of 0–100 %)

Do you think this patient will develop AKI stage 2 or 3 over the next 7 days?

YES	NO
-----	----

How confident do you feel about this prediction?

STRONG CONFIDENCE	MEDIUM CONFIDENCE	LOW CONFIDENCE
-------------------	-------------------	----------------

At what point in time are you making this prediction?

UPON ADMISSION	1ST MORNING IN ICU	AFTER 24 HOURS OF ICU
----------------	--------------------	-----------------------

Please specify date and time of this prediction (when you fill this questionnaire).

DATE	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	TIME	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
------	---	------	---

In order to avoid bias and to conduct sub-analyses on age, gender, degree of seniority and years of ICU experience, we would also like to collect some information about you. Please fill in your last name.

Figure 4.A.9 Prediction questionnaire.

Table 4.A.4 Patient characteristics and clinical outcomes for patients with predictions by physicians and AKIpredictor

	Admission cohort	Day1 cohort	Day1+ cohort
N	120	187	89
AKI-23 by SC and UO, n (%)	14 (12)	13 (7)	9 (10)
AKI-23 by SC, n (%)	11 (9)	10 (5)	8 (9)
Demographics			
Age, year	66 (51-74)	64 (52-74)	67 (53-74)
Male gender, n (%)	72 (60.0)	123 (65.8)	55 (61.8)
Height, cm	172 (165-178)	172 (165-179)	171 (163-178)
Weight, kg	75 (67-88)	76 (65-86)	74 (65-86)
Diabetic, n (%)	2 (1.7)	4 (2.1)	0 (0)
Baseline SC, mg/dL	0.87 (0.73-1.05)	0.88 (0.74-1.05)	0.92 (0.71-1.08)
Clinical parameters			
Elective admission, n (%)	75 (62.5)	122 (65.2)	60 (67.4)
Surgical category, n (%)			
Cardiac	53 (44.2)	81 (43.3)	43 (48.3)
Transplant	6 (5.0)	4 (2.1)	4 (4.5)
Others	45 (38.3)	61 (32.3)	28 (31.5)
Medical category, n (%)	15 (12.5)	41 (21.9)	14 (15.7)
Hemodynamic support at ICU admission, n (%)			
Pharmacological	81 (67.5)	131 (70.1)	66 (74.2)
Mechanical	1 (0.8)	3 (1.6)	2 (2.2)
Blood glucose at ICU admission, mg/dL	134 (115-158)	135 (116-153)	138 (113-159)
Sepsis upon ICU admission, n (%)	6 (5.0)	11 (5.8)	2 (2.2)
Maximum lactate on day 1, mg/dL	1.4 (1.0-2.2)	1.5 (1.1-2.3)	1.6 (1.1-2.4)
Bilirubin on day 1, mg/dL	0.55 (0.35-0.86)	0.54 (0.38-0.80)	0.49 (0.36-0.79)
APACHE II score on day 1	14 (10-17)	13 (10-17)	15 (11-17)
SOFA score on day 1	9 (4-11)	9 (5-11)	10 (7-12)
SC on day1, md/dL	0.88 (0.71-1.10)	0.87 (0.71-1.03)	0.89 (0.72-1.14)
Monitoring parameters^a			
Urine slope, ml/hour	-0.00014 (-0.0006 to 0.00046)	-0.00015 (-0.0005 to 0.00034)	-0.00021 (-0.00062 to 0.00031)
Total amount of urine, mL/hour	1130 (903-1537)	1035 (770-1458)	1077 (788-1515)
Blood pressure below 60 mmHg, min	9 (2-38)	10 (3-51)	12 (4-86)
Blood pressure above average, min	647 (569-707)	647 (561-695)	658 (585-708)
Dose of vasopressors, mg	2.7 (0-8.9)	3.5 (0-8.9)	4.3 (0-9.8)

Data is reported as median (IQR) unless otherwise indicated.

^a measured during first 24 hours of ICU stay.

BIBLIOGRAPHY

- [1] N. H. Lameire et al. Acute kidney injury: An increasing global concern. *Lancet* 382,9887 (2013), pp. 170–179.
- [2] J. G. Wilson, B. W. Butcher, and K. D. Liu. Evolving practices in critical care and their influence on acute kidney injury. *Curr. Opin. Crit. Care* 19,6 (2013), pp. 523–530.
- [3] E. A. J. Hoste et al. Epidemiology of acute kidney injury in critically ill patients: the multinational AKI-EPI study. *Intensive Care Med.* 41,8 (Aug. 2015), pp. 1411–1423.
- [4] M. Joannidis et al. Acute kidney injury in critically ill patients classified by AKIN versus RIFLE using the SAPS 3 database. *Intensive Care Med.* 35,10 (2009), pp. 1692–1702.
- [5] R. Bellomo, J. A. Kellum, and C. Ronco. Acute kidney injury. *Lancet* 380,9843 (Aug. 2012), pp. 756–766.
- [6] S. Nisula et al. 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, (2013), pp. 420–8.
- [7] Kidney Disease: Improving Global Outcomes (KDIGO) Acute Kidney Injury Work Group. KDIGO Clinical Practice Guideline for Acute Kidney Injury. *Kidney Int. Suppl.* 2,2 (June 2012), pp. 1–138.
- [8] N. Obermüller, H. Geiger, C. Weipert, and A. Urbschat. Current developments in early diagnosis of acute kidney injury. *Int. Urol. Nephrol.* 46,1 (Jan. 2014), pp. 1–7.
- [9] M. Ostermann and M. Joannidis. Acute kidney injury 2016: diagnosis and diagnostic workup. *Crit. Care* 20,1 (Dec. 2016), p. 299.
- [10] J. A. Kellum and J. R. Prowle. Paradigms of acute kidney injury in the intensive care setting. *Nat. Rev. Nephrol.* 14,4 (Jan. 2018), pp. 217–230.
- [11] K. Kashani, W. Cheungpasitporn, and C. Ronco. Biomarkers of acute kidney injury: the pathway from discovery to clinical adoption. *Clin. Chem. Lab. Med.* 55,8 (July 2017), pp. 1074–1089.
- [12] B. M. Beker, M. G. Corleto, C. Fieiras, and C. G. Musso. Novel acute kidney injury biomarkers: their characteristics, utility and concerns. *Int. Urol. Nephrol.* 50,4 (Apr. 2018), pp. 705–713.

- [13] S. M. Sutherland et al. Utilizing electronic health records to predict acute kidney injury risk and outcomes: workgroup statements from the 15th ADQI Consensus Conference. *Can. J. Kidney Heal. Dis.* 3,1 (Dec. 2016), p. 11.
- [14] K. B. Kashani. Automated acute kidney injury alerts. *Kidney Int.* (May 2018), pp. 1–7.
- [15] L. G. Forni et al. Identifying the patient at risk of acute kidney injury: A predictive scoring system for the development of acute kidney injury in acute medical patients. *Nephron - Clin. Pract.* 123,3-4 (2013), pp. 143–150.
- [16] L. E. Hodgson, B. D. Dimitrov, P. J. Roderick, R. Venn, and L. G. Forni. Predicting AKI in emergency admissions: an external validation study of the acute kidney injury prediction score (APS). *BMJ Open* 7,3 (Mar. 2017), e013511.
- [17] R. W. Haines et al. Acute Kidney Injury in Trauma Patients Admitted to Critical Care: Development and Validation of a Diagnostic Prediction Model. *Sci. Rep.* 8,1 (2018), p. 3665.
- [18] M. Flechet et al. 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.* 43,6 (2017), pp. 764–773.
- [19] M. P. Casaer et al. Early versus Late Parenteral Nutrition in Critically Ill Adults. *N. Engl. J. Med.* 365,6 (Aug. 2011), pp. 506–517.
- [20] H. R. H. de Geus, M. G. Betjes, and J. Bakker. Biomarkers for the prediction of acute kidney injury: a narrative review on current status and future challenges. *Clin. Kidney J.* 5,2 (Apr. 2012), pp. 102–108.
- [21] D. Shemin and L. D. Dworkin. Neutrophil gelatinase-associated lipocalin (NGAL) as a Biomarker for Early Acute Kidney Injury. *Crit. Care Clin.* 27,2 (Apr. 2011), pp. 379–389.
- [22] M. Darmon, M. Ostermann, and M. Joannidis. Predictions are difficult. . . especially about AKI. *Intensive Care Med.* 43,6 (2017), pp. 932–934.
- [23] R. Bellomo, S. T. Vaara, and J. A. Kellum. How to improve the care of patients with acute kidney injury. *Intensive Care Med.* 43,6 (2017), pp. 727–729.
- [24] S. Bailly, G. Meyfroidt, and J.-F. Timsit. What's new in ICU in 2050: big data and machine learning. *Intensive Care Med.* (Dec. 2017), pp. 1–4.

- [25] L. E. Hodgson, A. Sarnowski, P. J. Roderick, B. D. Dimitrov, R. M. Venn, and L. G. Forni. Systematic review of prognostic prediction models for acute kidney injury (AKI) in general hospital populations. *BMJ Open* 7,9 (Sept. 2017), e016591.
- [26] A. S. Levey and J. P. Bosch. A more accurate method to estimate glomerular filtration rate from serum creatinine: A new prediction equation. *Ann. Intern. Med.* 130,6 (1999), p. 461.
- [27] E. von Elm, D. G. Altman, M. Egger, S. J. Pocock, P. C. Gøtzsche, and J. P. Vandembroucke. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. *Ann. Intern. Med.* 147,8 (Oct. 2007), p. 573.
- [28] E. W. Steyerberg et al. Assessing the Performance of Prediction Models. *Epidemiology* 21,1 (Jan. 2010), pp. 128–138.
- [29] E. W. Steyerberg and Y. Vergouwe. Towards better clinical prediction models: Seven steps for development and an ABCD for validation. *Eur. Heart J.* 35,29 (2014), pp. 1925–1931.
- [30] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics* 44,3 (Sept. 1988), p. 837.
- [31] X. Robin et al. pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 12, (2011).
- [32] G. Nattino, S. Finazzi, and G. Bertolini. A new calibration test and a reappraisal of the calibration belt for the assessment of prediction models based on dichotomous outcomes. *Stat. Med.* 33,14 (2014), pp. 2390–2407.
- [33] A. J. Vickers and E. B. Elkin. Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. *Med. Decis. Mak.* 26,6 (Nov. 2006), pp. 565–574.
- [34] M. Fitzgerald, B. R. Saviile, and R. J. Lewis. Decision Curve Analysis. *JAMA* 313,4 (Jan. 2015), pp. 409–410.
- [35] B. Van Calster and A. J. Vickers. Calibration of risk prediction models: impact on decision-analytic performance. *Med. Decis. Making* 35,2 (2015), pp. 162–9.
- [36] J. A. Kellum, F. E. Sileanu, R. Murugan, N. Lucko, A. D. Shaw, and G. Clermont. Classifying AKI by Urine Output versus Serum Creatinine Level. *J. Am. Soc. Nephrol.* 26,9 (Sept. 2015), pp. 2231–2238.

NEAR-INFRARED CEREBRAL OXIMETRY TO PREDICT
OUTCOME AFTER PEDIATRIC CARDIAC SURGERY:
A PROSPECTIVE OBSERVATIONAL STUDY

Adapted from: **Flechet M** et al. Near-Infrared Cerebral Oximetry to Predict Outcome After Pediatric Cardiac Surgery. *Pediatr Crit Care Med.*2018;19(5):433-441. doi:10.1097/PCC.0000000000001495

Presented as

- Poster at the 30th Annual Congress of the European Society of Intensive Care Medicine (*ESICM17*). Vienna, Austria, October 2017.

ABSTRACT

OBJECTIVES To assess whether near-infrared cerebral tissue oxygen saturation, measured with the FORESIGHT cerebral oximeter (CAS Medical Systems, Branford, CT, USA) predicts pediatric intensive care unit length of stay, duration of invasive mechanical ventilation, and mortality in critically ill children after pediatric cardiac surgery.

DESIGN Single-center prospective, observational study.

SETTING Twelve-bed pediatric intensive care unit of a tertiary academic hospital.

PATIENTS Critically ill children and infants with congenital heart disease, younger than 12 years old, admitted to the pediatric intensive care unit between October 2012 and November 2015. Children were monitored with the FORESIGHT cerebral oximeter from pediatric intensive care unit admission until they were weaned off mechanical ventilation. Clinicians were blinded to cerebral tissue oxygen saturation data.

MEASUREMENTS AND MAIN RESULTS Primary outcome was the predictive value of the first 24 hours of postoperative cerebral tissue oxygen saturation for duration of pediatric intensive care unit stay (median [95% confidence interval (CI)], 4 days [3-8 d]) and duration of mechanical ventilation (median [95% CI], 111.3 hours (69.3-190.4 hr)). We calculated predictors on the first 24 hours of cerebral tissue oxygen saturation monitoring. The association of each individual cerebral tissue oxygen saturation predictor and of a combination of predictors were assessed using univariable and multivariable bootstrap analyses, adjusting for age, weight, gender, Pediatric Index of Mortality 2, Risk Adjustment in Congenital Heart Surgery 1, cyanotic heart defect, and time prior to cerebral tissue oxygen saturation monitoring. The most important risk factors associated with worst outcomes were an increased standard deviation of a smoothed cerebral tissue oxygen saturation signal and an elevated cerebral tissue oxygen saturation desaturation score.

CONCLUSIONS Increased standard deviation of a smoothed cerebral tissue oxygen saturation signal and increased depth and duration of desaturation below the 50% saturation threshold were associated with longer pediatric intensive care unit and hospital stays and with longer duration of mechanical ventilation after pediatric cardiac surgery.

5.1 INTRODUCTION

Although the outcome of children with congenital heart disease undergoing cardiac surgery has improved over the past years, significant morbidity and mortality still exist [1–4]. Inadequate tissue perfusion and oxygenation are often associated with poor prognosis [5]. To monitor this balance between tissue oxygen delivery and consumption, techniques such as central or mixed venous oxygen saturation, jugular or superior vena cava saturations, or arterial saturation can be used [6]. Because many of these techniques require invasive blood sampling, they are not suited for continuous monitoring; hence, hemodynamic instability or critical events are possibly missed.

Near-infrared spectroscopy (NIRS) has been proposed as a technique to monitor cerebral tissue oxygen saturation (SctO₂) noninvasively and continuously. NIRS projects near-infrared light into the brain using disposable sensors placed on the patient's forehead. From the transmitted and reflected light, NIRS-based cerebral oximeters calculate the levels of cerebral tissue oxygen. Several studies have reported the potential of NIRS oximetry to improve patient care during cardiac surgery. In a prospective study, Murkin et al [7] randomized coronary artery bypass patients to an active SctO₂ display and treatment intervention protocol or to blinded SctO₂ monitoring. They have shown that monitoring SctO₂ avoids profound cerebral desaturation and is associated with a significantly lower frequency of organ dysfunction. In children undergoing surgery for congenital heart disease, low cerebral saturation measured using NIRS cerebral oximetry has been shown to be associated with worse (neurodevelopmental) outcomes [8–10].

However, the usefulness of NIRS oximetry remains unclear in pediatric postoperative care. Only a limited number of retrospective studies have investigated its association with outcomes in this setting. Spaeder et al [11] showed that a reduced postoperative SctO₂ variability in neonatal survivors of congenital heart disease surgery is associated with poor neurodevelopmental outcomes. In a study by Phelps et al [12], low regional cerebral oxygen saturation by NIRS in the first 48 hours after the Norwood procedure was associated with adverse outcome. Vida et al [13] reported that a postoperative SctO₂ desaturation score below 50% was associated with major postoperative morbidities. Whether these findings apply to patients with a cyanotic heart defect who are likely to have lower SctO₂ during their stay remains currently unknown. Hence, despite its increasing acceptance, high uncertainty remains in the use of NIRS oximetry for critical care decision-making [14].

We hypothesized that NIRS-based cerebral oximetry in the postoperative care of infants and children after cardiac surgery is predictive for adverse

acute outcomes, such as prolonged pediatric intensive care unit (PICU) stay, prolonged duration of mechanical ventilation, and mortality. The interaction between cyanotic heart defect and $SCTO_2$ was analyzed separately.

5.2 MATERIALS AND METHODS

This prospective blinded observational study was performed between October 2012 and November 2015, in the PICU of the Leuven University Hospitals, Leuven, Belgium. The Institutional Review Board approved the enrollment and clinical data collection protocol, including a waiver of parental consent for study participation. The study is registered at ClinicalTrials.gov (NCT01706497).

5.2.1 *Study population*

Children after cardiac surgery, younger than 12 years old, with an arterial catheter in place, mechanically ventilated upon PICU admission or intubated after admission, and expected to stay at least 24 hours in the PICU, were eligible for the study. Patients were excluded if they had actual or potential brain damage, such as patients with traumatic brain injury, brain tumors, or patients after cardiopulmonary resuscitation. Patients were also excluded if they had a condition or a wound that prohibited the placement of the forehead sensors.

The following patient characteristics were prospectively collected: age, weight, gender, the Pediatric Index of Mortality (PIM) 2 score [15], cyanotic heart defect pre and post surgery, hemoglobin levels, arterial oxygen saturation (SAO_2) and central venous oxygen saturation (SvO_2) measured using intermittent blood sampling, and treatment with extra corporal membrane oxygenation (ECMO). Data from surgery were retrieved from hospital records and included univentricular circulation, cardiopulmonary bypass time, aortic clamp time, and deep hypothermic circulatory therapy. Risk Adjustment in Congenital Heart Surgery (RACHS)-1 score was calculated for all patients [16]. One patient had cerebral NIRS monitoring after implantation of a Levitronix CentriMag Left Ventricular Assist Device (Levitronix LLC, Waltham, MA), which we attributed a score of 6.

5.2.2 *Cerebral NIRS monitoring*

Cerebral tissue oxygen saturation was measured continuously with NIRS, using the FORESIGHT cerebral oximeter (CAS Medical Systems, Branford, CT). All eligible patients were monitored with bilateral sensors applied to

the frontotemporal area, from PICU admission until they were weaned off mechanical ventilation. Patients are generally admitted to the PICU intubated and progress quickly toward extubation in the first 2–12 hours following admission, provided that no major hemodynamic instabilities are still present. The monitor screens were blinded to the bedside clinicians with a sealed screen cover (Appendix, Figure 5.A.1), and monitoring data were stored with a minute-by-minute time resolution in the Patient Data Management System (MetaVision; iMD-Soft, Needham, MA). Clinicians did not have access to the NIRS data in order not to influence the independent predictive value of the signal.

5.2.3 Endpoints

The primary endpoint was the predictive performance of NIRS cerebral oximetry to predict the PICU length of stay (LOS [d]). Secondary endpoints were predictive performances of NIRS cerebral oximetry to predict hospital LOS, duration of invasive mechanical ventilation (hr), hospital mortality, and mortality at 90 days after admission to the PICU. Information on vital status at 90 days was obtained from the hospital information system. To account for death as a competing risk, duration of PICU and hospital stay and duration of mechanical ventilation were penalized in nonsurvivors to maximum duration plus 1 day or plus 1 hour, respectively. Continuous outcomes were log transformed because their distributions were positively skewed which resulted in heteroscedasticity in the residuals plot. For logistic regression analysis, the outcome variables were transformed into binary outcomes using the 75th percentile as arbitrary cutoff. This way, prolonged PICU stay was defined as longer than 8 days, prolonged hospital stay as longer than 21 days, and prolonged duration of mechanical ventilation as longer than 190 hours.

5.2.4 SCTO₂ predictors

The minute-by-minute SCTO₂ data were preprocessed as follows: first, the signal from the left and right electrodes was averaged; second, SCTO₂ values below 20% were considered as artifacts and removed prior to analysis; third, single missing values were imputed using linear interpolation. For each patient, the following summary statistics of the preprocessed minute-by-minute SCTO₂ were calculated and used as predictors: mean, linear trend (slope of a fitted trend line), standard deviation (SD) of the signal, and standard deviation of a smoothed signal (SD-s) created by taking the median of a rolling 20-sample window of the initial signal (Figure 5.1). Additionally, we calculated the

variability index defined by Spaeder et al [11], using the root mean of successive squared differences (RMSSD) of minute-by-minute NIRS cerebral SctO₂. Finally, we calculated the desaturation score introduced by Slater et al [17] as the area under the curve for commonly investigated saturation thresholds below 50% and 60% [17–21], for patient-specific saturation thresholds below the 25th and 50th percentiles, and for saturations thresholds above 70% and 80%. We also calculated the time the saturation was below or above these thresholds. We used both percentage of area and percentage of time to account for the variation of duration in the first 24 hours of monitoring between patients.

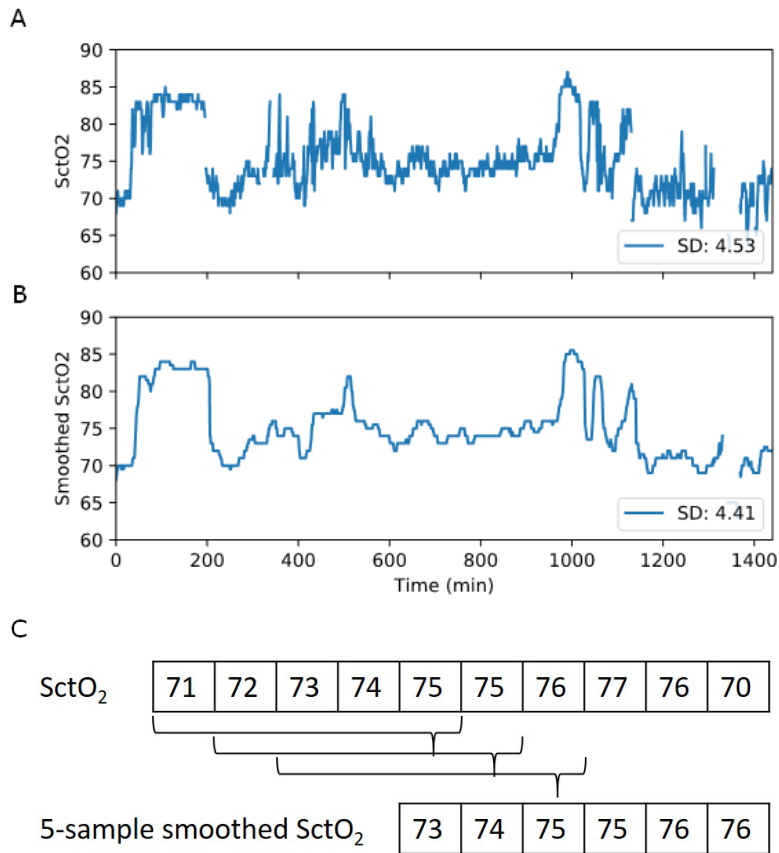


Figure 5.1 Illustration of SD-s

Example of **(A)** 24-hour standard deviation of SctO₂ (SD) and **(B)** standard deviation of a smoothed SctO₂ signal (SD-s), created by taking the median of a rolling 20-sample window of the initial signal. **(C)** Example of the creation of a 5-sample smoothed signal, created by taking the median of a rolling 5-sample window of the initial signal.

5.2.5 Association between SctO₂ predictors and outcomes

Figure 5.2 describes the steps undertaken to analyze the associations between the SctO₂ predictors and outcomes.

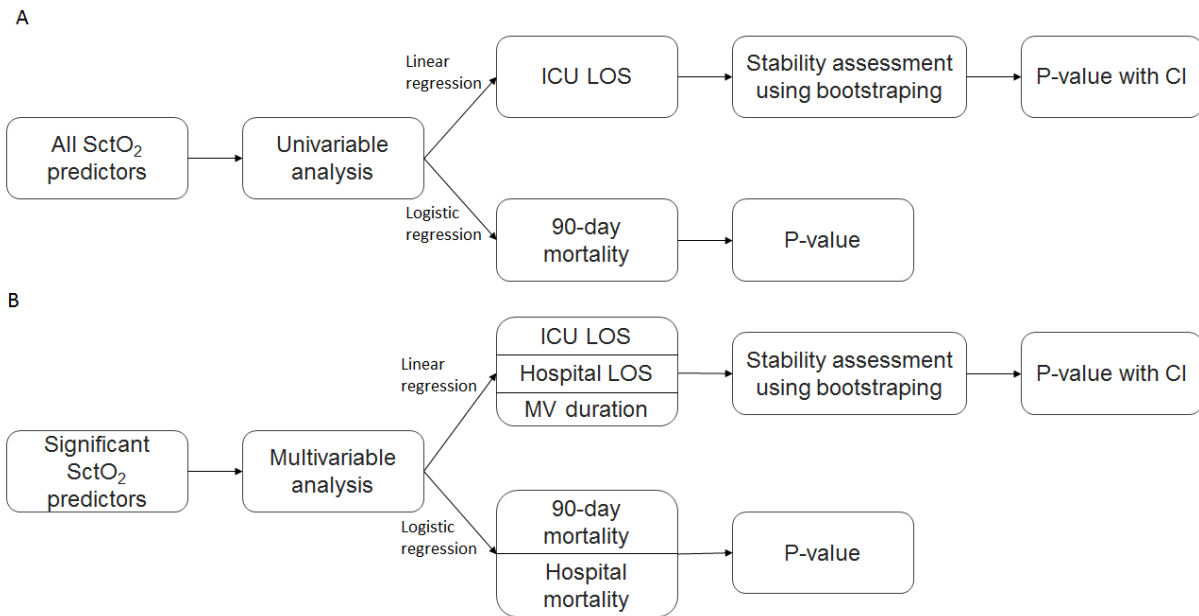


Figure 5.2 Scheme of analysis.

(A) Selection of significant predictors through univariable analysis. **(B)** For continuous outcomes, multivariable analysis were corrected for age, gender, weight, cyanotic heart defect, PIM2 score, RACHS-1 score, and time prior to SctO₂ monitoring. For binary outcome, multivariable analysis were corrected for age and PIM2 score.

First, univariable linear regression was used to investigate the association of each SctO₂-derived predictor separately, with PICU and hospital LOS, and with duration of invasive mechanical ventilation (Figure 5.2A). We used a bootstrap approach (5000 bootstraps) to create CIs and evaluate the stability of each model [22]. The SctO₂ predictors that were found to be independently and robustly associated with worse outcome were combined in a multivariable-adjusted bootstrap linear regression to assess their combined contribution for each outcome (Figure 5.2B). In the multivariable linear regression, models were adjusted for the time interval between admission and start of NIRS monitoring and for duration of mechanical ventilation prior to the start of NIRS monitoring, respectively. Models were additionally adjusted for age, weight, gender, postoperative cyanotic versus noncyanotic heart defect, PIM2 probability of death [15], and RACHS-1 score. As patients with a cyanotic heart defect post surgery are likely to have lower SctO₂ during their stay and to have less favorable outcomes, interaction terms were included in the multivariable model to understand how the SctO₂ signal is affected by the

cyanotic cardiopathy. Finally, receiver operating characteristic (ROC) curves (and area under the receiver operating characteristic curve (AUROC)) were used to assess the ability of the individual and combined predictors to discriminate the prolonged clinical outcomes.

Second, univariable (Figure 5.2A) and multivariable (Figure 5.2B) logistic regression was used to investigate the association of each independent SCTO₂-derived predictor with hospital and 90-day mortality. The low mortality prevalence precluded the use of the bootstrap approach and the adjustment for the complete list of aforementioned confounders and interaction terms. Instead, models were only adjusted for age and the PIM2 probability of death, which are strong predictors of adverse outcomes [15].

5.2.6 Statistical analysis

Data are presented as means and SDs, medians and interquartile ranges (IQRs), and numbers and proportions, where appropriate. Differences between perioperative characteristics, oxygen saturation signals, and outcomes of patients with and without cyanotic heart defect were compared with analysis of variance for continuous variables and Fisher exact test for categorical variables. Statistical significance was set at P-value of less than 0.05. All analyses were performed using Python version 2.7.13 (Python Software Foundation, <http://www.python.org>), Scipy version 0.18.1 (SciPy.org).

5.3 RESULTS

5.3.1 Study population

One hundred seventy-seven patients were included in the study. Demographics data, perioperative risk factors, and outcomes are presented in Table 5.1. Median (IQR) age at PICU admission was 4 months (1-14 mo). Median (IQR) weight upon PICU admission was 5.2 kg (3.8-8.0 kg). Overall, 107 (60.5%) were male, 45 (25.4%) had a univentricular physiology, 11 (6.2%) underwent deep hypothermic circulatory arrest, 114 (64.4%) had a cyanotic heart defect pre surgery, 70 (39.5%) had a cyanotic heart defect post surgery, and 8 (4.5%) were treated with ECMO during their stay. Median cardiopulmonary bypass duration was 84.5 minutes (56.3-113.8 min). Median aortic clamp time was 54.5 minutes (33.0-76.8 min). Median (IQR) PICU and hospital LOS were 4 days (3-8 d) and 10 days (6-21 d), respectively. Median (IQR) duration of mechanical ventilation was 111.3 hours (69.3-190.4 hr). Hospital and 90-day mortality were 5.1%. Median (IQR) duration of NIRS monitoring during the first 24 hours was 21.4 hours (9.3-

24 hr), with a median (IQR) delay between PICU admission and start of NIRS monitoring of 1.1 hours (0.5-3.8 hr). The median (IQR) delay between start of mechanical ventilation and start of NIRS monitoring was 0.5 hour (0.1-2.1 hr).

During the first 24 hours of monitoring, mean (SD) hemoglobin was 11.4 (1.6) g/dL. Mean (SD) SAO_2 , SvO_2 , and $SCTO_2$ were 93% (9%), 58% (18%), and 70% (8%), respectively. Patients with a postoperative cyanotic heart defect had more often a univentricular physiology ($p < 0.0001$), significantly higher RACHS-1 score ($p = 0.0002$), higher hemoglobin levels ($p = 0.0001$), lower SAO_2 ($p < 0.0001$), lower $SCTO_2$ ($p < 0.0001$), longer PICU ($p = 0.002$) and hospital stays ($p < 0.0001$), longer duration of mechanical ventilation ($p = 0.002$), and higher hospital mortality ($p = 0.03$) compared to patients with acyanotic heart lesions (Table 5.1).

5.3.2 Association between $SCTO_2$ predictors and PICU LOS, hospital LOS and duration of mechanical ventilation

Univariately, the $SCTO_2$ mean ($p = 0.0003$), the SD ($p < 0.0001$), the SD-s ($p < 0.0001$), the RMSSD ($p = 0.03$), the percentage of time below 50% ($p < 0.0001$) and 60% ($p < 0.0001$), above 70% ($p = 0.02$), and below the patient's median ($p = 0.002$), and the desaturation score below 50% ($p < 0.0001$), 60% ($p < 0.0001$), and above 70% ($p = 0.03$) were associated with PICU LOS (Appendix, Table 5.A.1). To reduce collinearity in the multivariable analysis, the significant predictors least associated with LOS (SD, and desaturation score and percentage of time below 60%, above 70%, and below the patient's median) were removed from further analysis.

Both the SD-s and the desaturation score below 50% remained associated with adverse clinical outcomes when considered individually (Appendix, Table 5.A.2) and combined (Table 5.2) in a multivariable analysis corrected for age, weight, gender, PIM2, RACHS-1 score, post-surgery cyanotic heart defect, and time+ prior to NIRS monitoring. The interaction analysis (Table 5.3) revealed that cyanotic cardiopathy influenced the association between $SCTO_2$ predictors and outcomes. In patients with acyanotic heart defect, both the desaturation score and the SD-s contributed to the adverse outcome, whereas in patients with a cyanotic heart defect, it was mainly the SD-s that contributes to the adverse outcome.

In a ROC curve analysis, the SD-s was more discriminant than the desaturation score below 50% (Appendix, Figures 5.A.2 and 5.A.3). Finally, the model combining the SD-s, the desaturation score below 50%, and corrected for confounders had excellent discrimination for all clinical outcomes (Figure 5.3) (AUROC [95% CI], 0.94 [0.94-0.94] for prolonged PICU stay; AUROC 0.93 [0.93-0.93]

Table 5.1 Baseline Characteristics and Outcomes

Characteristics	All	Cyanotic defect	Acyanotic defect	P-value
n	177	70	107	n.a.
Demographics				
Age, mo, median (IQR)	4 (1-14)	4.5 (1-11)	4 (1.5-17)	0.15
Weight, kg, median (IQR)	5.2 (3.8-8.0)	5.3 (3.6-7.9)	5.2 (3.8-9.3)	0.13
Male gender, n (%)	107 (60.5)	42 (60.0)	65 (60.7)	0.92
PIM2 probability of death, median (IQR)	0.11 (0.04-0.24)	0.21 (0.11-0.35)	0.05 (0.03-0.15)	<0.0001
Intraoperative				
Cyanotic heart defect before surgery, n (%)	114 (64.4)	70 (100.0)	44 (41.1)	<0.0001
RACHS-1 score, n (%)				0.0002
1	11 (6.2)	0 (0.0)	11 (10.3)	
2	73 (41.2)	26 (37.1)	47 (43.9)	
3	59 (33.3)	25 (35.7)	34 (31.8)	
4	20 (11.3)	8 (11.4)	12 (11.2)	
5	1 (0.6)	0 (0.0)	1 (1.0)	
6	13 (7.3)	11 (15.7)	2 (1.8)	
Univentricular physiology, n (%)	45 (25.4)	36 (51.4)	9 (8.4)	<0.0001
Cardiopulmonary bypass time ^b , min, median (IQR)	84.5 (56.3-113.8)	90 (58-132.5)	77 (55-110)	0.14
Aortic clamp duration ^b , min, median (IQR)	54.5 (33-76.8)	58 (0-74)	53 (37.5-77)	0.09
Deep hypothermic circulatory arrest ^c , n (%)	11 (6.2)	6 (8.6)	5 (4.7)	0.34
Postoperative				
Cyanotic heart defect post surgery, n (%)	70 (39.5)	70 (100.0)	0 (0.0)	n.a.
ECMO during PICU stay, n (%)	8 (4.5)	6 (8.6)	2 (1.9)	0.06
Hemoglobine ^d , g/dL, mean (SD)	11.4 (1.6)	11.9 (1.7)	10.9 (1.5)	0.0001
Arterial oxygen saturation ^d , %, mean (SD)	93.09 (9.0)	85.3 (9.8)	98.0 (2.8)	<0.0001
SvO ₂ ^{a,d} , d, %, mean (SD)	58.4 (18.4)	54.0 (19.9)	63.9 (14.5)	0.28
SCTO ₂ ^d , d, %, mean (SD)	70.2 (8.0)	66.1 (8.4)	73.0 (6.4)	<0.0001
Duration of SCTO ₂ monitoring ^d , hr, median (IQR)	19.5 (6.7-23.4)	21.8 (12.3-23.6)	16.6 (7.0-23.0)	0.03
Outcomes				
PICU LOS, d, median (IQR)	4.0 (3.0-8.0)	6.0 (4.0-14.0)	4.0 (2.0-6.5)	0.002
Hospital LOS, d, median (IQR)	10.0 (6.0-21.0)	16.0 (8.0-30.5)	8.0 (6.0-13.0)	<0.0001
Duration of invasive mechanical ventilation, hr, median (IQR)	111.3 (69.3-190.4)	143.0 (92.2-338.8)	93.9 (49.9-153.6)	0.002
90-day mortality, n (%)	9 (5.1)	6 (8.6)	3 (2.8)	0.16
Hospital mortality, n (%)	9 (5.1)	7 (10.0)	2 (1.9)	0.03

n.a., not applicable.

^a Measured in 18 patients.^b Available in 142 patients.^c Available in 140 patients.^d Measured during first 24 hr.

P-value is shown for comparison between patients with postoperative cyanotic versus acyanotic heart defect.

Table 5.2 Multivariable association between cerebral tissue oxygen saturation and continuous outcomes for all patients

		All patients (n = 177) – continuous outcomes		
	Outcomes	Coefficient	R-squared	P-value
SctO₂ predictors	PICU LOS			
	SD-s	0.12 (0.12–0.12)	0.609 (0.608–0.611)	0.009 (0.008–0.010)
	Desaturation score below 50%	0.43 (0.42–0.44)	0.609 (0.608–0.611)	0.003 (0.002–0.004)
SctO₂ predictors	Hospital LOS			
	SD-s	0.10 (0.10–0.10)	0.599 (0.597–0.600)	0.006 (0.005–0.007)
	Desaturation score below 50%	0.30 (0.29–0.30)	0.599 (0.597–0.600)	0.01 (0.01–0.01)
SctO₂ predictors	Duration of mechanical ventilation			
	SD-s	0.12 (0.11–0.12)	0.594 (0.592–0.595)	0.01 (0.01–0.02)
	Desaturation score below 50%	0.41 (0.40–0.41)	0.594 (0.592–0.595)	0.006 (0.005–0.008)

Association between SctO₂ and LOS or duration of mechanical ventilation was corrected for age, weight, gender, PIM2, RACHS-1, and cyanotic heart defect and time prior to NIRS monitoring.

Table 5.3 Association between cerebral tissue oxygen saturation and continuous outcomes including interactions terms

		All patients (n = 177) – continuous outcomes		
	Outcomes	Coefficient	R-squared	P-value
SctO₂ predictors	PICU LOS			
	Cyanotic	–0.43 (–0.44 to –0.42)	0.640 (0.639–0.642)	0.04 (0.04–0.05)
	SD-s * cyanotic	0.21 (0.21–0.21)	0.640 (0.639–0.642)	0.0003 (0.0003–0.0004)
	Desaturation score below 50% * cyanotic	–0.84 (–0.85 to –0.83)	0.640 (0.639–0.642)	0.004 (0.004–0.005)
	Desaturation score below 50% * SD-s	0.18 (0.18–0.18)	0.640 (0.639–0.642)	<0.0001 (<0.0001–<0.0001)
SctO₂ predictors	Hospital LOS			
	Cyanotic	–0.18 (–0.19 to –0.17)	0.624 (0.623–0.626)	0.26 (0.25–0.27)
	SD-s * cyanotic	0.16 (0.16–0.16)	0.624 (0.623–0.626)	0.0008 (0.0006–0.0009)
	Desaturation score below 50% * cyanotic	–0.66 (–0.67 to –0.66)	0.624 (0.623–0.626)	0.006 (0.005–0.007)
	Desaturation score below 50% * SD-s	0.13 (0.13–0.14)	0.624 (0.623–0.626)	0.0002 (0.0002–0.0003)
SctO₂ predictors	Duration of mechanical ventilation			
	Cyanotic	–0.44 (–0.45 to –0.43)	0.626 (0.625–0.627)	0.04 (0.04–0.05)
	SD-s * cyanotic	0.21 (0.21–0.21)	0.626 (0.625–0.627)	0.0004 (0.0004–0.0005)
	Desaturation score below 50% * cyanotic	–0.86 (–0.87 to –0.85)	0.626 (0.625–0.627)	0.004 (0.004–0.005)
	Desaturation score below 50% * SD-s	0.18 (0.18–0.18)	0.626 (0.625–0.627)	0.0001 (0.0001–0.0002)

Association between SctO₂ and LOS or duration of mechanical ventilation was corrected for age, weight, gender, PIM2, RACHS-1, and cyanotic heart defect and time prior to NIRS monitoring.

for prolonged hospital stay; AUROC 0.94 [0.94-0.94] for prolonged duration of mechanical ventilation).

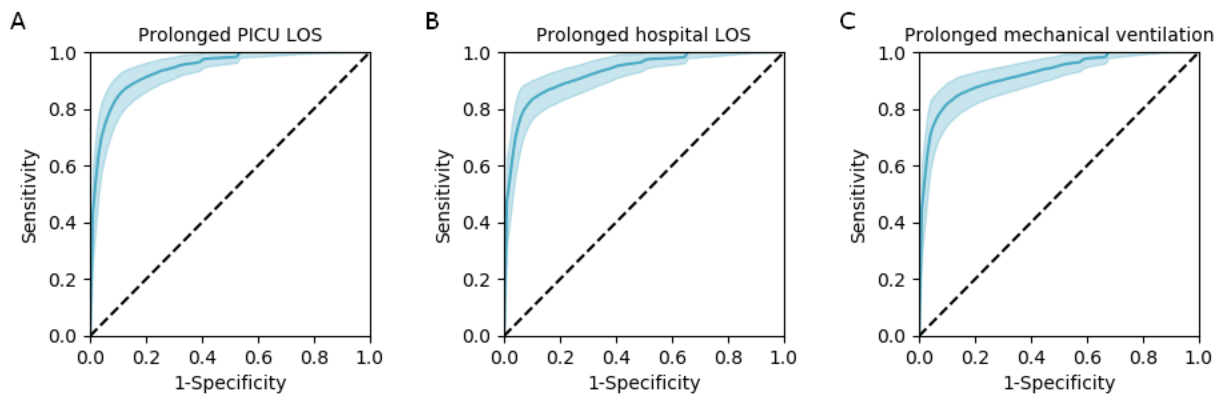


Figure 5.3 ROC curves and 95% CI of the bootstrap model combining the standard deviation of the smoothed $SctO_2$ signal (SD-s), the desaturation score below 50%, and corrected for confounders. (A) Prolonged PICU LOS (LOS > 8 d): AUROC (95% CI), 0.94 (0.94-0.94); (B) prolonged hospital LOS (LOS > 21 d): AUROC (95% CI), 0.93 (0.93-0.93); (C) prolonged duration of mechanical ventilation (duration > 190 hr): AUROC (95% CI), 0.94 (0.94-0.94).

5.3.3 Association between $SctO_2$ predictors and hospital and 90-day mortality

Univariately (Appendix, Table 5.A.3), and after correction for age and the PIM2 score (Appendix, Table 5.A.4), none of the investigated $SctO_2$ predictors were associated with increased risk of mortality (hospital or 90-day).

5.4 DISCUSSION

In this study, $SctO_2$ was monitored prospectively in the early postoperative period of a large pediatric cohort after cardiac surgery. We investigated the association between outcome and NIRS cerebral oximetry predictors identified in previous studies and several new metrics calculated from the NIRS time course. We found that an increased depth and duration of desaturation below the 50% saturation threshold and an increased SD-s in the first 24 hours after cardiac surgery are associated with longer PICU and hospital stays and with longer duration of mechanical ventilation, even after correction for several confounders.

Our findings support previous studies reporting that the $SctO_2$ desaturation score or low NIRS cerebral saturation is a predictor of adverse outcomes. Slater et al [17] introduced this desaturation score and showed that intraoperatively, an elevated desaturation score below 50% was associated with longer hospital

stay. Fischer et al [8] reported that an increased intraoperative desaturation score below 60% was associated with adverse outcome during aortic arch surgery. Vida et al [13] reported that the postoperative desaturation score below 50% was associated with major postoperative morbidities. In another study investigating postoperative NIRS after the Norwood procedure, Phelps et al [12] found that mean $SctO_2$ below 56% was a risk factor for subsequent adverse outcomes. Finally, Hansen et al [23] reported that the postoperative mean $SctO_2$ of patients with hypoplastic left heart syndrome was lower in patients with cardiac complications within 48 hours.

Expanding on these studies, we investigated alternative analytic approaches of the NIRS cerebral oxygen saturation time course. We found that an increased $SD-s$, which is a measure of variability of the low frequency components of the $SctO_2$ signal, was univariately associated with worse outcomes. Interestingly, this measure of variability remained an independent predictor in addition to the desaturation score and was a stronger predictor of poor outcome in patients with cyanotic heart lesions. Combining both resulted in excellent discrimination for prolonged stay and prolonged mechanical ventilation. The physiologic explanation for the association between an increased standard deviation and worse clinical outcomes is unclear and should be investigated further. It has been hypothesized that NIRS oximetry could be used to assess cerebrovascular autoregulation [11, 24]. In a recent study, Spaeder et al [11] introduced the cerebral tissue oxygenation index variability, which is a measure of variability of the high-frequency components of the $SctO_2$ signal [25], as the $RMSSD$ of averaged 1-minute cerebral tissue oxygenation index values for both the intraoperative and first 24 hours postoperative phases of monitoring. They found that a reduced cerebral tissue oxygenation index variability is associated with poor neurodevelopmental outcomes in neonates after congenital heart surgery and suggest that this reduced measure of variability might be a surrogate for impaired cerebral autoregulation. In our study, which included children on average older than the neonatal age and in which we have used a different cerebral oximeter, a reduced $RMSSD$ was not significantly associated with longer PICU or hospital stays, nor with longer duration of mechanical ventilation, after adjustment for confounders. Subsequent studies are necessary to confirm that the $SctO_2$ variability could be used as a surrogate of impaired autoregulation.

Evidently, patients with post-surgery cyanotic heart lesions had a lower mean SAO_2 and increased hemoglobin levels. Our study revealed significant differences in $SctO_2$ between patients with cyanotic versus acyanotic heart defect, a finding also reported in the pilot study by Tume and Arnold [26]. Notwithstanding this statistical difference, in spite of their lower SAO_2 , patients

with cyanotic heart lesions are still able to preserve brain tissue oxygenation as measured by NIRS, within what would be considered a clinically acceptable range. In these patients, the SD-s was a better predictor of adverse outcomes than the desaturation score. To our knowledge, this is the first large study investigating the interaction of cyanotic cardiopathy and SCTO₂. Our findings suggest that the pediatric population with cyanotic heart defect could benefit from different reference values of NIRS cerebral oxygen saturation.

The early recognition of children at risk after surgery for congenital heart defects could trigger interventions with the potential to improve outcome and avoid morbidity. In this study, we have discovered two novel metrics, the SD-s as a measure of SCTO₂ variability and the desaturation score below 50%, which have a robust association with outcome. Whether they provide additional and clinically relevant information in addition to conventional cardiovascular monitoring at the patient's bedside could not be assessed in this observational trial but remains an interesting hypothesis that requires further investigation. The physiologic meaning behind the association between an increased SCTO₂ variability (SD-s) and worse outcome, and whether it represents impaired cerebrovascular autoregulation, should be examined as well. The desaturation score below 50% might give the clinician an idea on the cumulative burden of low cerebral perfusion in a particular child and might be used as an early trigger to perform additional clinical, radiologic, or electrophysiologic tests to detect neurocognitive dysfunction. Pragmatic clinical trials can be set up to examine whether providing this new information to clinicians improves the outcome. For instance, the benefit of additional heart rate variability (HRV) monitoring in neonates was examined in a randomized clinical trial, where it could be demonstrated that displaying the HRV monitor was able to improve the outcomes of these infants, as compared to when the monitor was blinded [27]. In addition, the cost-effective potential of NIRS-based hemodynamic management after surgery for congenital heart defects should be addressed in future studies.

As the survival in children with congenital heart disease has improved, the focus has shifted to morbidity. The early recognition of patients at risk for poor outcome could trigger interventions to improve outcome because it allows the PICU staff to take early measures to stabilize the patient. As such, continuous display of the two metrics described in this study, the SD-s as a measure of SCTO₂ variability and the desaturation score below 50%, could provide extra information to the clinician in addition to conventional cardiovascular monitoring. However, the two metrics are currently not displayed by available cerebral oximeters, and it remains unclear how caregivers could improve or modify SCTO₂ variability. It remains unproven whether these prognostic

markers can be used as therapeutic targets. Pragmatic clinical trials can be set up to examine whether adding a new monitor or variable, and providing information to clinicians, improves the outcome. For instance, HRV monitoring has been shown to trigger earlier interventions for sepsis and improve the outcomes in neonates. SCTO₂ variability (SD-s) robustly associates with worse outcome, and we believe that the physiologic meaning behind this variable should be examined as well, and whether, for instance, it represents impaired cerebrovascular autoregulation. From a conceptual point of view, we hypothesize that the desaturation score below 50% might give the clinician an idea on the cumulative burden of low cerebral perfusion in a particular child and might be used as a trigger for clinical, radiologic, or electrophysiologic signs of neurocognitive dysfunction. In addition, the cost-effective potential of NIRS-based hemodynamic management after surgery for congenital heart defects is not known.

This study has several limitations. First, due to its single-center design, the findings highlighted in this study might not be generalizable to other populations and therefore should be validated in adequately powered prospective studies. Second, the typical low mortality prevalence of the pediatric cohort precluded any robust bootstrap analysis and might have underpowered our results related to hospital and 90-day mortality. This might explain why our results did not show that SCTO₂ provides additional prognostic value for mortality. Third, as the baseline SCTO₂ [28] was not registered prior to surgery, we could not assess trend changes compared with the patients' baseline. Nevertheless, this increases the applicability of our findings, as several centers might not register this cerebral oximeter baseline. Fourth, the calculation of the SCTO₂ standard deviation might be affected by the duration of cerebral oximetry monitoring which was shorter in patients with acyanotic heart defect. Finally, the observational design of the study prevents any conclusion on the clinical benefit of using SCTO₂ for postoperative management.

This study has several strengths. First, it is prospective in design and hence detailed in data collection. Second, it included a large patient population that enabled to perform robust statistical analysis with correction for confounders, most noticeably for cyanotic heart defect (except for the mortality analysis) and for PIM2. The analysis was based on bootstrapping, which also increased the robustness and generalizability of the findings. Third, because the clinicians were blinded to SCTO₂ data, treatment bias was excluded. Finally, this study is the first of its kind to study the interaction between NIRS cerebral oximetry and cyanotic heart defect in a critical care setting.

5.5 CONCLUSIONS

Increased desaturation below 50% and increased SctO₂ variability in the early postoperative period after cardiac surgery were found to be associated with longer PICU and hospital stays and with longer duration of mechanical ventilation, even after correction for several confounders. Furthermore, our study highlighted the difference in cerebral oxygen saturation between patients with cyanotic versus acyanotic heart defect. Hence, we recommend that future studies should be adequately powered to analyze these populations separately.

As this study was observational, our findings cannot support any conclusions regarding postoperative management of critically ill children after cardiac surgery. Opportunities for using the desaturation score and the variability of SctO₂ to drive therapeutic interventions remain to be investigated.

GRANT SUPPORT AND CONFLICT OF INTEREST

M. Flechet received funding from the Fonds Wetenschappelijk Onderzoek (FWO) as a PhD fellow (11Y1118N). Dr. Beckers disclosed off-label product use of the monitoring device used for observational purpose. Dr. Casaer received other support from a postdoctoral research grant and project grant by FWO and from the Clinical research Foundation of the University Hospitals Leuven (KOF). Dr. Van den Berghe received funding from Methusalem program of the Flemish Government (Belgium) and the European Research Council (ERC) advanced grant AdvG-2012-321670. Dr. Meyfroidt received funding from FWO as senior clinical investigator (1843118N).

The authors declare that they have no conflict of interest.

ACKNOWLEDGMENTS AND PERSONAL CONTRIBUTION

All analysis presented in this chapter were performed by Marine Flechet.

- Study concept and design: Meyfroidt, Vlasselaers, Desmet, Güiza, Van den Berghe
- Acquisition, analysis, or interpretation of data: **Flechet**, Güiza, Lamote, Delrue, Beckers, Casaer, Meyfroidt
- Drafting of the manuscript: **Flechet**
- Critical revision of the manuscript: **All authors**
- Statistical analysis: **Flechet**, Güiza, Meyfroidt
- Creation of figures: **Flechet**
- Administrative, technical, or material support: Wouters
- Study supervision: Meyfroidt

Part of the sensors and monitors used in this study were supplied by CAS Medical Systems Inc. We are grateful to the PICU nurses and residents for the patient care and setting up the near-infra- red spectroscopy monitoring, in particular to Koen Vanhonselbrouck, PICU head nurse, for the assistance in coordinating the monitoring setup; to Dr. Tom Fizez for patient management and signaling eventual technical problems; to Marc Denturck, Fredrik Hermans, and Jan Lauwers (biotechnology department University Hospitals Leuven (UZLEUVEN)) for blinding the FORESIGHT monitors.

5.A APPENDIX



Figure 5.A.1 Sealed-screen monitor.

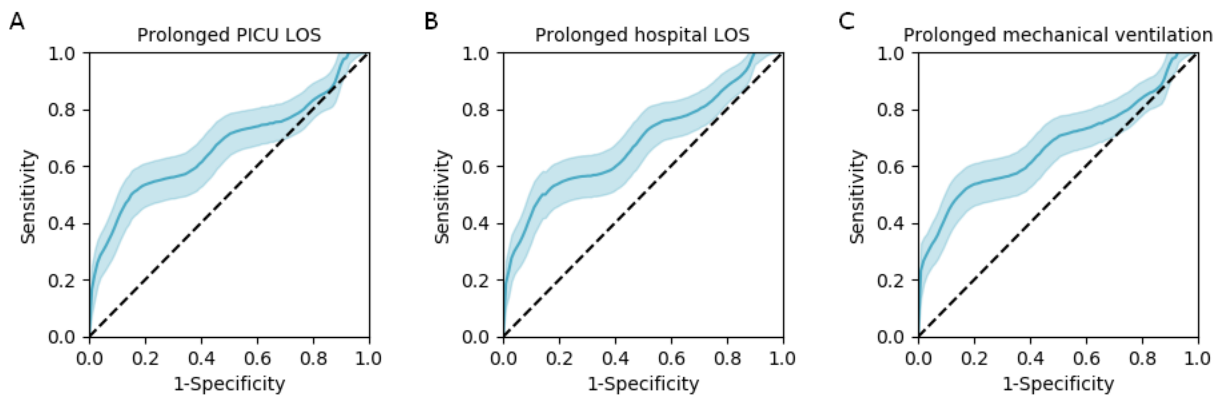


Figure 5.A.2 ROC curves and 95% CI of the bootstrap model including the standard deviation of the smoothed $SctO_2$ signal (SD-s).

(A) Prolonged PICU LOS (LOS > 8 days): AUROC (95% CI), 0.74 (0.74-0.74) (B) Prolonged hospital LOS (LOS > 21 days): AUROC (95% CI), 0.71 (0.71-0.71) , (C) Prolonged duration of mechanical ventilation (duration > 190 hours): AUROC (95% CI), 0.73 (0.73-0.73).

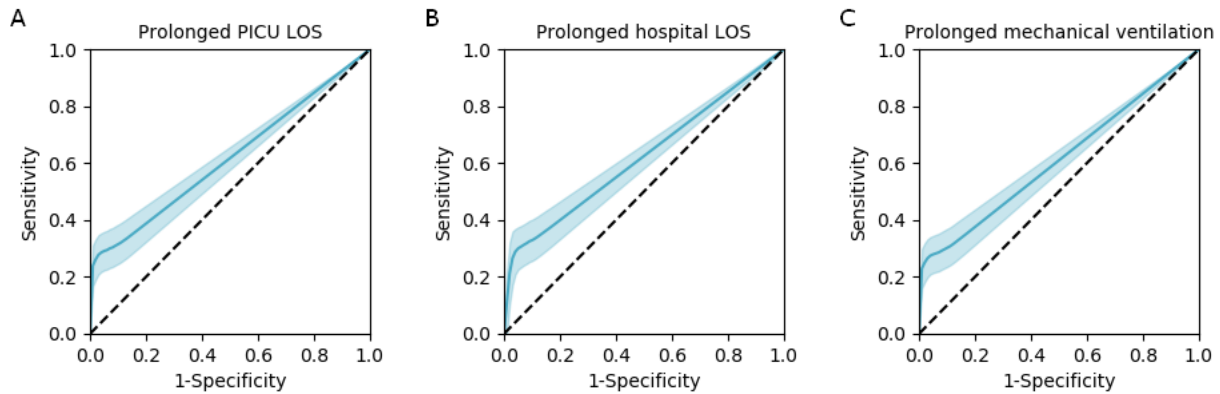


Figure 5.A.3 ROC curves and 95% CI of the bootstrap model including the desaturation score below 50%.

(A) Prolonged PICU LOS (LOS > 8 days): AUROC (95% CI), 0.62 (0.62-0.62), (B) Prolonged hospital LOS (LOS > 21 days): AUROC (95% CI), 0.61 (0.61-0.61), (C) Prolonged duration of mechanical ventilation (duration > 190 hours): AUROC (95% CI), 0.62 (0.62-0.62).

Table 5.A.1 Bootstrap univariable association between SctO₂ predictors and ICU LOS.

SctO ₂ predictors	Outcomes PICU LOS	Coefficient	R-squared	P-value
Mean		-0.04 (-0.04 to -0.04)	0.072 (0.070-0.074)	0.0003 (0.0003-0.0004)
Trend		5.64 (5.53-5.78)	0.009 (0.009-0.009)	0.21 (0.20-0.22)
SD		0.32 (0.31-0.32)	0.129 (0.127-0.130)	<0.0001 (<0.0001-<0.0001)
SD-s		0.30 (0.30-0.30)	0.131 (0.129-0.133)	<0.0001 (<0.0001-<0.0001)
RMSSD		-0.53 (-0.54 to -0.52)	0.028 (0.027-0.029)	0.03 (0.02-0.03)
% time below 50%		4.13 (4.08-4.19)	0.105 (0.104-0.107)	<0.0001 (<0.0001-<0.0001)
% time below 60%		1.56 (1.55-1.58)	0.112 (0.110-0.114)	<0.0001 (<0.0001-<0.0001)
% time below 25PCTL		2.91 (2.85-2.97)	0.013 (0.013-0.014)	0.13 (0.12-0.14)
% time below 50PCTL		3.67 (3.64-3.71)	0.051 (0.050-0.052)	0.002 (0.002-0.003)
% time above 70%		-0.48 (-0.48 to -0.47)	0.031 (0.031-0.032)	0.02 (0.02-0.02)
% time above 80%		0.19 (0.17-0.20)	0.003 (0.003-0.004)	0.44 (0.43-0.46)
% AUROC below 50%		0.88 (0.88-0.89)	0.109 (0.107-0.111)	<0.0001 (<0.0001-<0.0001)
% AUROC below 60%		0.88 (0.88-0.89)	0.109 (0.107-0.111)	<0.0001 (<0.0001-<0.0001)
% AUROC below 25PCTL		0.06 (0.06-0.07)	0.002 (0.002-0.002)	0.52 (0.51-0.53)
% AUROC below 50PCTL		-0.18 (-0.19 to -0.16)	0.002 (0.002-0.002)	0.56 (0.55-0.57)
% AUROC above 70%		0.07 (0.07-0.08)	0.025 (0.024-0.026)	0.03 (0.03-0.04)
% AUROC above 80%		0.005 (0.003-0.007)	0.003 (0.003-0.003)	0.49 (0.48-0.51)

Table 5.A.2 Multivariable association between SctO₂ predictors and continuous outcomes.

		All patients (n = 177) – continuous outcomes		
	Outcomes	Coefficient	R-squared	P-value
SctO₂ predictors	PICU LOS			
Mean		-0.02 (-0.02 to -0.02)	0.564 (0.562-0.565)	0.07 (0.07-0.07)
SD-s		0.15 (0.15-0.16)	0.582 (0.581-0.583)	0.0009 (0.0008-0.0011)
RMSSD		-0.17 (-0.17 to -0.16)	0.555 (0.553-0.557)	0.32 (0.31-0.33)
Desaturation score below 50%		0.52 (0.51-0.53)	0.589 (0.587-0.591)	0.0003 (0.0003-0.0004)
% time below 50%		2.18 (2.13-2.24)	0.582 (0.580-0.584)	0.002 (0.002-0.003)
SctO₂ predictors	Hospital LOS			
Mean		-0.02 (-0.02 to -0.02)	0.568 (0.567-0.570)	0.005 (0.005-0.006)
SD-s		0.13 (0.12-0.13)	0.578 (0.576-0.579)	0.0009 (0.0007-0.0010)
RMSSD		-0.01 (-0.01 to -0.01)	0.547 (0.545-0.549)	0.51 (0.50-0.53)
Desaturation score below 50%		0.37 (0.37-0.38)	0.577 (0.575-0.578)	0.002 (0.001-0.002)
% time below 50%		1.60 (1.56-1.64)	0.572 (0.570-0.574)	0.006 (0.005-0.007)
SctO₂ predictors	Duration of mechanical ventilation			
Mean		-0.01 (-0.01 to -0.01)	0.549 (0.547-0.550)	0.12 (0.11-0.13)
SD-s		0.15 (0.14-0.15)	0.568 (0.566-0.570)	0.002 (0.002-0.002)
RMSSD		-0.12 (-0.13 to -0.11)	0.541 (0.539-0.544)	0.42 (0.40-0.43)
Desaturation score below 50%		0.49 (0.49-0.50)	0.573 (0.571-0.575)	0.0009 (0.0007-0.0011)
% time below 50%		2.05 (2.00-2.10)	0.567 (0.565-0.569)	0.005 (0.004-0.006)

Association between SctO₂ and LOS or duration of mechanical ventilation was corrected for age, weight, gender, PIM2 score, RACHS-1 score, cyanotic heart defect and time prior to NIRS monitoring.

Table 5.A.3 Univariable association between SctO₂ predictors and 90-day mortality.

SctO ₂ predictors	Outcomes	Odds-ratio	AUROC	P-value
	90-day mortality			
Mean		0.99 (0.89-1.09)	0.52	0.77
Trend		4 ^e 30 (1 ^e -130_1 ^e 191)	0.67	0.71
SD		1.27 (0.72-2.22)	0.57	0.41
SD-s		1.67 (0.96-2.90)	0.74	0.07
RMSSD		1.01 (0.53-1.92)	0.33	0.97
% time below 50%		0.73 (2 ^e -5_-22705)	0.51	0.95
% time below 60%		1.56 (0.06-40.18)	0.46	0.79
% time below 25PCTL		10.97 (5 ^e -8_-2 ^e 9)	0.54	0.81
% time below 50PCTL		5 ^e 6 (0.00-6 ^e 15)	0.69	0.15
% time above 70%		1.83 (0.25-13.37)	0.53	0.55
% time above 80%		0.001 (1 ^e -11_-60953)	0.48	0.45
% AUC below 50%		0.67 (0.01-32.29)	0.49	0.84
% AUC below 60%		0.99 (0.47-2.07)	0.48	0.98
% AUC below 25PCTL		8 ^e 9 (0-10 ^e 200)	0.55	0.99
% AUC below 50PCTL		1.65 (0.05-49.61)	0.52	0.77
% AUC above 70%		0.73 (0.49-1.09)	0.68	0.12
% AUC above 80%		0.004 (3 ^e -10_-60533)	0.52	0.52

Table 5.A.4 Multivariable association between SctO₂ predictors and mortality.

All patients (n = 177) – categorical outcomes				
SctO ₂ predictors	Outcomes	Odds-ratio	AUROC	P-value
	Hospital mortality			
Mean		1.04 (0.94-1.15)	0.75	0.45
SD-s		1.28 (0.73-2.24)	0.86	0.40
RMSSD		0.43 (0.02-12.21)	0.81	0.62
Desaturation score below 50%		0.12 (0.00-26.56)	0.85	0.44
% time below 50%		0.00 (0.00-24402)	0.83	0.47
	90-day mortality			
Mean		1.03 (0.93-1.15)	0.80	0.54
SD-s		1.04 (0.57-1.88)	0.80	0.91
RMSSD		0.10 (0.00-4.73)	0.80	0.24
Desaturation score below 50%		0.13 (0.00-26.88)	0.81	0.46
% time below 50%		0.00 (0.00-28159)	0.79	0.49

Association between SctO₂ and mortality was corrected for age and PIM2 score.

BIBLIOGRAPHY

- [1] D. Mesotten et al. Neurocognitive development of children 4 years after critical illness and treatment with tight glucose control: a randomized controlled trial. *Jama* 308,16 (Oct. 2012), pp. 1641–50.
- [2] M. S. D. Agus et al. Tight glycemic control versus standard care after pediatric cardiac surgery. *N. Engl. J. Med.* 367,13 (2012), pp. 1208–19.
- [3] R. G. Ohye et al. Comparison of shunt types in the Norwood procedure for single-ventricle lesions. *N Engl J Med* 362,21 (2010), pp. 1980–1992.
- [4] T. Fivez et al. Early versus Late Parenteral Nutrition in Critically Ill Children. *N Engl J Med* 374,12 (2016), pp. 1111–1122.
- [5] G. V. Parr, E. H. Blackstone, and J. W. Kirklin. Cardiac performance and mortality early after intracardiac surgery in infants and young children. *Circulation* 51,5 (1975), pp. 867–874.
- [6] J. S. Tweddell et al. Mixed venous oxygen saturation monitoring after stage 1 palliation for hypoplastic left heart syndrome. *Ann. Thorac. Surg.* 84,4 (2007), 1301–1310, discussion 1310–1311.
- [7] J. M. Murkin et al. Monitoring brain oxygen saturation during coronary bypass surgery: A randomized, prospective study. *Anesth. Analg.* 104,1 (2007), pp. 51–58.
- [8] G. W. Fischer et al. Noninvasive cerebral oxygenation may predict outcome in patients undergoing aortic arch surgery. *J. Thorac. Cardiovasc. Surg.* 141,3 (2011), pp. 815–821.
- [9] J. P. Scott and G. M. Hoffman. Near-infrared spectroscopy: Exposing the dark (venous) side of the circulation. *Paediatr. Anaesth.* 24,1 (2014), pp. 74–88.
- [10] J. M. Murkin and M. Arango. Near-infrared spectroscopy as an index of brain and tissue oxygenation. *Br. J. Anaesth.* 103,Supplement 1 (2009), pp. i3–i13.
- [11] M. C. Spaeder, D. Klugman, K. Skurow-Todd, P. Glass, R. A. Jonas, and M. T. Donofrio. Perioperative Near-Infrared Spectroscopy Monitoring in Neonates With Congenital Heart Disease. *Pediatr. Crit. Care Med.* 18,3 (Mar. 2017), pp. 213–218.

- [12] H. M. Phelps et al. Postoperative cerebral oxygenation in hypoplastic left heart syndrome after the Norwood procedure. *Ann. Thorac. Surg.* 87,5 (May 2009), pp. 1490–4.
- [13] V. L. Vida et al. The Role of Regional Oxygen Saturation Using Near-Infrared Spectroscopy and Blood Lactate Levels as Early Predictors of Outcome After Pediatric Cardiac Surgery. *Can. J. Cardiol.* 32,8 (2015), pp. 970–977.
- [14] A. U. Hoskote et al. A Cross-Sectional Survey of Near-Infrared Spectroscopy Use in Pediatric Cardiac ICUs in the United Kingdom, Ireland, Italy, and Germany. *Pediatr. Crit. Care Med.* 17,1 (Jan. 2016), pp. 36–44.
- [15] A. Slater et al. PIM2: a revised version of the Paediatric Index of Mortality. *Intensive Care Med.* 29,2 (2003), pp. 278–285.
- [16] K. J. Jenkins and K. Gauvreau. Center-specific differences in mortality: Preliminary analyses using the Risk Adjustment in Congenital Heart Surgery (RACHS-1) method. *J. Thorac. Cardiovasc. Surg.* 124,1 (July 2002), pp. 97–104.
- [17] J. P. Slater et al. Cerebral Oxygen Desaturation Predicts Cognitive Decline and Longer Hospital Stay After Cardiac Surgery. *Ann. Thorac. Surg.* 87,1 (2009), pp. 36–45.
- [18] A. Casati et al. Continuous monitoring of cerebral oxygen saturation in elderly patients undergoing major abdominal surgery minimizes brain exposure to potential hypoxia. *Anesth. Analg.* 101,3 (2005), pp. 740–747.
- [19] T. Suemori, J. Skowno, S. Horton, S. Bottrell, W. Butt, and A. J. Davidson. Cerebral oxygen saturation and tissue hemoglobin concentration as predictive markers of early postoperative outcomes after pediatric cardiac surgery. *Paediatr. Anaesth.* 26,2 (2016), pp. 182–189.
- [20] F.-S. F. Yao, C.-C. a. Tseng, C.-Y. a. Ho, S. K. Levin, and P. Illner. Cerebral oxygen desaturation is associated with early postoperative neuropsychological dysfunction in patients undergoing cardiac surgery. *J. Cardiothorac. Vasc. Anesth.* 18,5 (2004), pp. 552–558.
- [21] G. M. Hoffman, C. L. Brosig, K. a. Mussatto, J. S. Tweddell, and N. S. Ghanayem. Perioperative cerebral oxygen saturation in neonates with hypoplastic left heart syndrome and childhood neurodevelopmental outcome. *J. Thorac. Cardiovasc. Surg.* 146,5 (2013), pp. 1153–64.
- [22] B. Efron and R. Tibshirani. Improvements on Cross-Validation: The 632+ Bootstrap Method. *J. Am. Stat. Assoc.* 92,438 (June 1997), pp. 548–560.

- [23] J. H. Hansen, J. Schlangen, S. Armbrust, O. Jung, J. Scheewe, and H. H. Kramer. Monitoring of regional tissue oxygenation with near-infrared spectroscopy during the early postoperative course after superior cavopulmonary anastomosis. *Eur. J. Cardio-thoracic Surg.* 43,2 (2013), pp. 37–43.
- [24] K. Brady et al. Real-time continuous monitoring of cerebral blood flow autoregulation using near-infrared spectroscopy in patients undergoing cardiopulmonary bypass. *Stroke* 41,9 (2010), pp. 1951–1956.
- [25] G. G. Berntson, D. L. Lozano, and Y. J. Chen. Filter properties of root mean square successive difference (RMSSD) for heart rate. *Psychophysiology* 42,2 (2005), pp. 246–252.
- [26] L. N. Tume and P. Arnold. Near-infrared spectroscopy after high-risk congenital heart surgery in the paediatric intensive care unit. *Cardiol. Young* 25,3 (2015), pp. 459–467.
- [27] J. R. Moorman et al. Mortality reduction by heart rate characteristic monitoring in very low birth weight neonates: A randomized trial. *J. Pediatr.* 159,6 (2011), 900–906.e1.
- [28] E. D. Sood, J. S. Benzaquen, R. R. Davies, E. Woodford, and C. Pizarro. Predictive value of perioperative near-infrared spectroscopy for neurodevelopmental outcomes after cardiac surgery in infancy. *J. Thorac. Cardiovasc. Surg.* 145,2 (Feb. 2013), 438–445.e1.

NEAR-INFRARED-BASED CEREBRAL OXIMETRY FOR
PREDICTION OF SEVERE ACUTE KIDNEY INJURY IN
CRITICALLY ILL CHILDREN AFTER CARDIAC SURGERY

Flechet M et al. Near-infrared-based cerebral oximetry for prediction of severe acute kidney injury in critically ill children after cardiac surgery (submitted for publication)

Partially presented as:

- Poster presentation at the 38th International Symposium on Intensive Care and Emergency Medicine (*ISICEM18*). Brussels, Belgium, March 2018.
- Poster presentation at the 31th Annual Congress of the European Society of Intensive Care Medicine (*ESICM18*). Paris, France, October 2018.

ABSTRACT

PURPOSE: Cerebral oximetry by near-infrared spectroscopy (NIRS) is used frequently in critically ill children but guidelines on its use for decision making in the pediatric intensive care unit (PICU) are lacking. We investigated cerebral NIRS oximetry in its ability to predict severe acute kidney injury (AKI) after pediatric cardiac surgery and assessed its additional predictive value to routinely collected data.

METHODS: Prospective blinded observational study performed between October 2012 and November 2015 in the PICU, University Hospitals Leuven, Belgium. Critically ill children with congenital heart disease, younger than 12 years old, were monitored with cerebral NIRS oximetry from PICU admission until they were successfully weaned off mechanical ventilation. The primary outcome was prediction of severe AKI 6 hours before its occurrence during the first week of intensive care. NIRS-derived predictors and routinely collected clinical data were compared and combined to assess added predictive value.

RESULTS: Of the 156 children included in the analysis, 55 (35%) developed severe AKI. The most discriminant NIRS-derived predictor was NIRS variability (area under the receiver operating characteristic curve (AUROC) 0.68; 95%confidence interval (CI), 0.67-0.68), but was outperformed by a clinical model including baseline serum creatinine, cyanotic cardiopathy pre-surgery, blood pressure and heart frequency (AUROC 0.75; 95%CI 0.75-0.75, $P < 0.001$). Combining clinical and NIRS information improved model performance (AUROC 0.79; 95%CI, 0.79-0.80; $P < 0.001$).

CONCLUSION: After pediatric cardiac surgery, NIRS variability combined with clinical information improved discriminability for AKI. Future studies are required to identify whether supplementary, timely clinical interventions at the bedside, based on NIRS variability analysis, could improve outcome.

6.1 INTRODUCTION

NIRS has gained popularity in the PICU, as it allows for continuous and non-invasive assessment of tissue oxygenation at the bedside [1, 2]. The technology is most frequently used on the patient forehead but can also be used on somatic sites to measure organ-specific tissue oxygen saturation [3, 4]. NIRS-detected cerebral hypoxia has been in relation to venous oximetry and anaerobic metabolism [3, 5, 6]. Therefore, cerebral NIRS monitoring is used as a hemodynamic monitor for early recognition of an inadequate global oxygen supply/demand relationship, and to adapt interventions to minimize the risk of secondary organ dysfunction.

Evidence on the usefulness of NIRS monitoring both for prognosis and to guide management decisions in pediatrics is limited, and has focused on cardiac surgery settings. A single study [7] suggests that NIRS-guided early detection of decreased tissue perfusion and oxygenation could allow initiation of prompt therapies to avoid organ damage and hence improve patient outcome. However, there is currently no consensus on which critical NIRS thresholds could guide patient care, trigger interventions, or be used for prognostication [8–10]. Given the relatively high cost of the monitoring sensors, it is crucial to determine specific settings in which NIRS monitoring could improve patient care.

One such setting, to which impaired global perfusion contributes significantly, is AKI. AKI is a very common complication after pediatric cardiac surgery, affecting 40% of these children [11, 12], and is associated with adverse outcomes, including prolonged duration of intensive care unit (ICU) and hospital stay and increased mortality [11–15]. A prior study has shown that decreased renal NIRS oxygen saturation was predictive of AKI after adult cardiac surgery [16]. Whether the more common cerebral site of NIRS monitoring is equally predictive of AKI is unclear.

In this study, we hypothesized that cerebral NIRS oximetry could adequately discriminate the risk of severe AKI in children following cardiac surgery and could be combined with routinely collected patient information to improve predictive performance.

6.2 METHODS

6.2.1 *Study design*

This prospective blinded observational study was performed between October 2012 and November 2015, in the PICU of the Leuven University Hospitals,

Leuven, Belgium [17]. Patient enrollment and clinical data collection protocol, including a waiver of parental consent for study participation, were approved by the Institutional Review Board. The study is registered at ClinicalTrials.gov (NCT01706497).

6.2.2 *Study population*

Children after cardiac surgery, younger than 12 years old, with an arterial line in place, mechanically ventilated upon PICU admission or intubated after admission, and expected to stay at least 24 hours in the PICU, were eligible for the study. Patients were excluded if they had actual or potential brain damage, such as traumatic brain injury, or after cardiopulmonary resuscitation. Patients were also excluded if they had a condition or a wound that prohibited the placement of the forehead NIRS sensors. In addition, we excluded patients who developed AKI within 6 hours after admission and patients in whom NIRS monitoring was initiated after AKI onset.

6.2.3 *Near-infrared spectroscopy monitoring*

Cerebral tissue oxygen saturation was continuously measured with NIRS, using the FORESIGHT cerebral oximeter (CAS Medical Systems Inc., Branford, CT, USA). All eligible patients were monitored with bilateral sensors applied to the frontotemporal area, from PICU admission until they were weaned off mechanical ventilation. The monitor screens were blinded to the bedside clinicians with a sealed screen-cover in order not to influence the independent predictive value of the signal.

6.2.4 *Prediction of acute kidney injury*

The primary outcome was prediction of severe AKI 6 hours before its occurrence during the first week of ICU stay. Severe AKI was defined as serum creatinine (serum creatinine (SC)) level ≥ 2 times the baseline level, or urine output (urine output (UO)) < 0.5 ml/kg/hour for ≥ 12 hours, or provision of dialysis (AKI stage 2 or 3) according to the Kidney Disease: Improving Global Outcome criteria [18]. Baseline SC was determined as the lowest level in the 3 months before admission. At least 1 daily SC measurement or at least 12 hours of UO in the first 7 days were required to assess AKI. The maximum AKI stage was used when the SC and UO criteria resulted in different stages.

Prediction of severe AKI was performed, first using the signal from the cerebral NIRS oximeter (NIRS) signal), second using routinely collected clinical

data, and third by combining NIRS and clinical information. Data were retrieved during the observation window (Appendix, Figure 6.A.1a-b), which lasted from admission until prediction. For patients with AKI, the observation window lasted until 6 hours before AKI onset. To reduce possible bias from monitoring duration, for patients without AKI, the observation window was determined by fitting a similar distribution of observation windows than for the patients with AKI (Appendix, Figure 6.A.1c).

6.2.4.1 NIRS model

A NIRS model was developed using the NIRS signal after preprocessing [17] and transformation to relevant predictors, including value-based metrics, variability metrics, frequency components, time and dose below various hypoxic thresholds (50%, 60%, patient-specific 25th percentile) [19, 20] and above various hyperoxic thresholds (80%, patient-specific 75th percentile). The complete list of investigated predictors is reported in Table 6.A.1 (Appendix). It is important to notice that up-to-date cerebral oximeters only display value-based metrics and the dose below a user-defined threshold.

6.2.4.2 Clinical model

A clinical prediction model was developed based on patient demographics and prospectively collected clinical information recorded during surgery, at ICU admission, and during ICU stay (Table 6.1). Monitoring information included minute-by-minute heart rate and systolic blood pressure, daily lactate levels, hemoglobin levels, arterial SAO₂ and central venous SvO₂ oxygen saturation measured using intermittent blood sampling, and treatment with extra corporal membrane oxygenation (ECMO). The median values and variability metrics (for minute-by-minute data) of monitoring information were used as clinical predictors.

6.2.4.3 Combining NIRS and clinical models

Predictors included in the NIRS model and in the clinical model were combined to assess the added predictive value of the NIRS signal as compared to routinely collected data.

6.2.5 Statistical analysis

Data are presented as means and standard deviation (SD), medians and interquartile range (IQR), and numbers and proportions, where appropriate. Differences between perioperative characteristics, NIRS signal, and outcomes

of patients with and without AKI were compared with one way ANOVA for continuous variables and Fisher exact test for categorical variables. Statistical significance was set at $P < 0.05$. All analyses were performed using Python version 2.7.13 (Python Software Foundation, <http://www.python.org>), Scipy version 0.18.1 (SciPy.org).

Reporting of the study was performed using the STROBE guidelines [21].

6.2.5.1 *Prediction model development*

Prediction models were developed using logistic regression. For each model, predictors with a univariable $P < 0.05$ were included in the models and backwards feature selection was used to identify the smallest and most accurate model, in order to reduce overfitting and optimize model generalizability. Model performance and stability were internally validated using bootstrapping (500 bootstrap replicas) [22].

6.2.5.2 *Diagnostic accuracy assessment*

To assess model performance, we reported discrimination, calibration and clinical usefulness [23]. Discrimination refers to how well the predictions allows to discriminate between patients with and without AKI. Discrimination was evaluated with the receiver operating characteristic (ROC) curve and the area under the ROC curve (AUROC). Calibration refers to the agreement between the observed frequency of AKI in the population and the model predictions. Calibration was assessed using calibration belts together with the distribution of patient numbers [24]. A statistically significant difference from perfect calibration is reported by a calibration test P -value < 0.05 [24]. Finally, the clinical usefulness of the model was assessed by the difference between the expected benefit and the expected harm associated with model classification of AKI. Clinical usefulness was visualized using decision curves and reported using ranges above *treat-all* and *treat-none* curves [25, 26].

6.3 RESULTS

6.3.1 *Study population*

A total of 177 patients met inclusion criteria and were monitored with cerebral NIRS oximetry. 21 patients were excluded of whom 4 had AKI within 6 hours after admission and 16 were monitored with NIRS after AKI onset (Figure 6.1). Additionally, NIRS monitoring was initiated later than 72 hours in 1 patient,

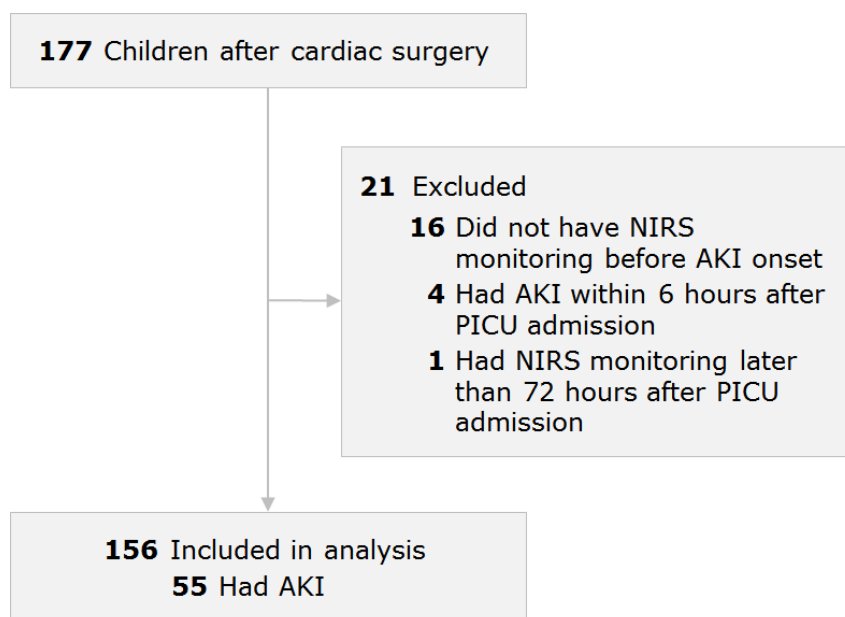


Figure 6.1 Flow diagram

which prevented the fit of prediction time for this patient. The remaining 156 patients were included in the analysis.

Demographics and outcomes of study participants are reported in Table 6.1. Fifty-five patients (35.3 %) developed severe AKI. Compared with patients without AKI, patients with AKI had a higher PICU and 90-day mortality risk ($P=0.02$ and $P=0.05$, respectively), longer PICU and hospital stay ($P<0.001$) and longer duration of mechanical ventilation ($P<0.001$).

6.3.2 Performance of NIRS model

Compared to patients without AKI, patients with AKI had a lower maximum value of NIRS signal (Appendix Table 6.A.1, $P=0.01$), lower mean ($P=0.007$), lower root-mean square of successive differences (root mean of successive squared differences (RMSSD)) 10 ($P<0.001$), and increased time and dose below 50% ($P=0.01$, $P=0.03$, respectively) and 60% ($P=0.001$). These predictors were calculated 6 hours prior to AKI diagnosis. However, given the low incidence of saturations below 50% and 60%, the robustness of these predictors could not be assessed with the bootstrapping approach.

Performance of each significant NIRS predictor for severe AKI is reported in Figure 6.2 and Table 6.A.2 (Appendix). Overall, the RMSSD, a measure of NIRS variability, had higher discriminability than the mean and the maximum values of NIRS (AUROC 0.68; 95%CI, 0.67-0.68, $P<0.001$), had wider (25-92%) and higher ranges of clinical benefit, and was well calibrated ($P=0.6$). Combining

the NIRS predictors slightly improved discriminability (Appendix, Figure 6.A.2 and Table 6.A.3, AUROC 0.69; 95%CI 0.69-0.70, $P < 0.001$). However, only RMSSD remained in the NIRS model after backwards feature selection.

Table 6.1 Patient characteristics

	All patients (n=156)	Patients with AKI (n=55)	Patients without AKI (n=101)	P-value
Demographics				
Age, month	4.0 (1.8-14.3)	2.0 (0.0-8.0)	6.0 (3.0-16.0)	0.35
Weight, kg	5.6 (3.8-8.7)	4.1 (3.3-7.4)	6.5 (4.5-8.8)	0.20
Height, cm	61.0 (53.0-74.5)	54.0 (50.0-68.0)	63.0 (57.0-76.0)	0.12
Male gender, No. (%)	97 (62.2)	37 (67.3)	60 (59.4)	0.33
Surgery data				
Univentricular circulation, No. (%)	40 (25.6)	15 (27.3)	25 (24.8)	0.85
Cardiopulmonary bypass, No. (%)	140 (89.7)	51 (92.7)	89 (88.1)	0.42
DHCA, No. (%)	8 (5.1)	5 (9.1)	3 (3.0)	0.09
Cyanotic heart defect pre-surgery, No. (%)	99 (63.5)	45 (81.8)	54 (53.5)	0.0004
Cardiopulmonary bypass duration, min	77.0 (48.5-105.0)	88.0 (55.5-111.5)	75.0 (45.0-98.0)	0.01
Aortic clamp duration, min	53.2 (36.0-72.3)	58.0 (45.5-78.5)	53.2 (34.0-68.0)	0.04
Minimum temperature, °C	34.6 (33.5-35.3)	34.4 (33.5-35.4)	34.6 (33.5-35.3)	0.14
Maximum lactate, mmol/L	2.2 (1.5-2.5)	2.5 (1.6-3.5)	2.0 (1.5-2.5)	0.0008
ICU admission data				
Pediatric Index of Mortality (PIM) ₂ probability of death, %	10.3 (3.5-22.0)	18.1 (5.1-33.2)	7.2 (3.2-15.5)	<0.0001
Baseline serum creatinine, mg/dL	0.31 (0.25-0.42)	0.39 (0.29-0.51)	0.29 (0.24-0.37)	<0.0001
Elective admission, No. (%)	138 (88.5)	46 (83.6)	92 (91.1)	0.19
Cyanotic heart defect post-surgery, No. (%)	58 (37.2)	22 (40.0)	36 (35.6)	0.60
Risk Adjustment in Congenital Heart Surgery (RACHS)-1 score, No. (%)				0.01
1	9 (5.8)	0 (0.0)	9 (8.9)	
2	67 (42.9)	19 (34.5)	48 (47.5)	
3	52 (33.3)	21 (38.2)	31 (30.7)	
4	19 (12.2)	12 (21.8)	7 (6.9)	
5	0 (0.0)	0 (0.0)	0	
6	9 (5.8)	3 (5.5)	6 (5.9)	
ECMO upon admission, No. (%)	6 (3.8)	5 (9.1)	1 (1.0)	0.02
Monitoring data				

ECMO during PICU stay, No. (%)	7 (4.5)	5 (9.1)	2 (2.0)	0.09
Heart frequency, beat/min	132 (118-145)	142 (131-150)	129 (116-139)	<0.0001
RMSSD of heart frequency, beat/min	2.61 (1.88-4.22)	2.69 (1.93-5.36)	2.60 (1.85-4.06)	0.10
Hemoglobine, g/dL	11.5 (10.2-12.4)	11.5 (10.4-12.6)	11.5 (10.2-12.3)	0.44
Systolic blood pressure, mm Hg	81.00 (70.75-90.00)	72.00 (60.50-84.00)	85.00 (75.00-93.00)	<0.0001
RMSSD of systolic blood pressure, mm Hg	3.30 (2.41-4.46)	3.07 (2.04-3.74)	3.52 (2.57-4.80)	0.005
Lactate, mmol/L	1.2 (0.9-1.8)	1.6 (1.0-3.0)	1.1 (0.9-1.5)	<0.0001
Maximum lactate, mmol/L	2.1 (1.5-3.1)	3.0 (1.8-3.8)	1.8 (1.4-2.6)	<0.0001
Arterial oxygen saturation, %, mean SD	93.6 (8.7)	93.7 (8.7)	93.6 (8.7)	0.93
Outcomes				
PICU LOS, days	4.0 (3.0-7.0)	7.0 (4.0-16.5)	3.0 (2.0-5.0)	0.0004
Hospital LOS, days	9.0 (6.0-18.0)	13.0 (8.0-31.5)	8.0 (6.0-14.0)	0.0003
Duration of mechanical ventilation, hours	97.7 (67.7-167.0)	164.2 (97.7-393.3)	74.7 (48.1-120.9)	0.0004
PICU Mortality, No. (%)	6 (3.8)	5 (9.1)	1 (1.0)	0.02
90-day mortality, No. (%)	5 (3.2)	4 (7.3)	1 (1.0)	0.05

Data are summarized using median (IQR) except where stated otherwise. P-value is shown for comparison between patients with and without AKI.

6.3.3 Performance of clinical model

Table 6.1 reports univariable differences in clinical variables between patients with and without AKI. Compared to patients without AKI, patients with AKI had a longer duration of cardiopulmonary bypass ($P=0.01$) and of aortic clamp ($P=0.04$), had higher maximum lactate levels during surgery ($P<0.001$), higher baseline SC ($P<0.001$), higher Risk Adjustment for Congenital Heart Surgery Score version 1 (RACHS-1) ($P=0.01$) and higher pediatric index of mortality (PIM2) probability of death ($P<0.001$). They were more likely to have a cyanotic heart defect prior to surgery ($P<0.001$), and to have received ECMO upon ICU admission ($P=0.02$). Regarding monitoring predictors, patients with AKI had higher heart frequency ($P<0.001$), higher lactate levels ($P<0.001$), lower systolic blood pressure ($P<0.001$) and less variable systolic blood pressure ($P=0.005$) than patients without AKI. After backwards feature selection, the remaining significant predictors were baseline SC, cyanotic heart defect prior to surgery, and median blood pressure and heart frequency (Appendix, Table 6.A.4). The multivariable clinical model achieved good discrimination (Figure 6.3, AUROC,

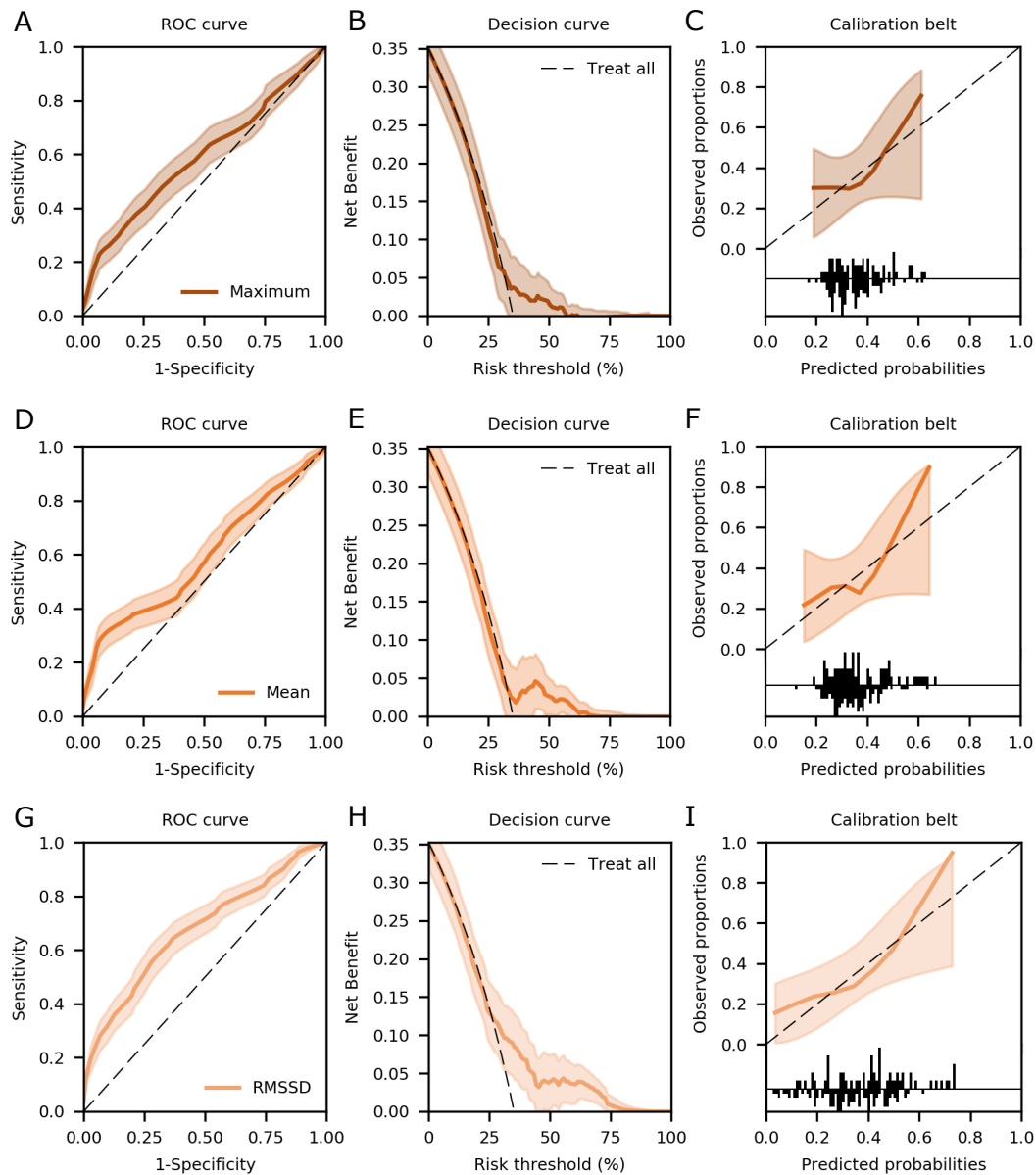


Figure 6.2 Performance of NIRS predictors.

Top row. Performance of maximum NIRS signal, **a.** ROC curve (AUROC, 0.59; 95% CI 0.59-0.59), **b.** Decision curve (clinical usefulness in ranges 32-58%), **c.** Calibration belts ($P=0.57$). *Middle row.* Performance of mean NIRS signal, **d.** ROC curve (AUROC, 0.59; 95% CI 0.58-0.59), **e.** Decision curve (clinical usefulness in ranges 34-72%), **f.** Calibration belt ($P=0.58$). *Bottom row.* Performance of RMSSD of NIRS signal, **g.** ROC curve (AUROC 0.68; 95% CI 0.67-0.68). **h.** Decision curve (clinical usefulness in ranges 27-94%). **i.** Calibration belt ($P=0.60$).

0.75; 95%CI 0.75-0.75), a wide range of clinical benefit (22-93%) and was well calibrated ($P=0.56$).

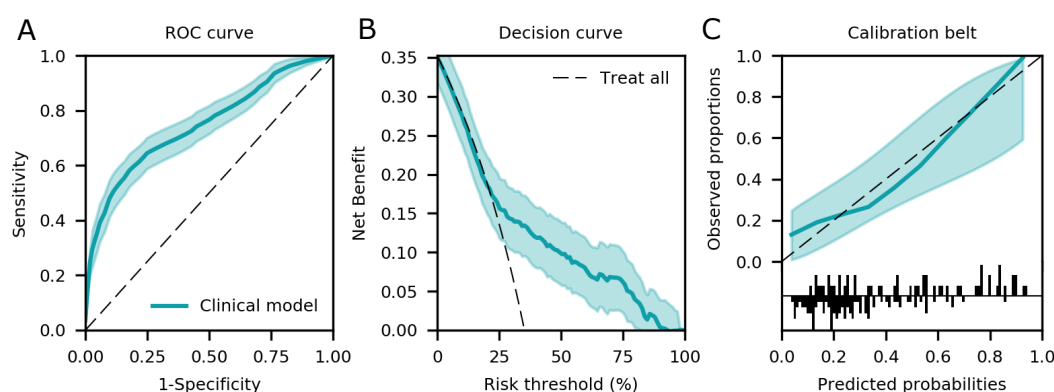


Figure 6.3 Performance of clinical model.

Performance of clinical model including baseline SC, cyanotic heart defect prior to surgery, heart rate and blood pressure; **a.** ROC curve (AUROC 0.75; 95% CI 0.75-0.75). **b.** Decision curve (clinical usefulness in ranges 22-93%). **c.** Calibration belt ($P=0.56$).

6.3.4 Comparison of NIRS and clinical models performance

The discriminative performance of NIRS RMSSD for severe AKI development was significantly lower than the performance of the clinical model ($P<0.001$). However, combining NIRS RMSSD with the clinical model improved model sensitivity which translated in significantly improved discrimination (Figure 6.4 and Table 6.A.5 (Appendix); AUROC, 0.79; 95% CI, 0.79-0.80; $P<0.001$), wider ranges of clinical benefit (14- 100%), and larger benefit in the 14-68% range compared with the clinical model (Appendix, Figure 6.A.3). For risk thresholds lower than 14%, the highest clinical benefit is achieved by considering that all patients have AKI. For risk thresholds comprised between 14% and 68%, the highest clinical benefit is achieved by combining NIRS RMSSD with the clinical model, as NIRS RMSSD improved model sensitivity (Appendix, Figure 6.A.3A). Finally, above 68%, the highest clinical benefit is achieved by using the clinical model only. Crucially at the prevalence of AKI in this cohort and for prevalences reported in the literature [11, 12], NIRS would provide benefit. These are the risk thresholds for which the tool is likely to have clinical impact. Depending on the intended use of the predictions, only if one would need a more sensitive model, would NIRS monitoring provide additional value

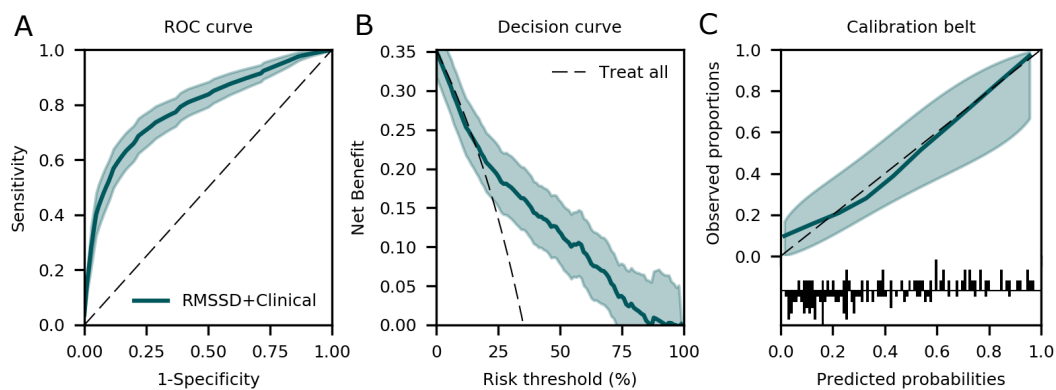


Figure 6.4 Performance of combined NIRS and clinical model.

Performance of model combining RMSSD of NIRS signal with clinical model, **a.** ROC curve (AUROC 0.79; 95% CI 0.79-0.80). **b.** Decision curve (clinical usefulness in ranges 16-100%). **c.** Calibration belt ($P=0.55$).

6.4 DISCUSSION

Currently, there is little evidence available to support the use of cerebral NIRS oximetry after pediatric cardiac surgery. In this study, we aimed to determine the potential of this cerebral oximeter to predict acute kidney injury by prospectively monitoring children after cardiac surgery for correction of a congenital heart disease. Overall, we found that the data currently displayed in cerebral oximeters have only fair discrimination for 6-hour-ahead prediction of severe AKI. Retrospective calculation of NIRS variability achieved better performance, although not sufficient to outperform a clinical model based on routinely collected patient information. However, combining NIRS variability with the clinical model significantly improved predictive capabilities, suggesting that the NIRS signal carries independent relevant information in addition to routinely collected clinical data.

Cerebral NIRS oximetry is appealing as it allows measuring brain oxygen levels non-invasively and continuously. In a pediatric perioperative setting where the use of invasive catheters is not feasible or available, such a monitor has high potential provided its clinical value is proven. Currently there is little evidence to support the use of cerebral oximetry at the pediatric patient bedside. Two randomized trials [7, 27] in coronary artery bypass patients or preterm infants used an active NIRS display and treatment intervention protocol based on NIRS desaturation. They have shown that NIRS monitoring avoids profound cerebral desaturation but were not powered to detect differences on mortality and morbidity. Other observational and retrospective studies have reported an association between (dose of) NIRS desaturation and worse (neurodevelopmental) outcomes [4, 28, 29]. A single study has investigated

the use of NIRS oximetry for AKI prediction in adults. The authors found that decreased renal NIRS oxygen saturation was predictive for AKI development after cardiac surgery [16]. In accordance with these studies, we found that NIRS-detected lower oxygen saturation was predictive for severe AKI. However, the mean and maximum NIRS value were only fair predictors of AKI, and although the time and dose of desaturation below 50% and 60% were predictive for AKI, they did not occur frequently enough for inclusion in the predictive model.

Interestingly, we found that decreased NIRS variability had more discriminative power than the NIRS value currently displayed at the patient bedside. Spaeder and colleagues were the first to investigate NIRS variability and found that decreased NIRS RMSSD, a notion of high frequency variability, was associated with poor neurodevelopmental outcomes in neonatal survivors of congenital heart disease surgery [30]. In a previous study [17], we found that the low frequency variability (SD-s) was associated with prolonged ICU stay and prolonged duration of mechanical ventilation. Here, as patients developed AKI early during ICU stay, the monitoring period before prediction was short which could have explain the lack of predictive capability from the SD-s. To our knowledge, this is the first study investigating the use of NIRS variability for acute clinical events, such as severe AKI. This study supports displaying such a metric at the patient bedside, as it may improve the clinical applicability of NIRS-based cerebral oximeters.

The current study not only reports an added predictive value of cerebral NIRS monitoring but also highlights the predictive performance of routinely collected patient information to predict severe AKI after pediatric cardiac surgery. AKI is an increasingly recognized concern in pediatric patients [31]. Therefore, it is essential to identify children whose kidney function will deteriorate to provide appropriate intervention to mitigate AKI [31]. For that purpose, several biomarkers have shown great diagnostic and predictive abilities [32, 33]. However, due to their high cost, serial measurements is not always applicable [34]. The model developed in the current study could be used continuously at the patient bedside, after translation to an online predictor such as similar models for adults [35] or encoded in electronic health records.

6.4.1 *Strengths and limitations*

Our study has several strengths: its prospective and blinded design that excludes treatment bias, its large sample size as compared to other studies investigating NIRS oximetry [3, 5, 6, 27, 30], and the thorough statistical analysis of the cerebral NIRS oxygen saturation which included not only value-based

metrics and dose below hypoxic and above hyperoxic thresholds as previous studies [6, 19], but also measures of variability and frequency.

Our study has several limitations. First, as no precise information on the cause of AKI was available, the clinical model did not include this information, nor did it include medication, while both could have influenced model performance. However, the clinical model included relevant surgical data that were associated with AKI. Second, we could not compare or combine the performance of the models with renal NIRS oximetry, which might be a more sensitive marker of kidney perfusion than cerebral NIRS oximetry [6, 16, 36]. However, major drawbacks limit the use of renal NIRS oximetry [37, 38]. First, renal oxygen saturation is influenced by the distance between the body surface and the kidney, which largely varies between patients. Additionally, the presence of subcutaneous fat alters the measurement of light absorption by hemoglobin. These limitations explain why the most common clinical application of the NIRS technology has been in assessing cerebral oxygen saturation, unaffected by these limitations. Third, as a single-center study, our findings might not generalize to other centers or to different (non-cardiac) populations. Although we performed internal validation via bootstrapping to improve generalizability, validation in external centers is required. Fourth, the lack of standardization between NIRS monitors additionally contributes to the difficulty of identifying NIRS critical thresholds in previous studies, and it is well known that the different devices have discordant values in certain ranges [39]; therefore, the findings of our study might not generalize to different cerebral oximeters. Finally, prospective validation of the models developed in this study is warranted to identify potential clinical benefit for their use at the patient bedside.

6.5 CONCLUSION

The predictive value of cerebral NIRS oximetry as displayed in the current monitors is limited for prediction of severe AKI after pediatric cardiac surgery. However, NIRS variability, in particular combined with routinely collected patient information, showed improved discrimination. Future studies are required to identify whether implementation of NIRS variability at the bedside could help identify children whose kidney function will deteriorate early enough to initiate mitigating interventions.

GRANT SUPPORT AND CONFLICT OF INTEREST

M. Flechet received funding from the Fonds Wetenschappelijk Onderzoek (FWO) as a PhD fellow (11Y1118N). Dr. Beckers disclosed off-label product use of the monitoring device used for observational purpose. Dr. Casaer received other support from a postdoctoral research grant and project grant by FWO and from the Clinical research Foundation of the University Hospitals Leuven (KOF). Dr. Van den Berghe received funding from Methusalem program of the Flemish Government (Belgium) and the European Research Council (ERC) advanced grant AdvG-2012-321670. Dr. Meyfroidt received funding from FWO as senior clinical investigator (1843118N).

The authors declare that they have no conflict of interest.

ACKNOWLEDGMENTS AND PERSONAL CONTRIBUTION

All analysis presented in this chapter were performed by Marine Flechet.

- Study concept and design: Meyfroidt, Vlasselaers, Desmet, Güiza, Van den Berghe
- Acquisition, analysis, or interpretation of data: **Flechet**, Güiza, Scharlaeken, Meyfroidt
- Drafting of the manuscript: **Flechet**
- Critical revision of the manuscript: **All authors**
- Statistical analysis: **Flechet**, Güiza, Meyfroidt
- Administrative, technical, or material support: -
- Study supervision: Meyfroidt

The FORESIGHT monitors and sensors used in the study were supplied partially by CAS Medical Systems Inc. We are grateful to the PICU nurses and residents for the patient care and setting up the near-infrared spectroscopy monitoring, in particular to Stoffel Lamote, MD, Heidi Delrue, MD, and Marc Beckers, MD, for patient management and data collection; to Pieter Wouters, MSc, for database exports; to Koen Vanhonsbrouck, PICU head nurse, for the assistance in coordinating the monitoring setup; to Tom Fivez, MD, for patient management and signaling eventual technical problems; to Marc Denturck, Fredrik Hermans, and Jan Lauwers (biotechnology department University Hospitals Leuven (UZLEUVEN)) for blinding the FORESIGHT monitors.

6.A APPENDIX

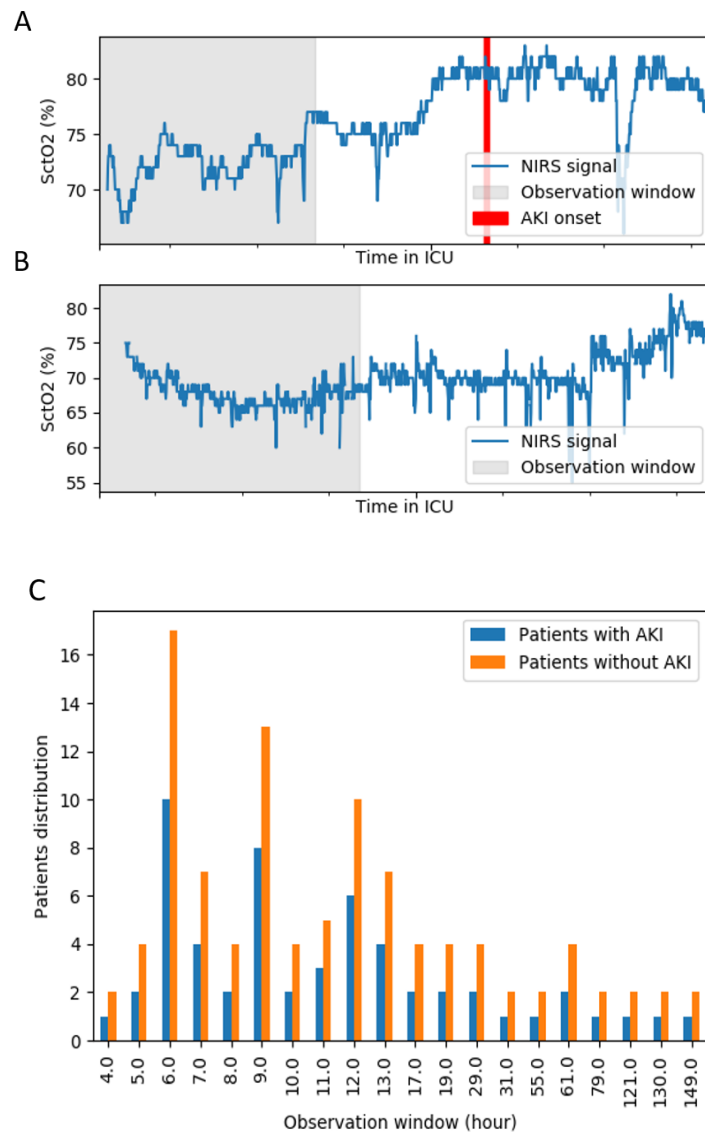


Figure 6.A.1 Distribution of observation windows.

Example of the observation window for a patient who (A) developed AKI during ICU stay, (B) did not develop AKI. For patient with AKI, the observation window corresponds to the time between admission and 6 hours prior to AKI onset. To reduce bias due to monitoring duration, the duration of the observation window for the non-AKI patients was chosen to match that of the AKI cases. (C) Bar chart of the distribution of observation windows for the patients with AKI together with the fitted distribution of observation windows for the patients without AKI.

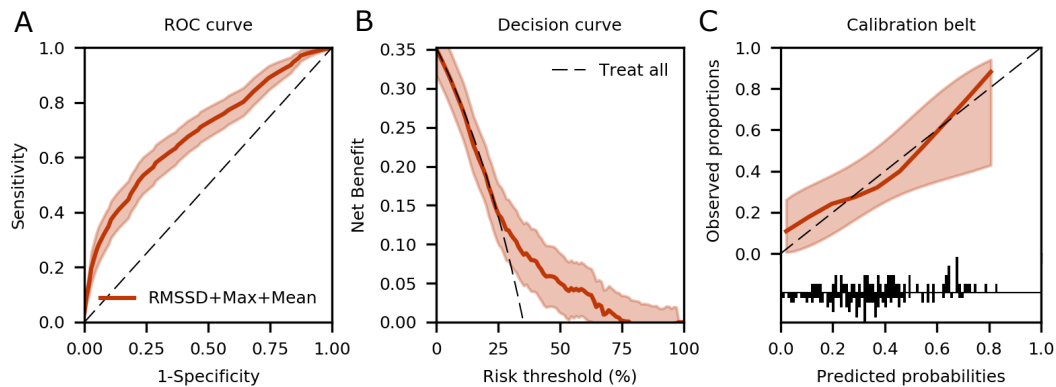


Figure 6.A.2 Performance of combined NIRS predictors.

A. ROC curve of model combining maximum, mean and RMSSD of NIRS signal (AUROC 0.69; 95% CI 0.69-0.70). **B.** Decision curve (clinical usefulness in ranges 25-79%). **C.** Calibration belt ($P=0.57$).

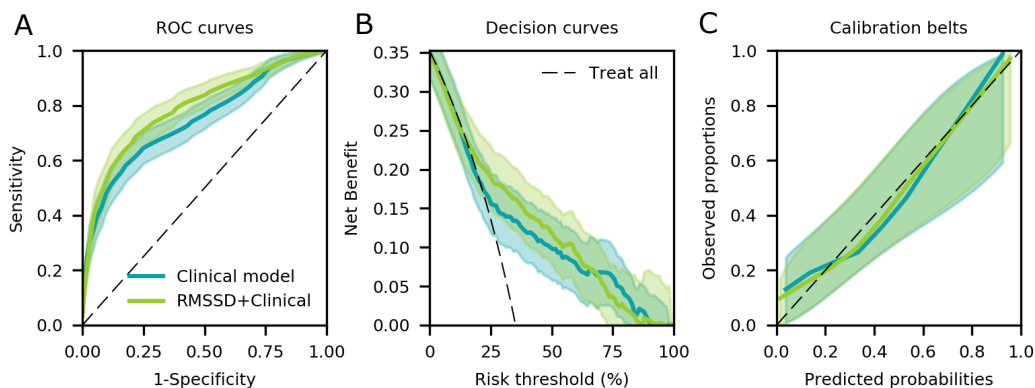


Figure 6.A.3 Comparison of performance of clinical model and combined NIRS + clinical model.

A. Comparison of ROC curves of clinical model (AUROC 0.75; 95% CI 0.75-0.75) and of combined NIRS RMSSD + clinical model (AUROC 0.79; 95% CI 0.79-0.80). **B.** Decision curve of clinical model (clinical usefulness in ranges 22-93%) and of combined NIRS RMSSD + clinical model (clinical usefulness in ranges 16-100%). Clinical benefit of combined model is improved by 2.2% as compared to only using the clinical model at a risk threshold of 20%; 2.4% at a risk threshold of 40%, and 1.3% at a risk threshold of 60%. **C.** Calibration belt of clinical model ($P=0.56$) and of combined NIRS RMSSD + clinical model ($P=0.55$).

Table 6.A.1 NIRS predictors.

	All patients (n=156)	AKI patients (n=55)	No AKI patients (n=101)	P-value
Value-based metrics				
Mean	70.2 (64.6-74.8)	69.8 (60.2-73.8)	70.4 (66.7-75.1)	0.007
Minimum	61 (53-66)	60 (51-67)	61 (54-66)	0.20
Maximum	78 (72-82)	76 (68-82)	79 (73-83)	0.01
Variability metrics				
Maximum-minimum	16 (12-21)	15 (11-19)	16 (12-21)	0.29
SD	2.9 (2.0-3.7)	3.2 (2.0-3.8)	2.9 (2.0-3.6)	0.69
SD-s	2.8 (1.8-3.6)	3.0 (1.8-3.7)	2.7 (1.8-2.8)	0.48
RMSSD	0.78 (0.63-0.95)	0.68 (0.56-0.87)	0.83 (0.69-0.98)	<0.0001
Slope	0.002 (-0.007-0.011)	0.005 (-0.006-0.014)	0.001 (-0.008-0.008)	0.28
Dose				
Dose < patient-specific 25 percentile	0.4 (0.3-0.6)	0.4 (0.2-0.6)	0.4 (0.3-0.6)	0.98
Dose < 50%	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.03
Dose < 60%	0.0 (0.0-0.2)	0.0 (0.0-0.7)	0.0 (0.0-0.0)	0.001
Dose > patient-specific 75 percentile	14.2 (10.7-16.5)	13.4 (10.5-15.8)	14.4 (11.3-16.7)	0.13
Dose > 80%	0.0 (0.0-0.9)	0.0 (0.0-0.5)	0.0 (0.0-1.1)	0.97
Time				
Time < patient-specific 25 percentile	20.0 (15.5-23.2)	20.0 (13.5-23.3)	20.0 (16.1-22.9)	0.41
Time < 50%	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.01
Time < 60%	0.0 (0.0-4.7)	0.0 (0.0-33.3)	0.0 (0.0-0.9)	0.001
Time > patient-specific 75 percentile	19.6 (16.0-22.4)	20.2 (15.8-21.7)	19.1 (16.2-23.2)	0.49
Time > 80%	0.0 (0.0-1.4)	0.0 (0.0-1.0)	0.0 (0.0-1.4)	0.95
Frequency components				
Magnitude of first largest frequency component	37277 (24816-58793)	39889 (20301-57287)	36063 (25421-59083)	0.93
Frequency of first largest frequency component, Hz	632 (420-1020)	625 (404-886)	660 (420-1020)	0.82
Magnitude of second largest frequency component	208 (114-469)	225 (125-537)	191 (110-464)	0.25
Frequency of second largest frequency component, Hz	225 (151-334)	203 (134-327)	225 (156-334)	0.81
Magnitude of third largest frequency component	208 (114-469)	225 (125-537)	191 (110-464)	0.25
Frequency of third largest frequency component, Hz	214 (132-347)	212 (127-301)	223 (132-347)	0.83

Data are reported as median IQR. Time and dose predictors are reported as time-weighted metrics. P-value is shown for differences between patients with and without AKI.

Table 6.A.2 Performance of individual NIRS predictors.

	AUROC (95% CI)	OR (95% CI)	P-value
RMSSD	0.68 (0.67-0.68)	0.056 (0.012-0.266)	<0.001
Maximum	0.59 (0.59-0.59)	0.951 (0.912-0.991)	0.06
Mean	0.59 (0.58-0.59)	0.944 (0.902-0.987)	0.04

Table 6.A.3 Performance of combined NIRS predictors.

	AUROC (95% CI)	OR (95% CI)	P-value
Combined model	0.69 (0.69-0.70)		
RMSSD		0.039 (0.007-0.234)	<0.001
Maximum		1.105 (0.982-1.243)	0.18
Mean		0.873 (0.775-0.984)	0.07

Table 6.A.4 Performance of clinical model.

	AUROC (95% CI)	OR (95% CI)	P-value
Clinical model	0.75 (0.75-0.75)		
Baseline scr!		472.71 (8.29-34002.48)	0.05
Cyanotic heart defect pre-surgery		2.68 (1.06-6.84)	0.12
Heart frequency		1.02 (1.00-1.05)	0.13
Blood pressure		0.97 (0.94-1.00)	0.12

Table 6.A.5 Performance of combined NIRS and clinical model.

	AUROC (95% CI)	OR (95% CI)	P-value
Combined model	0.79 (0.79-0.80)		
Baseline scr!		182.58 (2.19-29726.40)	0.11
Cyanotic heart defect pre-surgery		1.94 (0.72-5.26)	0.30
Heart frequency		1.02 (1.00-1.05)	0.16
Blood pressure		0.96 (0.92-1.00)	0.05
RMSSD		0.07 (0.01-0.44)	0.01

BIBLIOGRAPHY

- [1] A. Lima and J. Bakker. Noninvasive monitoring of peripheral perfusion. *Appl. Physiol. Intensive Care Med.* 2. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 39–49.
- [2] A. A. Garvey and E. M. Dempsey. Applications of near infrared spectroscopy in the neonate. *Curr. Opin. Pediatr.* 30,2 (Apr. 2018), pp. 209–215.
- [3] P. S. McQuillen et al. Regional and central venous oxygen saturation monitoring following pediatric cardiac surgery: Concordance and association with clinical variables. *Pediatr. Crit. Care Med.* 8,2 (2007), pp. 154–160.
- [4] J. M. Murkin and M. Arango. Near-infrared spectroscopy as an index of brain and tissue oxygenation. *Br. J. Anaesth.* 103, Supplement 1 (2009), pp. i3–i13.
- [5] J. Kaufman, M. C. Almodovar, J. Zuk, and R. H. Friesen. Correlation of abdominal site near-infrared spectroscopy with gastric tonometry in infants following surgery for congenital heart disease. *Pediatr. Crit. Care Med.* 9,1 (Jan. 2008), pp. 62–68.
- [6] S. B. Chakravarti, A. J. Mitnacht, J. C. Katz, K. Nguyen, U. Joashi, and S. Srivastava. Multisite Near-Infrared Spectroscopy Predicts Elevated Blood Lactate Level in Children After Cardiac Surgery. *J. Cardiothorac. Vasc. Anesth.* 23,5 (Oct. 2009), pp. 663–667.
- [7] J. M. Murkin et al. Monitoring brain oxygen saturation during coronary bypass surgery: A randomized, prospective study. *Anesth. Analg.* 104,1 (2007), pp. 51–58.
- [8] J. S. Tweddell, N. S. Ghanayem, and G. M. Hoffman. Pro: NIRS is "Standard of Care" for postoperative management. *Semin. Thorac. Cardiovasc. Surg. Pediatr. Card. Surg. Annu.* 13,1 (2010), pp. 44–50.
- [9] J. C. Hirsch, J. R. Charpie, R. G. Ohye, and J. G. Gurney. Near infrared spectroscopy (NIRS) should not be standard of care for postoperative management. *Semin. Thorac. Cardiovasc. Surg. Pediatr. Card. Surg. Annu.* 13,1 (2010), pp. 51–54.
- [10] A. Aneman et al. Advances in critical care management of patients undergoing cardiac surgery. *Intensive Care Med.* (Apr. 2018).

- [11] S. Singh. Acute kidney injury after pediatric cardiac surgery. *Ann. Card. Anaesth.* 19,2 (2016), p. 306.
- [12] J. J. Blinder et al. Acute Kidney Injury after Pediatric Cardiac Surgery: A Secondary Analysis of the Safe Pediatric Euglycemia after Cardiac Surgery Trial. *Pediatr. Crit. Care Med.* 18,7 (2017), pp. 638–646.
- [13] S. Li et al. Incidence, risk factors, and outcomes of acute kidney injury after pediatric cardiac surgery: A prospective multicenter study. *Crit. Care Med.* 39,6 (June 2011), pp. 1493–1499.
- [14] C. J. Morgan et al. Risk Factors for and Outcomes of Acute Kidney Injury in Neonates Undergoing Complex Cardiac Surgery. *J. Pediatr.* 162,1 (Jan. 2013), 120–127.e1.
- [15] J. H. Greenberg et al. Biomarkers of AKI Progression after Pediatric Cardiac Surgery. *J. Am. Soc. Nephrol.* (2018), ASN.2017090989.
- [16] D. K. Choi et al. Intraoperative renal regional oxygen desaturation can be a predictor for acute kidney injury after cardiac surgery. *J. Cardiothorac. Vasc. Anesth.* 28,3 (2014), pp. 564–571.
- [17] M. Flechet et al. Near-Infrared Cerebral Oximetry to Predict Outcome After Pediatric Cardiac Surgery. *Pediatr. Crit. Care Med.* 19,5 (May 2018), pp. 433–441.
- [18] Kidney Disease: Improving Global Outcomes (KDIGO) Acute Kidney Injury Work Group. KDIGO Clinical Practice Guideline for Acute Kidney Injury. *Kidney Int. Suppl.* 2,2 (June 2012), pp. 1–138.
- [19] J. P. Slater et al. Cerebral Oxygen Desaturation Predicts Cognitive Decline and Longer Hospital Stay After Cardiac Surgery. *Ann. Thorac. Surg.* 87,1 (2009), pp. 36–45.
- [20] F.-S. F. Yao, C.-C. a. Tseng, C.-Y. a. Ho, S. K. Levin, and P. Illner. Cerebral oxygen desaturation is associated with early postoperative neuropsychological dysfunction in patients undergoing cardiac surgery. *J. Cardiothorac. Vasc. Anesth.* 18,5 (2004), pp. 552–558.
- [21] E. von Elm, D. G. Altman, M. Egger, S. J. Pocock, P. C. Gøtzsche, and J. P. Vandembroucke. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. *Ann. Intern. Med.* 147,8 (Oct. 2007), p. 573.
- [22] B. Efron and R. Tibshirani. Improvements on Cross-Validation: The 632+ Bootstrap Method. *J. Am. Stat. Assoc.* 92,438 (June 1997), pp. 548–560.
- [23] E. W. Steyerberg et al. Assessing the Performance of Prediction Models. *Epidemiology* 21,1 (Jan. 2010), pp. 128–138.

- [24] G. Nattino, S. Finazzi, and G. Bertolini. A new calibration test and a reappraisal of the calibration belt for the assessment of prediction models based on dichotomous outcomes. *Stat. Med.* 33,14 (2014), pp. 2390–2407.
- [25] A. J. Vickers and E. B. Elkin. Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. *Med. Decis. Mak.* 26,6 (Nov. 2006), pp. 565–574.
- [26] M. Fitzgerald, B. R. Saville, and R. J. Lewis. Decision Curve Analysis. *JAMA* 313,4 (Jan. 2015), pp. 409–410.
- [27] S. Hyttel-Sorensen et al. Cerebral near infrared spectroscopy oximetry in extremely preterm infants: phase II randomised clinical trial. *BMJ* 350,January (2015), g7635.
- [28] G. W. Fischer et al. Noninvasive cerebral oxygenation may predict outcome in patients undergoing aortic arch surgery. *J. Thorac. Cardiovasc. Surg.* 141,3 (2011), pp. 815–821.
- [29] J. P. Scott and G. M. Hoffman. Near-infrared spectroscopy: Exposing the dark (venous) side of the circulation. *Paediatr. Anaesth.* 24,1 (2014), pp. 74–88.
- [30] M. C. Spaeder, D. Klugman, K. Skurow-Todd, P. Glass, R. A. Jonas, and M. T. Donofrio. Perioperative Near-Infrared Spectroscopy Monitoring in Neonates With Congenital Heart Disease. *Pediatr. Crit. Care Med.* 18,3 (Mar. 2017), pp. 213–218.
- [31] J. L. Jefferies and P. Devarajan. Early detection of acute kidney injury after pediatric cardiac surgery. *Prog. Pediatr. Cardiol.* 41, (2016), pp. 9–16.
- [32] L. Dong, Q. Ma, M. Bennett, and P. Devarajan. Urinary biomarkers of cell cycle arrest are delayed predictors of acute kidney injury after pediatric cardiopulmonary bypass. *Pediatr. Nephrol.* 32,12 (Dec. 2017), pp. 2351–2360.
- [33] C. R. Parikh et al. Postoperative Biomarkers Predict Acute Kidney Injury and Poor Outcomes after Pediatric Cardiac Surgery. *J. Am. Soc. Nephrol.* 22,9 (Sept. 2011), pp. 1737–1747.
- [34] H. R. H. de Geus, M. G. Betjes, and J. Bakker. Biomarkers for the prediction of acute kidney injury: a narrative review on current status and future challenges. *Clin. Kidney J.* 5,2 (Apr. 2012), pp. 102–108.
- [35] M. Flechet et al. 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.* 43,6 (2017), pp. 764–773.

- [36] G. M. Hoffman, N. S. Ghanayem, J. P. Scott, J. S. Tweddell, M. E. Mitchell, and K. A. Mussatto. Postoperative Cerebral and Somatic Near-Infrared Spectroscopy Saturations and Outcome in Hypoplastic Left Heart Syndrome. *Ann. Thorac. Surg.* 103,5 (May 2017), pp. 1527–1535.
- [37] J. Steppan and C. W. Hogue. Cerebral and tissue oximetry. *Best Pract. Res. Clin. Anaesthesiol.* 28,4 (2014), pp. 429–439.
- [38] K. M. Gist et al. A Decline in Intraoperative Renal Near-Infrared Spectroscopy Is Associated With Adverse Outcomes in Children Following Cardiac Surgery. *Pediatr. Crit. Care Med.* 17,4 (Apr. 2016), pp. 342–349.
- [39] C. Schmidt et al. The effects of systemic oxygenation on cerebral oxygen saturation and its relationship to mixed venous oxygen saturation: A prospective observational study comparison of the INVOS and ForeSight Elite cerebral oximeters. *Can. J. Anesth. Can. d'anesthésie* 83, (Feb. 2018).

VISUALIZING CEREBROVASCULAR AUTOREGULATION
INSULTS AND THEIR ASSOCIATION WITH OUTCOME IN
ADULT AND PEDIATRIC TRAUMATIC BRAIN INJURY

Adapted from: **Flechet M** et al. Visualizing cerebrovascular autoregulation insults and their association with outcome in adult and pediatric traumatic brain injury. *Acta Neurochirurgica Supplementum*. Vol 126. 2018:291-295. [doi:10.1007/978-3-319-65798-1-57](https://doi.org/10.1007/978-3-319-65798-1-57)

Presented as

- Abstract at the 16th International Symposium on Intracranial Pressure and Neuromonitoring (*ICP2016*). Boston, USA, June 2016.

Awarded the 3rd prize of the **Young Investigator Award**. Boston, USA, June 2016.

ABSTRACT

OBJECTIVE: The aim of this study is to visually assess the impact of duration and intensity of cerebrovascular autoregulation insults on 6-month neurological outcome in severe traumatic brain injury.

MATERIAL AND METHODS: Retrospective analysis of prospectively collected minute-by-minute intracranial pressure and mean arterial blood pressure data of 261 adult and 99 pediatric traumatic brain injury patients from multiple European centers. The relationship of 6-month Glasgow Outcome Scale with cerebrovascular autoregulation insults (defined as the low-frequency autoregulation index above a certain threshold during a certain time) was visualized in a color-coded plot. The analysis was performed separately for autoregulation insults occurring with cerebral perfusion pressure below 50 mmHg, with intracranial pressure above 25 mmHg and for the subset of adult patients that did not undergo decompressive craniectomy.

RESULTS: The color-coded plots showed a time-intensity dependent association with outcome for cerebrovascular autoregulation insults in adult and pediatric traumatic brain injury patients. Insults with a low-frequency autoregulation index above 0.2 were associated with worse outcome and below -0.6 with better outcome, with an approximately exponentially decreasing transition curve between the two intensity thresholds. All insults were associated with worse outcome when cerebral perfusion pressure was below 50 mmHg or intracranial pressure above 25 mmHg.

CONCLUSIONS: The color-coded plots indicate that cerebrovascular autoregulation is disturbed in a dynamic manner, such that duration and intensity play a role in the determination of a zone associated with better neurological outcome.

7.1 INTRODUCTION

Traumatic brain injury (TBI) is one of the most important health care problems worldwide [1, 2]. The management of severe TBI is primarily aimed at avoiding secondary brain damage, which mainly manifests as brain ischemia.

Cerebral pressure autoregulation (CAR) is the capacity of the cerebral vasculature to maintain a constant cerebral blood flow (CBF) through varying cerebral perfusion pressure (CPP). It is well known that autoregulation is often deficient in severe TBI [3], although the degree and range of this dysfunction can vary among patients, and in time within the same patient [4]. Figaji et al [5] have demonstrated the validity of the autoregulation concept in children with TBI.

Continuous monitoring of cerebrovascular autoregulation through parameters such as the Pressure Reactivity Index (PRX) [6] or Low Frequency Autoregulation Index (LAX) [7] has enabled the identification of CPP ranges in which autoregulation is more active. In retrospective analyses higher percentages of time of actual CPP contained within these ranges were associated with better outcomes [7–9].

PRX and LAX can be used to continuously identify episodes of potentially impaired CBF in TBI patients. A cut-off value for each index differentiates between episodes of active and disturbed autoregulation [10, 11]. However, as was demonstrated for intracranial pressure (ICP) [11], it is unlikely that a static threshold would capture the complexity of the association between autoregulation and outcome.

The aim of the present study is to assess the effect of CAR insults, according to varying definitions of intensity and duration, on functional outcome at six months based on prospectively collected data from continuously monitored adult and pediatric patients with severe TBI. In addition, the impact of ICP, CPP and decompressive craniectomy (DC) on the capacity to tolerate the CAR insults is investigated.

7.2 METHODS

7.2.1 *Study population*

The adult cohort consisted of 259 patients with severe TBI aged 16 years and older: 164 patients were included from the Brain-IT database [12], which collected data from 22 centers data between March 2003 and July 2005. The Multi-Centre Research Ethics Committee for Scotland (MREC/02/0/9) granted

the use of these data for scientific purposes on February 14, 2002. The data of the remaining 95 adult patients were collected from 4 centers: 38 from the San Gerardo Hospital in Monza, Italy, between March 2010 and April 2013; 25 from the University Hospitals Leuven, Belgium, between September 2010 and September 2013; 20 from the University Hospital Antwerp, Belgium, as part of the Individualized Targeted Monitoring in Neurocritical Care (NEMO) project [13], between May 2010 and June 2013; and 12 from the University Hospital Tübingen, Germany, between February and December 2009. Local Ethics Committee approval to use the anonymized data for this analysis was obtained in all centers.

The pediatric cohort consisted of 99 TBI patients, aged between 2 and 16 years: 81 patients were part of a study on TBI in children, recruited during 62 non-consecutive months up to July 2003, from two pediatric centers in Edinburgh and Newcastle, United Kingdom (UK) [Chambers2006a]. The study had Local Ethics Committee and management approval in both centers and informed consent was obtained before enrolment. The remaining 18 pediatric patients were part of the Brain-IT database.

7.2.2 Patient management

Patients were managed according to Brain Trauma Foundation (BTF) guidelines. Data collection included baseline risk factors (age, gender, admission Glasgow Coma Scale (GCS), admission pupil reactivity), minute-by-minute ICP and mean arterial blood pressure (MAP) monitoring data, and Glasgow Outcome Score (GOS) at 6 months. For the pediatric patients, a modified GOS was used, as described in the original paper [Chambers2006a]. Monitoring data in the NEMO database was recorded and stored every second; the median value of each minute interval was taken to obtain a minute-by-minute value. Signals from all datasets were reviewed independently by two senior clinicians in Leuven, and obvious artefacts at visual inspection were removed. The arterial blood pressure was calibrated to the level of the right atrium in Leuven, Antwerp and in some BrainIT centers, and to the level of the tragus in the other centers. A correction of the CPP values for arterial blood pressure transducer height was made based on the information that was obtained on the center-specific protocol. For patients from centers where the transducer was at atrium level and who were nursed with the head of the bed elevated at 30° , 10mmHg was subtracted from the registered CPP.

7.2.3 Visualization method

The method for visualizing the univariate association between insult and outcome used to assess the pressure and time burden of intracranial hypertension in Güiza et al [11] was applied in the current analysis to investigate the relationship between CAR and outcome. The LAX [7] was calculated for every minute during the monitoring period as a moving Pearson correlation coefficient between ICP and MAP. CAR insults were defined as a LAX value, i.e. a Pearson correlation coefficient, exceeding a certain intensity for a certain duration of time. The Pearson correlation coefficient between the average number of insults of a certain intensity and duration and GOS was calculated where supported data was available and was expressed by a graded color code: negative correlations in red and positive correlations in blue. The contour for zero correlation was highlighted in black and defined as the 'transition curve' as in the original study.

The relationship between CAR insults and outcome was visualized separately for insults for which CPP < 50 mmHg, those for which ICP > 25 mmHg and for the subset of adult patients that did not undergo DC (n=214).

All analyses were done in Matlab 2014b[®] (The MathWorks, Natick, MA, USA).

7.3 RESULTS

Demographics and outcomes of the studied cohorts are presented in Table 1.

The color-coded plots visualizing the correlations between GOS at 6 months and the average number of different types of CAR insults are shown in Figure 7.1A-C. In each plot, two clear overall regions emerge: one with negative correlations (blue), indicating types of CAR insults that occur more frequently in patients with higher GOS; and one with positive correlations (red), indicating types of CAR insults that occur more frequently in patients with lower GOS. The transition curves between the two zones are approximately exponential in all cohorts: for higher insult intensities, the transition occurs at shorter insult durations and, conversely, for lower insult intensities the transition occurs at longer insult durations. In all cases, irrespective of duration, CAR insults of LAX above 0.2 were associated with worse outcome and below -0.6 with better outcome. Figure 7.1D shows the overlaid transition curves for adults, adults without decompressive craniectomy and children, respectively; insults of LAX above 0 could be tolerated for 13, 19 and 35 minutes respectively. The plots for insults where CPP < 50 mmHg or for ICP > 25 mmHg were uniformly associated with worse outcome (colored red) for all studied cohorts (data not shown).

Table 7.1 Demographic and outcome data

	Adult cohort	Adult cohort without DC	Pediatric cohort
Number of patients (n)	259	214	99
LOS days, median (IQR)	15 (7-24.25)	14 (7-23)	4 (2-6.75)
Age, median (IQR)	42 (26-58)	42 (26-58)	11.4 (7.9-14.88)
Male gender (%)	79.9	80.4	74.7
Pupil reactivity (%)			
None	12.7	11.2	7.1
One	11.2	11.2	11.1
Two	70.7	71.5	74.7
Unknown, untestable or missing	5.4	6.1	7.1
GCS total, median (IQR)	7 (4-10)	7(4-10)	7 (5-8)
Unknown, untestable or missing (%)	6.6	6.1	1.0
GCS motor, median (IQR)	4 (1-5)	4 (1.5-5)	4 (2-5)
Unknown, untestable or missing (%)	4.3	2.8	0.0
DC (%)	17.4	0.0	Unknown
GOS, median (IQR)	4 (3-5)	4 (3-5)	4 (4-5)
GOS, n (%)			
1: death	46 (17.8)	36 (16.8)	12 (12.1)
2: vegetative	10 (3.9)	7 (3.3)	0 (0.0)
3: severe disability	70 (27)	58 (27.1)	7 (7.1)
4: moderate disability	49 (18.9)	44 (20.6)	39 (39.4)
5: low disability	84 (32.4)	69 (32.2)	41 (41.4)

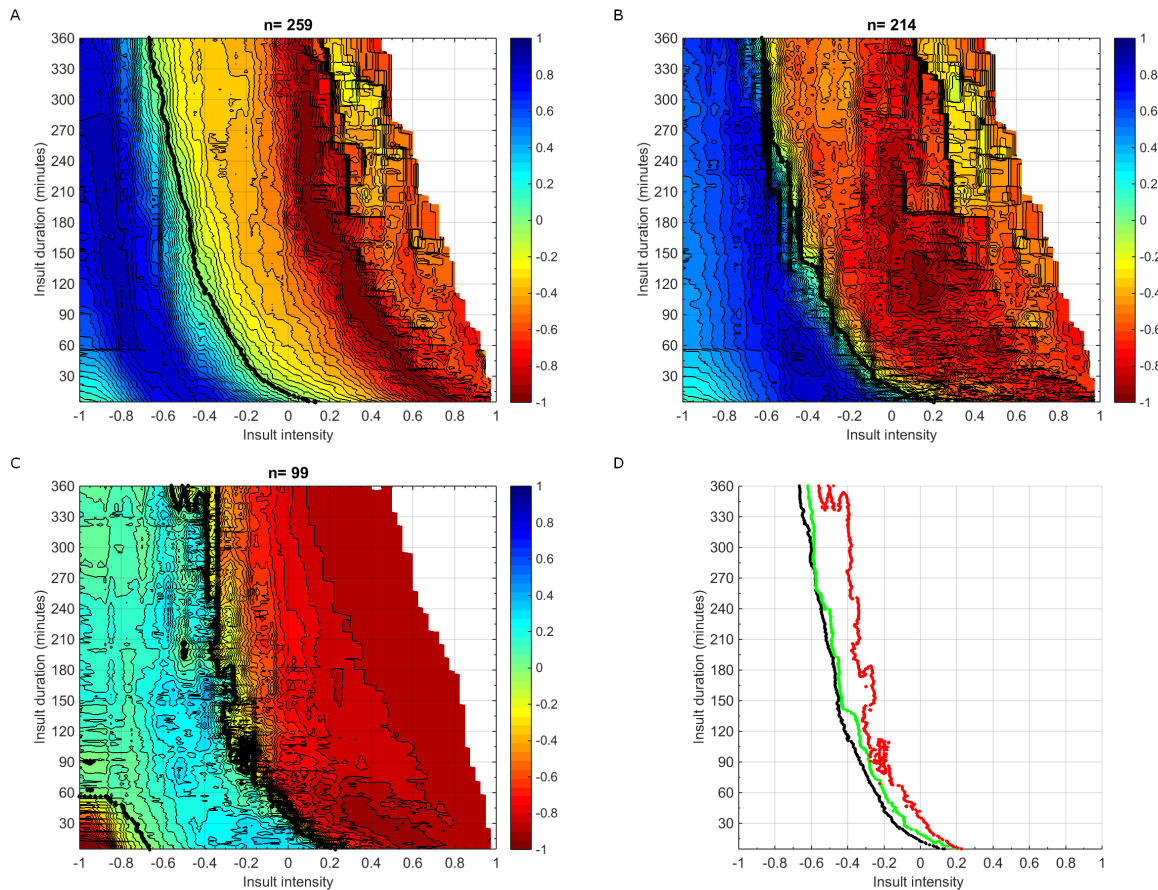


Figure 7.1 Visualization of correlation between GOS and average number of CAR insults per GOS category

(A) Adults ($n=259$) **(B)** Adults without decompressive craniectomy ($n=214$) **(C)** Children ($n=99$). Each colour-coded point in the plot refers to a number of CAR insults, defined by a LAX intensity threshold (X-axis), and a duration threshold (Y-axis). The univariate correlation of each CAR insult (characterized by LAX intensity and duration thresholds) with outcome is colour-coded according to the scale ranging from -1 to 1. Negative numbers indicate CAR insults associated with worse outcome (lower GOS categories); positive numbers indicate CAR insults associated with better outcome (higher GOS categories). The contour of zero correlation is highlighted in black, and is called the transition curve. **(D)** Overlaid transition curves for adults (black), adults without decompressive craniectomy (green) and children (red).

7.4 DISCUSSION

In this study, the univariate relationship between 6-month neurological outcome and CAR insults is summarized in color-coded plots. These plots do not represent the cumulative time/pressure dose per patient, but per type of insult, characterized by duration and intensity. The main finding is the emergence of regions of positive and negative association between CAR insults and outcome that were separated by transition curves, which had an exponential course similar to that seen in [7] for ICP. The higher the LAX, the shorter the time this insult type could be tolerated. Although the association appears stronger for the case of adults without decompressive craniectomy, remarkably the transition curves for the 3 studied cohorts lie very close together, hinting at the universal applicability of the CAR insult concept, irrespective of age. Insults when CPP < 50 mmHg or ICP > 25mmHg were associated with poor outcome regardless of CAR status, which mirrors the findings of Güiza et al [11].

This study supports previous work reporting on the dynamic aspect of autoregulation impairment in TBI [4, 6]. Additionally, this study elaborates on the concept of grey zone that has been introduced with other indices available at the bedside such as PRX or mean flow index (MX). Critical thresholds of autoregulation have been introduced for these indices [10, 14]. For both indices, the association with outcome is unclear between 0-0.05 and 0.3. In the present study, we showed that between -0.6 and 0.2, the LAX has a time-dependent relationship with outcome, which might explain why a clear association with outcome, that disregards the duration component, could not be defined. Further studies are needed to assess whether the proposed methodology could redefine the grey zone of PRX and MX. Furthermore, together with the previous analysis on ICP insults [11] the current study further advocates the need to bring the patient in a state of functional autoregulation in adult and pediatric TBI management.

This study has the following limitations. First, the sample size remains relatively small, noticeably so for the pediatric cohort, where the lack of data likely precluded the generation of a smoother transition curve. Second, as the data was available in minute-by-minute resolution only LAX could be studied to define CAR insults. Third, the data incorporate therapeutic influences, which cannot be removed. Fourth, artefacts in the monitoring data were manually removed by two clinical experts, and we cannot exclude that some artefacts went unnoticed. Lastly, we cannot exclude an influence on results from confounders that were not analyzed.

7.5 CONCLUSION

Following TBI, CAR is disturbed in a dynamic manner, such that duration and intensity play a role in the determination of a safe CAR zone. Insults of impaired CAR can only be sustained provided that they are of short duration. Hence, episodes of disturbed CAR should be considered as brain endangering secondary insults in their own right, irrespective of ICP and CPP. The current findings need to be validated with other CAR indexes and the relative weight of ICP and CAR insults needs to be further explored.

GRANT SUPPORT AND CONFLICT OF INTEREST

M.Flechet receives funding from the Fonds Wetenschappelijk Onderzoek (FWO) as a PhD. fellow (11Y1116N). G.Meyfroidt receives funding from FWO as senior clinical investigator (1846113N). G.Van den Berghe, through the KULeuven, receives long-term research financing via the Flemish government Methusalem programme. Brain-IT was funded by the European Framework Program (FP5-QRLI-2000-00454, QLGT-2002-00160 and FP7-IST-2007-217049). The NEMO project in the University Hospital Antwerp was funded by the Flemish Government Agency for Innovation by Science and Technology (IWT).

The authors declare that they have no conflict of interest.

ACKNOWLEDGMENTS AND PERSONAL CONTRIBUTION

All analysis presented in this chapter were performed by Marine Flechet and Fabian Güiza.

- Study concept and design: Güiza, Meyfroidt, Depreitere, Van den Berghe
- Acquisition, analysis, or interpretation of data: **Flechet**, Güiza, Meyfroidt
- Drafting of the manuscript: Güiza
- Critical revision of the manuscript: **All authors**
- Statistical analysis: **Flechet**, Güiza, Meyfroidt
- Administrative, technical, or material support: -
- Study supervision: Güiza

The authors wish to acknowledge the non-author steering group members of Brain-IT: Barbara Gregson, Tim Howells, Karl Kiening, Arminas Ragauskas, Juan Sahuquillo and Jan Oliver Neumann.

BIBLIOGRAPHY

- [1] A. A. Hyder, C. A. Wunderlich, P. Puvanachandra, G. Gururaj, and O. C. Kobusingye. The impact of traumatic brain injuries: a global perspective. *NeuroRehabilitation* 22,5 (2007), pp. 341–353.
- [2] W. Peeters et al. Epidemiology of traumatic brain injury in Europe. *Acta Neurochir. (Wien)*. 157,10 (Oct. 2015), pp. 1683–1696.
- [3] G. J. Bouma, J. P. Muizelaar, K. Bando, and A. Marmarou. Blood pressure and intracranial pressure-volume dynamics in severe head injury: relationship with cerebral blood flow. *J. Neurosurg.* 77,1 (July 1992), pp. 15–19.
- [4] G. E. Sviri, R. Aaslid, C. M. Douville, A. Moore, and D. W. Newell. Time course for autoregulation recovery following severe traumatic brain injury. *J. Neurosurg.* 111,4 (Oct. 2009), pp. 695–700.
- [5] A. A. Figaji et al. Pressure autoregulation, intracranial pressure, and brain tissue oxygenation in children with severe traumatic brain injury. *J. Neurosurg. Pediatr.* 4,5 (Nov. 2009), pp. 420–428.
- [6] M. Czosnyka, P. Smielewski, P. Kirkpatrick, R. J. Laing, D. Menon, and J. D. Pickard. Continuous Assessment of the Cerebral Vasomotor Reactivity in head injury. *Neurosurgery* 41,July (1997), pp. 11–19.
- [7] B. Depreitere et al. Pressure autoregulation monitoring and cerebral perfusion pressure target recommendation in patients with severe traumatic brain injury based on minute-by-minute monitoring data. *J. Neurosurg.* 120,6 (June 2014), pp. 1451–1457.
- [8] M. J. H. Aries et al. Continuous determination of optimal cerebral perfusion pressure in traumatic brain injury*. *Crit. Care Med.* 40,8 (Aug. 2012), pp. 2456–2463.
- [9] F. Güiza, G. Meyfroidt, T. Y. M. Lo, P. A. Jones, G. Van Den Berghe, and B. Depreitere. Continuous optimal CPP based on minute-by-minute monitoring data: A study of a pediatric population. *Acta Neurochir. Suppl.* 122, (2016), pp. 187–191.
- [10] E. Sorrentino et al. Critical Thresholds for Cerebrovascular Reactivity After Traumatic Brain Injury. *Neurocrit. Care* 16,2 (Apr. 2012), pp. 258–266.

- [11] F. Güiza et al. Visualizing the pressure and time burden of intracranial hypertension in adult and paediatric traumatic brain injury. *Intensive Care Med.* 41,6 (June 2015), pp. 1067–1076.
- [12] I. Piper et al. The BrainIT group: concept and core dataset definition. *Acta Neurochir. (Wien)*. 145,8 (Jan. 2003), pp. 615–629.
- [13] B. F. E. Feyen, S. Sener, P. G. Jorens, T. Menovsky, and A. I. R. Maas. Neuromonitoring in traumatic brain injury. *Minerva Anesthesiol.* 78,8 (2012), pp. 949–958.
- [14] E. Sorrentino et al. Critical thresholds for transcranial doppler indices of cerebral autoregulation in traumatic brain injury. *Neurocrit. Care* 14,2 (2011), pp. 188–193.

DEVELOPMENT OF THE PROTOTYPE OF A MONITOR TO VISUALIZE THE CUMULATIVE BURDEN OF SECONDARY BRAIN INJURY DYNAMICALLY

Partially presented as

- Abstract at the 16th International Symposium on Intracranial Pressure and Neuromonitoring (*ICP2016*). Boston, USA, June 2016.
- Abstract at the Annual Neurocritical Care Society meeting (*NCS2016*). Washington, USA, September 2016.
- Oral presentation at the meeting of the Belgian Society of Neurosurgery (*BSN17*). Brussels, Belgium, March 2017.
- Oral presentation at the 10th Euro Neuro meeting (*EURONEURO2018*). Brussels, Belgium, November 2018.

Awarded

- 3rd prize of the **Young Investigator Award** from *ICP2016*. Boston, USA, June 2016.
- 1st prize of the **Best Poster Award** from *EURONEURO2018*. Brussels, Belgium, November 2018.

ABSTRACT

PURPOSE: Clinical trials aiming at aggressively treating intracranial pressure (ICP) below fixed, generic thresholds have given disappointing results. The cumulative time-intensity burden (dose) of elevated ICP and of impaired cerebrovascular pressure autoregulation (CAR) are associated with worse outcome in patients with severe traumatic brain injury (TBI). At the population level, this dose can be represented graphically, to identify the time-intensity burden associated with good or poor outcome. We hypothesized that the course of individual patients could be visualized both in terms of ICP and CAR in the population-based plots. For that purpose, we aim to develop the prototype of a bedside monitor, to calculate and display the Low Frequency Autoregulation Index (LAX) as a measure of CAR, and to visualize patient-specific secondary brain insults in terms of ICP and of CAR, in a continuous way.

MATERIAL AND METHODS: Retrospective analysis of minute-by-minute ICP and CAR monitoring data from a large multicenter database of TBI patients. We visualized in individual patients the current dose and the cumulative dose, from the past 6 hours, and since intensive care unit (ICU) admission, both for ICP and CAR. Doses were overlaid on population-based color-coded plots that visualize the association with 6-month neurological outcome. Additionally, we calculated the percentage of time spent in the zone associated with poor outcome (red zone) and compared it to the time spent above the traditional threshold of 20 mmHg.

RESULTS: A prototype software was created to visualize the current and cumulative burden of ICP and CAR of individual patients. We present the clinical course of 2 patients, 1 with good outcome, and 1 who died.

CONCLUSIONS: The proposed method visualizes the current and cumulative time and pressure burden of ICP and CAR for individual patients, which could help a neuro-intensivist in identifying when a patient is currently in a state of potentially harmful elevated ICP, impaired autoregulation or when the outcome is at a turning point.

8.1 INTRODUCTION

Traumatic brain injury (TBI) is one of the most important health care problems worldwide [1, 2]. In the hours and days following the initial traumatic insult (or *primary injury*), destructive and self-propagating biological changes in the brain can lead to subsequent additional damage (or *secondary injury*) [3]. Prevention and treatment of secondary injury is the main goal of neuro-critical care in patients with TBI.

A subtype of secondary brain injury is intracranial hypertension [4]. Such events of elevated ICP have been shown to be associated with worse outcome [5]. International guidelines recommended ICP lowering therapies at a threshold of 20 or 22 mmHg [6, 7]. Following the intuition that universal thresholds are likely suboptimal and that duration of elevated ICP plays a role [8], Güiza et al plotted the association between neurological outcome and the dose of ICP, defined as a combination of intensity and duration [9]. Two distinct regions associated with good or poor neurological outcomes emerged from the plot, suggesting the importance of the time-intensity relationship of ICP for patient outcome. Additionally, these two regions were affected by the cerebrovascular pressure autoregulation (CAR) capability of the patient.

CAR is the capacity of the cerebral vasculature to maintain a constant cerebral blood flow (CBF) through varying cerebral perfusion pressure (CPP), the driving pressure gradient for blood to enter the skull. CAR is often deficient in severe TBI [10], although the degree and range of this dysfunction can vary among patients, and in time within the same patient [11]. Currently, international guidelines state that evidence is insufficient to support recommendation for the management of patients with severe TBI regarding CAR [6]. A potential culprit is that continuous monitoring of CAR is challenging [12], largely because it is usually based on indices reflecting the correlation between ICP and mean arterial blood pressure (MAP) when these signals are captured at very high frequencies, which restricts its use to specialized research centers [13]. As such, the LAX is a promising index developed by our research group [14], but has never been evaluated prospectively. In a previous study [15] (Chapter 7), we have shown that both the duration and intensity of deficient CAR, measured with LAX, play a role in the determination of zones associated with good or poor neurological outcome.

The population-based plots visualizing the associations of the doses of ICP and CAR with neurological outcome [9, 15] open up avenues for novel ways of managing patients with severe TBI. We hypothesized that the course of individual patients could be visualized both in terms of ICP and CAR in the population-based plots. For that purpose, we aim to develop the prototype of

a bedside monitor, to calculate and display the LAX, and to visualize patient-specific secondary brain insults in terms of ICP and of CAR, in a continuous way.

8.2 METHODS

8.2.1 *Patient population*

Retrospective analysis of minute-by-minute ICP and MAP monitoring data from patients with severe TBI included in the large multicenter Brain-IT database [16], which collected data from 22 centers data between March 2003 and July 2005. The use of these data for scientific purposes was granted by the Multi-Centre Research Ethics Committee for Scotland (MREC/02/0/9, February 14, 2002).

8.2.2 *Cerebral autoregulation*

CAR was assessed using the LAX [14], calculated every minute of the monitoring period, as a moving Pearson correlation coefficient between ICP and MAP.

8.2.3 *Visualization of patient-specific ICP and CAR dose*

The dose visualization concept was designed both for the ICP and LAX signals. As illustrated for the ICP signal in Figure 8.1, the ICP trace is shown together with the patient-specific ICP dose, overlaid on the population plot from [9]. The *current ICP dose* represents the combination of intensity and duration of ICP currently experienced by the patient. Additionally, the *6h cumulative ICP dose*, the worst ICP dose during the previous 6 hours, is represented to visualize whether the patient is recovering or worsening as compared to its state within the past 6 hours. Finally, the *cumulative ICP dose* shows the worst ICP dose the patient has experienced since admission to the ICU.

8.2.4 *Software development*

A prototype of a standalone software was developed to visualize the aforementioned concept dynamically at the patient bedside.

The software development was based on the Model-View-Controller software architectural pattern, using the Kivy Framework [17] from Python version 3.5 (Python Software Foundation, <http://www.python.org>). A scheme of the prototype software is shown in Figure 8.2. The user interacts with the controller, which communicates with the application to collect patient information from

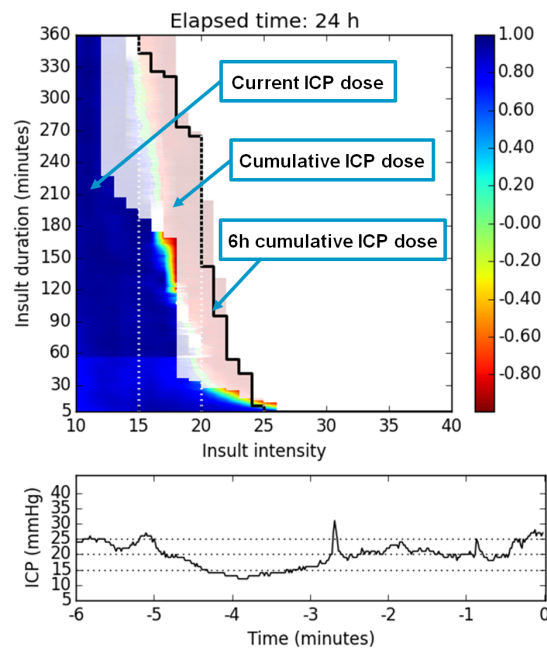


Figure 8.1 Snapshot of the patient-specific dose of ICP.

The region of ICP episodes currently experienced by the patient is plotted in opaque colors (*current ICP dose*); the cumulative ICP burden of the previous 6 hours, corresponding to the worst state experienced by the patient during the past 6 hours, is plotted with a black line (*6h cumulative ICP dose*); and the cumulative ICP burden since ICU admission is plotted using transparent colors (*cumulative ICP dose*).

Regions in blue correspond to doses associated with good 6-month neurological outcome and in red, with poor 6-month neurological outcome.

databases or the patient monitor, stores information and implements the logic. The application sends the information that needs to be visualized on the interface through the Kivy Language.

At the ICU of the University Hospitals Leuven (UZLEUVEN), patient data from the various monitors and devices are captured on a minute-by-minute basis in a Patient Data Management System (PDMS) (MetaVision, iMD-Soft, Needham, MA, USA). The PDMS contains several structured query language (SQL) databases, one of which is anonymized for educational use. This database was used to access patient data.

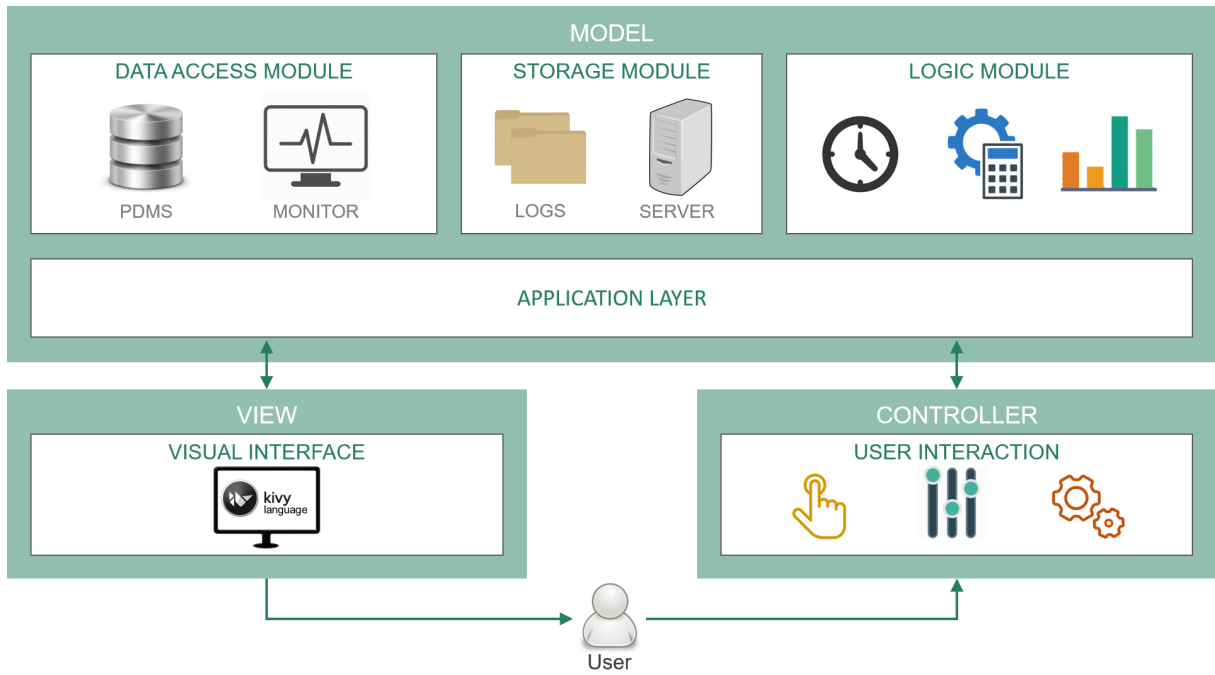


Figure 8.2 Scheme of software.

The user interacts with the controller, which communicates with the application to collect patient information from databases or the patient monitor, stores information and implements the logic. The application sends the information that needs to be visualized on the interface through the Kivy Language.

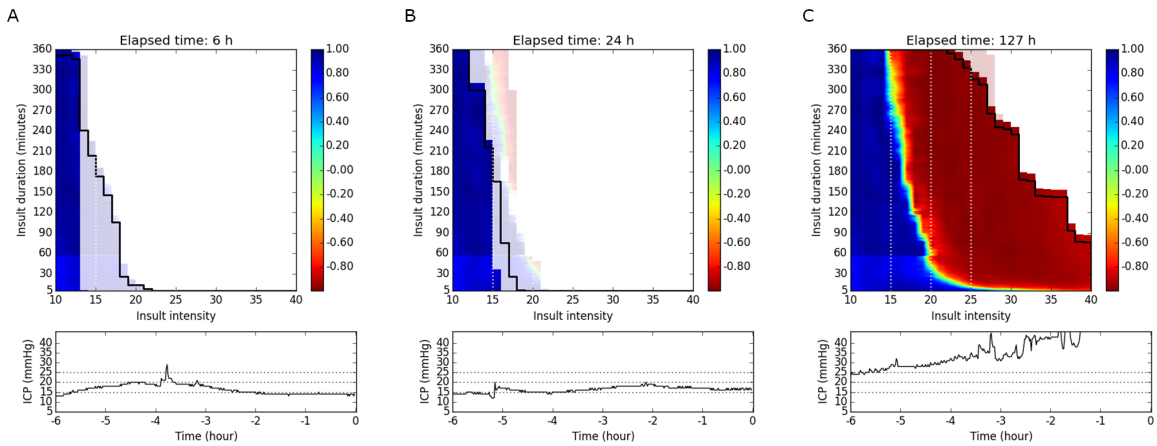


Figure 8.3 Visualization of patient-specific ICP dose in a patient with Glasgow Outcome Score (GOS) 1

A. Snapshot after 6 hours of ICP monitoring **B.** Snapshot after 24 hours of ICP monitoring **C.** Snapshot at the end of ICU stay and consequent patient death (127 hours of ICP monitoring).

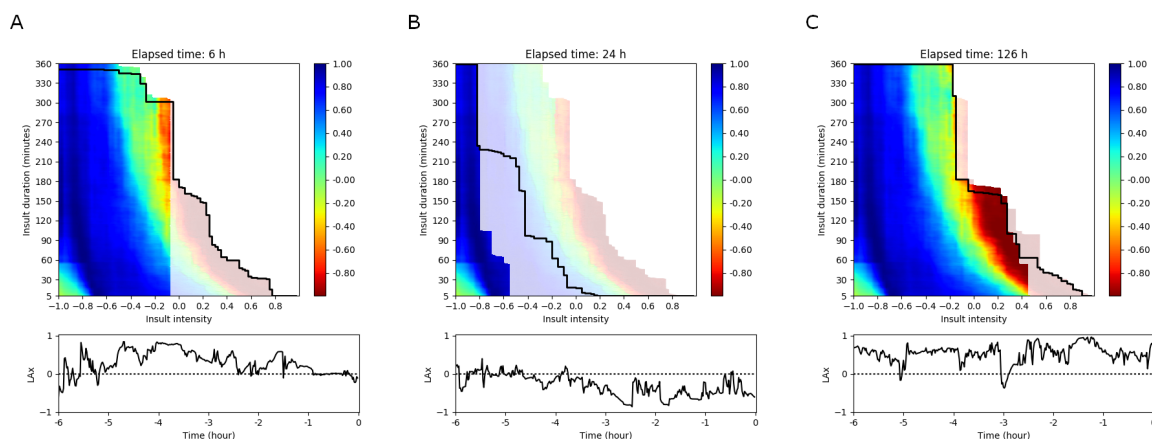


Figure 8.4 Visualization of patient-specific CAR dose in a patient with GOS 1
A. Snapshot after 6 hours of LAX monitoring **B.** Snapshot after 24 hours of LAX monitoring **C.** Snapshot at the end of ICU stay and consequent patient death (126 hours of LAX monitoring).

8.3 RESULTS

8.3.1 Visualization of patient-specific ICP and CAR dose

Figures 8.3 and 8.4 show three snapshots of the visualization of the ICP and CAR dose in a 23-year-old man with a Glasgow Coma Scale (GCS) of 11. His median [interquartile range (IQR)] ICP was 20.0 [17.0-25.0] mmHg. The patient's ICP was above 20 mmHg for 3616 minutes (47.3% of total monitoring time), and was in the red zone for the majority of the monitoring time (6420 minutes (84.0% of total monitoring time)). Regarding autoregulation, the patient's LAX was passive (below 0) 52.3% of the monitoring time and in the red zone 33.5% of the monitoring time. The patient died after 127 hours of ICP monitoring (GOS 1).

Figures 8.5 and 8.6 shows three snapshots of the visualization of the ICP and CAR dose in a 17-year-old man with a GCS of 6. His median [IQR] ICP was 16.0 [14.0-17.0] mmHg. The patient's ICP was above 20 mmHg for 874 minutes (9.6% of total monitoring time). During the majority of the monitoring time, the patient's ICP episodes remained in the blue zone. Only 170 minutes were spent in the red zone (1.9% of total monitoring time). Controversially, regarding autoregulation, the patient's LAX was passive (below 0) 78.8% of the monitoring time and in the red zone 76.7% of the monitoring time. The patient had good 6-month neurological outcome (GOS 5).

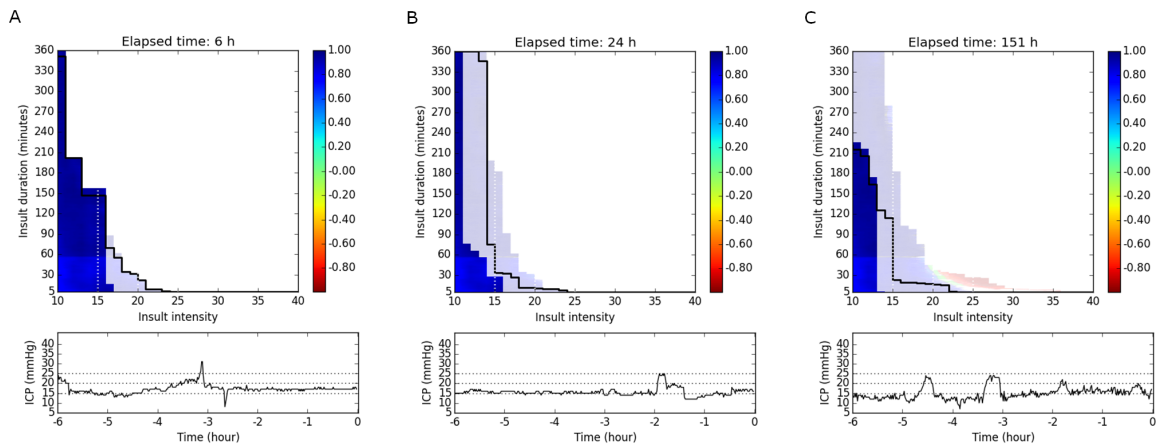


Figure 8.5 Visualization of patient-specific ICP dose in a patient with GOS 5
A. Snapshot after 6 hours of ICP monitoring **B.** Snapshot after 24 hours of ICP monitoring
C. Snapshot at the end of ICU stay (151 hours of ICP monitoring), after which the patient was discharged from the ICU.

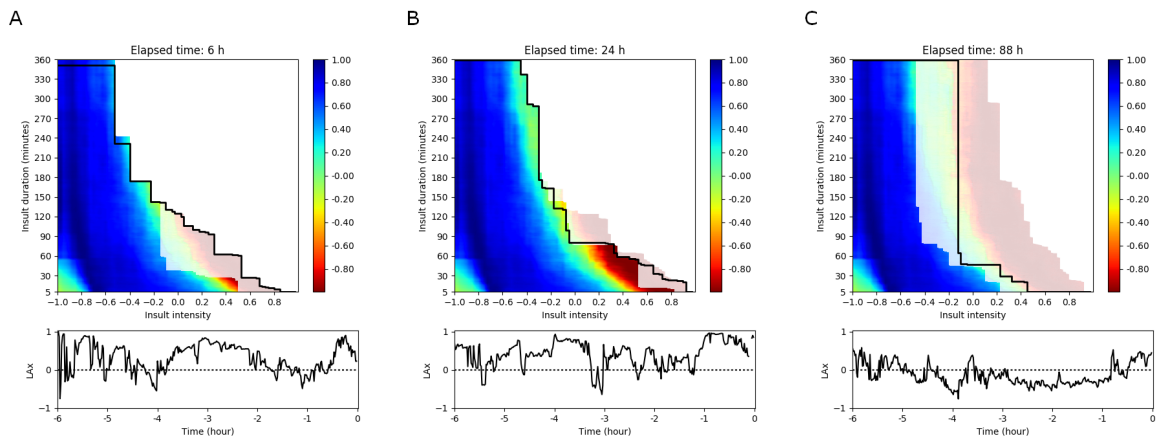


Figure 8.6 Visualization of patient-specific LAX dose in a patient with GOS 5
A. Snapshot after 6 hours of LAX monitoring **B.** Snapshot after 24 hours of LAX monitoring
C. Snapshot at the end of ICU stay (88 hours of LAX monitoring), after which the patient was discharged from the ICU.

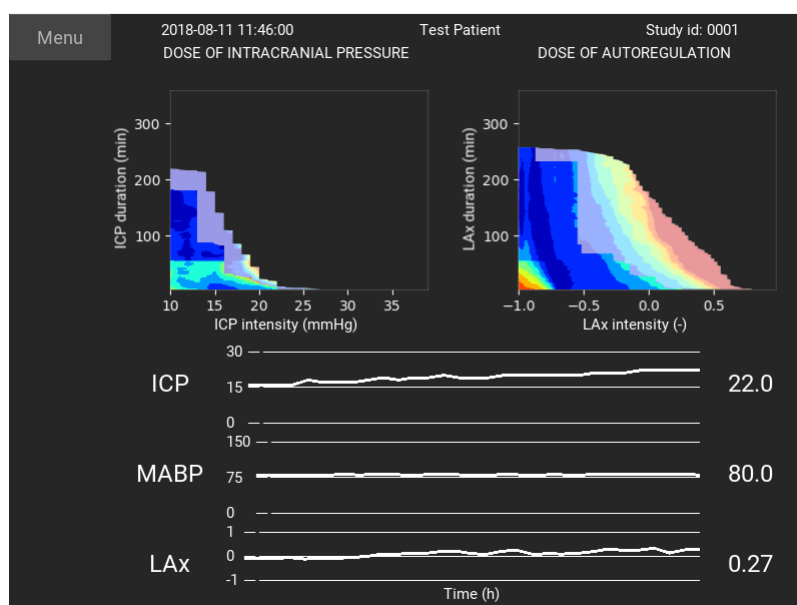


Figure 8.7 Snapshot of the main screen of the prototype software.

The clinical staff should enter the hospital number of the patient to start the software. The main screen displays the traces of ICP, MAP and LAX together with the dose of ICP and the dose of CAR.

8.3.2 Prototype software

The interface of the developed software is shown in Figure 8.7.

To start the software, the nurse enters the patient identification number. Every minute, the prototype queries the values of ICP and MAP signals with their related time stamp from the PDMS database. Following reception of data, the LAX index is calculated and signals are processed to compute the ICP and CAR doses. After calculation, the software refreshes its window to visualize the newly calculated logic.

8.4 DISCUSSION

In this study, we have demonstrated that patient-specific ICP and CAR doses can be visualized on population-based plots representing the univariate relationship between 6-month neurological outcome and ICP or CAR insults [9, 15]. Additionally, we presented a prototype of a monitor with two novel patient-specific parameters, the ICP *dose* and the CAR *dose*.

The developed prototype is ready to be tested in a prospective observational blinded study, to establish its feasibility to perform real-time analyses. The study is currently being set up and will investigate the correct functioning of

the software in real-time. Upon study completion, the software will be ready to be tested in an interventional study.

Currently, guidelines on management of patients with severe TBI recommend the administration of therapies based on generic and undifferentiated thresholds [6]. Both the dose of ICP and LAX have been associated with worse outcome in patients with severe TBI [9, 15]. Therefore, displaying patient-specific dose of ICP and CAR at the bedside could be a first step towards patient-specific management [18, 19]. They might provide information for neuro-prognostication, and therefore help family counseling. They might also help clinical decision making, such as administering a more aggressive therapy or assess the effect of current therapy [20]. A prospective randomized controlled trial (RCT) will be necessary to assess the potential clinical benefit of the software for the management of patients with severe TBI.

In the meantime, to facilitate the use of the prototype at the bedside, several challenges have to be addressed. First, an intrinsic risk from the integration of different monitored signals might arise. Indeed, different information sources might produce contradictory diagnostics as they evaluate different aspects of the health-state of the patient. Although at first, this might seem an unexpected situation; this is actually the reality that a clinician must confront daily. Additionally, software usability related to the design but also to the clinical requirements has to be examined. For that purpose, usability evaluations should be conducted with the medical team during feedback sessions in order to identify potential flaws or different clinical needs. For instance, in a long-stay patient, it might be more appropriate to show the cumulative dose of the previous 24 hours instead since ICU admission. After feedback, the prototype should be redesigned iteratively to ensure its effectiveness, efficiency and being satisfying to use. For that purpose, tests should be conducted by an interdisciplinary team of potential users including neuro-surgeons, neurosurgical trainee, intensivists, intensivist trainee, and nurses.

8.5 CONCLUSION

In this study, we have developed a prototype of a bedside monitor to visualize the current and cumulative ICP and CAR doses in patients with severe TBI. Opportunities for using these metrics as decision-support for the management of patients with severe TBI remains to be investigated prospectively.

GRANT SUPPORT AND CONFLICT OF INTEREST

M.Flechet receives funding from the Fonds Wetenschappelijk Onderzoek (FWO) as a PhD. fellow (11Y1118N). G.Meyfroidt receives funding from FWO as senior clinical investigator (1843118N). G. Van den Berghe, via the University of Leuven, receives structural research financing via the Methusalem program, funded by the Flemish Government (METHo8/07) and holds an European Research Council (ERC) Advanced Grant (AdvG-2012-321670).

The authors declare that they have no conflict of interest.

ACKNOWLEDGMENTS AND PERSONAL CONTRIBUTION

All analysis presented in this chapter have been performed by Marine Flechet.

- Study concept and design: Güiza, Meyfroidt, Depreitere, **Flechet**
- Acquisition, analysis, or interpretation of data: **Flechet**, Güiza, Meyfroidt, Depreitere, Brain-IT centers
- Drafting of the chapter: **Flechet**
- Critical revision of the chapter: **All authors**
- Creation of figures: **Flechet**
- Development of software: **Flechet**
- Study supervision: Güiza

The authors would like to thank Dominiek Cotteem, Fredrik Hermans, and Bram Vercammen, for their logistic help.

BIBLIOGRAPHY

- [1] A. A. Hyder, C. A. Wunderlich, P. Puvanachandra, G. Gururaj, and O. C. Kobusingye. The impact of traumatic brain injuries: a global perspective. *NeuroRehabilitation* 22,5 (2007), pp. 341–353.
- [2] W. Peeters et al. Epidemiology of traumatic brain injury in Europe. *Acta Neurochir. (Wien)*. 157,10 (Oct. 2015), pp. 1683–1696.
- [3] K. N. Corps, T. L. Roth, and D. B. McGavern. Inflammation and Neuroprotection in Traumatic Brain Injury. *JAMA Neurol* (Jan. 2015).
- [4] N. Stocchetti, T. Zoerle, and M. Carbonara. Intracranial pressure management in patients with traumatic brain injury. *Curr. Opin. Crit. Care* 23,2 (2017), pp. 110–114.
- [5] C. Lazaridis et al. Patient-specific thresholds of intracranial pressure in severe traumatic brain injury. *J. Neurosurg.* 120,4 (2014), pp. 893–900.
- [6] Brain Trauma Foundation. Guidelines for the Management of Severe Traumatic Brain Injury 4th Edition. September (2016).
- [7] Brain Trauma Foundation. Guidelines for the Management of Severe Traumatic Brain Injury 3rd Edition. *J. Neurotrauma* 24,212 (2007), pp. 1–106.
- [8] A. Vik et al. Relationship of “dose” of intracranial hypertension to outcome in severe traumatic brain injury. *J. Neurosurg.* 109,4 (Oct. 2008), pp. 678–84.
- [9] F. Güiza et al. Visualizing the pressure and time burden of intracranial hypertension in adult and paediatric traumatic brain injury. *Intensive Care Med.* 41,6 (Apr. 2015), pp. 1067–1076.
- [10] G. J. Bouma, J. P. Muizelaar, K. Bandoh, and A. Marmarou. Blood pressure and intracranial pressure-volume dynamics in severe head injury: relationship with cerebral blood flow. *J. Neurosurg.* 77,1 (July 1992), pp. 15–19.
- [11] G. E. Svirid, R. Aaslid, C. M. Douville, A. Moore, and D. W. Newell. Time course for autoregulation recovery following severe traumatic brain injury. *J. Neurosurg.* 111,4 (Oct. 2009), pp. 695–700.
- [12] L. Rivera-Lara, A. Zorrilla-Vaca, R. G. Geocadin, R. J. Healy, W. Ziai, and M. A. Mirski. Cerebral Autoregulation-oriented Therapy at the Bedside. *Anesthesiology* 126,6 (2017), pp. 1187–1199.

- [13] M. Czosnyka, P. Smielewski, P. Kirkpatrick, R. J. Laing, D. Menon, and J. D. Pickard. Continuous Assessment of the Cerebral Vasomotor Reactivity in head injury. *Neurosurgery* 41, July (1997), pp. 11–19.
- [14] B. Depreitere et al. Pressure autoregulation monitoring and cerebral perfusion pressure target recommendation in patients with severe traumatic brain injury based on minute-by-minute monitoring data. *J. Neurosurg.* 120,6 (June 2014), pp. 1451–1457.
- [15] M. Flechet et al. Visualizing Cerebrovascular Autoregulation Insults and Their Association with Outcome in Adult and Paediatric Traumatic Brain Injury. *Intracranial Press. Neuromonitoring XVI*. Ed. by T. Heldt. Cham: Springer International Publishing, 2018, pp. 291–295.
- [16] I. Piper et al. The BrainIT group: concept and core dataset definition. *Acta Neurochir. (Wien)*. 145,8 (Jan. 2003), pp. 615–629.
- [17] M. Virbel, T. E. Hansen, and O. Lobunets. Kivy-A Framework for Rapid Creation of Innovative User Interfaces. *Mensch Comput. Work.* 2011, pp. 69–73.
- [18] C. Lazaridis, C. G. Rusin, and C. S. Robertson. Secondary brain injury: Predicting and preventing insults. *Neuropharmacology* (June 2018), pp. 1–8.
- [19] U. Johnson, A. Lewén, E. Ronne-Engström, T. Howells, and P. Enblad. Should the Neurointensive Care Management of Traumatic Brain Injury Patients be Individualized According to Autoregulation Status and Injury Subtype? *Neurocrit. Care* 21,2 (2014), pp. 259–265.
- [20] N. Stocchetti and A. I. Maas. Traumatic intracranial hypertension. *N. Engl. J. Med.* 370,22 (May 2014), pp. 2121–2130.

DISCUSSION

Critically ill patients are admitted to the intensive care unit (ICU) with acute life-threatening conditions. The diagnostic-therapeutic cycle in these patients is short as their clinical situation may vary rapidly. It is therefore of great interest to detect those patients most vulnerable to specific organ deterioration as early as possible. Prediction plays a crucial role in the management of critically ill patients.

Due to the massive quantities of data generated at the patient bedside, clinicians are facing the problem of information overload. It is virtually impossible to process all the data from the various sources at the same time and over time. Big data analytics, a term that encompasses the application of data-driven advanced analytics such as machine learning techniques, are able to process large amount of data to perform pattern recognition, predictions, or generate data-driven hypothesis.

In this thesis, we aimed to apply big data analytics to gain novel insights in critical illness and to develop and validate decision support applications to help clinicians in the management of critically ill patients. We focused on three objectives, with potential to improve patient care, namely 1) early detection of acute kidney injury, 2) assessment of the utility of near-infrared-based cerebral oximetry, and 3) detection and refinement of secondary brain injuries.

9.1 EARLY DETECTION OF ACUTE KIDNEY INJURY

9.1.1 *Main findings*

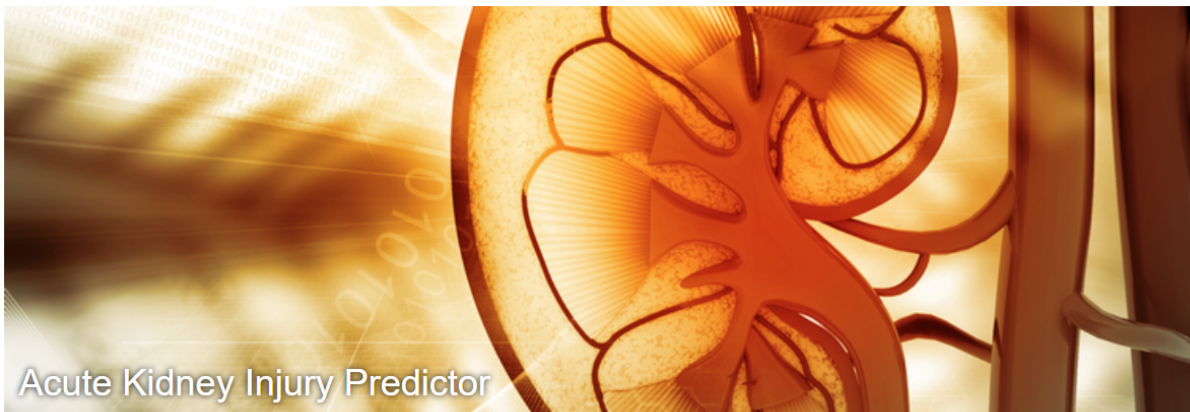
Acute kidney injury (AKI) affects approximately 40% of critically ill patients and is associated with increased risks of morbidity and mortality, and with high financial costs [1–4]. AKI is diagnosed according to an increase in serum creatinine from its baseline value or a low rate of urine output [5]. Early diagnosis of AKI remains a major clinical challenge in critical care [6].

Its imprecise early identification could partially explain why the search for strategies and interventions to mitigate its course has been unsuccessful [7–12].

In **Chapter 3**, we developed and validated machine-learning-based prediction models (jointly referred to as AKIpredictor) for the development of AKI during the first week of ICU stay in a general ICU population. Four models were developed to mimic the sequential availability of clinical data at the bedside: (1) before and (2) upon ICU admission, (3) on the morning of the first day, and (4) after 24 hours of ICU stay. All models were based on routinely collected patient information, and the fourth model added information extracted from time series of monitoring data. The models were developed both for the prediction of any stage and of the most severe stages of AKI. The models' performance was good to very good, and gradually improved with data availability. Additionally, this study was the largest study of the most researched AKI biomarker, which, when measured at ICU admission, was outperformed by the AKIpredictor. The models were made available online via an open-access web application (Figure 9.1, www.akipredictor.com).

Although this study is a major step forward in the global aim of recognizing AKI early, several limitations limit its broad application. First, AKI was only diagnosed based on the serum creatinine criterion, as hourly measurements of urine output were not prospectively collected. However, a high proportion of patients develop AKI according to the urine output criterion only [13]. Second, prospective external validation is required to assess its generalizability and adoption outside the research area. Indeed, it is unclear how such models should be used in clinical practice and how they compare to bedside physicians.

To answer this question, in **Chapter 4**, we conducted a prospective observational blinded study to compare the predictive performance of AKIpredictor and physicians for the development of the most severe stages of AKI during the first week of ICU stay. Interestingly, AKIpredictor retained good predictive performance when AKI was classified according to both urine output and serum creatinine criteria. Physicians predicted AKI with good discrimination but with an over-estimated risk. Although achieving a slightly lower discrimination, AKIpredictor showed improved calibration and net benefit. This suggests an added value to physicians' predictions, in particular for junior physicians, when physicians were not confident in their predictions, or when it is of interest to gather predictions at fixed time points such as inclusion in clinical trials.



Will your patient develop AKI?

Acute Kidney Injury (AKI) occurs frequently in critically ill patients. AKI is associated with increased morbidity and mortality, and a high financial cost. The Kidney Disease/Improving Global Outcomes (KDIGO) working group has defined and classified AKI in three stages of increasing severity, according to serum creatinine levels and urine output. Although no effective treatment currently exists that could attenuate the course of AKI, early prediction of AKI to detect the patients at risk could be a first step in the discovery and assessment of new therapies. For that purpose, we propose an AKI predictor for patients admitted to an intensive care unit (ICU). Using routinely available patient characteristics, the AKI predictor is able to predict if an adult patient will develop any stage of AKI (as defined by the KDIGO serum creatinine criteria) during the first week of ICU stay.

AKI predictor

Critically ill patient information

Pre-admission information	ICU admission information	Day 1 information
<p>Age *</p> <input type="text" value="50"/> <p>18-99 years</p>	<p>Blood glucose upon ICU admission *</p> <input type="text" value="130"/> <p>25-750 mg/dl</p>	<p>Serum creatinine measured on day 1 *</p> <input type="text"/> <p>0-7 mg/dl</p>
<p>Baseline serum creatinine *</p> <input type="text" value="0.5"/> <p>0-7 mg/dl</p>	<p>Suspected sepsis upon ICU admission *</p> <input type="text" value="No"/>	<p>Apache II score on day 1 *</p> <input type="text"/> <p>3-50</p>
<p>Patient is diabetic *</p> <input type="text" value="No"/>	<p>Hemodynamic support upon ICU admission *</p> <input type="text" value="Pharmacological"/>	<p>Maximum lactate measured on day 1 *</p> <input type="text"/> <p>0-24 mg/dl</p>
<p>Planned/Unplanned admission *</p> <input type="text" value="Unplanned admission"/>	<p>Compute prediction</p>	<p>Bilirubin measured on day 1 *</p> <input type="text"/> <p>0-23.5 mg/dl</p>
<p>Surgical/Medical category *</p> <input type="text" value="Cardiovascular surgery (non transplant)"/>	<p>Empty form</p>	<p>Hours in ICU *</p> <input type="text"/> <p>0-24 hours</p>
<p>Add admission information</p>		<p>Compute prediction</p>

Figure 9.1 Snapshot of AKIPredictor website, www.akipredictor.com.

9.1.2 *Current impact of research*

Risk assessment and early detection of AKI have been listed as one of the main steps to improve the outcome of patients with AKI [14]. The open-access website provides opportunities to use the AKIpredictor for research, such as comparison with new AKI biomarkers or selection of high risk patients, and for clinical purposes worldwide. Between its release in January 2017 and August 2018, the website was accessed more than 3200 times by more than 2200 users from 95 countries (Figure 9.2). The majority of the users accessed the website from the United States, France, Mexico, Belgium and the United Kingdom. Approximately 15% of the users access the website recurrently.

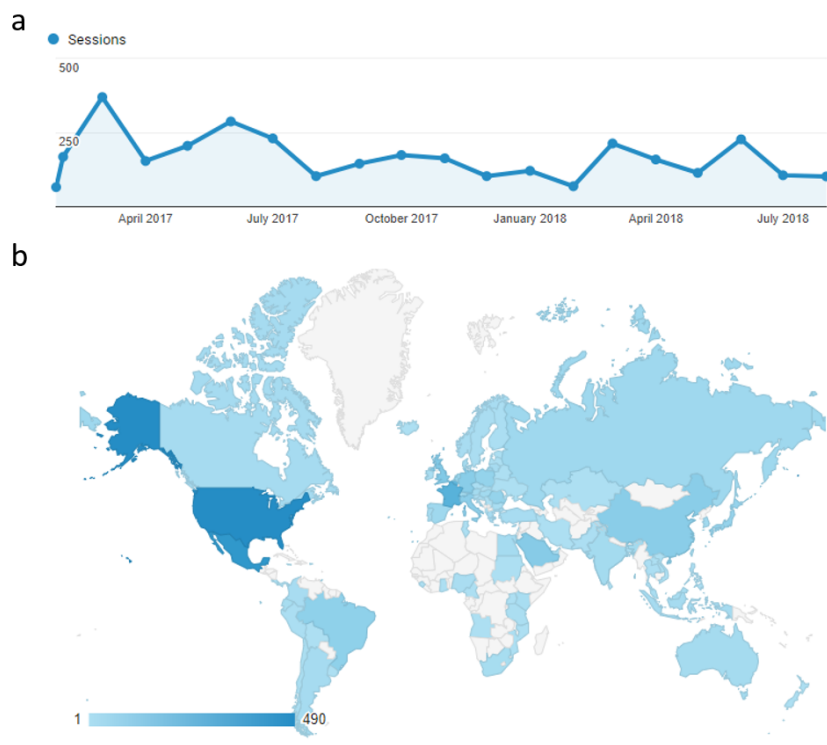


Figure 9.2 Usage statistics of the AKIPredictor website.

a. Monthly number of sessions between January 2017 and August 2018, **b.** Geographic coverage: color saturation according to number of sessions per country.

The prospective study, although of limited sample size, strengthens the generalizability of AKIpredictor to predict AKI with good performance in similar populations. Our findings suggest a potential for overall improvement of care with the concurrent use of physicians' expertise and accurate prediction models like the AKIpredictor. To the best of our knowledge, this study is the

first to report performance of physicians to predict AKI, laying ground for benchmarking opportunities.

Previous research has focused on biomarkers for early detection of AKI but their performance has not been compared against physicians. Nevertheless, biomarkers and prediction models such as AKIpredictor have several advantages. Physicians are not always available at the bedside and we have shown that clinical expertise and confidence about their prediction affected physicians' predictive performance. Biomarkers and prediction models, on the other hand, would offer a systematic and similar stratification within clinical teams and across centers. However, biomarker assessment has an additional cost. All these aspects have to be considered to make the optimal choice for AKI early detection.

To be clinically useful, biomarkers should add value beyond routinely collected data. Investigating new AKI biomarkers is often started in small sample-sized studies. The AKIpredictor web-application could be used as a benchmark in such studies to assess the benefit of biomarkers as compared to available clinical data.

9.1.3 *Future perspectives*

The current AKIpredictor web-application requires manual data entry. To promote the use of such models, automatic data extraction from electronic health record (EHR) could improve the user friendliness. Additionally, automatic data extraction is required to implement the model based on patient monitoring.

Although the prospective validation of the AKIpredictor (discussed in Chapter 4) strengthens its generalizability, large multi-center and multi-continental prospective validation is required to ensure validity worldwide. Such validation could highlight a potential need for recalibration or identify subpopulations for which the model is not optimal. Such validation is warranted in particular in North and South America, where there is an apparent interest in the predictor but where the population might be different as compared to Belgian critically ill patients.

A potential next step towards personalized medicine would combine such clinical prediction model with biomarker assessment (Figure 9.3). Decision curve analyses showed that AKIpredictor could be used to identify the high-risk patients who would benefit most from incurring the additional cost of a biomarker evaluation, which would reduce unnecessary and expensive testing. Additional clinical benefit of this combination has to be examined prospectively. Potential promising biomarker candidates are the urinary cell cycle arrest

biomarkers, tissue inhibitor of metalloproteinase-2 (TIMP-2), and insulin-like growth factor-binding protein 7 (IGFBP-7). TIMP-2 and IGFBP-7 could be specific early markers of AKI, as they are expressed to prevent cells from dividing in case of cellular stress. They achieved good performance in predicting AKI across different populations of critically ill [15–17] and the combination of the two biomarkers is available on the market as a diagnostic test (NephroCheck, Astute Medical, San Diego, CA, USA). However, the exact role of this biomarker in clinical practice is still unclear [18].

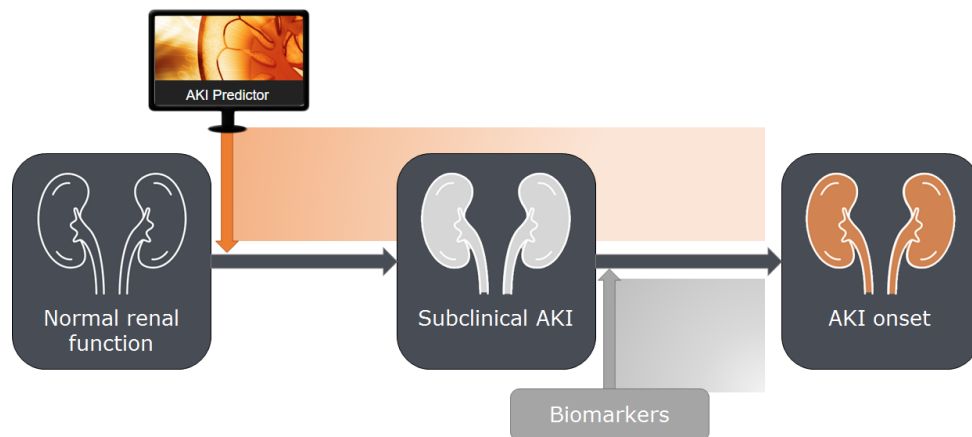


Figure 9.3 Potential diagnostic tool for early detection of AKI [19]. Biomarkers could be measured in high risk patients stratified by AKI predictor.

Finally, in order to investigate the impact of using the AKI predictor at the bedside, prospective randomized studies are warranted, focusing on patient outcome such as mortality or long-term renal function. Upon success, this would open up potential valorization avenues such as licensing the algorithm to companies involved in biomarkers or drug development or in electronic health records.

9.2 ASSESSMENT OF THE UTILITY OF NIRS-BASED CEREBRAL OXIMETRY

9.2.1 *Main findings*

Cerebral oximetry by near-infrared spectroscopy (NIRS) is used frequently in critically ill children to measure cerebral tissue oxygen saturation (SCTO₂) [20, 21] but guidelines on its use for decision making are lacking. There is currently no consensus on which critical NIRS thresholds could guide patient care, trigger interventions, or be used for prognostication [22–24]. Given the relatively high

cost of the monitoring sensors, it is crucial to determine specific settings in which NIRS monitoring could improve patient care.

With the aim of gaining insights into the usefulness of cerebral NIRS oximetry, a prospective observational blinded study was conducted at the pediatric intensive care unit (PICU) of the University Hospitals Leuven (UZLEUVEN) between October 2012 and November 2015. Critically ill children and infants with congenital heart disease, younger than 12 years old, were monitored with the FORESIGHT cerebral oximeter (CAS Medical Systems, Branford, CT) from PICU admission until they were weaned off mechanical ventilation. Physicians were blinded to the cerebral oximeter signal.

In **Chapter 5**, we present results of the observational study. We assessed whether $SctO_2$ predicts ICU and hospital outcomes, in particular PICU length of stay, duration of invasive mechanical ventilation, and mortality in critically ill children after pediatric cardiac surgery. We calculated predictors on the first 24 hours of $SctO_2$ monitoring, using advanced time-series analysis. Of all the investigated predictors and after correction for several confounders, both a measure of variability (standard deviation of a smoothed $SctO_2$) and the dose of desaturation (depth and duration below the 50% saturation threshold) remained significant predictors. Hence, we found that an increased $SctO_2$ variability and an increased dose of desaturation were associated with longer PICU and hospital stays and with longer duration of mechanical ventilation after pediatric cardiac surgery. Furthermore, our study highlighted the difference in $SctO_2$ between patients with cyanotic versus acyanotic heart defect. Hence, we recommend that future studies should be adequately powered to analyze these populations separately.

In pediatric perioperative settings, the use of invasive catheters is not feasible or available. Therefore, cerebral NIRS oximetry is also used as a hemodynamic monitor for early recognition of an inadequate global oxygen supply/demand relationship, and to adapt interventions to minimize the risk of secondary organ dysfunction [25]. Hemodynamic instabilities contribute significantly to the development of AKI, which is very common after pediatric cardiac surgery and associated with adverse outcomes [26, 27]. Therefore, in **Chapter 6**, we investigated the ability of cerebral NIRS oximetry to predict severe AKI after pediatric cardiac surgery and assessed its additional predictive value to routinely collected data. Overall, we found that the data currently displayed in cerebral oximeters have only fair discriminability for 6-hour-ahead prediction of severe AKI. Retrospective calculation of NIRS variability (measured using the root mean square of successive differences) achieved better performance, although not sufficient to outperform a clinical model based on routinely collected patient information. However, combining NIRS

variability with the clinical model significantly improved predictive capabilities, suggesting that the NIRS signal carries independent relevant information in addition to routinely collected clinical data.

9.2.2 *Current impact of research*

Current cerebral oximeters only display the value of cerebral tissue oxygen saturation and its trace over the past hours at the bedside. We did not find any association between S_{ctO_2} value and the investigated patient outcomes. Some versions of cerebral oximeters allows the user to calculate the desaturation dose below a user-specific threshold (Root with O₃ Regional Oximetry, Masimo, CA, USA). However, studies showing the benefit of using such metric to improve patient outcomes are lacking. The two analysis performed in this thesis suggest that NIRS variability is associated with patient outcome and might provide additional information to routinely collected patient information. As this study was observational, our findings cannot support any conclusions regarding the postoperative management of critically ill children after cardiac surgery. Future studies are required to identify whether the implementation of NIRS variability at the bedside could help drive therapeutic interventions and improve patient outcome.

9.2.3 *Future perspectives*

The physiologic meaning behind NIRS variability is unclear. It has been hypothesized that NIRS oximetry could be used to assess cerebrovascular autoregulation (11, 24). Future studies are necessary to assess whether NIRS variability is related to impaired cerebral autoregulation. Additionally, studies are warranted to assess how caregivers could improve or modify NIRS variability [28]. These steps will be necessary in order to design a NIRS-based intervention to assess whether corrective interventions can effectively reverse the desaturation or the decreased or increased variability, and whether these interventions can lead to improved patient outcome.

Another future perspective is to analyze whether cerebral NIRS monitoring could add value for other outcomes than the ones already investigated. During ICU stay, critically ill infants and children are treated for life-threatening conditions during crucial phases of development which makes them particularly vulnerable for impaired longer-term neurocognitive functioning [29–31]. Indeed, 4 years after admission to an ICU, children who had been critically ill revealed a severe intelligence deficit of 15 points on the intelligence quotient (IQ) and impaired performance on other tests for neurocognitive

development as compared with healthy siblings and matched control children [32]. To date, the exact cause(s) of the longer-term neurocognitive deficit observed in children who have been critically ill remain unclear. The most likely culprit is low brain perfusion due to low cardiac output and/or low blood pressure, which would deprive the brain of the normal supply of nutrients and oxygen. Cerebral NIRS monitoring could be an early indicator of critical changes in brain perfusion.

For that purpose, children involved in the current observational study have been seen at follow-up 2 years after ICU stay. A battery of tests, similar to the ones performed in [33], were performed to assess children neurocognitive function. The neurocognitive testing has been finalized recently, and the analysis will be performed in a near future.

9.3 DETECTION AND REFINEMENT OF SECONDARY BRAIN INJURIES

9.3.1 *Main findings*

Traumatic brain injury (TBI) is one of the most important health care problems worldwide, affecting approximately 2.5 million people each year and yielding 75 000 deaths [34]. In the United States only, the annual burden of TBI is estimated more than \$75 billion [35]. TBI is caused by the application of mechanical force to the head, inducing different types of brain damage, such as brain parenchyma contusions, diffuse lesions to axonal fibers due to stretching and tearing, and epidural or subdural hematomas [36]. In addition, TBI may disrupt the integrity of the blood-brain barrier and damage the neurovascular unit.

In the hours and days following the initial traumatic insult (or *primary injury*), destructive and self-propagating biological changes in the brain can lead to subsequent additional damage (or *secondary injury*) [37]. Prevention and treatment of secondary injury is the main goal of neuro-critical care in patients with TBI. Following the failure of several randomized controlled trials to improve the outcome of patients with severe TBI [38–41], there is an urgent need for better definitions of secondary insults [42–44].

A subtype of secondary brain injuries is elevated intracranial pressure (ICP), which has been shown to be associated with worse outcome [45]. Based on this observational study, international guidelines recommended ICP lowering therapies at a threshold of 22 mmHg [46]. Following the intuition that universal thresholds are likely suboptimal and that duration of elevated ICP plays a role [47], Güiza et al [48] plotted the association between neurological outcome and the dose ICP, defined as a combination of intensity and duration. Two

distinct regions associated with good or poor neurological outcomes emerged from the plot, suggesting the importance of the time-intensity relationship of ICP for patient outcome. Additionally, these two regions were affected by the cerebrovascular pressure autoregulation (CAR) capability of the patient.

Another promising avenue in the refinement of secondary brain injury is the assessment of CAR. Currently, international guidelines state that evidence is insufficient to support a recommendation regarding CAR-based management in TBI patients [46]. In **Chapter 7**, we investigated whether episodes of deficient CAR are associated with worse clinical outcomes, and can be defined as an additional type of acute secondary brain injury. Using a similar methodology to Güiza et al [48], we investigated the association between 6-month neurological outcome and the dose of cerebral autoregulation deficiency, in a large database of patients with severe TBI. The resulting color-coded plots showed a time-intensity association with outcome for CAR insults in adult and pediatric patients with severe TBI. For both cohorts, two distinct regions associated with good or poor outcomes were found, separated by an exponential transition curve. This indicates that CAR is disturbed in a dynamic manner, such that duration and intensity play a role in the determination of a zone associated with better neurological outcome.

The two population-based color-coded plots open up avenues for novel ways of managing patients with severe TBI. Indeed, the course of individual patients could be visualized both in terms of ICP and CAR in the population-based plots, and patient-specific thresholds can be inferred from the plots to adapt therapy. Additionally, continuous monitoring of CAR is challenging and usually based on indices reflecting the correlation between ICP and blood pressure signals captured at high frequency [49, 50]. As such, the Low Frequency Autoregulation Index (LAX) is a promising index developed by our research group [50], which has never been evaluated prospectively. For that purpose, in **Chapter 8**, we developed the prototype of a bedside monitor. The prototype is able to calculate and display the LAX as a measure of CAR, and to visualize patient-specific secondary brain injuries in terms of ICP and of CAR, in a continuous way (Figure 9.4).

9.3.2 *Current impact of research*

The proof-of-concept has been presented at several neuro-related conferences and meetings (ICP2016, NCS2016, BSN2017) and has raised the audience interest. Potential valorization avenues have been considered, and licensing or the creation of a spin-off company to promote the software at a larger scale will be examined.

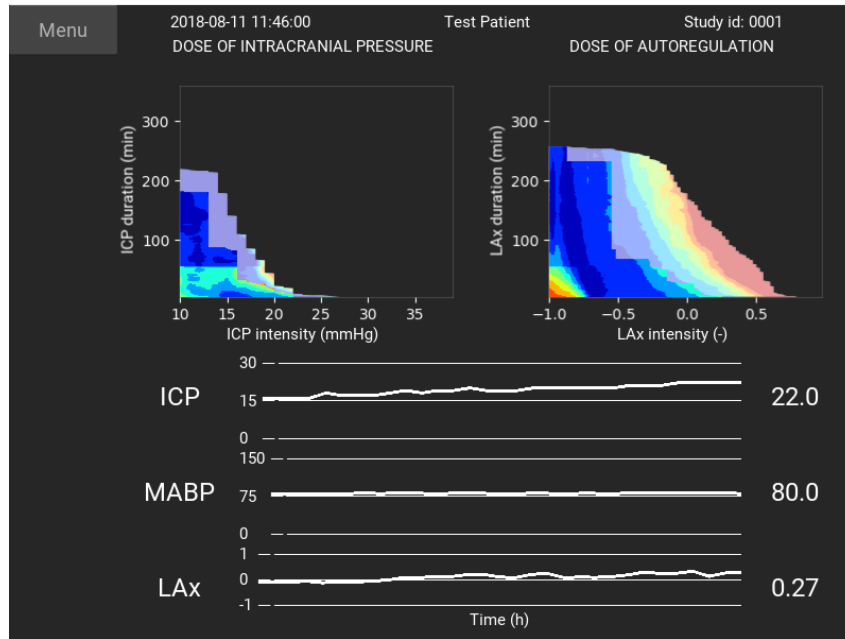


Figure 9.4 Snapshot of the prototype interface.

9.3.3 Future perspectives

The global framework International Initiative for Traumatic Brain Injury Research (INTBIR) has projects to advance TBI research in Europe, the United States and Canada [51]. Under the umbrella of INTBIR, CENTER-TBI is a large European project that aims to improve the care of patients with TBI [52]. CENTER-TBI is currently recruiting 5400 patients from 60 centers (including UZLEUVEN) across 20 countries, of which 1800 will be critically ill. The creation of this large and unique database will offer new opportunities to validate the findings presented in this thesis.

A prospective observational blinded study to evaluate the prototype is currently being set up. The study will investigate the correct functioning of the software in real-time. Upon study completion, the software will be ready to be tested in an interventional study. Such study will evaluate whether the outcomes of patients with severe TBI can be improved by using the prototype for patient management. To prevent lack of conclusion [53], a randomized controlled trial with three arms could be considered where the first two arms would be managed according to standard of care, respectively blinded and non-blinded to the software, and the last arm would be managed according to a particular carefully-designed therapeutic protocol.

9.4 IMPLEMENTATION DETAILS

This final section focuses on the specific implementation details that have been used during this thesis. Programming was performed using Python version 2.7 and 3.5, from the Anaconda suite, using PyCharm as integrated development environment (IDE). Python is a free, open-source, widely used, and high level programming language. Its applications are diverse and include web development, statistics, software development, graphical user interface development and system administration. Its diversity, open-access, ease of implementation and large community motivated the choice of Python as programming language.

Statistical analysis were performed using Numpy, Pandas and Statsmodel Python libraries. Machine learning models were developed using the scikit-learn Python library [54], and Weka 3.7 [55]. Weka provides a free, open-source machine learning toolbox via a graphical interface, hence not requiring any programming knowledge. The AKIpredictor website was developed using the Django 1.8 framework [56]. Django tackles the majority of the web interface to encourage rapid development and design. Its use was motivated by the necessity of a sklearn- and python-compatible website to run the AKIpredictor. Hosting of the website is tackled by Webfaction on a Nuxit-registered domain. The prototype software was developed using the Kivy framework [57]. The use of Kivy was motivated by its free, open-access, simplicity of programming, and cross platform compatibilities aspects.

Github was used as a version control software for on-going projects, to track changes in codes and documents. Finalized or submitted projects were uploaded to Gitlab where I created a common repository for our laboratory (<https://gitlab.com/licm>), to promote code sharing between PhD students and to ensure tracking and transferability of finalized projects and of future directions for submitted projects.

Large databases of clinical data were managed using the structured query language (SQL) in the MYSQL and MSSQL (for interface with hospital system) management system.

9.5 OVERALL CHALLENGES

Several challenges have been encountered during this thesis. First and foremost, data preprocessing is a necessary step before any analysis. Often underestimated, data preprocessing was one the most important and time-consuming step in this thesis. It involved artefacts removal and identification of missing, incorrect, or discrepancies within the data. It was necessary to acquire

various medical knowledge on each of the studied topics, and it involved many discussions with clinical and IT experts.

Second, when designing prediction models, a right balance had to be found between using a certain number of features to improve model performance but prevent overfitting. This step was performed with feature selection, which required an in-depth review of literature to identify predictors of interest in the studied fields. To promote the applicability of the developed models, only routinely collected data were used in the models. Afterwards, the optimal learning algorithm had to be identified. The use of random forest in Chapter 3 was motivated by its ease of interpretation and robustness. In Chapters 5 and 6, simple logistic regression were used due to the limited sample size of the study. Particular care was given towards findings generalizability using interval validation techniques via bootstrapping, and external validation when sample size was sufficient.

Finally, the translation of the developed analytics towards patient bedside should not be underestimated. First, we had to identify the optimal translation channel. An open-access website was designed for the AKI predictor as the models were sparse and based on routinely collected information. Although the last model involved monitoring data and therefore, software implementation within the Patient Data Management System (PDMS) would be necessary to allow its use. A software was designed for the visualization of secondary brain injuries, as it requires processing of monitoring patient data. To promote clinical use, we had to think about software usability, which will be improved using usability test by end users in a near future. Attaining the required knowledge for web and software design was a serious challenge.

9.6 GENERAL CONCLUSION

In this thesis, we successfully applied big data analytics to gain novel insights into critical care medicine, through an extensive collaboration between clinical experts and engineers. First, we developed and validated machine-learning-based prediction models, which outperformed one of the most studied biomarkers, for early detection of acute kidney injury in critically ill adults. We made these models accessible online via a web-application that is currently used worldwide. To understand their clinical utility, the models were prospectively compared to physicians. Not only did the models maintain their original performance, but they also achieved an increased clinical benefit and calibration compared to ICU physicians. Second, we investigated novel indices that highlight the potential benefit of a different application of near-infrared-based cerebral oximetry, for prognostication of ICU and hospital outcomes,

and for the prediction of acute kidney injury, in critically ill children after cardiac surgery. Finally, in patients with severe traumatic brain injury, we showed that the dose of impaired cerebral autoregulation is associated with adverse outcomes. To translate this concept to the bedside, we developed the prototype of a monitor that visualizes dynamically patient-specific secondary brain injuries in terms of intracranial pressure and cerebral autoregulation. This prototype lays the ground for a potential adaptation of the management of patients with severe traumatic brain injury. Overall, the findings obtained in the three investigated domains of critical illness open up avenues for prospective interventional clinical studies to assess whether the developed tools or the novel insights can improve process of care and patient outcome.

BIBLIOGRAPHY

- [1] R. Bellomo, J. A. Kellum, and C. Ronco. Acute kidney injury. *Lancet* 380,9843 (Aug. 2012), pp. 756–766.
- [2] J. Gunst et al. Impact of Early Parenteral Nutrition on Metabolism and Kidney Injury. *J. Am. Soc. Nephrol.* 24,6 (June 2013), pp. 995–1005.
- [3] M. Joannidis et al. Acute kidney injury in critically ill patients classified by AKIN versus RIFLE using the SAPS 3 database. *Intensive Care Med.* 35,10 (2009), pp. 1692–1702.
- [4] S. Nisula et al. 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, (2013), pp. 420–8.
- [5] Kidney Disease: Improving Global Outcomes (KDIGO) Acute Kidney Injury Work Group. KDIGO Clinical Practice Guideline for Acute Kidney Injury. *Kidney Int. Suppl.* 2,2 (June 2012), pp. 1–138.
- [6] S. M. Sutherland et al. Utilizing electronic health records to predict acute kidney injury risk and outcomes: workgroup statements from the 15th ADQI Consensus Conference. *Can. J. Kidney Heal. Dis.* 3,1 (Dec. 2016), p. 11.
- [7] L. S. Chawla, P. W. Eggers, R. A. Star, and P. L. Kimmel. Acute Kidney Injury and Chronic Kidney Disease as Interconnected Syndromes. *N. Engl. J. Med.* 371,1 (July 2014), pp. 58–66.
- [8] M. Ostermann and M. Joannidis. Biomarkers for AKI improve clinical practice: no. *Intensive Care Med.* 41,4 (Apr. 2015), pp. 618–622.
- [9] F. T. Billings et al. High-Dose Perioperative Atorvastatin and Acute Kidney Injury Following Cardiac Surgery. *JAMA* 315,9 (Mar. 2016), pp. 877–888.
- [10] A. X. Garg et al. Perioperative Aspirin and Clonidine and Risk of Acute Kidney Injury: A Randomized Clinical Trial. *JAMA* 312,21 (Dec. 2014), pp. 2254–2264.
- [11] P. Young et al. Effect of a Buffered Crystalloid Solution vs Saline on Acute Kidney Injury Among Patients in the Intensive Care Unit. *JAMA* 314,16 (Oct. 2015), pp. 1701–1710.

- [12] F. P. Wilson et al. Automated, electronic alerts for acute kidney injury: a single-blind, parallel-group, randomised controlled trial. *Lancet* 385,9981 (May 2015), pp. 1966–1974.
- [13] J. A. Kellum, F. E. Sileanu, R. Murugan, N. Lucko, A. D. Shaw, and G. Clermont. Classifying AKI by Urine Output versus Serum Creatinine Level. *J. Am. Soc. Nephrol.* 26,9 (Sept. 2015), pp. 2231–2238.
- [14] J. A. Kellum, R. Bellomo, and C. Ronco. Progress in Prevention and Treatment of Acute Kidney Injury. *JAMA* 320,5 (Aug. 2018), p. 437.
- [15] J. L. Koyner et al. Tissue Inhibitor Metalloproteinase-2 (TIMP-2)*IGF-Binding Protein-7 (IGFBP7) Levels Are Associated with Adverse Long-Term Outcomes in Patients with AKI. *J. Am. Soc. Nephrol.* 2, (Dec. 2014), pp. 1–8.
- [16] I. Gocze et al. Urinary biomarkers TIMP-2 and IGFBP7 early predict acute kidney injury after major surgery. *PLoS One* 10,3 (Jan. 2015), e0120863.
- [17] L. Dong, Q. Ma, M. Bennett, and P. Devarajan. Urinary biomarkers of cell cycle arrest are delayed predictors of acute kidney injury after pediatric cardiopulmonary bypass. *Pediatr. Nephrol.* 32,12 (Dec. 2017), pp. 2351–2360.
- [18] A. Vijayan et al. Clinical Use of the Urine Biomarker [TIMP-2] × [IGFBP7] for Acute Kidney Injury Risk Assessment. *Am. J. Kidney Dis.* 68,1 (July 2016), pp. 19–28.
- [19] M. Darmon, M. Ostermann, and M. Joannidis. Predictions are difficult... especially about AKI. *Intensive Care Med.* 43,6 (2017), pp. 932–934.
- [20] A. Lima and J. Bakker. Noninvasive monitoring of peripheral perfusion. *Appl. Physiol. Intensive Care Med.* 2. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 39–49.
- [21] A. A. Garvey and E. M. Dempsey. Applications of near infrared spectroscopy in the neonate. *Curr. Opin. Pediatr.* 30,2 (Apr. 2018), pp. 209–215.
- [22] J. S. Tweddell, N. S. Ghanayem, and G. M. Hoffman. Pro: NIRS is “Standard of Care” for postoperative management. *Semin. Thorac. Cardiovasc. Surg. Pediatr. Card. Surg. Annu.* 13,1 (2010), pp. 44–50.
- [23] J. C. Hirsch, J. R. Charpie, R. G. Ohye, and J. G. Gurney. Near infrared spectroscopy (NIRS) should not be standard of care for postoperative management. *Semin. Thorac. Cardiovasc. Surg. Pediatr. Card. Surg. Annu.* 13,1 (2010), pp. 51–54.

- [24] A. Aneman et al. Advances in critical care management of patients undergoing cardiac surgery. *Intensive Care Med.* (Apr. 2018).
- [25] N. S. Ghanayem and G. M. Hoffman. Near Infrared Spectroscopy as a Hemodynamic Monitor in Critical Illness. *Pediatr. Crit. Care Med.* 17, (2016), S201–S206.
- [26] S. Singh. Acute kidney injury after pediatric cardiac surgery. *Ann. Card. Anaesth.* 19,2 (2016), p. 306.
- [27] J. J. Blinder et al. Acute Kidney Injury after Pediatric Cardiac Surgery: A Secondary Analysis of the Safe Pediatric Euglycemia after Cardiac Surgery Trial. *Pediatr. Crit. Care Med.* 18,7 (2017), pp. 638–646.
- [28] M. C. Spaeder. Near-Infrared Spectroscopy Monitoring After Pediatric Cardiac Surgery: Time for an Intervention? *Pediatr. Crit. Care Med.* 19,5 (2018), pp. 496–497.
- [29] M. J. Madderom et al. Neurodevelopmental, educational and behavioral outcome at 8 years after neonatal ECMO: a nationwide multicenter study. *Intensive Care Med* 39,9 (Sept. 2013), pp. 1584–1593.
- [30] G. Garcia Guerra et al. Quality of life 4 years after complex heart surgery in infancy. *J. Thorac. Cardiovasc. Surg.* 145,2 (Feb. 2013), 482–488.e2.
- [31] T. Tlaskal et al. Long-term results after correction of persistent truncus arteriosus in 83 patients. *Eur J Cardiothorac Surg* 37,6 (June 2010), pp. 1278–1284.
- [32] D. Mesotten et al. Neurocognitive development of children 4 years after critical illness and treatment with tight glucose control: a randomized controlled trial. *Jama* 308,16 (Oct. 2012), pp. 1641–50.
- [33] S. Verstraete et al. Long-term developmental effects of withholding parenteral nutrition for one week in the PICU: a 2-year follow-up of the international randomised controlled PEPaNIC-trial. *Lancet Respir Med* (2018 (In press)).
- [34] A. I. Maas et al. Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI). *Neurosurgery* 76,1 (2015), pp. 67–80.
- [35] V. G. Coronado et al. Trends in traumatic brain injury in the US and the public health response: 1995–2009. *J. Safety Res.* 43,4 (2012), pp. 299–307.
- [36] G. T. Manley and A. I. R. Maas. Traumatic Brain Injury: An International Knowledge-Based Approach. *Jama* 310,5 (2013), pp. 473–4.
- [37] K. N. Corps, T. L. Roth, and D. B. McGavern. Inflammation and Neuroprotection in Traumatic Brain Injury. *JAMA Neurol* (Jan. 2015).

- [38] P. J. Andrews et al. Hypothermia for Intracranial Hypertension after Traumatic Brain Injury. *N. Engl. J. Med.* 373,25 (2015), pp. 2403–2412.
- [39] D. Cooper and J. Rosenfeld. Decompressive craniectomy in diffuse traumatic brain injury. *N. Engl. J. Med.* 364, (2011), pp. 1493–1502.
- [40] R. Chesnut, W. Videtta, P. Vespa, and P. Le Roux. Intracranial Pressure Monitoring: Fundamental Considerations and Rationale for Monitoring. *Neurocrit. Care* 21,S2 (2014), pp. 64–84.
- [41] P. J. Hutchinson et al. Trial of Decompressive Craniectomy for Traumatic Intracranial Hypertension. *N. Engl. J. Med.* 375,12 (Sept. 2016), pp. 1119–1130.
- [42] P. Le Roux. Intracranial pressure after the BEST TRIP trial: a call for more monitoring. *Curr. Opin. Crit. Care* 20,2 (Apr. 2014), pp. 141–7.
- [43] N. Stocchetti and A. I. Maas. Traumatic intracranial hypertension. *N. Engl. J. Med.* 370,22 (May 2014), pp. 2121–2130.
- [44] C. Lazaridis, C. G. Rusin, and C. S. Robertson. Secondary brain injury: Predicting and preventing insults. *Neuropharmacology* (June 2018), pp. 1–8.
- [45] C. Lazaridis et al. Patient-specific thresholds of intracranial pressure in severe traumatic brain injury. *J. Neurosurg.* 120,4 (2014), pp. 893–900.
- [46] Brain Trauma Foundation. Guidelines for the Management of Severe Traumatic Brain Injury 4th Edition. September (2016).
- [47] A. Vik et al. Relationship of “dose” of intracranial hypertension to outcome in severe traumatic brain injury. *J. Neurosurg.* 109,4 (Oct. 2008), pp. 678–84.
- [48] F. Güiza et al. Visualizing the pressure and time burden of intracranial hypertension in adult and paediatric traumatic brain injury. *Intensive Care Med.* 41,6 (Apr. 2015), pp. 1067–1076.
- [49] M. Czosnyka, P. Smielewski, P. Kirkpatrick, R. J. Laing, D. Menon, and J. D. Pickard. Continuous Assessment of the Cerebral Vasomotor Reactivity in head injury. *Neurosurgery* 41,July (1997), pp. 11–19.
- [50] B. Depreitere et al. Pressure autoregulation monitoring and cerebral perfusion pressure target recommendation in patients with severe traumatic brain injury based on minute-by-minute monitoring data. *J. Neurosurg.* 120,6 (June 2014), pp. 1451–1457.
- [51] International Initiative for Traumatic Brain Injury Research. <http://intbir.nih.gov>.
- [52] CENTER-TBI. <https://www.center-tbi.eu/>.

- [53] P. Le Roux et al. The International Multidisciplinary Consensus Conference on Multimodality Monitoring in Neurocritical Care: Evidentiary Tables : A Statement for Healthcare Professionals from the Neurocritical Care Society and the European Society of Intensive Care Medici. *Neurocrit. Care* (Jan. 2015).
- [54] F. Pedregosa et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 12, (2011), pp. 2825–2830.
- [55] E. Frank, M. Hall, L. Trigg, G. Holmes, and I. H. Witten. Data mining in bioinformatics using Weka. *Bioinformatics* 20,15 (2004), pp. 2479–2481.
- [56] J. Forcier, P. Bissex, and W. J. Chun. *Python web development with Django*. Addison-Wesley Professional, 2008.
- [57] M. Virbel, T. E. Hansen, and O. Lobunets. Kivy-A Framework for Rapid Creation of Innovative User Interfaces. *Mensch Comput. Work.* 2011, pp. 69–73.

SUMMARY

Critically ill patients are admitted to the intensive care unit (ICU) with acute life-threatening conditions. The diagnostic-therapeutic cycle in these patients is short as their clinical situation may vary rapidly. It is therefore of great interest to detect those patients most vulnerable to specific organ deterioration as early as possible. Due to the massive quantities of data generated at the patient bedside, clinicians are facing the problem of information overload. It is virtually impossible to process all the data from the various sources at the same time and over time. Big data analytics, a term that encompasses the application of data-driven advanced analytics such as machine learning techniques, are able to process large amount of data to perform pattern recognition, predictions, or generate data-driven hypothesis.

In this thesis, we aimed to apply big data analytics to gain novel insights in critical illness and to develop and validate decision support applications to help clinicians in the management of critically ill patients. We focused on three highly prevalent conditions, with important clinical consequences.

In the **first part**, we focused on early detection of acute kidney injury (AKI). AKI affects approximately 40% of critically ill patients and is associated with increased risks of morbidity and mortality, and with high financial costs. Early diagnosis of AKI is a major clinical challenge, which partially explains why the search for AKI therapies has been unsuccessful. Using the large multicenter EPANIC database, we developed and validated machine-learning-based prediction models (jointly referred to as AKIpredictor) for the development of AKI during the first week of ICU stay. Four models were developed to mimic the sequential availability of clinical data at the bedside. All models were based on routinely collected patient information. The models' performance was fair to good, and gradually improved with data availability. Additionally, this study was the largest study of an AKI biomarker, which, when measured at ICU admission, was outperformed by the AKIpredictor. The models were made available online via an open-access web application (www.akipredictor.com), which has been used worldwide by more than 2000 users.

To understand how these models should be used in clinical practice and how they compare to bedside physicians, we conducted a prospective observational blinded study in 252 critically ill adults. Using structured questionnaires,

physicians were asked upon admission, on the first morning of ICU stay, and after 24 hours, to predict the development of AKI. Predictions were compared against those made by the AKIpredictor. Physicians predicted AKI with good discrimination but with an over-estimated risk. Although achieving a slightly lower discrimination than physicians, the AKIpredictor showed improved calibration and net benefit, suggesting an added value to physicians' predictions, in particular for junior clinicians, when physicians were not confident in their predictions, or when it is of interest to gather predictions at fixed time points such as at inclusion in clinical trials.

In the **second part**, we focused on the assessment of the utility of near-infrared-based cerebral oximetry. Cerebral oximetry by near-infrared spectroscopy (NIRS) is used frequently in critically ill children to measure oxygen saturation levels in the frontal lobe. However, guidelines on how it should be used to guide patient care, trigger interventions, or for prognostication are lacking. Using the data from a prospective observational study where children after cardiac surgery were monitored with a blinded NIRS cerebral oximeter, we assessed association between the NIRS signal and ICU and hospital outcomes. Using advanced time-series analysis, we calculated predictors on the first 24 hours of NIRS monitoring, and found that both a measure of variability and the dose of desaturation below 50% were significant predictors of adverse outcomes. Furthermore, our study highlighted the difference in the NIRS signal between patients with cyanotic versus acyanotic heart defect. Hence, we recommend that future studies should be adequately powered to analyze these populations separately.

Cerebral NIRS oximetry is also used as a hemodynamic monitor for early recognition of an inadequate global oxygen supply/demand relationship in critically ill children, as the use of invasive catheters is not feasible or available. Hemodynamic instabilities contribute significantly to the development of AKI, a common complication after pediatric cardiac surgery. We investigated the ability of cerebral NIRS oximetry to predict severe AKI after pediatric cardiac surgery and assessed its additional predictive value to routinely collected data. Overall, we found that the data currently displayed in cerebral oximeters have only fair discriminability for 6-hour-ahead prediction of severe AKI. Retrospective calculation of NIRS variability achieved better performance, although not sufficient to outperform a clinical model based on routinely collected patient information. However, combining NIRS variability with the clinical model significantly improved predictive capabilities, suggesting that the NIRS signal carries independent relevant information in addition to routinely collected clinical data.

Future prospective interventional studies are required to identify whether the implementation of NIRS variability at the bedside could help drive therapeutic interventions and improve patient outcome, or for prognostication.

In the **last part** of this thesis, we focused on the detection of secondary brain injuries in patients with severe traumatic brain injury (TBI). TBI is one of the most important health care problems worldwide. In the hours and days following the initial traumatic insult (or *primary injury*), destructive and self-propagating biological changes in the brain can lead to subsequent additional damages (or *secondary injury*). Prevention and treatment of secondary injury is the main goal of neuro-critical care in patients with TBI.

A subtype of secondary brain injuries is elevated intracranial pressure (ICP), which is associated with worse outcome. The combination of the duration and the intensity of ICP defines a dose, and the association between various ICP doses and outcomes can be visualized using large cohorts of patients with severe TBI. Two distinct regions associated with good or poor neurological outcomes emerged from the plot, and were affected by the cerebrovascular pressure autoregulation (CAR) capability of the patient. CAR is the physiological mechanism by which the brain is able to maintain a constant cerebral blood flow during changes in cerebral perfusion pressure, and is often impaired in patients with severe TBI.

We investigated whether doses of deficient CAR are associated with worse clinical outcomes, and can be defined as a subtype of acute secondary brain injury. We analyzed the association between 6-month neurological outcome and the dose of CAR, in a large database of patients with severe TBI. The resulting color-coded plots showed a time-intensity association with outcome for CAR insults in adult and pediatric patients with severe TBI. For both cohorts, two distinct regions associated with good or poor outcomes were found, separated by an exponential transition curve. This indicates that CAR is disturbed in a dynamic manner, such that duration and intensity play a role in the determination of a zone associated with better neurological outcome.

Currently, international guidelines recommended ICP lowering therapies at a universal threshold of 22 mmHg and no recommendation exist regarding management of CAR. The two population-based color-coded plots open up avenues for novel ways of managing patients with severe TBI. Indeed, the course of individual patients could be visualized both in terms of ICP and CAR in the population-based plots, and patient-specific thresholds can be inferred from the plots to adapt therapy. Additionally, continuous monitoring of CAR is challenging and usually based on indices reflecting the correlation between ICP and blood pressure captured at high frequency. As such, the Low Frequency Autoregulation Index (LAX) is a promising index developed by our research

group, which has never been evaluated prospectively. Therefore, we developed the prototype of a bedside monitor, to calculate and display the LAX as a measure of CAR, and to visualize patient-specific secondary brain injuries in terms of ICP and of CAR, in a continuous way. A prospective observational blinded study to evaluate the correct functioning of the prototype in real-time is currently being set up. Upon study completion, the software will be ready to be tested in an interventional study. Such study will evaluate whether the outcomes of patients with severe TBI can be improved by using the prototype to guide patient management.

Overall, this thesis focused on the application of big data analytics in different fields of intensive care medicine. Several applications have been developed, ranging from prediction models, to advanced statistical analysis of times-series of medical signals, and knowledge discovery. We generated new knowledge from routinely collected data or from monitoring devices for which the usefulness has not been proven. We successfully translated two of these applications towards patient bedside, which open up avenues for prospective interventional clinical studies to assess whether the developed tools can improve process of care and patient outcome.

Marine Flechet

Biomedical Engineer

✉ flechet.marine@gmail.com 🏠 Rue Tour l'Evêque 16C2
+32/475 52 70 45 4600 Visé

*Passionate, hardworking and creative engineer with strong interest in **data mining techniques** applied in biomedical-related fields, which aim to **improve patient care**.*

Research experience

Katholieke Universiteit Leuven, Belgium

PhD student Clinical **prediction models** in critical illness:
Oct 2014 – Nov 2018 from computer to bedside

🏆 PhD Fellow - Research Foundation Flanders (FWO)

🏆 Young Investigator Award - ICP2016

🏆 Best Poster Award - EURONEURO2018

✓ Data mining ✓ Clinical research

University of Liège, Belgium & University of Canterbury, New Zealand

Master's thesis Safety, performance and compliance: clinical
Feb 2014 - Jun 2014 and in silico evaluation and optimisation of
glycemic control in three countries.

🏆 Best master's thesis in Biomedical engineering - AIM

🏆 Best master's thesis award - AILg

✓ Statistics ✓ Machine learning

Ecole Polytechnique fédérale de Lausanne (EPFL), Switzerland

Lab Immersion Modeling feedback between the circadian clock
Sep 2013 - Jan 2014 and the cell cycle using dynamical system tools.

✓ System modelling ✓ Biology

Experience

Student monitor Systems Modeling and Analysis, University of Liege

Student monitor Digital Electronics, University of Liege.

Student representative Biomedical Engineering, University of Liege.

Education

B. Engineering Sciences
2009 - 2012

- **Cum laude overall**, Summa cum laude in 3rd year
- Graduated from the University of Liège, Belgium

M.Sc. Biomedical Engineering
2012 - 2014

- **Summa cum laude**
- Graduated from the University of Liège, Belgium
- 6-month **Erasmus program** at the EPFL, **Switzerland**



Skills

Languages

French

English

German

Dutch

Soft skills

- ✓ Organization
- ✓ Problem solving
- ✓ Presentation skills
- ✓ Team working

Industry knowledge

- ✓ Research
- ✓ Data mining - Statistics
- ✓ Programming
- ✓ Clinical study

Informatics

- ✓ Python
- ✓ SQL
- ✓ Matlab
- ✓ Java

Interests

Skiing • Reading
Squash • Running

Publications

JOURNAL ARTICLES

Flechet M, Grandas FG, Meyfroidt G. (2016). Informatics in neurocritical care: new ideas for Big Data. CURRENT OPINION IN CRITICAL CARE, 22 (2), 87-93. doi: 10.1097/MCC.000000000000287. (citations: 4) (Impact factor: 3.32).

Flechet M*, Guiza F*, Schetz M, Wouters P, Vanhorebeek I, Derese I, Gunst J, Spriet I, Casaer M, Van den Berghe G, Meyfroidt G. (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 MEDICINE, 43 (6), 764-773. doi: 10.1007/s00134-017-4678-3. (citations: 18) (Impact factor: 12.02). * *contributed equally*

Flechet M, Guiza F, Vlasselaers D, Desmet L, Lamote S, Delrue H, Beckers M, Casaer MP, Wouters P, Van den Berghe G, Meyfroidt G. (2018). Near-Infrared Cerebral Oximetry to Predict Outcome After Pediatric Cardiac Surgery: A Prospective Observational Study. PEDIATRIC CRITICAL CARE MEDICINE, 19 (5), 433-441. doi: 10.1097/PCC.0000000000001495. (citations: 2) (Impact factor: 3.50).

Dickson JL, Stewart KW, Pretty CG, **Flechet M**, Desai T, Penning S, Lambermont BC, Benyo B, Shaw GM, Chase JG. (2018). Generalisability of a Virtual Trials Method for Glycaemic Control in Intensive Care. IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, 65 (7), 1543-1553. doi: 10.1109/TBME.2017.2686432. (citations: 13) (Impact factor: 3.58).

Jacobs A*, **Flechet M***, Vanhorebeek I, Verstraete S, Ingels C, Casaer MP, Soto-Campos G, Verbruggen SV, Joosten KF, Güiza F, Van den Berghe G. Performance of pediatric mortality prediction scores for pediatric intensive care unit mortality and 90-day mortality. PEDIATRIC CRITICAL CARE MEDICINE. Online first. doi: 10.1097/PCC.0000000000001764. (Impact factor: 3.50) * *contributed equally*

Flechet M, Güiza F, Scharlaeken I, Vlasselaers D, Desmet L, Van den Berghe G, Meyfroidt G. Near-Infrared-based cerebral oximetry for prediction of severe acute kidney injury in critically ill children after cardiac surgery. Submitted for publication.

Flechet M*, Falini S*, Bonetti C, Güiza F, Van den Berghe G, Meyfroidt G. Prediction of severe acute kidney injury in critically ill adults: a comparison between clinicians and AKIpredictor, a prospective observational study. Submitted for publication. * *contributed equally*

CONFERENCE PROCEEDINGS

Flechet M, Meyfroidt G, Piper I, Citerio G, Chambers I, Jones PA, Lo T-YM, Enblad P, Nilsson P, Feyen B, Jorens P, Maas A, Schuhmann MU, Donald R, Moss L, Van den Berghe G, Depreitere B, Güiza F. (2018). Visualizing Cerebrovascular Autoregulation Insults and Their Association with Outcome in Adult and Paediatric Traumatic Brain Injury. Vol 126.; 291-295. doi: 10.1007/978-3-319-65798-1_57

Scientific outreach

Flechet M, Meyfroidt G. (2017). The AKIpredictor an online calculator to predict acute kidney injury, ICU Management & Practice.

Meeting abstracts at international conferences

Flechet M, Güiza F, Schetz M, Wouters P, Vanhorebeek I, Derese I, Gunst J, Van den Berghe G, Meyfroidt G. Early detection of acute kidney injury during the first week of the ICU. In: Critical Care: vol. 19 (Abstract No. P285). doi: 10.1186/cc14365. Presented at ISICEM, Brussels, Mar 2015.

Flechet M, Güiza F, Schetz M, Wouters P, Vanhorebeek I, Derese I, Gunst J, Van den Berghe G, Meyfroidt G. Data mining in intensive care: a focus on early detection of acute kidney injury Presented at IDA, Saint Etienne, Oct 2015



Flechet M, Güiza F, Schetz M, Wouters P, Vanhorebeek I, Gunst J, Van den Berghe G, Meyfroidt G. Continuous Monitoring of Urine Output And Hemodynamic Disturbances Improves Early Detection of Acute Kidney Injury During First Week of ICU Stay. In: Intensive care medicine experimental: vol. 3 (Suppl 1). doi: 10.1186/2197-425X-3-S1-A13. Presented at ESICM, Berlin, Oct 2015.

Flechet M, Güiza F, Schetz M, Wouters P, Vanhorebeek I, Derese I, Gunst J, Casaer MP, Van den Berghe G, Meyfroidt G. Computerized Prediction of AKI Outperforms Plasma NGAL: a Retrospective Analysis Presented at AKI&CRRT, San Diego, Mar 2016



Flechet M, Meyfroidt G, Piper I, Citerio G, Chambers I, Jones PA, Lo T-YM, Enblad P, Nillson P, Feyen B, et al. Visualizing autoregulation insults and their association with outcome in adult traumatic brain injury. Presented at ICP, Boston, Jun 2016. *Awarded 3rd price Young Investigator Award.*



Meyfroidt G*, **Flechet M***, Depreitere B, Van den Berghe G, Güiza F. Real-time visualization of the pressure and time burden of ICP in individual patients Presented at ICP, Boston, Jun 2016 * *contributed equally. Awarded 3rd price Young Investigator Award.*

Meyfroidt G*, **Flechet M***, Depreitere B, Van den Berghe G, Güiza F. Real-time visualization of the cumulative pressure and time dose of intracranial pressure in individual traumatic brain injured patients Presented at NCS, Washington, Sep 2016 * *contributed equally*

Flechet M, Güiza F, Vlasselaers D, Desmet L, Lamote S, Delrue H, Beckers MJ, Casaer MP, Wouters P, Van den Berghe G, Meyfroidt G. Predictive value of cerebral oximetry for acute outcomes after pediatric cardiac surgery: a prospective observational study. In: Intensive Care Medicine Experimental: vol. 5 (2) (Abstract No. 1056) (pp.528-529). doi:10.1186/2197-425X-3-S1-A13. Presented at ESICM, Vienna, Sep 2017.

Scharlaeken I*, **Flechet M***, Vlasselaers D, Desmet L, Van den Berghe G, Güiza F, Meyfroidt G. A clinical prediction model for severe AKI after pediatric cardiac surgery Presented at ISICEM, Brussels, Mar 2018. * *contributed equally*



Falini S*, **Flechet M***, Scharlaeken I, Güiza F, Van den Berghe G, Meyfroidt G. Systolic blood pressure and lactate levels can predict severe pediatric AKI after cardiac surgery: a machine learning approach. Presented at ESICM, Paris, Oct 2018. * *contributed equally*



Flechet M, Meyfroidt G, Depreitere B, Van den Berghe G, Güiza F. Real-time visualization of the dose of intracranial pressure and cerebral autoregulation in traumatic brain injured patients. Presented at EuroNeuro, Brussels, Nov 2018. *Awarded the 1st price of the Best Poster Award.*



Certificates

2018	IEC 62304: medical device software	Allanta
2018	Project management for researchers	KULeuven
2018	Introduction to Leadership	KULeuven
2018	Theory and practice of questionnaire construction and analysis	FLAMES
2017	Creative thinking for PhD students	KULeuven
2017	Essentials of image editing	KULeuven & Somersault18:24
2017	Advanced Python	GNode
2017	Time and Self-management	KULeuven
2017	Workshop Intellectual Property Rights, translation and exploitation of Research	Human Brain Project
2017	Networking	KULeuven
2017	Exploitation of Research – Technology and Knowledge transfer	KULeuven
2017	Creating effective poster	Principiae
2017	High Performance Python	KULeuven
2017	Insights into your strengths and ambitions	KULeuven
2016	Tools for time series	FLAMES
2015	Writing for a medical journal	KULeuven
2015	Writing to a general audience	KULeuven
2015	Scientific writing for biomedical sciences	KULeuven
2015	Presentation and Seminar Skills	KULeuven
2015	Scientific Integrity for starting PhDs	KULeuven

The following authors from [Flaticon](#) have to be acknowledged for the use of graphics in this thesis:

- [Smartline](#)
- [Becris](#)
- [Prettycons](#)
- [Freepik](#)
- [Itim2101](#)

Icons were licensed by [CC 3.0 BY](#).

