Predictive modeling and severity scoring after cardiac arrest in patients treated with targeted temperature management

Johnsson, Jesper

2021

Document Version:
Publisher's PDF, also known as Version of record

Predictive modeling and severity scoring after cardiac arrest in patients treated with targeted temperature management

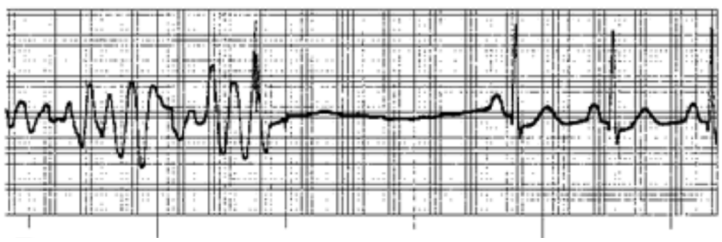
JESPER JOHNSSON
DEPARTMENT OF CLINICAL SCIENCES | LUND UNIVERSITY

To predict the unpredictable

The ability to accurately predict future outcome in critically ill patients may be considered a holy grail in modern intensive care. It is a challenging, yet important task for the clinician to evaluate and reliably assess the individual patient's opportunity for an acceptable neurological and functional recovery after severe illness such as cardiac arrest, with only limited information available. Information that will establish the foundation to decide whether to continue the treatment and observation in the intensive care unit (ICU), or withdraw life-sustaining therapy if prolonged intensive care is not considered being in the patient's best interest. There is a need to better stratify analyses of clinical trials including cardiac arrest patients, and an ambition to enhance tailored post-resuscitation care in the future.

This thesis aims partly to demonstrate the benefits of specific severity scoring models designed for cardiac arrest patients undergoing temperature intervention, but mainly to take a step forward towards improved alternative models to reliably predict outcome for those patients who survive the initial resuscitation.

Jesper Johnsson is a senior consultant in anaesthesiology and intensive care, and also a specialist in emergency medicine working in Helsingborg Hospital, Sweden. His research focuses on cardiac arrest patients and post-resuscitation care in the ICU.



An electrocardiogram (ECG) showing a ventricular fibrillation (VF) successfully defibrillated into a sinus rhythm (SR) and hopefully the return of spontaneous circulation (ROSC) with delivery of oxygen to the central nervous system (CNS). And so it begins...

Department of Clinical Sciences Lund
Anaesthesia and Intensive Care
Helsingborg, Sweden

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2021:33
ISBN 978-91-8021-039-3
ISSN 1652-8220

LUND UNIVERSITY

FACULTY OF MEDICINE

Printed by Media-Tryck, Lund 2021 NORDIC SWAN ECOLABEL 3041 0903

Predictive modeling and severity scoring after cardiac arrest in patients treated with targeted temperature management

Jesper Johnsson

LUND
UNIVERSITY

DOCTORAL DISSERTATION
by due permission of the Faculty of Medicine, Lund University, Sweden.
To be defended in Helsingborg on May 7th 2021 09:00.

Faculty opponent
Therese Djärv

Supervisor
Niklas Nielsen

Co-Supervisors
Martin Annborn, Attila Frigyesi & Sten Walther

Organization LUND UNIVERSITY Department of clinical sciences Anaestesiology and Intensive care **Author** Jesper Johnsson	**Document name** DOCTORAL DISSERTATION **Date of issue** May 7th 2021 Sponsoring organization

Title and subtitle Predictive modeling and severity scoring after cardiac arrest in patients treated with targeted temperature management

Abstract

Background: Primary cardiovascular disease (CVD) leading to cardiac arrest, is a common cause for critical care admission globally. The risk of severe ischaemic brain injury or death is still considerable. Several factors are known to influence later functional outcome in comatose, adult out-of-hospital (OHCA) patients with return of spontaneous circulation (ROSC) and treated with modern post-resuscitation care. Targeted temperature management (TTM) is still recommended in international guidelines after cardiac arrest. Registries of TTM patients are a valuable source of data to facilitate comparison between groups of OHCA patients with different baseline characteristics and thereby identify which variables are important for outcome. It remains, however, challenging to accurately and reliably predict outcome in this heterogeneous group of patients. Although earlier studies have shown that patients' background factors, prehospital circumstances and arrest characteristics are strongly associated with outcome, none of these variables are taken into account in the current recommended multimodal neurological prognostication algorithm (ERC/ESICM) of cardiac arrest patients treated and observed in intensive care units (ICU).

Aim: 1) To analyse if varying levels of TTM after OHCA were associated with later functional outcome in an international observational registry (study I). **2)** To investigate which variables and clinical information that carry the most predictive information in an OHCA population (study II). **3)** To investigate if a supervised machine learning algorithm called artificial neural network (ANN) could create reliable predictions of outcome for comatose OHCA patients using variables available already on hospital admission (study II), and during the first 3 days of ICU observation, using cumulative added information including biomarkers of brain injury (study III). **4)** To compare the predictive performance of an ANN with another supervised machine learning model called XGBoost, and to investigate the generalisability of each model (study IV). **5)** To demonstrate the hazard of adjustment for SAPS 3 scores in outcome studies on temperature intervention (Letter to the Editor).

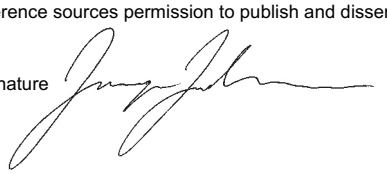
Methods: OHCA patients from the INTCAR 1.0 and 2.0-registries, and the TTM trial were included for data analysis. Background and prehospital data, clinical variables available on admission to hospital and cumulative information collected from the first three days of intesive care (including different levels of biomarkers for brain injury) were used. Logistic regression, as well as two supervised machine learning models (ANN and XGBoost), were used for the analyses. Patient outcome was the dichotomised Cerebral Performance Category scale (CPC) where CPC 1–2 denoted a good functional outcome and CPC 3–5 denoted a poor functional outcome, respectively.

Results: 1) There was no significant association between temperature and outcome (p=0.35) in OHCA patients included in INTCAR 2.0 **2)** ANN predicted outcome with an AUC of 0.89 using 54 clinical variables available on admission to hospital and outperformed a model based on logistic regression (p=0.029). **3)** Incorporating biomarkers such as NSE improved the AUROC over the first 3 days of intensvie care. When adding NFL the prognostic performance was excellent from day 1. **4)** ANN and XGB predicted outcome with equal performance (AUROC of 0.86) (p=0.64). Internal validation showed similar performance in both models, whereas external validation performed well, but with significantly lower precision (p=0.04). **5)** The temperature component used to calculate SAPS 3 score during the first hour following admission to intensive care, greatly influenced the predicted hospital mortality rate in a model of simulated cardiac arrest patients undergoing different levels of temperature intervention.

Conclusion: 1) No significant difference in outcome at hospital discharge was found in patients receiving lower- vs higher TTM, supporting the findings from the TTM trial. **2)** ANN predicted long-term functional outcome on hospital admission well and factors related to the prehospital setting carried most predictive information. **3)** Clinically accessible biomarkers (NSE) and research-grade biomarkers (NFL) increased the prognostic performance in our ANN. **4)** ANN and XGB performed equally well when predicting outcome at hospital discharge using early variables only. When externally validated, both models performed well, but with lower descrimination. **5)** TTM-adjusted severity scoring models would probably improve the assessment of mortality in TTM treated cardiac arrest patients.

Key words cardiac arrest, out-of-hospital, functional outcome, targeted temperature management, TTM, machine learning, artificial neural networks, Cerebral performance category, prediction, prognostication, ICU, SAPS 3

Supplementary bibliographical information Lund University, Faculty of Medicine Doctoral Dissertation Series 2020:59		**Language** English
ISSN and key title 1652-8220		**ISBN** 978-91-8021-039-3
Recipient's notes	**Number of pages** 139	Price
	Security classification	

Signature  Date 2021-03-30

Predictive modeling and severity scoring after cardiac arrest in patients treated with targeted temperature management

Jesper Johnsson

LUND
UNIVERSITY

Cover photo from iStock

Lund University, Faculty of Medicine
Department of Clinical Sciences Lund
Anaesthesia & Intensive Care, Helsingborg Hospital, Sweden

ISBN 978-91-8021-039-3
ISSN 1652-8220

Printed in Sweden by Media-Tryck, Lund University
Lund 2021

I know we could live tomorrow,
but I don't think we should wait

Lahle Pourkarim

Jag talar till mina söner
Ni är jägare, krigare, liksom jag
Ni kommer snart att ge er ut
för att nedlägga villebråd
Ni kommer att gå ut i strid
Ingen
kan jaga
bara för dödandets skull
Ingen
kan vinna
bara för vinnandets skull
Någon måste kunna ta emot vad ni nedlagt
Ni måste ha någon att dela bytet med
Ni måste ha något att försvara
Ensamma är ni inget
Ni kanske kommer att vandra ensamma
leva ensamma
men ni kommer alltid att leta efter någon
som ni kan jaga åt
och ni kommer alltid
att vilja ha något att försvara
Lägg inte bytet i bankfack
Försvara aldrig det oförsvarliga
dumheten, hyckleriet, främlingshatet
Bli män på era egna villkor
och ifrågasätt alltid makten
En dag ska ni inte längre höra koltrasten
eller känna doften av hägg, av snö
Jag talar till mina söner
Bli vänner med livet
med kvinnan
Ni har mycket att lära
och ingen blir någonsin fullärd
Gå inte sysslolösa
Bäst sover man efter ett bra dagsverke
Tacka inte nej till en kopp kaffe
och bjud alltid på en själva
Håll Döden på avstånd och ljug inte bort era känslor
Välj och ta ansvar för valet och lita på Gud
Var inte till lags
Lyd inte
Gå egna vägar
Jag talar till mina söner
Större än allt annat är kärleken och friheten
Sjung om kärleken
Sjung om friheten
Gör varje dag till en fest för livet
Jag talar till mina söner

Ulf Lundell ’91.

Table of Contents

List of publications

This thesis is based on the following papers and letter to the Editor, which will be referred to in the text by their Roman numerals.

I. **Johnsson J**, Wahlström J, Dankiewicz J, Annborn M, Agarval S, Dupont A, Forsberg S, Friberg H, Hand R, Hirsch KG, May T, McPherson JA, Mooney MR, Patel N, Riker RR, Stammet P, Søreide E, Seder DB, Nielsen N. Functional outcomes associated with varying levels of targeted temperature management after out-of-hospital cardiac arrest – An INTCAR2 registry analysis. *Resuscitation*. 2020 Jan 1;146:229-236.

II. **Johnsson J**, Björnsson O, Andersson P, Jakobsson A, Cronberg T, Lilja G, Friberg H, Hassager C, Wise M, Nielsen N*, Frigyesi A*. Artificial neural networks improve early outcome prediction and risk classification in out-of-hospital cardiac arrest patients admitted to intensive care. *Crit Care.* 2020;24:474.

III. Andersson P, **Johnsson J**, Björnsson O, Cronberg T, Hassager C, Zetterberg H, Stammet P, Undén J, Kjaergaard J, Friberg H, Blennow K, Lilja G, Wise M, Dankiewicz J, Nielsen N*, Frigyesi A*. Predicting neurological outcome after out-of-hospital cardiac arrest with cumulative information; development and internal validation of an artificial neural network algorithm. *Crit Care.* 2021;25:83

IV. **Johnsson J**, Björnsson O, Andersson P, Dankiewicz J, Walther S, Annborn M, Frigyesi A, Nielsen N. Early prediction of functional outcome after out-of-hospital cardiac arrest using artificial neural networks and eXtreme gradient boost models: development and external validation. Unpublished manuscript.

V. **Johnsson J**, Dankiewicz J, Walther S, Nielsen N. Hazards of adjustment in studies on temperature interventions. Letter to the Editor. *Resuscitation.* 2021 Feb 4;160:140-141

Papers I-III are published open access. Paper IV is an unpublished manuscript. Publication V is a Letter to the Editor. *Niklas Nielsen and Attila Frigyesi contributed equally in paper II and III.

Abbreviations

AI	Artificial intelligence
AHA	American Heart Association
ALS	Advanced life support
AMI	Acute myocardial infarction
ANN	Artificial neural network
APACHE	Acute Physiology and Chronic Health Evaluation
AUC	Area under the curve
AUROC	Area under the receiver operating characteristic curve
BLS	Basic life support
BNP	N-terminal pro–B-type natriuretic peptide
CI	Confidence interval
CNS	Central nervous system
CPC	Cerebral Performance Category Scale
CPR	Cardiopulmonary resuscitation
CT	Computed tomography
CVD	Cardiovascular disease
ECG	Electrocardiogram
ECMO	Extracorporeal membrane oxygenation
EEG	Electroencephalography
EMR	Estimated mortality rate
EMS	Emergency medical services
ERC	European Resuscitation Council
ESICM	European Society of Intensive Care Medicine
FNR	False-negative rate
FPR	False-positive rate
GCS	Glasgow Coma Scale
GCS-M	Glasgow Coma Scale motor score
GFAP	Glial fibrillary acidic protein
ICU	Intensive care unit

IHCA	In-hospital cardiac arrest
ILCOR	International Liaison Committee on Resuscitation
IL-6	Interleukin-6
INTCAR	International Cardiac Arrest Registry
IQR	Interquartile range
MRI	Magnetic resonance imaging
mRS	Modified Ranking Scale
NFL	Neurofilament light chain protein
NSE	Neuron-specific enolase
OHCA	Out-of-hospital cardiac arrest
OMR	Observed mortality rate
PCAS	Post-cardiac arrest syndrome
PEA	Pulseless electric activity
PCT	Procalcitonin
PMR	Predicted mortality rate
ROC	Receiver operating characteristic curve
ROSC	Return of spontaneous circulation
SAPS 3	Simplified Acute Physiology Score 3
SHAP	Shapley additive explanations
SMR	Standardised mortality ratio
SOFA	Sequential Organ Failure Dysfunction score
SSEP	Short-latency somatosensory evoked potentials
S100B	S100 calcium-binding protein B
Tau	Tau protein
TPR	True-positive rate
TNR	True-negative rate
TTM	Targeted temperature management
TTM trial	Target temperature management after out-of-hospital cardiac arrest trial
TTM2 trial	Targeted hypothermia versus targeted normothermia after out-of-hospital cardiac arrest trial

UCH-L1	Ubiquitin carboxy-terminal hydrolase L1
VF	Ventricular fibrillation
VT	Ventricular tachycardia
WLST	Withdrawal of life-sustaining therapy

Introduction

To predict the unpredictable

Primary cardiac disease leading to sudden cardiac arrest is a common cause of critical care admission globally. Despite the last decades of continual progress in emergency medicine, invasive cardiac interventions and advanced intensive care, the risk of severe ischaemic brain injury or death is still considerable. Approximately 50% of the initially resuscitated cardiac arrest patients admitted to intensive care will never be discharged from hospital due to extensive neurological damage, and in the majority of these patients, death occurs as the consequence of a decision to withdraw life-sustaining therapy. But how can we be certain, when making such a decision, which patients will survive with a good chance of acceptable functional outcome, and which patients have suffered too severe neurological damage and thus will not be able to recover?

Several factors are known to influence functional neurological outcome in the comatose, adult out-of-hospital cardiac arrest (OHCA) patients with return of spontaneous circulation (ROSC) and who are treated with modern post-resuscitation care including targeted temperature management (TTM) as recommended in international guidelines. Patient registries are a valuable source of data to facilitate comparison between groups of OHCA patients with different baseline characteristics and thereby identify which variables are important for outcome. However, it remains challenging to accurately and reliably predict functional outcome in this highly heterogeneous group of patients, especially on an individual level. Although previous studies have shown that patients' background factors, prehospital circumstances and arrest characteristics are strongly associated with later outcome, none of these variables are taken into account in the current recommended multimodal neurological prognostication algorithm (from ERC/ESICM) of cardiac arrest patients treated and observed in intensive care units (ICU).

This thesis aims partly to demonstrate the benefits of specific severity scoring models designed for cardiac arrest patients undergoing temperature intervention, but mainly to take a step forward towards improved alternative models, to reliably predict functional outcome for those who survive the initial resuscitation.

Jesper Johnsson

Background

Cardiac arrest

Cardiac arrest is the cessation of mechanical cardiac activity, which if left untreated will inevitably, lead to death. Without intervention a cardiac arrest in many cases is the natural end of the biological process of aging, and should therefore be left without further resuscitation efforts. However, sometimes the arrest is premature, sudden and unexpected, and prompt performance of cardiopulmonary resuscitation (CPR) must be started in order to restore spontaneous cardiac activity and meanwhile, supply the organs and tissues with oxygenated blood. If the cardiac arrest is successfully reversed and ROSC is restored, the arrest may nevertheless have caused anoxic damage to the body. Especially the brain is susceptible to anoxia due to its low energy stores and high metabolic rate. Varying degree of neurological damage or even death is therefore sometimes the unavoidable result of a prolonged cardiac arrest. According to Swedish law and many other legal systems globally, death has occurred when all brain functions are totally and irreversibly lost[1] which can be determined via 'direct criteria', including protocolised neurological examination of the cranial nerves and apnoea test, in patients treated with invasive ventilatory support in an ICU. But in most cases 'indirect criteria', including confirmed absence of pulse and spontaneous breathing for a sufficient time, are enough to assume that no brain activity remains and the patient can be declared deceased.

Physiology and types of cardiac arrest

Aetiology

The majority of all cardiac arrest are of cardiac origin and often the result of an underlying ischaemic heart disease causing a malignant cardiac arrythmia. Such arrhythmias include pulseless ventricular tachycardia (VT) or ventricular fibrillation (VF), both inhibiting an effective pump function of the heart. The sudden onset of arrythmia can be secondary to an acute plaque rupture and thrombus formation in a coronary artery causing acute myocardial infarction (MI), or following reperfusion.[2] Previous studies on OHCA have documented acute MI in more than 50% of the patients.[3,4] Cardiac arrhythmia can also be primary and arise from areas of chronic myocardial scarring after previous MI. This is particularly evident in combination

with chronic heart failure and decreased ejection fraction of the left ventriculus.[5] Other, more rare causes of malignant arrhythmia, are cardiomyopathies and genetic ion-channel abnormalities.[5] In a minority of cardiac arrest cases, the origin is non-cardiac. In this group arrests related to drowning or other causes of hypoxia, intoxications, trauma, sepsis, anaphylaxis, pulmonary embolus or primary cerebral events are found. Arrests of non-cardiac aetiology represent approximately 18-25% of all cardiac arrests.[6-8] Even if the word 'sudden' often is used as a prefix to describe the onset of cardiac arrests, studies have shown that more than 50% of all cardiac arrest patients experience various levels of warning symptoms during the weeks prior to the arrest. Such symptoms include heart palpitations, dyspnoea, syncope and chest pain.[9]

Primary cardiac arrest rhythms

Shockable rhythms

Arrhythmias triggered by acute myocardial ischaemia and infarction or derived from an arrhythmogenic substrate in myocardial scar tissue, is often classified as shockable rhythms, including pulseless VT or VF. As the classification indicates, these rhythms can be converted into sinus rhythm by depolarising the myocardium simultaneous with an electric shock (defibrillation). Even if cardiac arrest implies an inactive heart with lack of contractility, an electrocardiogram (ECG) recording on a monitor or on the defibrillator shows otherwise. A typical VT presents with wide regular complexes present on the ECG, often with a heart rate over 100 beats per minute. Despite an ongoing VT, the heart is sometimes still able to produce a cardiac output and thereby maintain a sufficient circulation with blood flow to essential organs. But in cases with pre-existing heart disease or in very fast tachycardia, the cardiac output is dramatically reduced and a circulatory collapse is inevitable (pulseless). A VF on the other hand, typically results in a sudden loss of pump function and a total forward failure of the heart. This situation results in a subsequent unconsciousness due to the immediate discontinuation of oxygen supply to the brain. An ECG reading would show an erratic trace from disorganised and turbulent electrical impulses stimulating uncoordinated myocardial twitching. In approximately 20-25% of cardiac arrest cases the first recorded rhythm is a VF.[10-12]

Non-shockable rhythms

Rhythms classified as non-shockable include pulseless electrical activity (PEA) and asystole and has become the predominant primary arrest rhythm, whereas the incidence of VF has continued to decline during the last four decades.[13] A PEA, also known as electromechanical dissociation, is characterised by an impaired cardiac output with an insufficient perfusion or no perfusion at all, despite normal and regular electric activity on the ECG due to sufficient coordinated electrical discharges. PEA does not necessarily mean the lack of mechanical activity. There can be ventricular contractions and detectable pressures in the aorta, which are

referred to as pseudo-PEA.[14] A pseudo-PEA could possibly contribute to a state of low-flow perfusion to some extent, but may be difficult to differentiate from a preserved cardiac activity in severely shocked patients without palpable pulses.[15] Asystole is defined by undetectable electrical activity on the ECG and the absence of a palpable pulse, indicating a state of non-existing circulation. PEA and asystole are more common in cardiac arrest of a non-cardiac aetiology (massive pulmonary embolus, trauma, hypovolemia, intoxications, electrolyte disturbances, hypothermia etc.) and represent the majority of first recorded rhythms following a circulatory collapse.[10] Neither PEA nor asystole can be converted into sinus rhythm by defibrillation. However, if a non-shockable rhythm can be converted into a shockable, due to effective high-quality CPR and other medical treatment, this is associated with a better probability of survival.[16] The prognosis of asystole is often worse and more associated with a poor outcome compared to VT/VF. This association could be related to the fact that asystole sometimes develops from a VF and thus represents a later stage in the cardiac arrest process.[17]

This indicates that rhythm is not a constant variable throughout the cardiac arrest situation, but can change as a result of defibrillation or due to other interventions such as chest compressions and drug administration. Studies have shown that initial rhythm is strongly associated with possible ROSC and long-term survival, and thus has evident prognostic implications.[18] However, it is important to recognise that the initial rhythm often is represented by the first rhythm registered and might differ from the true initial rhythm. Another way to commonly stratify cardiac arrests is by the location of the arrest; in-hospital cardiac arrest (IHCA) or out-of-hospital cardiac arrest (OHCA), respectively.

In-hospital cardiac arrest

IHCA denotes all cardiac arrests that occur within hospital walls including the emergency department (ED) where the patient might have been brought prior to the arrest. IHCA are more often related to non-cardiac origins such as septic shock, hypoxia, hypovolemia and bleeding. Only approximately 20% of IHCA patients present with a shockable rhythm, but in these cases a rapid defibrillation is associated with improved outcome.[19] Traditionally, IHCA has been considered a condition with utterly poor prognosis. Despite the in-hospital location with rapid access to medical service, the patient could be critically ill with the arrest secondary to other severe illness. Although outcome still remains poor, IHCA is associated with improved overall survival over the last two decades according to recent data,[12,20] which also supports a better 30-day survival compared to OHCA.[21]

Out-of-hospital cardiac arrest

OHCA differs somewhat from IHCA, and the two groups are not entirely comparable in terms of aetiology, age, arrest characteristics and bystander proficiency as well as bystander efforts. These patients more often present with an arrest of cardiac origin and with a shockable rhythm.[22] The overall survival rate after OHCA continues to improve due to a variety of factors including advanced

prehospital care provided by emergency medical services (EMS),[23,24] access to automated external defibrillators (AED),[25] laypersons awareness of cardiac arrest and bystander-CPR,[26-28] along with standardised intensive care bundles and more early cardiac intervention.[29,30]

The studies and analyses included in this thesis will focus on the functional outcome of comatose patients resuscitated from OHCA and with arrests of a presumed cardiac cause.

Epidemiology and survival

Cardiovascular disease (CVD) is the leading cause of death in the world, and the absolute global burden of CVD is continuously increasing since the millennium.[31] Approximately 50% of all CVD-related deaths is the result of a cardiac arrest.[12] Incidence as well as survival and quality of cardiac arrest documentation varies between reporting countries globally. The world-wide incidence of OHCA patients treated by EMS was estimated to 62 cases per 100.000 population annually in a systematic review published in 2010. Approximately 75-85% of these patients presented with an arrest of a primary cardiac cause.[32] Ten years later the same estimation was 30-97.1 individuals per 100.000 population across the world as presented in the first report from ILCOR.[33] In Europe, the corresponding number was 56 cases per 100.000 population as reported in the EuReCa TWO study,[34] whereas the average numbers were higher (73 cases per 100.000 population) in the United States according to the 'Resuscitation outcome consortium network' (ROC)-registry.[12] Even if the total burden of CVD is increasing throughout the world, both overall 30-day survival rates and survival to hospital discharge simultaneously continue to improve, although the results differ considerably between countries as displayed in figure 1.

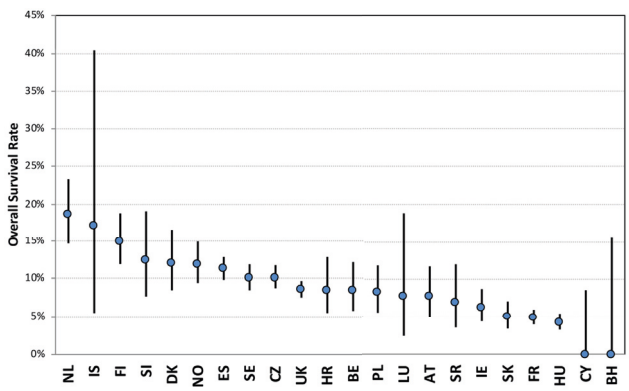


Figure 1. Overall survival to hospital discharge after OHCA in 21 European countries. Gräsner et al. Survival after out-of-hospital cardiac arrest in Europe - Results from the EuReCa TWO study. Resuscitation 2020 Mar 1;148: 218-226. Reprinted with permission.

Epidemiology and survival in Sweden

In 2019 approximately 5900 cases of OHCA were treated by EMS which is relatively unchanged since 2017.[35] Previous reports have shown an evident increase in incidence for about a decade. The majority of patients was male (70%), but the proportion of women is steadily increasing. Median age was 71 years and age distribution has been relatively stable over time (unchanged for three decades). Ischaemic heart disease is the most common cause of OHCA in the age group of adults over 40 years. OHCA in younger patients is more frequently of non-cardiac origin including trauma, suicide, intoxication, pulmonary embolus and near-drowning, whilst 'sudden infant death syndrome' was one of the main causes in infants.

The most common location of arrest was at home, and in the majority of cases first recorded rhythm was non-shockable (PEA or asystole) which is associated with a worse outcome.[36] This might be related to an increasing proportion of unwitnessed arrest in the population in general. Shockable rhythms (pulseless VT and VF) tended to be less frequent in the 2019 report. This applied to both witnessed and unwitnessed OHCA. The increase in incidence of non-shockable rhythms was also evident in arrests with an obvious cardiac origin. Despite this, the average survival after OHCA has increased and nearly doubled during the last two decades, from 4,5% in 2000 to 11% in 2019. The reason for this improvement is multifactorial including increased availability of acute coronary interventions and modern intensive care managing the post-cardiac arrest syndrome as later described in this thesis. But the factors with most favourable influence on outcome are probably the presence of high-qualitative CPR and early defibrillation performed by bystanders.

Cardiopulmonary resuscitation

History

There are numerous examples throughout history of more or less successful, yet legendary, efforts to restore the sudden cessation of breathing and circulation in both animals and human beings. However, the birth of modern CPR in humans took place in the 1960's when mouth-to-mouth ventilation in combination with closed chest compressions was introduced as an effective method of resuscitation. The closed chest compression technique was described already in 1878 by Boehm, as a method to circulate blood in the human body in emergency situations. The concept was then rediscovered in the 1950's, and in 1960 "The cardiac pump theory" was released by Kouwenhoven and colleagues.[37] In 1966 the first CPR guidelines were published by the American Heart Association,[38] which in many ways resemble the guidelines of modern resuscitation. The possibility for treatment with early defibrillation was introduced in prehospital settings after the implementation of portable defibrillators in the mid 1960's followed by extensive availability of public AEDs to be used by laypeople in the beginning of the 21st century.[39]

Chain of survival

The concept of "Chain of survival" was first proposed by Cummins et al. in 1991,[40] and updated in 2005 with a modification to increase focus on interventions and in-hospital critical care (post resuscitation care) as well as prevention of cardiac arrest by early recognition of angina and other risk factors.[41] Early CPR with focus on chest compression to buy time, and early defibrillation to restart the heart are the very foundation of modern resuscitation algorithms. The chain symbolically refers to the fact that successful resuscitation is only as strong as its weakest link and that survival from cardiac arrest depends on a sequence of interventions (figure 2). The chain-concept is a well-established part of both resuscitation science and CPR-training, and fully adapted in the ERC guidelines and education programs.

Figure 2. The ERC chain of survival. CPR, cardiopulmonary resuscitation. © European Resuscitation Council – www.erc.edu. Reprinted with permission.

The International Liaison Committee on Resuscitation (ILCOR) was formed in 1992 and has since then provided a forum for major international organisations of resuscitation to collaborate and develop guidelines and recommendations which are renewed every fifth years, with the latest version delayed one year due to the Covid-19 pandemic and therefore released in 2021.[42]

Out-of-hospital cardiac arrest care primarily aims to minimise the circulatory no-flow time which is defined as the time from arrest until start of BLS. The no-flow time is kept as short as possible by means of an immediate onset of basic CPR by bystanders as soon as a person is found unconsciousness and not breathing normally. It is prioritised to alert the emergency dispatch centre for EMS activation, and to localise a nearby AED for early defibrillation, but without interruption of the ongoing CPR. The time from onset of CPR (basic CPR or advanced CPR by EMS) until ROSC is referred to as the low-flow time, and is known to be an important variable associated with later functional outcome and survival in OHCA patients.[43,44] An uninterrupted basic CPR is an essential part of resuscitation. In recent years there has even been a shift in paradigm from prioritising the maintenance of a free airway and repetitive ventilation (mouth-to-mouth or with a bag valve mask), to focus on early defibrillation and uninterrupted effective chest compressions, which are associated with a more favourable outcome.[45]

The current recommended 2021 ERC algorithm for advanced life support (ALS) in adults is displayed in figure 3. As soon as a person is found unresponsive and not breathing normally, uninterrupted CPR should be started with a repeating cycle of 30 chest compressions and 2 breaths (ventilation). A defibrillator should be attached in order to assess the rhythm while chest compressions temporarily pause. Depending on whether the initial rhythm is shockable or non-shockable, the algorithm recommends defibrillation and continued CPR, or continued CPR only. The rhythm should be reassessed every two minutes throughout the cycle. According to the algorithm, it is essential to minimise all interruption of the ongoing CPR, and every defibrillation should be followed by another two-minute CPR-cycle before rhythm reassessment. If the patient should regain ROSC during CPR, measures to induce immediate post-resuscitation care should be commenced and are described in further details later in this chapter. The benefits of administering drugs during resuscitation to improve outcome have long been discussed but so far there is no convincing evidence regarding type of drug, dose or time of administration. In the current guidelines, a standard dose of epinephrine (1 mg) repeated every 3-5 minutes during CPR is recommended. After a third defibrillation in a shock-resistant VF, amiodarone (a membrane stabilising antiarrhythmic drug) can be administered once. Drugs included in the algorithm might improve outcome in a short-term perspective, but evidence is lacking regarding any long-term benefits on functional outcome.[46]

In addition, the algorithm contains different measures and recommendations to consider, including advanced airway management, mechanical chest compressions, identification and treatment of reversable causes (4 H's and 4 T's), ultrasound imaging, coronary angiography/PCI or extracorporeal CPR (ECMO) if indicated and available.

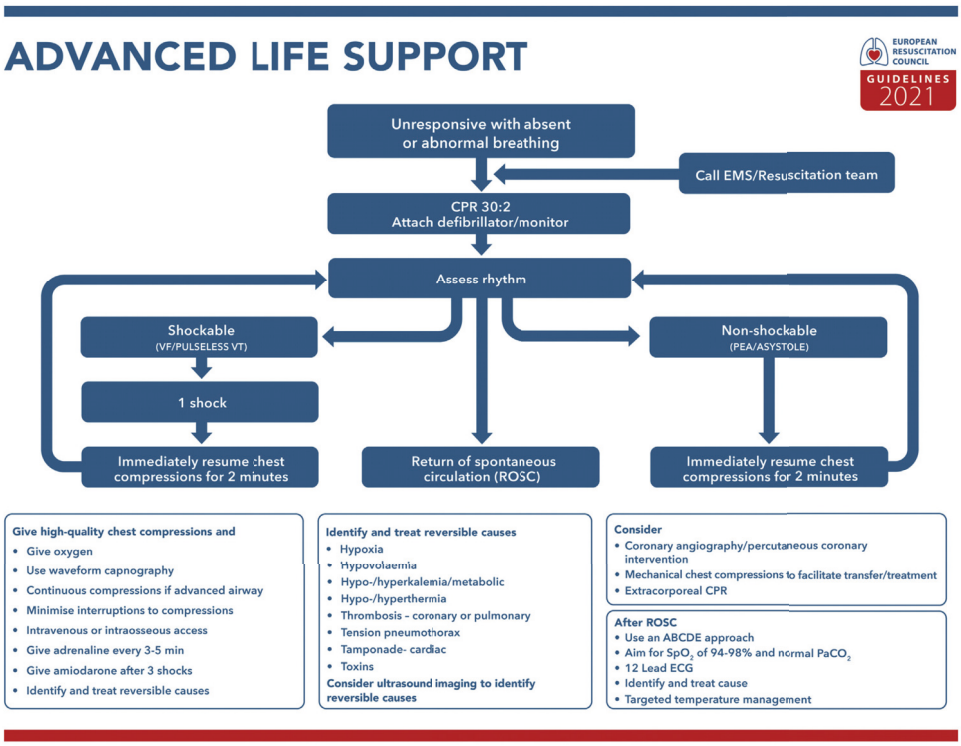


Figure 3. The Advanced Life Support (ALS) algorithm in adults. EMS, emergency medical services; CPR, cardiopulmonary resuscitation; VF/Pulseless VT, ventricular firbrillation/pulseless ventricular tachycardia; PEA, pulseless electrical activity; ROSC, return of spontaneous circulation; ABCDE, Airway Breathing Circulation Disability Exposure; SpO_2, peripheral capillary oxygen saturation; $PaCO_2$, partial pressure of carbon dioxide in arterial blood; ECG, electrocardiogram. © European Resuscitation Council 2021 – www.erc.edu. Reprinted with permission.

Post-cardiac arrest syndrome

Patients who regain ROSC after successful resuscitation face the consequences of whole-body ischaemia secondary to the arrest. Injuries that arise during periods of no-flow (before CPR) and low-flow (during CPR) are in most cases not directly reversed by the re-establishment of circulation to organs and tissues. The reperfusion itself may even cause additional damage through a range of complex inflammatory mechanisms, which are addressed as the post-cardiac arrest syndrome (PACS).[47,48] PACS is used to describe the processes and symptoms related to the post-cardiac arrest situation and that may occur during the first hours or days after ROSC. The four components of PACS are defined as:

- Post-cardiac arrest brain injury
- Post-cardiac arrest myocardial dysfunction
- Systemic ischaemia/reperfusion response
- Persistent precipitating pathology

Post-cardiac arrest brain injury

This is the most typical manifestation of PACS and includes in the immediate situation symptoms including impaired consciousness or coma and seizures-myoclonus due to the brain's susceptible vulnerability and low tolerance for the exposure to both ischaemia and reperfusion. The ischaemic-reperfusion response creates a variety of reactions on a cellular level with activation of cell-death signalling pathways. Mechanisms include among other, the formation and release of free radicals, disrupted electrolyte homeostasis, immediate cellular depletion of energy substrate with a switch to anaerobic metabolism and a subsequent intracellular acidosis. This will macroscopically manifest in an impaired cerebrovascular autoregulation, failure of cerebral microcirculation, oedema due to disruption of the blood-brain barrier and post-ischaemic neurodegeneration. Post-cardiac arrest brain injuries largely contribute to the high mortality rate in OHCA patients admitted to ICUs, due to a 'withdrawal of lift sustaining therapy' (WLST) decision during the prognostication process.[49]

Post-cardiac arrest myocardial dysfunction

Another component that contributes to lower survival rates after OHCA, is post-cardiac arrest myocardial dysfunction.[50] Due to myocardial stunning and global hypokinesia, both systolic and diastolic function is impaired, and further diminished by dysrhythmias. Severe cases could evolve into a cardiogenic shock situation including pronounced reduction of cardiac output, irreversible hypotension with hypoperfusion, and a state of general cardiovascular collapse. In cardiac arrests related to acute coronary syndrome, early revascularisation of myocardial infarction could be essential for outcome.[51] However, studies have shown that the clinical manifestations are more related to stunning than to permanent myocardial damage, indicating that the dysfunction is transient and full recovery can occur.[50]

Systemic ischaemia/reperfusion response

The global ischaemia and subsequent reperfusion after ROSC activate a systemic inflammatory response which resembles the clinical manifestations seen in septic shock. The hypoxic stress and disruption of metabolic substrate removal initiate a state of severe shock with endothelial activation and adrenal suppression as well as activation of coagulation and immunological pathways, similar to the physiological mechanisms seen in septic shock patients.[52] This pan-systemic response appears within the first hours after ROSC and is clinically manifested with cardiovascular insufficiency, impaired vasoregulation, tissue hypoxia/ischaemia, high fever, susceptibility to infections and in severe cases multi organ failure. Mostly, these pathologies are reversible with early and goal-directed intensive care interventions.

Persistent precipitating pathology

The post-cardiac arrest syndrome is further complicated and aggravated in the presence of persisting pathology that contributed to, or even caused the arrest itself. Such pathologies include cardiomyopathies, acute coronary syndrome (ACS), thromboembolic disorders (pulmonary embolism), chronic obstructive pulmonary disorder (COPD), sepsis, various toxidromes and trauma-related haemorrhagic cardiac arrest. All of which manifest clinically according to aetiology and therefore require different treatment depending on their contribution to the resulting PACS.

Management of PACS

For cardiac arrest patients who do not achieve ROSC very rapidly during initial resuscitation, and thereby remain comatose, adequate measures should be initiated to reduce the impact of PACS pathophysiology. Survival to hospital discharge is in general significantly lower than survival to hospital admission,[53] indicating that these patients are critically ill and requires extensive supportive therapy including intensive care provided in an appropriate setting. Such specific therapeutic strategies with protocolised interventions for optimisation and extensive monitoring can be summarised in the concept of post-resuscitation care.

Post-resuscitation care

The majority of resuscitated OHCA patients require advanced post-resuscitation care after ROSC, which means ICU admission and intensive care. However, the patient spectrum is wide within the heterogenic OHCA population and treatment strategies must accommodate a variety of illness severity with different risk profile related to patient's background (age, comorbidities, medical history), arrest characteristics, prehospital circumstances and the impact of PACS components. The recommended post-resuscitation care algorithm is displayed in figure 4.

Due to post-cardiac arrest brain injury, the patients might be deeply unconscious on hospital admission or present with generalised seizures or myoclonus. This situation commonly demands immediate intubation to secure the airway, enable mechanical ventilation and to ensure adequate gas exchange. Sedation and analgesia might be necessary to control seizures and to lower the sympathetic stress and overall cerebral metabolism. Ventilation strategies should aim for normoventilation with a peripheral saturation level of 94-98% and with normal values on arterial blood gases including normocarbia to avoid harmful cerebral vasoconstriction.[54] To secure sufficient oxygen delivery and reduce oxygen demands, early goal-directed therapy and meticulous haemodynamic optimisation is essential and needs to be closely monitored with ECG, invasive blood pressure, urinary output, echocardiography and metabolic monitoring.[55] Circulatory instability is common due to myocardial dysfunction and the systemic ischaemic/reperfusion state earlier described. Haemodynamic support includes intravenous fluids, vasopressors, inotropes and occasionally the use of circulatory support devices. Dysrhythmias should be treated with medication or cardioversion and with narrow preventive normalisation of any electrolyte disturbance. Since coronary artery disease is present in the majority of OHCA patients and the arrest is often caused by an acute myocardial infarction, the diagnosis and management of any underlying acute coronary syndrome is of crucial importance.[6] All resuscitated cardiac arrest patients with suspected ST-elevation myocardial infarction should undergo immediate coronary angiography with percutaneous coronary intervention (PCI) if indicated, or with thrombolytic therapy if PCI is not available.

Due to the multiplied increase in cerebral metabolism during ongoing seizures, anti-convulsive and sedative drugs should frequently be used to prevent further cerebral injury caused by status epilepticus.[56] However, the diagnosis might be difficult due to sedation and muscle relaxation therapy. The ERC recommends intermittent EEG recordings (to detect epileptic activity) and continuous simplified amplitude EEG as well as pharmacological treatment to control seizure. Myoclonus is common after cardiac arrest, especially in patients who remain comatose. It is associated with a poor outcome if early onset and it might be particularly difficult to treat. In addition to providing standard intensive care in resuscitated comatose OHCA patients with ventilatory, circulatory and cerebral support, the use of targeted temperature management (TTM) to prevent secondary brain injury is an implemented but somewhat controversial intervention in the post-resuscitation care.

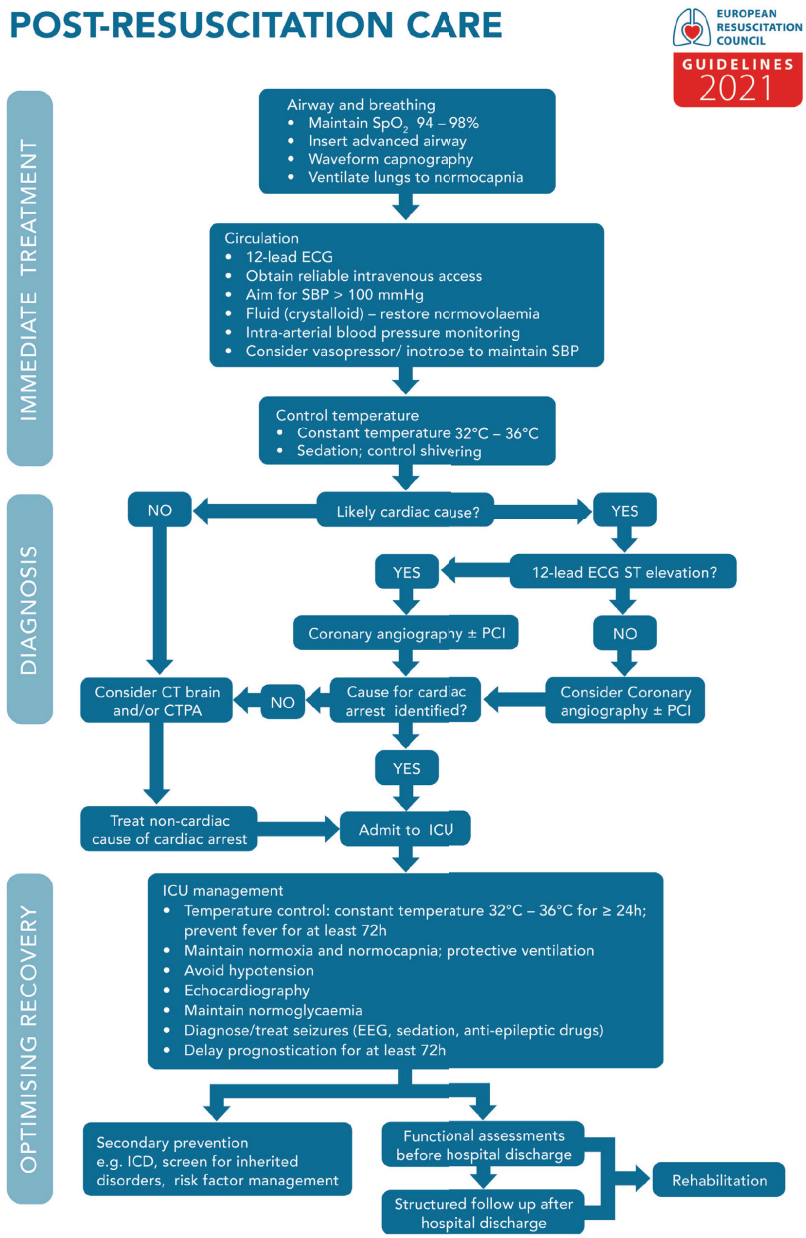


Figure 4. The post-resuscitation care algorithm. SpO2, peripheral capillary oxygen saturation; ECG, electrocardiogram; SBP, systolic blood pressure; PCI, percutaneous coronary intervention; CT, computed tomography; CTPA, computed tomography pulmonary angiogram; ICU, intensive care unit; EEG, electroencephalography; ICD, implanted cardioverter defibrillator. © European Resuscitation Council guidelines 2021 – www.erc.edu. Reprinted with permission.

Targeted temperature management

Reduction of body core temperature in animal experimental models is known to influence the pathophysiological pathways involved in ischaemic brain damage resulting in a reduction of neurological injury.[57,58] The exact mechanisms are however not completely understood. From retrospective clinical studies we know the converse relation, where hyperthermia is shown to be associated with more unfavourable neurological outcome due to an increase in brain metabolism and cerebral excitotoxity.[59,60] This scientific background combined with several case reports describing cardiac arrest victims surviving prolonged resuscitations with accidental hypothermia,[61,62] has led to a belief in the beneficial neuroprotective properties of induced hypothermia to reduce brain injury in OHCA patients.

In 2002 two smaller randomised controlled trials reported improvement in mortality and neurological function with mild induced hypothermia after cardiac arrest.[63,64] Both studies included OHCA patients with initial shockable rhythm and provided mild induced hypothermia (MIH) with targeted temperatures between 32-34°C for 12-24h. Treatment with hypothermia as an intervention for neuroprotection was then widely adopted and recommended in international guidelines in 2005, even if the evidence was of low quality according to GRADE-classification.[65] The Target Temperature Management (TTM) trial was published in 2013 and showed no difference in mortality or long-term neurological outcome between the two temperature groups compared (33°C vs 36°C) and is further described in the methods section.[66] The results from the TTM trial have later been confirmed in several sub-studies following the publication of the original study.[67-72]

In the 2021 ERC post-resuscitation care guidelines, a constant targeted temperature management (TTM) between 32°C-36°C for at least 24h is recommend,[56] as well as avoidance of fever (≥37.6°C) which is common within 48h following cardiac arrest. TTM is a potent treatment that may prevent secondary brain damage, but many questions remain to be answered. The optimal post-resuscitation care, including controlling body temperature, still remains controversial and questions regarding which temperature to target, how long to deliver temperature control, the optimal way of rewarming and whether different target temperatures might be more appropriate for different patients or in different risk-groups are still unanswered. In the near future we will have the results from the TTM2 trail, which started in December 2017 and has randomised 1900 OHCA patients with a presumed cardiac or unknown cause, to compare TTM at 33°C with targeted normothermia and early treatment of fever (defined as ≥37.8°C).[73]

Prediction of ICU patients in general

The ability to reliably, and as early as possible, predict future outcome in patients may be considered a holy grail and a cornerstone of modern medicine. In an ICU perspective, the matter is more complicated with patients that are often unconscious, and with a wide spectrum of comorbidities and varying clinical status. The case-mix (the composition of the ICU cohort) of admitted patients is highly heterogeneous and challenging in terms of assessment, treatment and prediction of later outcome. The problem involving case-mix heterogeneity was recognised already in the 1980's, when it became obvious that hospital mortality in ICU patients, varied considerably between hospitals. These differences and variations were of course multifactorial and could have many explanations including level of training in ICU-staff, different infrastructure (i.e. availability of advanced medical equipment or ICU design) and aspects of organisation (i.e. management strategies or patient-nurse ratios). Many of these variables and their possible association to quality and outcome are still important to identify due to the fact that they can be influenced and changes in the overall quality of care can be achieved. But even more important, from an outcome assessment and evaluating perspective, is the possibility of adjustment for differences in case-mix in order to ensure that cohorts with different characteristics and attributes are correctly compared with each other.

A variety of different scoring models have been developed over the years to predict outcome in critically ill patients and the most common measurement of outcome is the risk of in-hospital mortality. Other purposes can be to characterise illness severity and level of organ dysfunction, or to assess resource use. The first ICU scoring systems available in a scientific context enabled correction for differences in cohort characteristics (gender, age, time spent in hospital before ICU admission etc.), physiological measures (body temperature, vital parameters etc.) and selected biomarkers in blood tests (lactate, electrolytes, blood-gas analyses etc.) and they were carefully chosen using rigid statistical processes. Eventually, combinations of variables with high impact on outcome (survival) were identified and collected in standardised sets to be used clinically. This standardised way of describing ICU cohorts has gained general acceptance, and most ICUs today use one or more scoring systems to describe their patients as well as estimating mortality. The term often used in these scores, to describe predicted probability of dying, is the estimated mortality rate (EMR), whereas the corresponding outcome is called the observed mortality rate (OMR). The standardised mortality ratio (SMR) is the ratio between OMR and EMR (or the predicted mortality rate - PMR).[74]

The Acute Physiology and Chronic Health Evaluation (APACHE), the Mortality Probability Model (MPM) and the Simplified Acute Physiology Score (SAPS) are all examples of implemented and widely used scoring systems to group patients according to severity of illness,[75] whereas the Sequential Organ Failure Dysfunction (SOFA) score, is designed to describe organ dysfunction.[76]

Over the years, the scoring systems have continued to develop and continuously been re-evaluated by recurrent calibrations. The rationale behind the need for calibration, is changes in case-mix of ICU patients over time and the adoption of new treatment alternatives that may influence outcome. Modern implemented scores do not only collect physiological variables, but also use data presented on hospital and ICU admission, including patient's background information, previous and existing comorbidities, clinical status on admission etc.

The different scores have different characteristics and collect data at different time points before or during the ICU stay. The SAPS score is commonly used for ICU prognostication in adults (≥16 years of age) admitted to a general ICU. The first version of SAPS was designed already in 1984 and has since then undergone transitions and updates using larger cohorts from various geographical regions to calibrate the score and better reflect the current ICU population. In the latest version – SAPS 3, three different categories or boxes of data, are collected: patient characteristics before ICU admission (box 1), circumstances of ICU admission (box 2) and presence and degree of physiologic derangement at ICU admission (box 3)[77] as shown in figure 5.

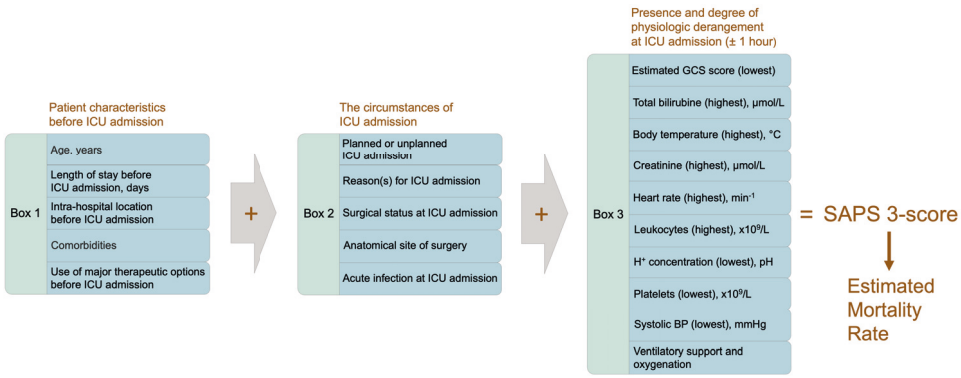


Figure 5. The Simplified Acute Physiology Score version 3 (SAPS 3). SAPS 3 consists of three boxes with sets of clinical variables related to ICU admission. Each variable is transformed into a numeric value and summed up to a SAPS 3 score and then transformed into a probability of death (EMR – Estimated Mortality Rate). Ventilatory support includes invasive ventilation and non-invasive mechanical ventilation. Oxygenation is derived from the ratio beween the partial pressure of oxygen in arterial blood and the fraction of inhaled oxygen (PaO_2/FiO_2). ICU, intensive care unit; GCS, Glasgow coma scale; BP, blood pressure.

Each variable from the three boxes is transformed into a numeric value and summed up to a final SAPS 3 score and then transformed into EMR. In Sweden, the SAPS 3 model has been calibrated to predict '30-day mortality', i.e. the probability to die within 30 days after ICU admission, instead of 'in-hospital mortality', as in the original SAPS 3 model. This model has shown a good capability to discriminate with an area under the receiver operating characteristic curve (AUROC) of 85%,[78-80] and continues to perform well when calibrated for prediction of long-term mortality (90-days and 180-days, respectively). Model performance and the interpretation of AUROC is further described in the Methods section of this thesis.

However, there are challenges when using severity scoring models in clinical practice. They can still not be recommended for bedside predictions to guide medical decision, due the model's uncertainties on the individual patients level.[81] Situations where the current models perform poorly include conditions with very high mortality, and where mortality prediction based on variables available already at admission could optimise treatment decisions. Such conditions include OHCA in patients regaining ROSC and who are admitted to ICU for post-resuscitation care with neurological prognostication. Mortality in these patients appears to be independently associated to factors and variables not taken into account in the APACHE or in the SAPS 3 scores, indicating the need for alternative scores adapted for cardiac arrest patients to improve prediction.[82-84]

Knowledge and understanding of case-mix are crucial when assessing and comparing ICU cohorts and their outcome. The use of a reliable and well-calibrated prognostic model is important for description, design and analysis of observational studies in ICU patients. The now established generic severity scoring models are not ideal for mortality prediction of OHCA patients undergoing TTM, due to pitfalls related to the temperature variable that might require special considerations. In the modern era of big data and increased computational power, it may be worth reconsidering the current prognostic scores and develop new models with the ability to handle large amounts of routinely collected clinical information.

Prognostication of post-cardiac arrest patients in the ICU

Cardiac arrest patients are challenging to prognosticate and require special consideration during outcome assessment due to heterogeneity. The most common severity scoring models, such as APACHE and SAPS 3, tend to offer only moderate accuracy when predicting outcome.[82-84] Instead, there are other instruments recommended to assess functional neurologic prognosis in comatose survivors after cardiac arrest treated in the ICU. These instruments include (with variation between hospitals) four main modalities: clinical neurological examinations, prognostic serum bio markers of brain injury, neurophysiology (EEG and SSEPs) and neuroimaging (CT and MRI). Combined they constitute a multimodal neurological prognostication tool-box which aims to evaluate the extent of brain injury by assessing different aspects of the pathophysiological scenario (figure 6).

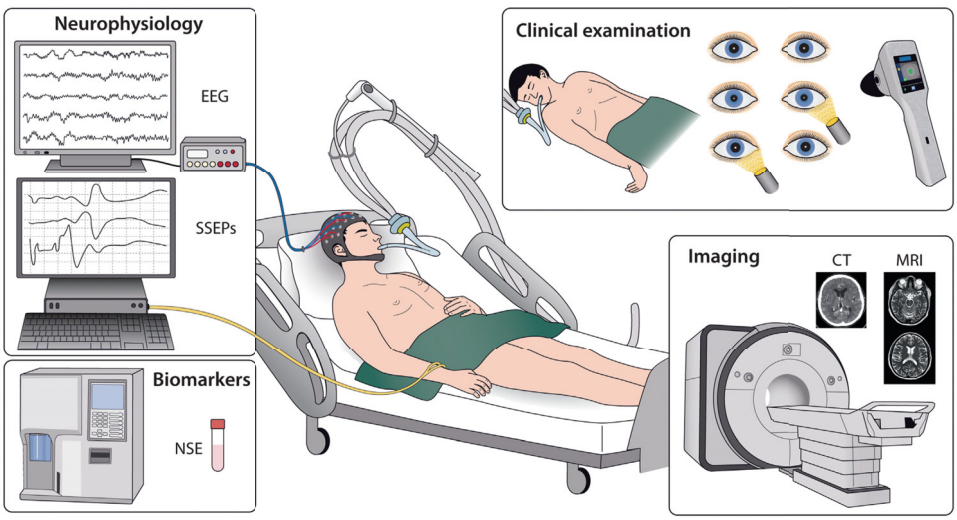


Figure 6. Multimodal neurological prognostication in comatose survivors after cardiac arrest. EEG, electroencephalography; SSEPs, somatosensory evoked potentials; NSE, Neuron-specific enolase: CT, computed tomography; MRI, magnetic resonance imaging. © European Resuscitation Council guidelines 2021 – www.erc.edu. Reprinted with permission.

Albeit demographic and clinical variables carry important prognostic information, none of them are included in the current recommended algorithms.[85] The multimodal prognostication should be performed not earlier than 72h after the arrest to avoid any confounding impact of prolonged sedation, hypothermia or metabolic derangement following TTM treatment during the acute phase immediate post-arrest.

Neurological prognostication primarily aims to avoid false-positive predictions (predicted poor outcome, observed good outcome), and secondarily to minimise false-negative predictions (predicted good outcome, observed poor outcome). The strategy for prognostication has constantly evolved since the first systematic approach published in 2006 by the American Academy of Neurology (AAN).[86] The initial model was based on studies from before the TTM-era and differs extensively from the present ERC/ESICM algorithm which is implemented in clinical practice and presented in figure 7. Prognostic tools, including chemical biomarkers, electrophysiological tests and neuroimaging, have improved greatly since the first algorithms and the present multimodal protocolised approach has improved the overall post-resuscitation care.

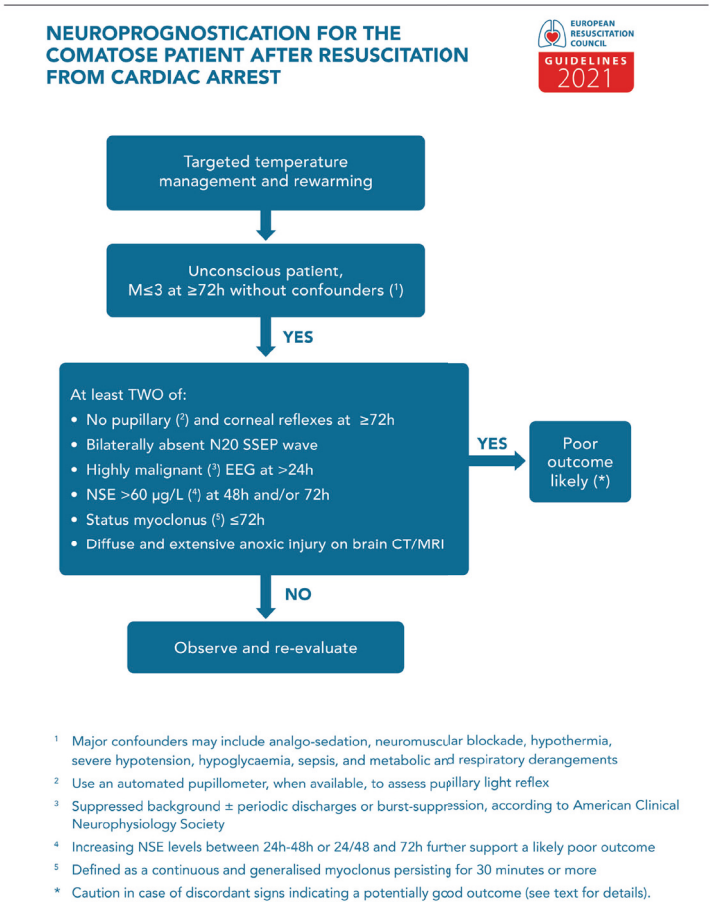


Figure 7. Neuroprognostication algorithm for the comatose patient after resuscitation from cardiac arrest. The algorithm is entered ≥72h after ROSC if the patient remains unconscious with a Glasgow coma motor score ≤3 and confounders are excluded. ROSC, return of spontaneous circualtion; M, motor response score on the Glasgow coma scale; SSEP, somatosensory evoked potentials; EEG, electroencephalography; NSE, Neuron-specific enolase; CT, computed tomography; MRI, magnetic resonance imaging. © European Resuscitation Council guidelines 2021 – www.erc.edu. Reprinted with permission.

Clinical neurological examination

Evaluation of consciousness

Coma or various levels of decreased consciousness despite ROSC, is a common manifestation in cardiac arrets patients and an indication for post-resuscitation care in the ICU. The Glasgow coma scale (GCS) is a commonly used quantitative scale that assesses decreased consciousness of various aetiology including the post-cardiac arrest scenario. The GCS-scale consists of three subcategories (eye, motor and verbal response) which are scored depending on how the patient reacts and responds to stimuli (table 1). A fully awake, alert and lucid patient would be assigned the top score of 15, whereas a deeply unconscious and unresponsive patient would gain the lowest score of 3. In the updated ERC prognostication guidelines, a GCS motor response ≤3 after 72h post ROSC indicates a likely poor outcome.[56]

A prolonged comatose state is a marker of poor outcome and a major basis for unfavourable neurological prognostication. However, ICU treatment with TTM includes the use of continuous sedation and sometimes neuromuscular blocking agents for several days, which will affect evaluation of consciousness if prognostication is performed too early after the discontinuation of these drugs. Metabolic derangements and other organ dysfunctions may also complicate the assessment and need to be excluded as confounders. The majority of post-cardiac arrest survivors will wake up within the first week,[87] but late awakenings are not uncommon and may involve the risk of an inaccurate false prediction of a poor outcome.[88]

Table 1. The Glasgow Coma Scale (GCS).

Subcategory	Points	Criteria
Eye response (E)	4	Eye open spontaneously
	3	Eye opening to verbal command
	2	Eye opening to pain
	1	No eye opening
Motor response (M)	6	Obeys commands
	5	Localising pain
	4	Withdrawal from pain
	3	Flexion response to pain
	2	Extension response to pain
	1	No motor response
Verbal response (V)	5	Oriented
	4	Confused
	3	Inappropriate words
	2	Incomprehensible words
	1	No verbal respons

Brainstem reflexes

The brainstem is the least susceptible area to anoxic injury and affected brainstem reflexes have potential to recover. Patients with poor outcome may present with a fully or partially intact brainstem, whereas clinical signs of brainstem dysfunction commonly reflect more severe brain injuries. To assess brainstem functions in unconsciousness patients pupillary (light) reflexes and corneal reflexes are examined routinely. The pupillary reflex is used to investigate sensory and motor functions of the eye, and the corneal reflex is used to test the fifth and seventh cranial nerve (the trigeminal and facial nerves), respectively. Sedatives and neuromuscular blockage may falsely reflect absent brainstem reflexes in patients with a good outcome.[89,90] However, absent corneal reflexes combined with bilaterally absent pupillary reflexes, predicted poor functional outcome with no false positive predictions and with narrow 95% confidence intervals in a TTM-material.[90] These examinations are therefore recommended in the prognostication algorithm.

Presence of post-anoxic clinical seizures

About one third of cardiac patients present with clinical seizures,[91] varying from generalised tonic-clonic convulsions to very subtle twitching of extremities or small parts of the face. Clinical seizures may be masked by sedation or neuromuscular blockage and thereby requiring neurophysiological tests to be detected. The same applies for non-convulsive seizures and status epilepticus (SE). Clinical seizures can often be treated with increased doses of sedatives or with anti-convulsive medication, including benzodiazepines and specific antiepileptic drugs. Electrographic seizures can be associated to a more unfavourable outcome,[91] but if readily responsive to treatment the prognosis for recovery is improved.[92] The presence of electrographic SE, and also its treatment, may contribute to prolonged unconsciousness and thereby constitute a confounding factor in cardiac arrest patients with late awakening.

Myoclonus

Myoclonus are brief, frequent, repetitive and involuntary muscle jerks or twitching, often located in extremities, chest musculature or small parts of the face,[93-96] and with an early onset after anoxic cerebral insults. The exact pathophysiological mechanism is still unclear. Status myoclonus indicates generalised myoclonic seizures for >30 minutes independent of electrographic findings and are associated with a poor neurological outcome if occurring within ≤48h post-arrest.[93,95,96] Myoclonus can be either epileptic or non-epileptic depending on the localisation of brain injury. A specific form of persistent action myoclonus is the Lance-Adams syndrome which is rare condition sometimes seen after hypoxic brain injury and is associated with overall good neurological recovery.[96,97] This condition emphasises the importance of a multimodal approach in post-cardiac arrest prognostication.

Prognostic serum biomarkers of brain injury

Prognostic biomarkers, not only used for the specific assessment of anoxic brain injury, are a growing field in modern medicine. Biomarkers can be used to detect and monitor the course of illness, but also to evaluate treatment effect and to tailor further therapy for the individual patient. Biomarkers are biochemical components from the cells, often released after cell damage to the blood, where they can be quantified and measured with serological tests. The absolute concentration of brain injury biomarkers can also be measured directly in the cerebrospinal fluid, where the concentrations may also be higher compared to serum levels. However, liquor analyses require lumbar puncture (spinal tap), which is an invasive and far more complicated procedure compared to peripheral blood tests.

Biomarkers reflecting the entire spectra of post-cardiac arrest syndrome include not only biomarkers related to brain injury, but also to cardiac injury and to systemic inflammation (table 2). Each biomarker varies in predictive capability and little is known about their prognostic values when combined.[98] For prognostication of neurological outcome in cardiac arrest patients, only one serum biomarker related to brain injury (NSE) is recommended to be used in combination with other prognostication modalities according to current guidelines.[56] The quality and characteristics of the different 'clinically accessible' and 'research-grade' prognostic biomarkers of brain injury, will be briefly described later in this section.

Table 2. Biomarkers reflecting the post-cardiac arrest syndrome and categorised after their main origin.

Brain injury	Cardiac injury	Systemic inflammation
NSE (Neuron-specific enolase)	**TnT** (Troponin T)	**PCT** (Procalcitonin)
NFL (Neurofilament light)	**BNP** (N-terminal pro-B-type natriuretic protein)	**IL-6** (Interleukin-6)
S100B (S100 calcium-binding protein B)	**Copeptin** (CT-proAVP)	
Tau (Tau protein/MAPT))		
GFAP (Glial fibrillary acidic protein)		
UCH-L1 (Ubiquitin carboxy-terminal hydrolase L1)		

Biomarkers of neurological injury are unaffected by sedative and muscle relaxant drugs, but has a major disadvantage in the lack of distinct and clear definitions of normal ranges (by age if applicable) and cut-off values related to outcome. Another important limitation is that cardiac arrest patients may die from other causes than brain injury, which are not always identified and will influence the association between biomarker levels and outcome. Efforts have been made to also add biomarkers associated with cardiac injury and systemic inflammation to the prognostication process, but none of these are included in the current guidelines at this point.[56,99]

Biomarkers of brain injury can be categorised according to their neuro-anatomical origin (neuronal body, axon and glial cell) as illustrated in figure 8, or according to their temporal appearance-profile in serum related to ROSC following OHCA (acute, delayed and sustained) as illustrated in figure 9.

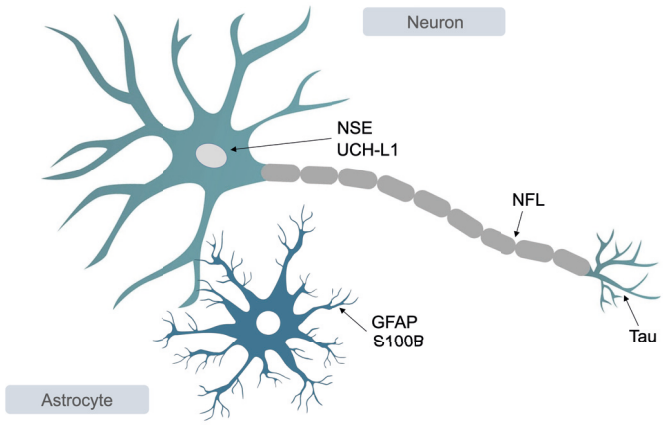


Figure 8. Biomarkers of brain injury. The figure is a schematic illustration of a nuron and an astrocyte (glial cell) with the primary structural location of biomarkers related to brain injury. Some of these are implemented as clinically accessible biomarkers whereas others are considered research-grade biomarkers, as described in paper III. Tau, Tau protein; NF-L, Neurofilament light; NSE, Neuron-specific enolase; UCH-L1, Ubiquitin carboxy-terminal hydrolase L1; GFAP, Glial fibrillary acidic protein; S100B, S100 calcium-binding protein B.

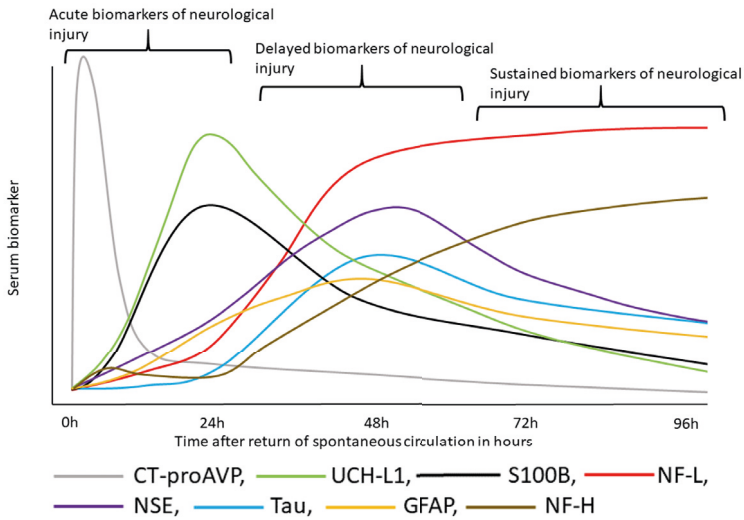


Figure 9. Schematic temporal profile of biomarkers of brain injury after OHCA. The picture is a schematic overview of the temporal trend of selected biomarkers devided in acute, delayed or sustained, by their relation to time after ROSC. CT-proAVP, Copeptin; UCH-L1, Ubiquitin carboxy-terminal hydrolase L1; S100B, S100 calcium-binding protein B, NF-L, Neurofilament light; NSE, Neuron-specific enolase; Tau, Tau protein; GFAP, Glial fibrillary acidic protein; NF-H, Neurofilament heavy; Illustration by Florian Ebner. Reprinted with permission.

NSE

Neuron-specific enolase is a 78 kDa intracellular glycolytic enzyme with a variety of neuroprotective properties.[100] It is present in neurons and neuroendocrine cells, but also in oligodendrocytes, erythrocytes and platelets. This is a source of error and an important limitation in situations of haemolysis, which is common in cardiac arrest patients, and a reason for the sample to be contaminated.[101] The same applies to numerous cancer types where NES serve as a marker, including small-cell lung cancer, renal cancer, liver cancer and colorectal cancer. This complicates the use of NSE as a prognostic marker for brain injury, even if several studies have indicated that elevated serum levels of NSE from 24-48h post arrest is associated with poor neurological outcome.[101,102] NSE is currently the only serum biomarker recommended to aid neurological prognostication after cardiac arrest according to ERC/ESICM guidelines. However, specific cut-off values are not defined and should therefore be established locally due to a lack of a calibration standards.[56]

NFL

Neurofilaments are CNS specific structural proteins highly expressed in large-calibre myelinated axons and may also play a role in intracellular transport to axons and dendrites.[101] Five different neurofilaments are identified and three of them are labeled after their molecular weight – heavy chain, medium chain and light chain (NFL, 68 kDa). After neuroaxonal injury, neurofilaments are released into the extracellular space and subsequently into the bloodstream, where it can be quantified and measured. Other locations for NFL include peripheral nerves and it can also be expressed in renal cancer. NFL has shown promising results in detecting the level of axonal injury in various neurological disorders.[103] It is a novel biomarker of neuronal injury and a robust predictor of poor outcome after cardiac arrest already at 24h.[104] As a single biomarker NFL present the most promising predictive capability among currently available biomarkers with an excellent discriminating capability when predicting poor neurological outcome already at 24h post-cardiac arrest.[103,105] In a post hoc analysis of the COMACARE trial, NFL predicted poor outcome with an AUROC of 0.98 at 24h, and continued to perform similarly at 48h and 72h, respectively.[105] In a prospective study evaluating data from the TTM trial, NFL was significantly elevated at 24, 48 and 72h in patients with a poor functional outcome following cardiac arrest and in the ROC analysis NFL was superior to serum tau, NSE and S100B.[104]

S100B

S100 proteins are a multigenetic family of several human calcium binding subtypes, including S100B which has been extensively studied as a prognostic biomarker of cardiac arrest and traumatic brain injury. These proteins are not CNS specific but present in most tissues with a large variety of different tissue-specific regulating effects. S100B are located in astrocytes, Schwann cells and dendritic cells, and acts as an extracellular signalling substance and a intracellular regulator.[106] S100 was elevated in patients with poor outcome after cardiac arrest compared to those with

good outcome, until approximately 72h post-arrest.[107-110] Highest prognostic accuracy was at 24-48h. S100B is not part of the ERC/ESICM recommendations for prognostication but is mentioned in the AHA's guidelines for post-resuscitation care as a possible predictor of functional outcome. Cut-off values are however not defined in these guidelines.[111]

Tau

Tau protein is a microtubule-associated stabilising marker with a weight varying between 48-67 kDa and is predominantly localised in the plasma membrane and nucleus of unmyelinated axons of the cortex, astrocytes and oligodendrocytes. Other locations include kidneys, liver, testis, muscles and peripheral nerves. Tau is enhanced in numerous cancers including glioma, breast cancer, renal cancer and prostate cancer. It is also expressed in a variety of different neurodegenerative disorders including parkinsonism and dementia as well as in traumatic brain injury. Increased serum tau values have been found to be significantly associated with poor outcome after OHCA. Tau has shown to be more robust to haemolysis than NSE, and intervention with TTM does not affect the accuracy of outcome prediction.[112]

GFAP

Glial fibrillary acidic protein is a 50kDa monomeric intermediate-filament component of the astrocytic cytoskeleton mainly expressed in the CNS. Other locations include Schwann cells of the peripheral nervous system, osteocytes, keratinocytes. Chondrocytes and Leydig cells of testis. GFAP has been extensively studied solitarily, and in combination with UCH-L1, as a prognostic marker of traumatic brain injury.[113] GFAP is also known to increase in both paediatric and adult post-arrest populations, but with varying predictive ability.[114-116]

UCH-L1

Ubiquitin carboxy-terminal hydrolase-L1 is a 26kDa neuronal deubiquitinase mainly localised in the neuron's cell body (in nucleoplasm and cytoplasm), in axons and synapses, where it is important for neuron-axonal stability and repair after brain damage. Other locations include neuroendocrine cells, the peripheral nervous system, endothelial and smooth muscle cells. It is enhanced in various cancer forms like urothelial and endothelial cancer. UCH-L1 combined with GFAP has been studied after traumatic brain injury, as described above. In terms of a prognostic biomarker following cardiac arrest, the scientific substrate is rather sparse. A pilot-study found the association between elevated UCH-L1 and poor outcome in a paediatric cardiac arrest population,[114] and in adult cardiac arrest patients UCH-L1 is not investigated.

Biomarkers in a future perspective

Prognostic serum biomarkers of brain injury possess enormous potential to revolutionise future post-resuscitation management, and especially to improve the prognostication process in the ICU. However, the expectations of an optimal biomarker related to cerebral injury for clinical use, are demanding: it must be easily accessible, be objectively measurable with clear definitions of range by age and gender, be stable ex-vivo, be able to discriminate between various degrees of brain injury with high accuracy and distinct cut-off values for outcome prediction, be highly specific for the CNS, be unaffected by haemolysis and finally, be cheap to analyse. So far, no available biomarker fulfils all of these criteria.

Future serum biomarkers of brain injury should not only be able to detect or to stratify ongoing neurological injury, but also its recovery and repair. They should provide valid and actionable real-time information to the clinician that reflects both injury progression and intervention response, in short-term and long-term perspectives. In this way, appropriate initial treatment can be defined, as well as tailored and modifiable therapy during critical post-resuscitation care and in later neurorehabilitation.

Neurophysiology

Electroencephalography

EEG is a non-invasive real-time recording of cortical electrical activity, and is the most commonly used method for prognostication after cardiac arrest.[117] It can be performed as an isolated examination, typically 20-30 minutes of recording, including full-montage of multichannel scalp electrodes (routine-EEG). Or alternatively as a continuous recording over time, but with fewer channels (cEEG). The full-montage offers a more detailed registration of the cerebral cortex in terms of background activity, discharges and electrographic reactivity from stimuli, but is to be considered a snapshot of brain activity. Recordings with cEEG provide information with lower resolution, but over a longer time period and the ability to detect intermittent findings and to assess the effect of sedation or antiepileptics.

EEG is strongly indicated for seizure detection in critical ill patients and is also recommended for prognostication in comatose survivors after cardiac arrest.[56] Comatose cardiac arrest patients often display a supressed (flat) background activity with low voltage pattern on the EEG, which can persist for hours or days after ROSC. During the recovery process intermittent discontinuous cortical activity is followed by more continuous activity. 'Highly malignant patterns' including burst-suppression background with or without discharges, suppressed background without discharges and suppressed background with continuous periodic discharges, are all reliable predictors of poor functional outcome.[118] The ERC/ESICM guidelines include two EEG patterns considered indicative of poor neurological outcome; 'unreactive burst-suppression' and 'unreactive status epilepticus'.[56]

The main limitation of EEG as a prognostic tool, is that it is influenced by sedative drugs commonly used during intensive care. There is also significant inter-rater variability between interpreters when assessing EEG recordings.[119] Historically, there have been significant challenges to report EEGs in both clinical practice and in prognostication studies, due to the lack of international standardised nomenclature. However, the suggested classification and terminology from the American Clinical Neuro-physiology Society (ACNS) has increasingly been used for critical care patients in recent years.[120]

A study by Attia et al. published in Lancet 2019 showed that an ECG algorithm based on machine learning could be trained to identify atrial fibrillation in patients with normal sinus rhythm with excellent accuracy.[121] This methodology could possibly be used in future clinical studies to detect associations between complex patterns in high-resolution EEG-recordings and later outcome.

EEG-recordings were not included as a predictive variable in this thesis as this modality was performed only on indication and in a minority of the analysed patients from the TTM trial, and thereby not optimal for machine learning models.

Short-latency somatosensory evoked potentials

SSEPs include bedside non-invasive tests with an electrical stimulus of an afferent sensor nerve, typically the median, and the registering of the responses that propagates to the contralateral somatosensory cerebral cortex. The bilateral absence of a cortical response indicates loss of integrity of thalamocortical projections (N20)[122] and is a reliable predictor of a poor neurological outcome with very high specificity (0-2% FPR). SSEP is recommended for post-cardiac arrest prognostication according to the ERC/ESICM guidelines.[56] Sensitivity is however low and reduces the method's utility.[85] In contrast to EEG, SSEP is not influenced by sedative drugs in moderate doses and is reliable regardless of TTM.

SSEP were not included as a predictive variable in this thesis as this modality was performed only on indication and in a minority of the analysed patients from the TTM trial, and thereby not optimal for machine learning models.

Neuroimaging

Computed tomography

CT is an easily accessible and commonly used modality for imaging in clinical practice. It is often performed early after hospital admission to exclude major intracranial haemorrhage, extensive ischaemic stroke and other cerebral pathologies leading to unconsciousness. A CT scanning of the brain within 24h after cardiac arrest is recommended by the ERC/ESICM guidelines.[56] Global cerebral oedema post-arrest is an indicator of hypoxic ischaemic brain injury and associated with poor neurological outcome.[85] The level of evidence is however low. Generalised oedema on CT is manifested as decreased cortical grey matter attenuation with a loss of normal grey-white differentiation and decreased attenuation in the bilateral basal ganglia. There is also effacement of the cerebrospinal fluid-containing spaces and of cortical sulci as seen in figure 10.

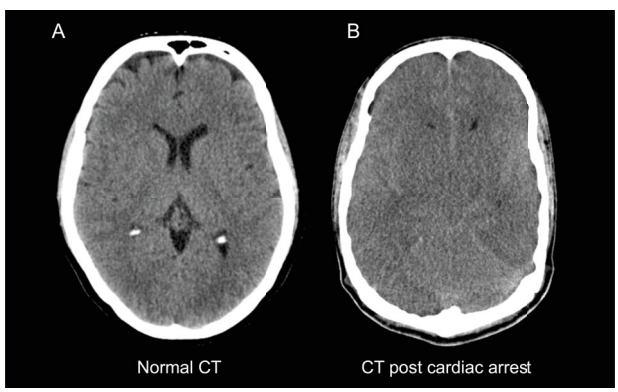


Figure 10. Computed tomography of the brain illustrating generalised cerebral oedema after cardiac arrest.

Magnetic resonance imaging

MRI has the advantage over CT in that it provides more detailed images with better spatial definition and with structured information regarding oedema and other manifestations of anoxic brain injury. Disadvantages include the need for logistical planning and that cardiac arrest patients may be severely unstable and therefore not able to undergo the investigation. Current ERC/ESICM guidelines suggest MRI within 2 to 5 days, in cardiac arrest patients who remain comatose and where a CT scanning did not reveal any significant injury.[56] The apparent diffusion coefficient (ADC) is a measure of the magnitude of diffusion (of water molecules) within tissues and its values are known to be significant predictors of poor outcome with very high specificity after cardiac arrest.[123] CT and MRI are not influenced by sedative medication or by TTM, but the presence of electrolyte derangements (hyponatremia), hypoglycaemia or epileptic seizures may mimic cytotoxic oedema on MRI.[124]

Neither CT nor MRI were included as predictive variables in this thesis as they were performed only on indication and in a minority of the analysed patients from the TTM trial, and thereby not optimal for machine learning models. However, machine learning is a powerful tool with the ability to analyse large amounts of data and also detect patterns in complex images. It should therefore be an ideal technique to develop reliable computer-assisted tools to aid MRI interpretation in future outcome studies as well as in clinical practice.

Assessment of cerebral functional outcome

Survival including long term functional neurological status is the most common outcome measure used to assess comatose OHCA patients who survive the initial resuscitation and treated with post-resuscitation care in an ICU. Neurological sequelae in OHCA survivors range from completely asymptomatic to an unresponsive vegetative state. Neither the optimal time to perform the assessment, nor the ideal reliably modalities to use, have been established in clinical practice. Neurological recovery caused by anoxic brain injury often continue for at least 6 months, even if the most evident recovery occur within the initial 3 months after the arrest.[125] However, for the individual patient the long-term effect and later potential improvement in functional outcome is unknown.[126,127] One recently published study indicated neurological improvement until 18 months post-OHCA, even if most recovery occurred within the initial 6 months.[128]

The Glasgow-Pittsburgh Cerebral Performance Category (CPC) scale is a widely used clinician-rated tool, designed to classify neurological status following brain damage (table 3).[129] It is commonly used to assess functional status after cardiac arrest. The CPC scale is divided into five categories of performance, in which CPC 1 indicates good cerebral performance with normal function or minor disability; CPC 2 indicates moderate cerebral disability, independent in activities of daily life; CPC 3 indicates severe cerebral disability and dependent on others for daily activities; CPC 4 indicates a patient in a coma or a vegetative state; and CPC 5 indicates death according to the Utstein guidelines.[130,131] The CPC scale can be transformed into a dichotomised CPC scale, where CPC1-2 is regarded as 'good' cerebral functional outcome including patients independent for daily activities but may have a minor disability, and CPC 3-5 is regarded as 'poor' cerebral functional outcome including patients dependent on others, in a coma, a vegetative state or with unresponsive wakefulness syndrome, and dead. This binary classification simplifies both outcome reporting and statistical analysis but also results in a loss of granularity. The dichotomised CPC-scale was used as primary outcome measure for the studies included in this thesis.

Table 3. The Cerebral Performance Category Scale.

Definition of binary outcome in this thesis	Cerebral Performance Category Scale
Good outcome	**CPC 1 – Good cerebral performance.** Might have mild neurologic or psychologic deficit. Conscious, alert, able to work.
	CPC 2 – Moderate cerebral disability. Conscious, sufficient cerebral function for independent activities of daily life. Able to work in a sheltered environment.
Poor outcome	**CPC 3 – Severe cerebral disability:** Wide range of neurolgical conditions where the patient is dependent on others in activities of daily life.
	CPC 4 – Coma or vegetative state: Any degree of coma without the presence of all brain death criteria. No interaction with the environment. Unresponsive wakefulness syndrome (may have spontaneous eye-opening and sleep/awake cycles).
	CPC 5 – Brain death: Certified brain dead or death by traditional criteria.

The CPC-scale is relatively simple to use in a clinical context but have been criticised for being poorly defined and for using subjective criteria not suitable for current clinical environment.[132] The scale has a relative insensitivity for mild disabilities with a subsequent tendency for survivors to be categorised as CPC 1 (the ceiling effect). An alternative scale used for assessment of cerebral functional outcome in cardiac arrest patients is the modified Rankin Scale (mRS) which was originally developed for stroke. The mRS is more detailed than the CPC-scale, it assesses both body functions and activity/participation, and is considered more suitable for the post-OHCA population (table 4).

Table 4. The Modified Rankin Scale.

Definition of binary outcome in this thesis	Modified Rankin Scale
Good outcome	**mRS 0 – No symptoms**
	mRS 1 – No significant disability. Able to carry out all usual activities desoite som symptoms.
	mRS 2 – Slight disability. Able to look after affairs without assistance, but unabel to carry out all previous activities.
	mRS 3 – Moderate disability. Requires some help, but able to walk unassisted.
Poor outcome	**mRS 4 – Moderately severe disability.** Unable to walk without assistance and unable to attend own bodily needs without assistance.
	mRS 4 – Severe disability. Bedridden, incontinent and requring constant nursing care and attention.
	mRS 5 – Dead

Determination of CPC and mRS categories can be done by clinical examination, face-to-face interviews or by patient chart review using standardised instruments.[133] Variation in inter- and intra-reviewer agreement may however influence reliability and generalisability of these scores.[134] There have also been discussions whether these scales are valid in both short-term and long-term follow-up. However, the TTM trial showed that the difference in neurological function between hospital survival and 180-day survival was limited.[66] Other studies have also indicated that CPC at hospital discharge is a useful surrogate measure of long-term survival.[135] Survival and the recovery to an acceptable neurological status is important aspects of outcome, from both a medical profession's, and the individual patient's perspective. However, impairment in cognitive and executive functions, including attention, memory, abstract reasoning, verbal fluency, motor and process skills, is common in TTM-treated OHCA patients.[136-138] These cognitive dysfunctions, as well as fatigue, are not taken into account in crude scales such as CPC or mRS, but may affect patient's quality of life and societal participation.[137,139] The extent of these cognitive disabilities is predominantly mild, but may affect the activities of daily living and further complicate the rehabilitation process. Nevertheless, many OHCA survivors report an acceptable quality of life 12 months post-arrest, particularly in comparison with population norms.[140]

Anxiety and depressive symptoms are reported in approximately 25% of cardiac arrest survivors,[141] and this percentage is doubled when patients are assessed by more detailed neuropsychological investigations.[142] These are circumstances with potential, not only to negatively influence the course of patient's rehabilitation and recovery, but to reduce their quality of life[143] and increase caregiver strains.[144] There is a significant association between cognitive impairment and lower participation, together with depression, fatigue and restricted mobility in OHCA survivors.[145]

In accordance with ILCOR's recommendations regarding evaluation of post-cardiac arrest outcome, different aspects should be considered including survival, neurological function and health related quality of life measures.[146] This indicates the need for even more nuanced outcome measures designed for the post-cardiac arrest population.

Stratification of illness severity in OHCA patients

The post-cardiac arrest syndrome is a heterogeneous condition with involvement of multiple organ systems, which causes substantial morbidity and mortality even after successful CPR.[47] The neurological injury, in terms of primary anoxic brain damage and secondary post-cardiac arrest brain injuries, is a significant determinant of later functional outcome.[147] It contributes not only to a substantial portion of in-hospital morbidity and death,[47] but also to a significant impact on quality of life for those who survive and can be discharged from hospital,[137,139] as earlier described in this chapter. The idea that patients with different severities of neurologic injury require different types of treatment may appear evident, but there have only been sparse efforts to overcome this problem in post-resuscitation care. Possibly as a remnant from an era where there was very little or no therapy to offer these patients. Beside the brain damage itself, there are other aspects of cardiac arrest to address including the link between systemic pathophysiologic post-arrest changes and neurological recovery.

Accurate prognostication post-cardiac arrest is crucial to avoid pursuing futile intensive care interventions when poor outcome is inevitable, but primarily to avoid inappropriate WLST in patients who may otherwise have a chance of achieving meaningful and acceptable neurological recovery. Risk stratification can also be used in the OHCA population for purposes other than bedside outcome prediction including selection of patients for early coronary angiography,[148] rapid transport to cardiac arrest centres[149] or to guide decision-making to intervene with more advanced high-risk invasive therapies, such as ECMO.

Numerous attempts have been made to create robust, yet straight forward scoring systems to classify severity of illness in OHCA patients already on admission to hospital or the ICU, and with greater precision than the traditional classification models (APACHE and SAPS 3) which are known to underperform in this population.[82-84] Several models using logistic regression statistics have been proposed and typically show moderate to good accuracy, including the 'CAHP (Cardiac Arrest Hospital Prognosis) risk score',[150] the 'RACA (ROSC after cardiac arrest) score',[151] the 'OHCA risk score',[152] the 'MIRACLE$_2$ score',[153] and a prediction model published by Eertmans et al. in 2018.[154] The 'TTM-risk score' based on data from the TTM trial, using ten independent predictors associated with a poor outcome including death at six months after OHCA, managed to achieve good discrimination of outcome[43] and was compared to an alternative model based on machine learning and included in this thesis (paper II). So far, none of these models have been precise enough to be used for the prediction of individual patients.

The application for an ideal scoring system would be to serve as a clinical tool to improve risk-adjusted outcome prediction and reliably assess individual risk of a poor functional outcome. This could have valid clinical implications for early allocation to specific interventions (tailored therapy) and later in the clinical course to inform prognosis and optimise continued life support. Risk-classification also has

advantages in a strictly scientific perspective, not least when performing clinical outcome-studies in OHCA cohorts. Inaccurate neurological prognostication leading to WLST and premature deaths, may significantly bias clinical studies with a subsequent failure in detecting the true study outcomes. Validated cardiac arrest-specific severity scoring model could also facilitate comparison between groups with different baseline characteristics.

Large pragmatic clinical trials have often been criticised for being too heterogeneous and possibly dilute any intervention effect that theoretically may be relevant for subgroups of patients.[155] The recent rapid development of computational power and the increasing interest in machine learning-based algorithms as an alternative to conventional regression models, have shown benefits when analysing large high-resolution data registers.[156,157] A machine learning algorithm can be developed and trained to capture non-linear feature correlations and patterns in big-data collections where a classic statistical approach would fail or at least have more difficulties in detecting dependencies between variables. To be able to detect subgroups in an OHCA population with an increased risk of a poor outcome, or subgroups that may benefit from a specific intervention or need extensive rehabilitation, further studies based on large high-resolution data sets from heterogeneous clinical trials are necessary to demonstrate significant associations.

This may be an important area of cardiac arrest research and a future utility for machine learning to create reliable scoring models for early classification of illness severity and thereby reinforce those currently used in clinical practice and clinical trials. The results from one of the studies (paper II) included in this thesis showed that a supervised machine learning model could be used as a statistical tool to stratify a heterogenous trial population in risk classes and help determine intervention effect across subgroups. Artificial intelligence and the use of machine learning models will be further described in the following section of this chapter.

Artificial neural networks (ANN)

Artificial intelligence (AI) is a branch of computer science devoted to the performance of tasks that normally require human intelligence. The theory is complex and will not be fully explained in this thesis. A major subbranch of AI is machine learning in which computers learn to perform tasks by analysing data rather than requiring specific programming instructions, i.e. they generate their own decision-making algorithms.[158]

Most machine learning algorithms used in medicine are trained by means of a process called supervised learning which have given rise to the concept of supervised machine learning. This means that the computer is presented with 'examples' that have been labelled using an external standard that serves as the "ground truth". During development of the machine learning models used in our studies, this labelling was achieved by exposing the computer to numerous of 'examples' from OHCA registries, i.e. patients with different combinations of variable values and the corresponding outcome of each patient. By repeating this training process over and over in order to learn, and punishing the computer when making incorrect predictions, an algorithm is eventually developed with an ability to categorise between good or poor outcome. As described in further detail in the 'Methods' section, the training set (a patient registry) is split and a randomly chosen test set is set aside for later testing. The process when a trained algorithm is exposed to the test set is called internal validation, and is crucial to assess accuracy of the algorithm but also its generalisability. The trained algorithm can also be exposed to a completely separate test set (a different patient registry) which is called external validation, and provides further assessment of the algorithm's generalisability.

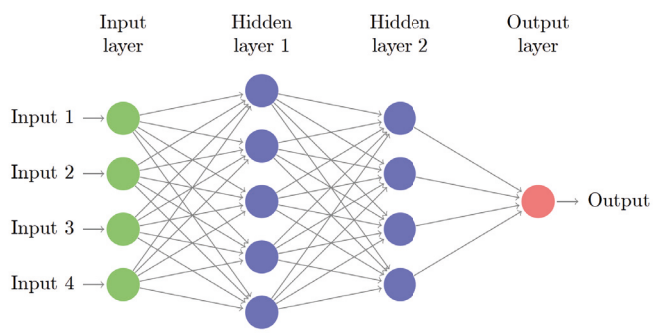


Figure 11. Artificial neural network (ANN). A schematic ANN with an input layer with four nodes, two hidden layers with five and four nodes and an output layer with one node. All nodes are connected to the previous and next layers by weights which resemble the connection between neurons in the human central nervous system. In total, an ANN as illustrated above will have 54 weights when the bias nodes (not shown) are included. The input layer in our studies was clinical variables (background, prehospital, admission, biomarker-data) whereas the output layer was the binary outcome variable (dichotomised CPC-scale into good or poor outcome). CPC, Cerebral performance category.

ANNs is a type of supervised machine learning model, inspired by the biological network of connected and communicating neurons in the human brain as illustrated in figure 11. ANNs can be seen as learning algorithms that model input-output relationships and replicate the way humans learn. The algorithms are functions or approximators created through advanced statistical techniques and designed to acquire their own knowledge. By extracting useful patterns from a data set, the models are mapping inputs to outputs and model non-linear relationships between them. The ability to detect more complex dependencies and patterns between variables in data compared to conventional regression models, has led to the application of ANNs in pattern recognition and prediction. Due to the numerous connections (weights) between nodes in the network, they can receive, transform and send information forward throughout the layers. The ANN strives during training to minimise the difference between the prediction and the actual outcome (i.e. the error) by slightly adjusting the weights by means of two key elements, 'gradient descent' and 'back-propagation'. Gradient descent is an optimisation algorithm used to find the values of a function's parameters (coefficients) that minimise a cost function. First the initial parameter's values are defined and from there gradient descent uses calculus to iteratively adjust the values to minimise the error. One type of gradient descent is 'Stochastic gradient descent', an extension of the gradient descent optimisation algorithm for minimising a loss function of a predictive model on a training dataset. Stochastic gradient descent can be used to train (optimise) many different models, including linear regression and logistic regression, and is the most efficient algorithm discovered for ANNs. Back-propagation refers to the method for computing the gradient, for the weights in a neural network graph structure, starting from the output layer and moving backwards.

Fundamental for machine learning models is, that weights connecting the nodes are only adjusted by the algorithm itself and are not changed by the user. This auto-adjustment process is the way the computer recognises and learns patterns between the input and output variables. However, the user can control some parameters (hyperparameters) in the search for the best possible prediction model. Hyperparameters include; number of nodes in each layer, numbers of hidden layers, how each node handle information from the previous layer (activation function) and how many times the data set pass through the network (number of epochs) and more. To select the optimal hyperparameters is essential to create the best training conditions for the algorithm to perform and can be done manually or be automated by using a random search, a grid search or Bayesian optimisation. ANN was used in paper II-IV to predict functional outcome in registries of OHCA patients.

eXtreme gradient boost (XGBoost)

XGBoost is another machine learning algorithm based on gradient boosted decision trees which are designed and known for their speed and performance.[159] This highly effective and widely used model has shown to be very successful in outperforming other machine learning algorithms in a multitude of different types of problem-solving. This is especially evident when analysing small to medium-sized and tabulated data sets where ANN models might encounter difficulties and possibly perform worse. The XGBoost algorithm has recently been dominating applied machine learning and Kaggle competitions for structured or tabular data, and the algorithm goes by lots of different names such as gradient boosting, multiple additive regression trees, stochastic gradient boosting or gradient boosting machines. Boosting is an ensemble learning technique where new models are added to correct the errors made by existing models. Models are added sequentially until no further improvements can be made. A popular example is the AdaBoost algorithm that weights data points that are hard to predict. Gradient boosting is an approach where new models are created that predict the residuals or errors of prior models and then added together to make the final prediction. It is called gradient boosting because it uses a gradient descent algorithm to minimise the loss when adding new models. Sometimes, it may not be sufficient to rely upon the results of just one machine learning model. Ensemble learning offers a systematic solution to combine the predictive power of multiple learners. The resultant is a single model which gives the aggregated output from several models. The models that form the ensemble, also known as base learners, could be either from the same learning algorithm or different learning algorithms. Bagging and boosting are two widely used ensemble learners. Though these two techniques can be used with several statistical models, the most predominant usage has been with decision trees. XGBoost was used in paper IV to predict functional outcome in registries of OHCA patients and to compare its performance and accuracy to an ANN model.

Machine learning in modern medicine

During the last decade, the increased computational power and constantly improving algorithms have led to a renaissance for machine learning to analyse large data sets. Machine learning has also been found valuable in various medical areas and clinical settings including interpretation of ECG-patterns and MRI, detection of cardiac arrest in emergency calls or in the emergency department, prediction of outcome in traumatic brain injury and prediction of the need for critical care as an alternative to conventional triage and early warning scores.[121,160-163] It has also been suggested for risk and mortality prediction in patients admitted for intensive care as an alternative to SAPS 3.[157] Recently, machine learning models have been used to predict the outcome in OHCA cohorts with high accuracy early in the chain of survival, where overall mortality is above 80%,[164,165] but these models are not applicable to patients admitted to ICUs after OHCA.

Machine learning has a great deal of promise and is undoubtedly an enormous potential for future big data analysis. Models with ability to reliably stratify risk and classify illness severity in heterogenic patient groups to improve prognostic performance and enhance tailored care would be a great progress. These techniques will hopefully be implemented in future clinical contexts to increase the speed and consistency of diagnostic procedures and develop prediction models as the algorithms continue to improve in accuracy.

Ethical considerations

Patients admitted to hospital after being resuscitated from cardiac arrest are predominantly in a comatose state and unable to provide informed consent for intensive care treatment or approve participation in clinical trials.

It is a challenging, yet important task for the clinician to evaluate and reliably assess the individual patient's opportunity for an acceptable neurological and functional recovery after severe illness such as cardiac arrest, with only limited information available. There are also some important ethical considerations to address when treating comatose cardiac arrest patients in an ICU. These considerations combined with clinical information based on the multimodal prognostication process, will establish the foundation for a clinical decision-making to continue the treatment, or withdraw life-sustaining therapy if prolonged intensive care is not considered being in the patient's best interest.

Ethical considerations should be used, not only in strict clinical contexts during treatment of cardiac arrest patients or in decision-makings on level of care, but also when recruiting patients for clinical trials. They can be defined in four principles of biomechanical ethics to balance and regard during post-resuscitation care.[166]

- *Autonomy* – to obtain informed consent if possible and respect the patient´s expressed or presumed wishes.
- *Beneficence* – to act in the best interest of the patient and provide treatment for the possibility of a good neurological outcome.
- *Non-maleficence* – to not intentionally injure the patient or to provide life-saving treatment in patients with obvious risk of extensive brain injuries and poor quality of life.
- *Justice* – to avoid utilisation of public resources in the treatment of patients with no benefit of the provided care and to strive for a fair distribution of available hospital resources.

Accurate prognostication methods to be used for early reliable outcome prediction in the heterogenic cardiac patient population with high mortality rates, are crucial to identify patients who will benefit the most of resource-intensive, invasive and potential perilous interventions. From an ethical perspective, it is important to withdraw intensive care as soon as possible, in patients for whom further treatment is deemed futile. Robust prognostication methods as well as elaboration of international standardised processes for WLST, are important to minimise the risk of premature use of WLST as part of a self-fulfilling prophecy. The effect of early withdrawal of life sustaining therapies (eWLST) on outcome is considerable as reported in observational studies on comatose OHCA patients.[88,167-169]

Aims of thesis

The overall aim of this thesis is to take a step forward towards improved and more reliable prediction of functional outcome for out-of-hospital cardiac arrest patients who survive the initial resuscitation and is observed and treated with post-resuscitation care in the ICU during the prognostication phase. It also aims to demonstrate the need for specific severity scoring models designed for cardiac arrest patients undergoing temperature intervention to better stratify analyses of clinical trials and to enhance tailored post-resuscitation care in the future.

- To analyse if varying levels of TTM after OHCA were associated with later functional outcome in an international observational registry (INTCAR2.0).
- To investigate which variables and clinical data that carry the most predictive information in an OHCA population.
- To use ANNs to create a model for early prediction of long-term functional outcome in comatose OHCA patients admitted to ICU, and use this model to investigate the intervention effect in cardiac arrest patients treated with TTM.
- To investigate whether cumulative information obtained during the first three days of intensive care of comatose OHCA patients, increased the predictive performance of ANNs, with and without clinically accessible and research-grade biomarkers.
- To compare the predictive performance of an ANN with another machine learning model called eXtreme Gradient Boost (XGBoost).
- To investigate the generalisability of two different machine learning models, by internal validation on the development cohort, and external validation on a separate OHCA cohort.
- To demonstrate the hazard of adjustment for SAPS 3 scores in outcome-studies on temperature interventions.

Methods and materials

The first paper in this thesis is a retrospective analysis of prospective data in the International Cardiac Arrest Registry 2.0 (INTCAR 2.0) while the second and third papers are post hoc analyses of the Target Temperature Management trial (TTM trial).[66] The fourth paper are post hoc analyses of a merged data set based on the INTCAR 1.0 and 2.0, and the TTM trial data set. This chapter summarises the materials and methods described in papers I–IV. Finally, the last publication V (Letter to the Editor) is the result of an observation, based on the analysis of a simulated data set with created cardiac arrest cases. Detailed descriptions of the materials and methods used in the four studies are presented in each paper. The materials and methods used are summarised in table 5, which also includes the number of participants for the final analyses.

Table 5. Overview of the four studies included in the thesis.
Intensive care unit (ICU), Cerebral Performance Category scale (CPC). *Participants for final analysis. **The number of participants in study III varied based on the time point and biomarkers of interest.

Paper	I	II	III	IV
Design	A retrospective analysis of a prospective international registry	Post hoc analysis of an international randomised multicentre trial	Post hoc analysis of an international randomised multicentre trial	Post hoc analysis of an international cardiac arrest registry and an international randomised multicentre trial
Study population	Consecutive collected out-of-hospital cardiac arrest patients treated in the ICU	Specific ICU population: Comatose survivors of out-of-hospital cardiac arrest of a presumed cardiac cause	Specific ICU population: Comatose survivors of out-of-hospital cardiac arrest of a presumed cardiac cause	Consecutive collected out-of-hospital cardiac arrest patients and a specific ICU population with comatose survivors of out-of-hospital cardiac arrest from a presumed cardiac cause
	Adults	Adults	Adults	Adults
	2008–2017	2010–2013	2010–2013	2001–2017
Participants*	n = 1710*	n = 932*	n = 932**	n = 4431*, n = 931*
Variables	Background, pre-hospital, admission, in-hospital variables	Background, pre-hospital and admission variables	Similar to study II and at day 1, 2 and 3 after ICU admission	Variables available on hospital admission
Outcome	Binary hospital discharge functional outcome: CPC 1–2 or CPC 3–5	Binary 180 days functional outcome: CPC 1–2 or CPC 3–5	Binary 180 days functional outcome: CPC 1–2 or CPC 3–5	Binary hospital discharge functional outcome: CPC 1–2 or CPC 3–5
Method	Logistic regression	Artificial neural networks	Artificial neural networks	Artificial neural networks and eXtreme gradient boost

The Utstein-style reporting of cardiac arrest data

The term "Utstein-style" is synonymous with consensus reporting guidelines for resuscitation and originated from an international multidisciplinary meeting of the European Society of Cardiology, the European Academy of Anaesthesiology, the European Society for Intensive Care Medicine, and related national societies, held at the Utstein Abbey near Stavanger, Norway in June 1990. The purpose of this initial meeting was to develop, by consensus, uniform terms and definitions for OHCA resuscitation.[170] Since then, a standardised reporting style has also been introduced for IHCA and for paediatric cardiac arrest.[171,172] We chose to report data from the registries according to the Utstein-style template in this thesis. The Utstein reporting template establishes a standard framework for comparing systems of care for cardiac arrest. The Utstein template was completed in 2004 for both out-of-hospital and in-hospital cardiac arrest. The updated template for in-hospital cardiac arrest was developed by an international committee between 2013 and 2018 and follows the style of the prior 2015 update for out-of-hospital cardiac arrest. Essential elements were classified as core. Others were classified as supplemental. Core variables should enable reasonable comparisons between systems and are considered essential for quality improvement programs. Together with core variables, supplementary variables are considered useful for research. The six classes of data elements are hospital factors, patient variables, pre-event factors, cardiac arrest processes, post-resuscitation processes, and outcomes.

These Utstein-style template facilitate comparison of cardiac arrest registries throughout the world. However, in a recent validation study, the Utstein factors explained only 51% of the variation in survival to hospital discharge in an international material evaluating OHCA registries from 12 countries, for the period 1st of January 2006 through 31st of December 2011. The authors concluded that modifiable Utstein factors should be targeted to improve survival but that the observed variability in outcome is incompletely explained by the Utstein-style templates and that further studies are needed to identify the ideal constituents of cardiac arrest data registration.[173]

Registries and data sources

The International Cardiac Arrest Registry (INTCAR)

INTCAR is a multinational, internet-based registry of consecutive cardiac arrest patients (both in-hospital and out-of-hospital) admitted to large volume cardiac arrest centres and treated in intensive care unit (ICU) settings in the United States and Northern Europe.

The INTCAR 1.0 registry started in 2006 and was a collaborative effort between members of the North-American Neurocritical Care Society and a European network originated from the Hypothermia Network Registry.[174] In 2011 the registry was updated and revised into the INTCAR 2.0, which reflects more current practices than the original data set and it continued to enrol patients until 2017. INTCAR 2.0 contains even more "core" data points with additional resolution including demographic characteristics, pre-arrest condition, resuscitation characteristics, post-resuscitation therapies including targeted temperature management (TTM) and post-arrest outcomes including survival to discharge and long-term functional outcome after hospital discharge. It predominantly encompasses a prospectively registered sample of consecutive patients of all ages with cardiac arrest of both in-hospital and out-of-hospital origin admitted to intensive care, and where each participating centre treated patients according to local protocols. Data were derived from ambulance charts, admission journals, ICU observation charts and medical records from hospitals and rehabilitation centres. Prehospital data were defined according to the Utstein guidelines[175] and in-hospital data according to the extended Utstein guidelines for reporting post-resuscitation care.[176] Besides the main INTCAR 2.0 basic data set including patient and cardiac arrest characteristics, treatment methods and prognostication, the database also collected information regarding imaging, hemodynamic variables, cardiology studies and interventions, EEG and seizures, neurological short-term and long-term functional outcome as well as treatment methods and complications.

Ethics

The regional ethical review board in Lund, Sweden approved the registry (Protocol 2007/7 Dnr 272/2007) and local ethical approval was granted as per regulations of each participating hospital. Information about the study was provided to patients who regained consciousness or to next-of-kin if required by legal statute in each country.

The Target Temperature Management (TTM) trial registry

The TTM trial recruited a total of 950 patients from 2010 to 2013 in 36 ICUs in Europe and Australia. The trial included comatose (Glasgow Coma Scale (GCS) ≤ 8) adults (≥18 years of age) with a sustained return of spontaneous circulation (ROSC) after successful resuscitation from OHCA of presumed cardiac cause. Patients were admitted to ICUs and randomised to TTM at 33°C or 36°C using a web-based application.[66] Exclusion criteria were unwitnessed cardiac arrest with asystole on the initial ECG, ≥4h from ROSC to screening, limitations in therapy including do-not-resuscitate orders or known illness making survival to 180 days unlikely, previous bleeding diathesis, known or suspected intracranial haemorrhage or stroke, body temperature ≤30°C on admission, pregnancy, persistent cardiogenic shock despite medical interventions and mechanical assist, and pre-existing neurological disability (CPC3-4).[177] The primary outcome of the TTM trial was all-cause mortality until the end of the trial, defined as 180 days after the last participant was included (survival analysis). Secondary outcomes were a composite outcome of all-cause mortality and poor neurologic function according to CPC at hospital discharge and at 6 months, all-cause mortality at hospital discharge and at 180 days, neurologic function at hospital discharge and at 180 days, quality of life at 180 days, best neurologic outcome during the trial period, and safety measures.[66]

Figure 12 illustrates the trial study timeline from the cardiac arrest to assessment of neurological outcome after 180 days. The trial protocol was approved by ethical committees in each participating country, and informed consent was waived or obtained from all participants or relatives according to national legislation, in line with the Helsinki declaration.[177] Patient data were entered in an online electronic case record form and externally monitored on site.

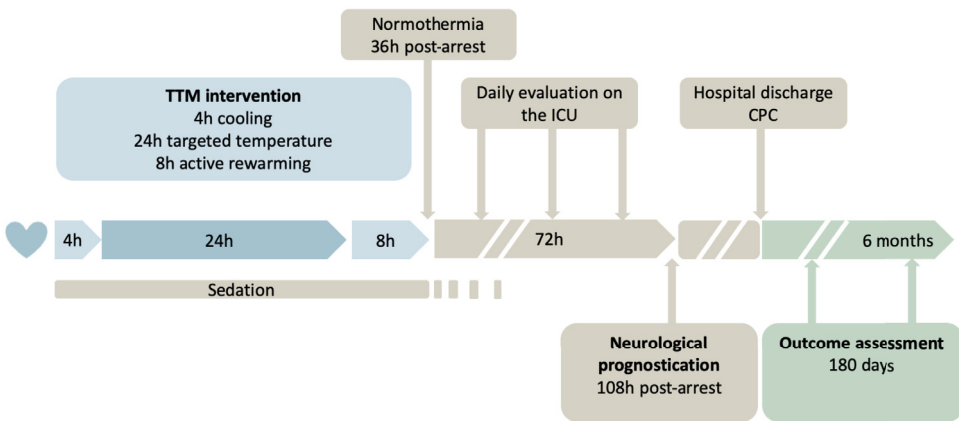


Figure 12. The TTM trial study timeline from cardiac arrest to assment of neurological outcome after 180 days. Illustration by Sofia Backman. Reprinted with permission.

The results of the main trial were subjected to sensitivity analyses for time, study centre, and other possible biases and have been elaborated in post-hoc analyses and sub-studies. All have shown similar outcomes in both temperature groups.[67,70,71,178] Therefore, the pooled TTM data set was used for the analysis in paper II, III and IV.

Ethics

Ethical consent was obtained by the ethics board of each participating country and followed the principles of the Declaration of Helsinki and its amendments.[177] The regional ethical review board in Lund, Sweden approved the registry (Protocol 2009/6 Dnr 324/2009). Informed consent was obtained from surviving patients. A patient's next of kind was informed of the inclusion in the study during the first contact with the hospital.

Variable selection and strategy

Paper I

In paper I we investigated whether the results from the TTM trial could be demonstrated in OHCA patients included in the INTCAR 2.0 registry containing cardiac arrest data where baseline variables allow for adjusted analyses. We chose to exclude variables with missing values of >5% and thus no imputation was performed.[179] INTCAR 2.0 data were derived from ambulance charts, admission journals, ICU observation charts and medical records from hospitals and rehabilitation centres. Prehospital data were defined according to the Utstein guidelines[175] and in-hospital data according to the extended Utstein guidelines for reporting post-resuscitation care.[176] Comorbidities were registered if they were pharmacologically or previously surgically treated, or subject to continuous monitoring at the time of the cardiac arrest. ROSC was defined as time from collapse until return of spontaneous circulation, leading to stable circulation without the need for CPR for at least 20 minutes. Temperature management was defined as an active attempt to keep the patient's body temperature within a prescribed target range. TTM at 32-34°C was defined as TTM-low and TTM at 35-37°C as TTM-high. Adverse events during ICU care were recorded according to a predefined protocol. The primary outcome variable was survival with good neurological function at hospital discharge, using the CPC-scale. The secondary outcome was adverse events related to TTM during ICU care.

Paper II

In paper II baseline comorbidities, demographics, prehospital data, arrest characteristics, physiological variables, as well as admission data, were systematically collected according to the Utstein criteria[175,176] and categorised as background-, prehospital and admission variables. Time from cardiac arrest to initiation of BLS and ALS was recorded. No-flow and low-flow times were defined as the time from cardiac arrest to the start of CPR (BLS or ALS) and the time from the start of CPR to ROSC, respectively. Time to ROSC was defined as the time from cardiac arrest to the first recorded time point of sustained (>20 min) spontaneous circulation. "No flow" (indicating the time from arrest until the start of cardiopulmonary resuscitation (CPR) and "low flow" (indicating the time from the start of CPR until the return of spontaneous circulation (ROSC)) are often used to describe the circumstances of the CPR treatment. Primary outcome of this study was 6 months functional outcome including survival using a dichotomised CPS-scale. Analyses were performed using all 54 variables available at admission to hospital to compare performance in predicting long-term outcome with a previous model based on logistic regression (the TTM risk score) as well as a compact 15-variable model designed to rank the 15 most important variables based on their individual importance.

Paper III

In paper III all variables from the TTM trial up to day three were included; background information, prehospital and hospital admission records along with data obtained at 24h (day 1), 48h (day 2), 72h (day 3). CT, MRI, EEG and SSEP were not included as these modalities were performed only on indication and in a minority of patients. Variables with >20% missing values were excluded. The TTM trial biobank collected blood samples from 29 of the 36 trial sites on day 1, 2 and 3 and comprised approximately 70% of the total TTM trial patient population. Biomarkers analysed in the biobank were grouped by whether they were considered clinically accessible or research-grade. Three models (A, B and C) were developed for each of the three days studied (a total of nine datasets):

- Level A: Clinical variables only
- Level B: A & clinically accessible biomarkers: NSE, S100B, TnT, BNP, and PCT
- Level C: B & research-grade biomarkers: NFL, copeptin, IL-6, tau, GFAP, and UCH-L1

To ensure that the prognostic value of the biomarkers would not be weakened by the imputation technique, we excluded patients with missing values corresponding to the exact day the data was missing for NSE and NFL in level B and C, respectively, which resulted in level B and C having approximately 30% fewer patients in each dataset. The outcome variable was a dichotomised CPC-scale graded by a blinded assessor after an interview face-to-face or by telephone at the six-month follow-up.[68]

Paper IV

In paper IV we used a combined data set from INTCAR 1.0 and INTCAR 2.0 to identify variables at the time of admission and registered in both data sets. To build a data set suitable for model training and to harmonise with the TTM data set, all patients under 18 years and those with arrest sites other than out-of-hospital were excluded. The variables registered in the INTCAR1+2 registry and in the TTM trial registry, slightly differed due to different registry structure. Some variables therefore had to be aligned and re-coded to enable the algorithms to perform the training and testing processes on both registries. The two final data sets included 29 comparable predicting variables available on time for hospital admission. The outcome variable was good or poor functional status at hospital discharge according to a dichotomised CPC-scale.

Statistics

Continuous variables are presented as mean with the corresponding standard deviation (SD) if normally distributed and as median and interquartile range (IQR) if non-normally distributed. Binary and categorical variables are presented as numbers and percentages. For all studies the null hypothesis (H_o) testing was performed using two tailed tests and a two-sided significance level of $p<0.05$ was considered statistically significant.

In paper I categorical data was compared using Chi-square tests, continuous normally distributed data were compared using Student's *t*-test and non-normally distributed data by the Wilcoxon-Mann-Whitney test. A univariate logistic regression was performed and presented as odds ratios (OR) with 95% confidence intervals (CI) indicating the association of the variable with a good outcome and OR-values >1 indicating a favourable association. A multivariate analysis was also performed using logistic regression with adjustment for important covariables with a potential to influence outcome after cardiac arrest. Missing values in selected covariates was below 5% and thus no imputation was performed. Goodness of fit was tested using the Hosmer-Lemeshow test. Propensity score analysis and sub-class matching to compensate for differences in background characteristics between intervention groups was not performed in study I due to unfulfilled positivity assumption. This is mandatory to remove treatment bias, but was not possible since the probability of receiving any of the two targeted temperatures was not equal due to country-specific treatment regime (treatment with 33°C more common in the US and 36°C more common in Europe).

In paper II-IV the Mann-Whitney *U* test was used for continuous data and Fisher´s exact test for categorical data when comparing groups. Missing values were imputed by using a simple mean or mode substitution (paper II and IV) and a median or mode imputation based on the statistics of the training sets (paper III). To evaluate performance of ANN and XGBoost models, we examined area under the receiver operating characteristics curve (AUROC) described in the 'To measure model performance' section of this chapter. Difference between AUC-curves and their corresponding CIs were calculated using bootstrapping and DeLong's method.[180,181]

Sensitivity and specificity

'Sensitivity' and 'specificity' are terms commonly used to demonstrate how well different tests or diagnostic methods perform in detecting disease in ill patients and ruling out disease in healthy patients. The corresponding rationale in this thesis is the ambition to create prediction models with best possible performance to identify variables associated with clinical outcome.

The meaning of sensitivity and specificity can sometimes be a bit confusing and is easily mixed up and misunderstood. The following scenarios are possible examples when a patient undergoes an examination or test for prognostication, and the corresponding result is related to later clinical outcome.

- A '*true negative'* assessment denotes a negative/normal assessment in a patient with a good outcome.
- A '*true positive'* assessment denotes a positive/pathological assessment in a patient with a poor outcome.
- A '*false negative'* assessment denotes a negative/normal assessment in a patient with a poor outcome.
- A 'false positive' assessment denotes a positive/pathological assessment in a patient with a good outcome.

Regarding neurological prognostication after cardiac arrest, it is crucial to conduct assessments with as high precision as possible. Primarily to eliminate the risk of withdrawing intensive care in patients with a possibility of later recovery, but also to minimise futile interventions in patients who are unrecoverable due to the extent of neurological injury. By combining various examinations and tests during the multimodal prognostication process, we try to rule out false positive predictions and avoid false negative predictions.

Sensitivity

The sensitivity of an assessment describes how well it identifies poor outcome patients, i.e. the probability of a positive/pathological assessment to result in patients with a true poor outcome. Only poor outcome patients are included when calculating sensitivity. For example, if the sensitivity of an assessment is 90%, it will identify 90% of patients with later poor outcome and miss 10%.

$$Sensitivity = \frac{True\ positives}{True\ positives + False\ negatives}$$

Specificity

The specificity of an assessment describes how well it identifies good outcome patients, i.e. the probability of a negative/normal assessment to result in patients with a true good outcome. Only good outcome patients are included when calculating specificity. For example, if the specificity of an assessment is 90%, it will misidentify 10% of patients with a later good outcome as unrecoverable.

$$Specificity = \frac{True\ negatives}{True\ negatives + False\ positives}$$

False positive rate (FPR) is sometimes used and denotes the rate of false positives among all cases that ought to be negative. This can also be expressed as 1–specificity. Occasionally the terms 'Positive predictive value' (PPV) and 'Negative predictive value' (NPV) are referred to in context of diagnostic tests.

Positive predictive value

PPV indicates the probability that the patient with a *positive/pathological* assessment truly *will* have a poor outcome. Only positive/pathological assessment results are included when calculating PPV. For example, if the assessment is positive/pathological with a PPV of 90%, there is on average 90% chance that outcome is poor and 10% that outcome is good.

$$PPV = \frac{True\ positives}{True\ positives + False\ positives}$$

Negative predictive value

NPV indicates the probability that the patient with a *negative* assessment truly *will not* have a poor outcome. Only negative/normal assessment results are included when calculating NPV. For example, if the assessment is negative/normal with a NPV of 90%, there is on average 90% chance that outcome is good and 10% that outcome is poor.

$$NPV = \frac{True\ negatives}{True\ negatives + False\ negatives}$$

Accuracy

Accuracy is the number of corrected assessments divided by all assessments.

$$Accuracy = \frac{True\ positives + True\ negatives}{All\ assessments}$$

To measure model performance

Prediction of two possible outcome is called binary classification. We used a dichotomised CPC-scale (good or poor functional status) as the binary outcome variable in all studies included in this thesis. The prediction models were trained to distinguish between the two outcomes, i.e. discrimination. The ability to perform in discrimination is measured using the AUROC and has a value between 0 and 1. The better the model can classify which outcome a patient will have, the closer the AUROC will come to 1 (maximum discrimination) and vice versa. An AUROC of 0.5 indicates a model with no ability to discriminate and not better than a completely random selection. There are no established definitions for the discriminatory accuracy of AUROC and the label of the ranges vary in the literature.[182] The following ranges were used in this thesis: AUROC 0.5–0.7 as poor performance, 0.7–0.8 as fair performance, 0.8–0.9 as good performance and 0.9–1.0 as excellent performance.

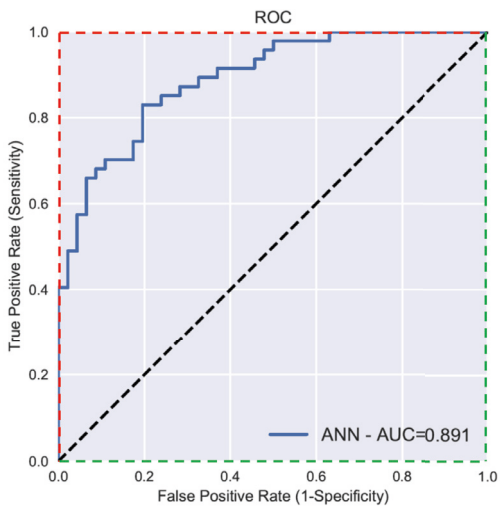


Figure 13. The receiver operating characteristic (ROC) curve visualised in blue color. The sensitivity is plotted on the *y*-axis (true-positive rate), and the 1-specificity (false-positive rate) is plotted on the *x*-axis. The area under the receiver operating characteristics curve (AUROC-curve) in this example is 0.891 (model from study II), i.e. an ANN model trained to a performance to predict functional outcome in cardiac arrest patients (destinguished between good or poor outcome) with an almost 90% accuracy. The dashed black line represents a model with an AUC of 0.5 indicating that the model will be right or wrong in 50% of cases which is not better than a random selection. The red and green dashed lines represent perfect tests with an AUC of 1 and -1, respectively, indicating that both models can descriminate between two outcomes in 100% of the cases. ANN, artificial neural network; AUC, area under the curve.

The ROC-curve is a technique for visualising, organising and selecting classifiers based on their performance. They are commonly used in medical decision making, and in recent years have been used increasingly in machine learning and data mining research.[183] The curve describes the model's capability to classify at different thresholds and offers a more comprehensive assessment including all possible

decision thresholds from a diagnostic test compared to the computing of sensitivity and specificity treating each value of the predictor as a possible cut-point. This is especially evident in the event of continuous or ordinal predictors. The true-positive rate (TPR) and the false-positive rate (FPR) are calculated at various thresholds and plotted as in figure 13. The TPR is also named sensitivity and is plotted on the *y*-axis. The FPR plotted on the *x*-axis is named 1 - specificity. In some models the 1-FPR is instead plotted on the *x*-axis corresponding to specificity. This presentation simplifies the visual interpretation of the interaction between specificity and sensitivity.

An overall ROC-curve is mostly useful in the early stages of evaluation of a new diagnostic test. Once the diagnostic ability of a test is established, only a portion of the ROC-curve (partial AUROC) is usually of interest, i.e. regions with high specificity (100-95%) and not the average specificity over all sensitivity values. This strategy is used to minimise the FPR in a specific diagnostic test and is highly prioritised in tests used for neurological prognostication since the result could lead to WLST and the death of a patient.[184] ROC-curves are also useful for assessing the predictive ability of two or more biomarkers for the same disease or for comparing performance between two prediction models as in paper II and IV. In general, the test with the higher AUC may be considered better, However, in cases where specific values of sensitivity and specificity are only clinically relevant for the comparison, then partial AUROCs are compared as described above.[184]

Both FPR and TPR are calculated based on the confusion matrix for the chosen threshold. The confusion matrix for a binary classification problem is a 2x2 table which displays the performance of a prediction model at a specific threshold (e.g. 0.5).

Table 6. The confusion matrix. The rows represent the predicted outcome, and the columns represent the observed outcome.

	Observed Positive	**Observed Negative**
Predicted Positive	True positives (TP)	False positives (FP)
Predicted Negative	False Negatives (FN)	True negative (TN)

As seen in table 6, each prediction can be either a true positive (TP), a false positive (FP), a false negative (FN) or a true negative (TN). These four measures are the foundation for calculating numerous performance measures of a model, such as the accuracy, precision (positive predictive value) and negative predictive value.

Calibration

Calibration describes the level of agreement between the predicted probability and the observed outcome in a model and is important to measure its performance. The calibration is often described in a calibration plot, with the predicted probability on the *x*-axis and the observed outcome on the *y*-axis. As the outcome is binary, the predictions are usually plotted by decile on the *x*-axis, with the corresponding observations on the *y*-axis.[185] A perfect calibration is represented by the diagonal of the calibration plot, where the prediction on the *x*-axis correlates perfectly with the corresponding observations on the *y*-axis. This measure was even further developed by Finazzi et al. when adding a calibration belt with confidence intervals (80% and 95%, respectively)[186] in the plot as seen in figure 14. The calibration belts also indicate if the prediction model is underestimating or overestimating the risk (over or under the bisector). To create precise calibration curves (narrow CIs), a minimum of 200 patients with and without the studied outcome is required.[187]

Other examples of calibration measures include goodness-of-fit tests, such as the Hosmer-Lemeshow test and Pearson's chi-squared test. A commonly used overall measure is the standardised mortality rate (SMR), which is the ratio between observed mortality rate (OMR) and the expected mortality rate (EMR).[74]

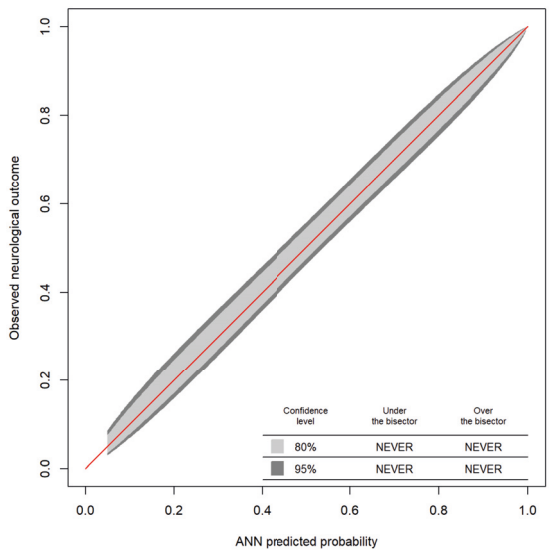


Figure 14. Calibration plot. An example of a calibration plot with the GiViTi calibration belts. This plot displays the ANN external validation model from study IV, which was well calibrated and with good level of agreement between predicted probability and observed outcome, i.e. the model did not overestimate or underestimate the risk of a poor outcome. The diagonal red line represents a perfect calibration. GiViTi, Italian Group for Evaluationof Interventions in Intensive Care Medicine; ANN, artificial neural network

Using machine learning to predict functional outcome

The very basics of machine learning and ANN is described in the 'Background' chapter. In this section the development of algorithms and the training/testing processes will be briefly explained and illustrated followed by detailed information regarding model development in study II-IV.

An ANN differs from logistic regression models by being highly adaptable when finding patterns, even in subsets of the data. The ANN can be trained to capture non-linear feature correlations in big-data collections where a classic statistical approach would fail, or at least have more difficulties in detecting dependencies between different variables. This adaptive attribute of the algorithms is a strength, but also a weakness due to the risk of overfitting. An overfitted model does not perform well when being tested in a different patient population or on a new data set, i.e. the generalisability of the model is low. There are several methods to avoid the problem of overfitting where splitting of the data set is a fundamental one.

The dataset is divided into a training set (used for model development) and randomly selected test set to provide an unbiased internal evaluation of the final model as illustrated in figure 15. A part of the training set can be allocated as a validation set to test the model performance during training. Changing the validation set during training based on predefined splits is called cross-validation (k-fold cross-validation). After finishing the training, the test set earlier removed, is used for internal validation of model performance but also to assess its generalisability. To implement the model as a prediction model, external validation should be performed in which a trained model is tested on a new data set as done in study IV.

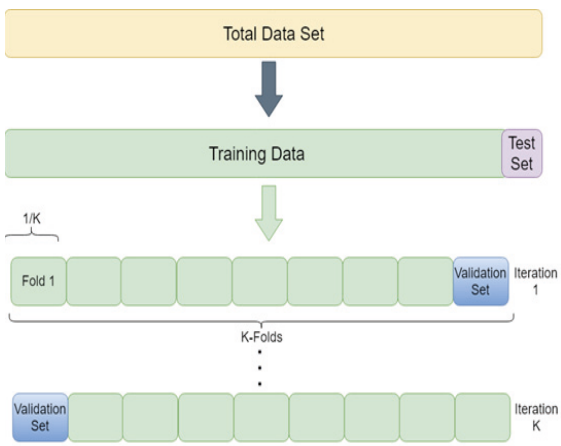


Figure 15. Training, validation and test sets. The figure illustrates how the data is split into training set and a test set (internal or external), and how the training set can be split into different validation sets (*k*-fold cross-validation). Illustration by Ola Björnsson. Reprinted with permission.

Model development

Study II

A prediction model was created to predict long-term functional outcome in comatose OHCA survivors using information available on ICU admission (background, prehospital and admission data), in 54 variables from the TTM trial registry. We also created a simplified prediction model by ranking all variables after their individual performance adding one variable at the time according to their relative importance.

A test set, corresponding to 10% of the data was randomly chosen and set aside to test the performance of the final ANN model. The remaining data (90%) was used for training and development. The training set was randomly divided into five equal-sized groups, to allow for cross-validation during model development.

Our ANN consisted of one input layer, a number of hidden layers and one output layer. A Bayesian optimisation approach, based on the Tree-structured Parzen Estimator (TPE), was used to find the best possible network architecture.[188] The search for optimal hyperparameters was performed with limits presented in table 7. The final model was chosen based on the AUROC of the cross-validations. The AUROC was reported using the test set data and was then compared to a logistic regression-based model's AUROC (after removing patients who originally had missing values)[43] using the method of DeLong's.[180,181]

Table 7. Hyperparameters during model development in study II. The predifined limits for hyperparameter tuning during development to find the best possible model using Bayesian optimisation.

Hyperparameters	Limits during model development
Number of hidden layers	1–4
Nodes in each layer	5–400
Batch size	1–128
Drop-out rate	0–0.3 for the input layer and 0–0.5 for the hidden layers
Norm regularisation	L_1, L_2 or Max-norm
Activation function for the hidden layers	Rectified linear unit (ReLU) or hyperbolic tangent function
Optimisation	Adam implementation of stochastic gradient descent or a slightly different version called Adam AMSgrad

The final model was also used to investigate the effect of TTM of 33°C vs 36°C based on patients' risk stratification and to see if any of the two temperatures was associated to outcome. Our aim was to investigate if TTM would be beneficial in any of the risk classes. This was presented as logarithmic diagnostic odds in risk class 0-20%, 20-40%, 40-60%, 60-80% and 80-100%, respectively.

Study III

Similar to study II a prediction model was created to predict long-term functional outcome in comatose OHCA survivors. However, in this study we used cumulative clinical variables along with clinically accessible and research-grade biomarkers collected during the first three days of ICU observation. We used biomarkers from the TTM trial biobank, which collected blood samples from 29 of the 36 trial sites, to create additional levels of biomarkers in addition to the clinical variables already available from the TTM trial registry.

An ANN model was developed for each of the nine datasets with three levels of biomarkers from 24h (day 1), 48h (day 2) and 72h (day 3) after ICU admission. This is further described in the 'Variable selection and strategy' section earlier in this chapter.

The data sets were randomly divided into a training set for model development (80%) and a test set for internal validation (20%). The randomisation key was created at the time of hospital admission; hence the split was the same for all models. The number of variables was reduced in each dataset by using a correlation threshold of 98%, a missing values threshold of 20%, a minimum incidence of 2% for unique binary variable events, and a wrapper variable selection method which combined the feature selection algorithm with Shapley values. As in study II a five-fold cross-validation was used and a Bayesian optimisation algorithm to find the optimal hyperparameter values as presented in table 8.

Table 8. Hyperparameters during model development in study III. The predifined limits for hyperparameter tuning during development to find the best possible model using Bayesian optimisation.

Hyperparameters	Limits during model development
Number of hidden layers	1–3
Nodes in each layer	5–250
Batch size	4–128
Drop-out rate	0–0.5 for the input layer and 0–0.5 for the hidden layers
Norm regularisation	L_1, L_2 or Max-norm
Activation function for the hidden layers	ReLU or hyperbolic tangent function

Model performance was reported by calculating the AUROC for the test set data and displaying the ROC for all models. To find the optimal probability threshold for cardiac arrest prognostication, we based the threshold on 100% specificity in the training set. The distribution of the confusion matrix of the test set was then reported.

Study IV

A prediction model was created to predict functional outcome on hospital discharge in comatose OHCA survivors using early information available on hospital admission from 29 selected variables in the merged INTCAR1+2 registry which . A random selection of 90% of the patients in the data sets was allocated for development and training of the models whereas the remaining 10% were used for internal testing to validate model performance. The TTM data set was used for independent external testing to validate the models' performance when exposed to a juvenile data set. Similar to study II and III the training set was randomly divided into five equal-sized groups, to allow for k-fold cross-validation during model development. Performing cross-validation entailed that each configuration of a model resulted in five "sub-models" trained on partially different data. This allowed for an ensemble-prediction on the test set and for this two metrics were used. The first metric denoted cross-validated (AUC) was used to determine the best model and was defined as the average of the AUC from each sub-model on the respective cross-validation set. The second metric denoted ensemble AUC was defined as the resulting AUC after averaging the prediction patient-wise from each sub-model on the test set. The ensemble AUC then was used to test the generalisability of the best model. A Bayesian optimisation approach was used to find the best possible hyperparameters as presented in table 9.

Table 9. Hyperparameters during ANN-model development in study IV. The predifined limits for hyperparameter tuning during development to find the best possible model using Bayesian optimisation.

Hyperparameters	Limits during model development
Number of hidden layers	1–4
Nodes in each layer	5–500
Batch size	1-256
Drop-out rate	0–0.5 for the input layer and 0–0.5 for the hidden layers
Norm regularisation	L_1, L_2, both $L_1$and L2 (elastic net) or Max-norm
Activation function for the hidden layers	Rectified linear unit (ReLU), exponential linear unit (ELU) or hyperbolic tangent function
Optimisation	Standard batched gradient descent, Adam implementation of stochastic gradient descent or a slightly different version called Adam AMSgrad

Similar to study II and III, the ANN in this study consisted of one input layer, a number of hidden layers and one output layer. The final model was chosen based on the AUROC of the cross-validations and the AUROC was reported using the test set data. The AUC was used as a performance measure.[183] The difference between AUC-curves and their corresponding CIs were calculated using both bootstrapping and DeLong´s method.[180,181]

In resemblance to ANN, the Bayesian optimisation approach was used to find the optimal hyperparameters for the XGBoost model as presented in table 10. Due to the abundance of parameters, we limited our search to some of the most impactful parameters and left the remaining as default.

Table 10. Hyperparameters during XGBoost-model development in study IV. In resemblance to ANN, the Bayesian optimisation approach was used to find the optimal hyperparameters for XGBoost. The table presents the predefined limits for hyperparameters during development to find the best possible model.

Hyperparameters	Limits during model development
Number of rounds	2–1000
Learning rate (eta)	0–1
Gamma	0–10
Max depth	1–50
Min child weights	0–10
Max delta step	0–10
Subsample	0.2–1
Colsample by tree	0.4–1
Limits for regularisation parameters lambda (L_2) and alpha (L_1)	0–5

We also performed an inter-model analysis between ANN and XGBoost to compare their performance in the internal and external validation. Finally a SHAP algorithm was applied to define individual variable importance.

Shapley additive explanation algorithm

Due to the complex nature of ANNs and other advanced machine learning algorithms there might be a understandable resistance to accept and implement these models in research studies as well as in a future clinical context. One way to enhance understanding for these type of prediction models and to highlight the contribution of each variable to the final prediction is to apply the SHAP algorithm based on Shapley's values[189,190] with roots in game theory. The algorithm explains how each variable either increase or decrease the risk of a predicted outcome as illustrated in figure 16. A weakness with SHAP includes the possibility to create misleading interpretations which can hide biases.[191]

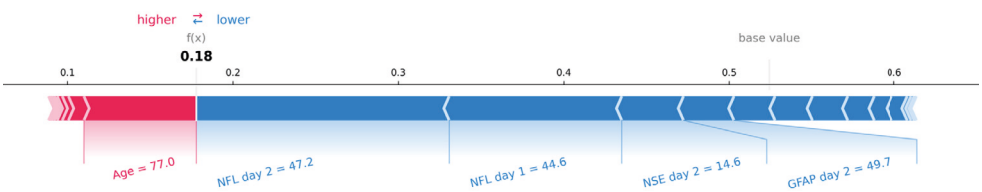


Figure 16. SHAP explanation algorithm. The figure illustrates an example of patient-specific prediction using the Shapley additive explanation algorithm (SHAP). In this case the patient was predicted to have a 18% risk of a poor outcome. Age contributed most to increasing risk (red mark) whereas the moderate levels of brain injury biomarkers contributed most to decreasing risk (blue mark). NFL, Neurofilament light; NSE, Neron-specific enolase; GFAP, Glial fibrillary acidic protein.

Software

The statistical analysis was performed using R, version 3.2.3–4.0.0 (R Foundation for Statistical Computing) and Python, version 3.6.4–3.8.3 (Python Software Foundation).[192,193] All ANN models in the thesis were developed using TensorFlow, an open-source machine learning framework developed by Google.[194]

The 'tableone' package was used to calculate the differences in the study populations.[195] The 'forest plot' package in R was used to display the odds ratio.[196] ROC curves and AUROC calculations were performed using the 'pROC' package in R.[197] The 'Optimalcutpoints' package in R was used for calculating thresholds for the confusion matrix in study III.[198] The 'Boruta–Shap' and 'shap' packages in Python were used for variable selection and explanation of the ANN model in study III.[199,200] The schematic ANN figure was created using the 'TikZ' package.[201] The XGBoost model was created using the Python package XGBoost.[202]

The merging and adjustment of INTCAR 1.0 and INTCAR 2.0 into INTCAR1+2 for study IV was performed in R (R Core Team, 2013).

Results

Paper I

Analysing functional outcome with varying levels of TTM after OHCA in patients selected from the INTCAR 2.0 registry

In paper I, a total of 1710 patients with any cause of cardiac arrest were included in the final analysis after excluding patients with arrest other than out-of-hospital, age below 18 years and missing data on outcome or targeted temperature according to the flow chart displayed in figure 17. Patients registered before 2013 were excluded to minimise treatment bias due to the change in treatment strategy of OHCA patients following publication of the TTM trial.[66]

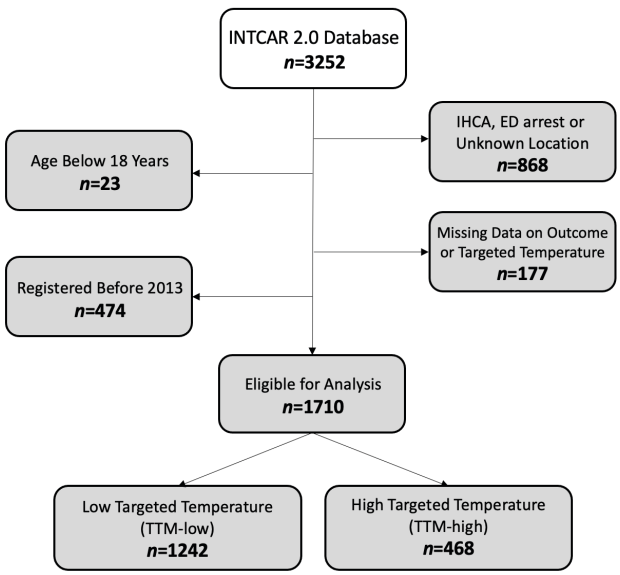


Figure 17. Flow chart. The flow chart displays cardiac arrest patients registered in the INTCAR 2.0 registry between October 2008 and November 2017. All patients eligible for analysis were grouped in TTM-low or TTM-high according to prescribed temperature treatment. TTM, targeted temperature management; INTCAR, International cardiac arrest registry; IHCA, in-hospital cardiac arrest; ED, emergency department. TTM-low denotes 32-34°C and TTM-high denotes 35-37°C.

The distribution of TTM-low vs TTM-high was not even between participating countries. Patients in the TTM-low group (*n*=1242) were predominately from the United States and Norway, while patients in the TTM-high (*n*=468) group were from Sweden and Luxembourg as shown in figure 18.

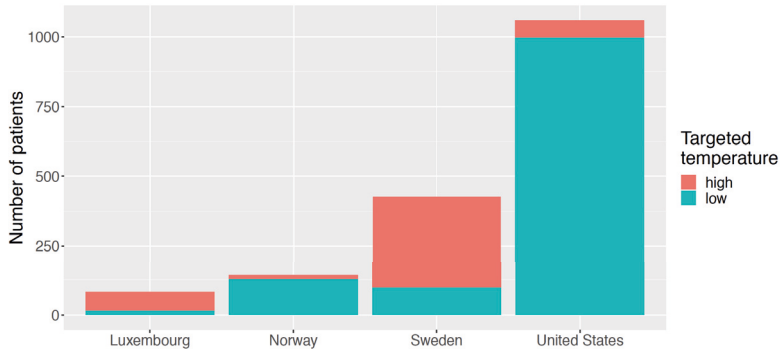


Figure 18. Targeted temperature by country. The majority of patients receiving TTM-high did so in Sweden, with Luxembourg contributing some patients. The United States almost exclusively used TTM-low. TTM-low denotes 32-34°C and TTM-high denotes 35-37°C. TTM, targeted temperature management.

The number of patients with good functional outcome at hospital discharge (CPC1-2) was 389 of 1242 (32%) in TTM-low and 135 of 468 (29%) in TTM-high, whereas the majority of patients in both temperature groups had poor functional outcome. Mortality (CPC 5) was similar in both groups, 59% (735 of 1242) in TTM-low and 62% in TTM-high (289 of 468) as seen in figure 19. A Chi-square test for temperature (TTM-low or TTM-high) vs outcome (good or poor) had a *p*-value of 0.352.

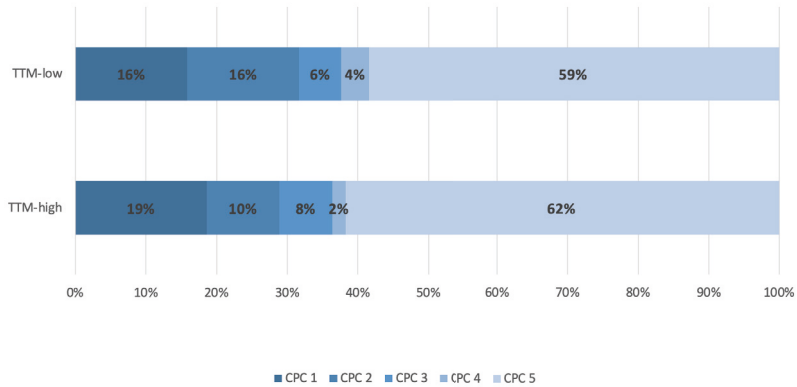


Figure 19. CPC distribution and comparison between the low- and high targeted temperature groups. Values are percentages of the total amount of patients in that group. CPC, cerebral performance category; CPC 1 indicates good cerebral performance with normal function or minor disability; CPC 2 indicates moderate cerebral disability, independent in activities of daily life; CPC 3 indicates severe cerebral disability and dependent on others for daily activities; CPC 4 indicates a patient in a coma or a vegetative state; and CPC 5 indicates death, according to the Utstein guidelines. TTM-low denotes 32-34°C and TTM-high denotes 35-37°C. TTM, targeted temperature management.

Table 11. Univariate and multivariate logistic regression analysis of baseline factors and their association with outcome. Odds ratios for good neurological outcome for the group in entirety where a value of >1 indicates each factor's beneficial influence on outcome. A *p*-value of <0.05 was considered significant. In the multivariate model adjustment for potential confounding factors previously known to influence outcome after out-of-hospital cardiac arrest (OHCA) such as age, gender, comorbidities, arrest characteristics, angiography and shock on admission was made. CI, confidence intervals; TTM, targeted temperature management; COPD, chronic obstructive pulmonary disease; VT, ventricular tachycardia; VF, ventricular fibrillation; AED, automated external defibrillator; ROSC, return of spontaneous circulation; CPR, cardiopulmonary resuscitation; EMS, emergency medical service.

	Univariate analysis Odds Ratio (95% CI)	*p*-value	Multivariate analysis Odds Ratio (95% CI)	*p*-value
TTM-low[a]	1.12 (0.89-1.42)	0.32	1.27 (0.94-1.73)	0.11
Age (per year)	0.97 (0.97-0.98)	<0.001	0.97 (0.96-0.98)	<0.001
Male sex	1.98 (1.56-2.53)	<0.001	1.33 (0.97-1.83)	0.08
Previous chronic heart failure	0.67 (0.50-0.90)	0.008	1.01 (0.68-1.47)	0.97
Previous myocardial infarction	1.00 (0.76-1.30)	0.983	1.10 (0.76-1.59)	0.60
Previous hypertension	0.61 (0.49-0.75)	<0.001	0.78 (0.58-1.04)	0.09
Previous insulin dependent diabetes	0.35 (0.23-0.52)	<0.001	0.45 (0.26-0.75)	<0.001
Previous non-insulin dependent diabetes	0.67 (0.48-0.93)	0.019	0.88 (0.57-1.36)	0.57
Previous COPD	0.30 (0.21-0.43)	<0.001	0.46 (0.27-0.74)	<0.001
Previous dementia/cognitive impairment	0.18 (0.06-0.42)	<0.001	0.23 (0.07-0.64)	0.01
Witnessed cardiac arrest	1.96 (1.50-2.58)	<0.001	1.80 (1.26-2.58)	<0.001
VT/VF or AED-advised shockable rhythm	6.31 (4.96-8.08)	<0.001	4.39 (3.23-6.01)	<0.001
Time to ROSC (per minute)[b]	0.98 (0.98-0.99)	<0.001	0.99 (0.98-0.99)	<0.001
Urgent angiography	2.74 (2.20-3.41)	<0.001	1.60 (1.21-2.13)	<0.001
Shock on admission[c]	0.45 (0.36-0.56)	<0.001	0.51 (0.39-0.67)	<0.001
Bystander CPR				
No	Ref	Ref	Ref	Ref
Yes	2.20 (1.73-2.82)	<0.001	1.43 (1.05-1.95)	0.02
Arrest with EMS present	1.06 (0.72-1.54)	1.000	1.47 (0.90-2.37)	0.12

[a]TTM-low denotes 32-34°C. [b]If unwitnessed arrest, time is calculated from emergency call. [c]Shock on admission is defined as a systolic blood pressure of less than 90 mmHg for more than 30 minutes or end-organ hypoperfusion unless vasoactives are administered.

Primary outcome in this study was functional outcome at hospital discharge dichotomised into good outcome (CPC 1-2) or poor outcome (CPC3-5) according to the description of the CPC-scale in the Methods section. The univariate logistic regression analysis showed no statistically significant differences in functional outcome at hospital discharge between the two target temperature groups (OR=1.12, 95% CI 0.89-1.42, p=0.32) and neither did the multivariate analysis (OR =1.27, 95% CI 0.94-1.73, p=0.11) as shown in table 11. Among covariables selected by their potential to influence outcome after cardiac arrest, the presence of a shockable rhythm had the strongest multivariate association with good outcome (OR =4.39, 95% CI 3.23-6.01, p<0.001).

For the predefined subgroup analysis defined by the design variables in the TTM trial,[66] a forest plot was created assessing the interaction of age (above or below 65 years), sex, time to ROSC (above or below 25 minutes), initial rhythm (shockable or non-shockable) and circulatory shock on admission, to investigate whether any of these groups would signal a positive association to either TTM-high or TTM-low. TTM-low was associated with a good outcome and favoured patients of male sex and with absence of circulatory shock on hospital admission as seen in figure 20.

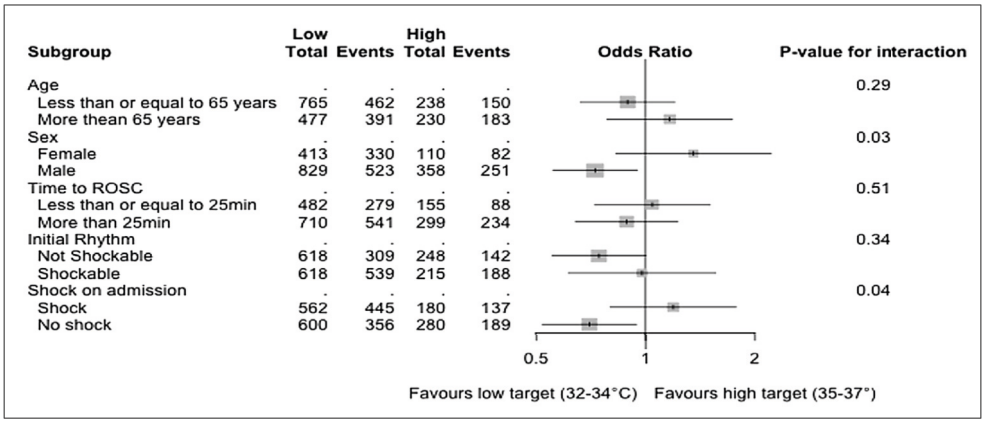


Figure 20. Odds ratio of outcome according to subgroups. The forest plot shows the odds ratio for five predefined subgroups in regard to whether these subgroups were favoured by a low- or a high targeted temperature at hospital discharge. The horizontal bars represent 95% confidence intervals (CI). *p*-values are for the tests of subgroup heterogeneity (tests of interactions) and a *p*-value of <0.05 was considered significant. For unwitnessed cardiac arrests the time to ROSC was calculated from time of emergency call. Shock on admission is defined as a systolic blood pressure of less than 90 mmHg for more than 30 minutes or end-organ hypoperfusion unless vasoactives are administered. 'Low target' denotes 32-34°C (TTM-low) and 'High target' denotes 35-37°C (TTM-high). ROSC, return of spontaneous circulation; TTM, targeted temperature management.

Secondary outcome in this study was adverse events related to TTM during ICU stay and is displayed in table 12. Haemodynamic instability leading to discontinued TTM was more common in TTM-low (*n*=58, 4.9% vs. *n*=8, 1.7%, *p*<0.001) and pneumonia was similarly common in both groups (*n*=435, 38,4% in TTM-low and *n*=170, 37.1% in TTM-high, *p*=0.67). There were no statistically significant differences in the frequency of adverse events regarding major bleeding (*n*=88, 7.8% in TTM-low vs. *n*=30, 6.6% in TTM-high, *p*=0.47), sepsis (*n*=3, 0.3% in TTM-low vs. *n*=0, 0% in TTM-high, *p*=0.66) or seizures (*n*=98, 8.5% in TTM-low vs. *n*=39, 8.4% in TTM-low, *p*=1.00).

Table 12. Advers events for total sample dichotomised into low- and high targeted teperature groups. Secondary outcome in the study was adverse events during the ICU stay. Data are presented as *n* (%) and *n* denotes the number of cases with valid data. The events were compared using Chi-square and a *p*-value of <0.05 was considered significant. TTM, targeted temperature management.

	n	TTM-low[a] (%)	TTM-high[b] (%)	*p*-value
Signs of seizure during TTM	1616	98 (8.5)	39 (8.4)	1.00
Pneumonia Clinical or Microbial diagnosis[c]	1590	435 (38.4)	170 (37.1)	0.67
Major bleeding[d]	1591	88 (7.8)	30 (6.6)	0.47
TTM discontinued - Haemodynamic instability	1649	58 (4.9)	8 (1.7)	<0.001
TTM discontinued - Severe sepsis/septic shock[e]	1649	3 (0.3)	0 (0)	0.66

[a]TTM-low denotes 32-34°C. [b]TTM-high denotes 35-37°C. [c]Pneumonia is defined as 3 of the following criteria: progressive or new infiltrates on chest X-ray (mandatory), fever above 38°C, leucocytosis and purulent mucus in tube. [d]Major bleeding is defined as cerebral bleeding or bleeding requiring transfusion.[e]Severe sepsis/septic shock is defined according to the criteria of the American College of Chest Physician and Society of Critical Care Medicine.

Paper II

Prediction of long-term functional outcome after OHCA with variables accessible on admission to hospital

In paper II, a total of 932 patients and 54 variables from the TTM trial data set were included for final analysis. We randomly selected 839 patients (90%) for model development (training set) and 93 patients (10%) to evaluate the model's prognostic performance (test set). The cross-validated AUROC (for the training set) was 85.2%, and the AUROC when evaluating performance using the test set was 89.1%.

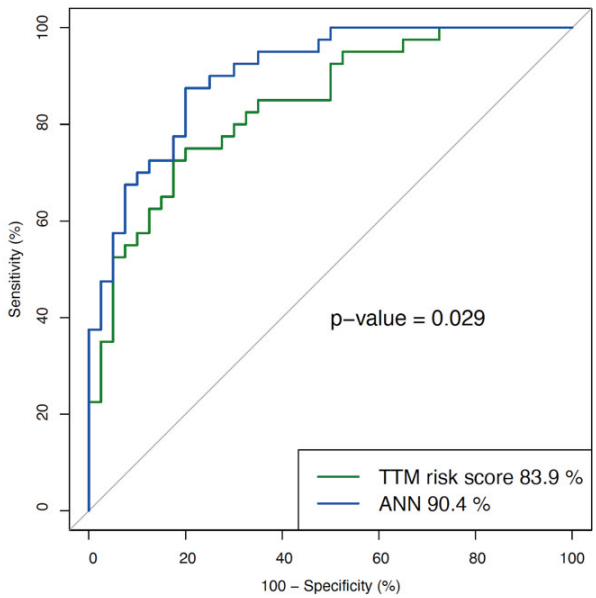


Figure 21. Performance comparison between the TTM risk score and the ANN model. Comparison between the performance of the TTM risk score and our ANN model based on the 80 patients in the test set. TTM, targeted temperature management; ANN, artificial neural network.

We compared the performance of our ANN model with that of a similar prediction model, the 'TTM risk score'[203] described in section 'Prognostication of post-cardiac arrest patients in the ICU' in the Background section. The ANN model performed significantly better (AUROC 90.4% vs. 83.9%, p=0.029), as seen in figure 21. A total of 13 patients were removed from the test set before comparing the two models, as the 'TTM risk score' model could not handle missing values. This change in the test set explains the difference between the ANN model AUROC determined here to 90.4% and to 89.1% as reported above.

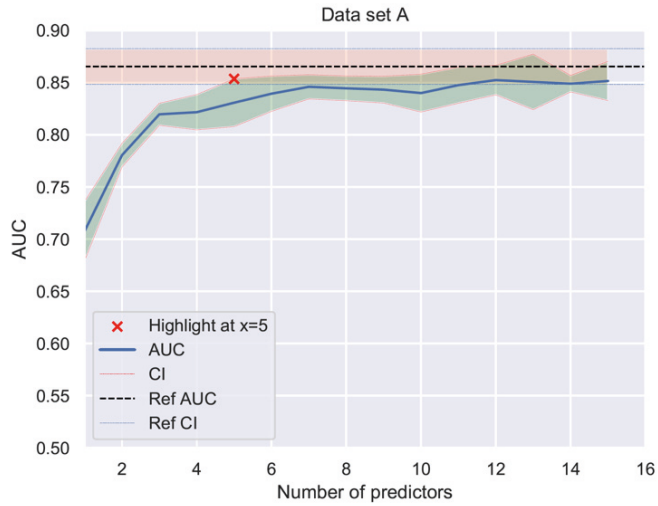


Figure 22. Model performance as variables are added. In the search for a simplified model, we ranked all input variables by subtracting one variable at a time from the developed model and calculating the AUROC. We then started with the most important variables from this ranking and added one variable at a time back to the model, calculating the AUROC based on the training set at each step.The AUROC (AUROC = AUC in figure 22) is based on the cross-validation (training), represented by the blue line with its corresponding CI. The best performing model, with 54 variables, is represented by the dotted line and its corresponding CI. The CIs overlapped after five variables had been added; this is marked with a red X to indicate the point after which no significant difference was found between the two models. AUROC, area under the receiver operating characteristic; CI, confidence interval.

Table 13. Model performance as variables are added. All variables were ranked and then added one at a time, starting with age, to build a model from the ground up The model performance is shown as the AUROC with the corresponding CI based on the cross-validation (training; see figure 22), and the corresponding test set performance is also shown. For comparison, the performance of the final model is shown as well. AUROC and AUC, area under the receiver operating characteristic; CI, confidence interval; ROSC, return of spontanous circulation; GCS, Glasgow Coma Scale; AMI, acute myocardial infarction; CV, cross-validation.

No. of variables	Variables	$AUROC_{CV}$	$AUROC_{test}$
1	Age	0.708 (±0.0286)	0.657
2	+ Time to ROSC	0.780 (±0.0113)	0.799
3	+ First monitored rhythm	0.820 (±0.0106)	0.852
4	+ Previous cardiac arrest	0.822 (±0.0169)	0.861
5	+ GCS motor score	0.832 (±0.0229)	0.863
6	+ Dose of adrenaline	0.839 (±0.0170)	0.826
7	+ Creatinine	0.846 (±0.0117)	0.837
8	+ Cardiac arrest location	0.854 (±0.0119)	0.857
9	+ Previous AMI	0.843 (±0.0129)	0.835
10	+ Diabetes	0.840 (±0.0182)	0.844
11	+ Length	0.848 (±0.0173)	0.869
12	+ Time to Advanced CPR	0.853 (±0.0142)	0.870
13	+ pH	0.851 (±0.0266)	0.880
14	+ Platelets	0.849 (±0.0079)	0.875
15	+ Bystander witnessed arrest	0.852 (±0.0188)	0.886
54	All variables	0.852 (±0.0172)	0.891

All 54 variables were ranked based on how much they influenced AUROC when removed from the model. One variable at a time was then added stepwise, starting with the most important variable based on the previous ranking, and calculated the AUROC based on the training set at each step. Each AUROC calculation for the 15 most important variables is seen in figure 22 and further described in detail in table 13.

After hospital admission, patients included in the TTM trial were randomised to a target temperature of 33°C or 36°C for 28 h (figure 12). By dividing the cohort into five risk groups based on predictions made by the ANN model, we could investigate whether one of the two temperature treatments would be beneficial to a certain risk group more than to others. No such specific risk group was shown to benefit from a specific targeted temperature strategy, as displayed in figure 23.

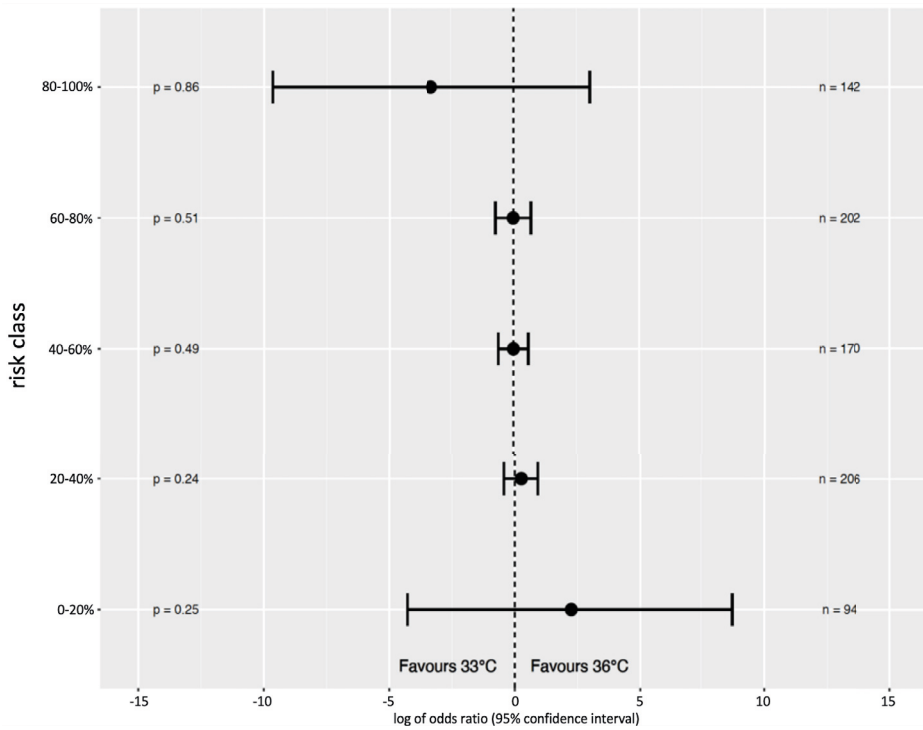


Figure 23. Treatment effect based on the ANN-model-stratified risk groups. The forest plot displays the logarithmic diagnostic odds ratio for five ANN-stratified groups of CPC score >2 and its association to treatment with TTM at 33°C and 36°C. A diagnostic odds ratio >1 implies a better functional outcome when treated with 36°C compared to when treated with 33°C and vice versa. CPC, cerebral performance category; TTM, targeted temperature management.

Paper III

Prognostication of long-term functional outcome after OHCA using cumulative added information during the first three days after ICU admission

The TTM trial registry used in paper II was also used for analyses in this study. Of the 932 patients included for further analysis, six patients were removed due to missing information on outcomes and one patient was removed due to extensive missing values. As displayed in figure 24, patients who had either awakened or died during the time windows 0–24h (day 1), 24–48h (day 2) and 48–72h (day 3), after ICU admission were removed accordingly.

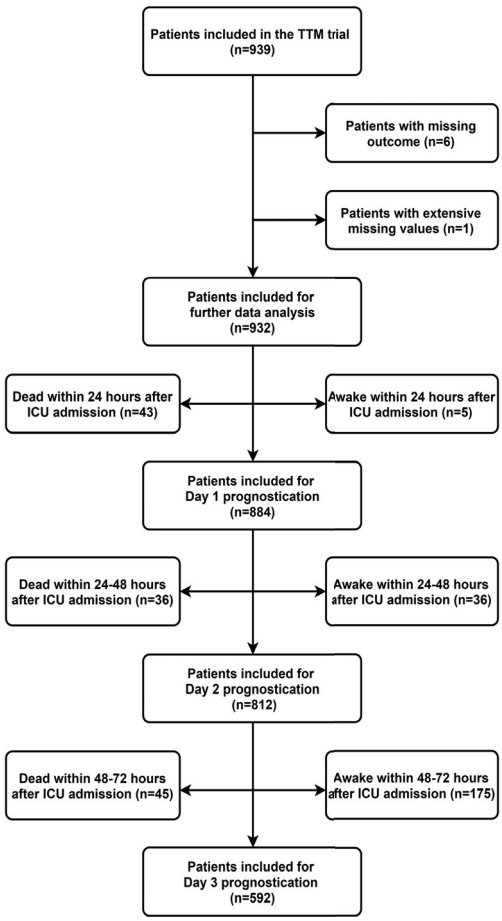


Figure 24. Flowchart. Flowchart for the study populations at day 1 (24h), 2 (48h) and 3 (72h) after ICU admission. After each day, patients who died or woke up were removed from further analysis to focus on the prognostication of comatose patients. Furthermore, on each day, three datasets were created that reflected the levels of biomarkers included in the model (not shown here). Population characteristics are based on 'Patients included for further data analysis (n = 932)' (see table 1A–D in supplementary material in paper III). TTM, targeted temperature management; ICU, intensive care unit.

Table 14. Overview and prognostic performance. Overview and prognostic performance of the ANN models during the first three days after ICU admission. In level A, we used all available data from the TTM trial. In level B, we added clinically accessible biomarkers and for level C, we added research-grade biomarkers as well. The prognostic performance is displayed as the area under the receiver operating characteristics curve (AUROC) and by using a confusion matrix. Note the threshold for the confusing matrix was based on the threshold for 100% specificity in the training set. TN, true negative; TP, true positive; FN, false negative; TP, true positive.

Timeline	Type of data	Number of patients			Number of variables		Model performance		Confusion matrix (test set)				
		Total (n)	Train set (n)	Test set (n)	Before variable selection (n)	After variable selection (n)	Training set (cross validation) AUROC (%) (CI 95 %)	Test set AUROC (%) (CI 95 %)	Probability thres-hold	TN (n)	FP (n)	FN (n)	TP (n)
Day 1 (24h)	Level A	884	702	182	120	22	85.7 (83.2-88.6)	81.9 (75.9-87.9)	0.981	99	0	82	1
	Level B	638	502	136	125	21	89.9 (86.8-92.2)	81.8 (74.9-88.6)	0.983	70	0	57	9
	Level C	690	545	145	131	22	96.6 (94.7-97.5)	95.2 (91.9-98.4)	0.887	76	1	21	47
Day 2 (48h)	Level A	812	645	167	174	21	85.8 (82.9-88.5)	78.1 (71.1-85.0)	0.935	88	2	70	7
	Level B	578	460	118	184	17	93.7 (91.7-95.9)	89.7 (84.3-95.0)	0.963	59	0	44	15
	Level C	624	495	129	196	21	96.6 (95.2-97.9)	96.1 (93.4-98.8)	0.986	67	0	31	31
Day 3 (72h)	Level A	592	469	107	228	17	84.9 (80.6-87.9)	86.7 (80.4-93.1)	0.920	60	0	22	25
	Level B	415	328	87	243	12	92.7 (89.8-95.1)	94.1 (89.4-98.8)	0.885	40	0	22	25
	Level C	442	347	95	261	23	96.6 (94.9-98.1)	94.7 (90.6-98.7)	0.820	45	0	16	34

As described in the Methods section, for each day three data sets were created including three different biomarker-levels. The number of included patients varied between the data sets based on the number of patients with missing values. The number of variables, before and after variable selection, and the number of patients in each dataset are presented in table 14.

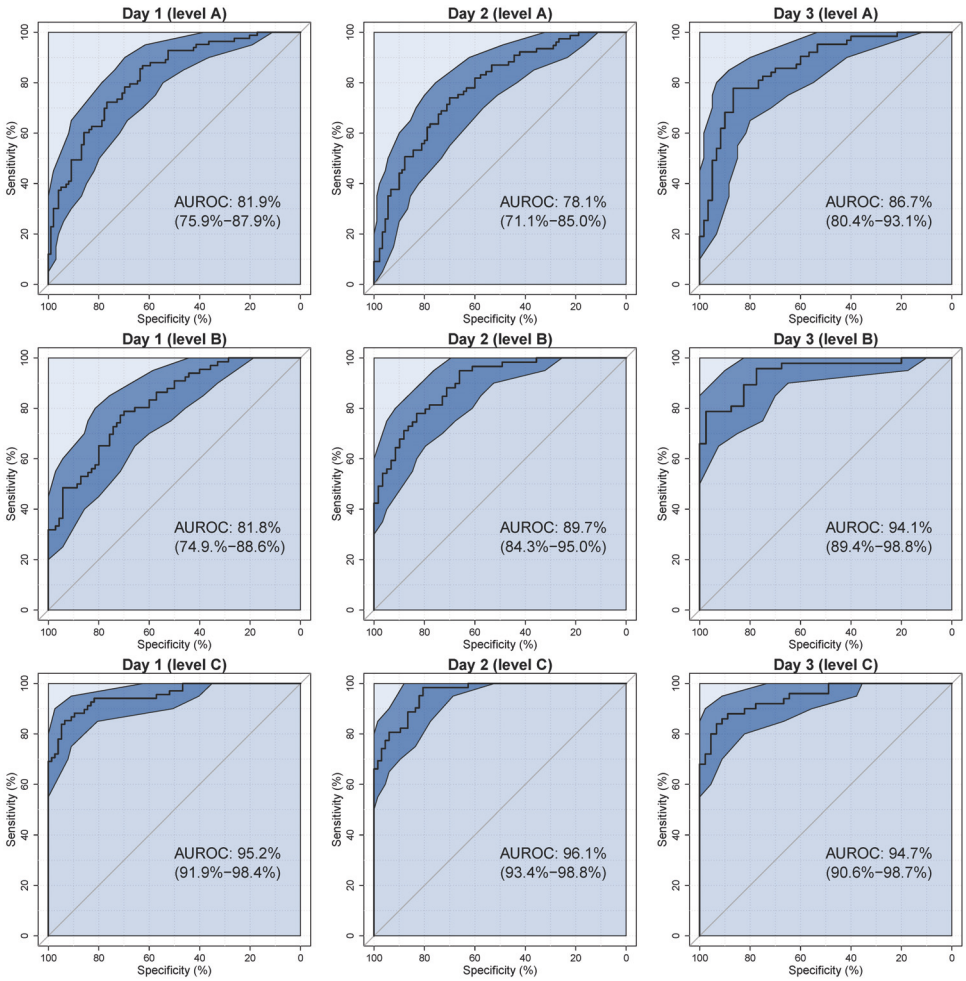


Figure 25. Prognostic performance for all nine models. Model performance predicting poor neurological outcome after 6 months based on the corresponding test set. The columns represent the timeline after ICU admission. The rows represent the different levels of biomarkers added to the available clinical variables from the TTM trial: none (level A), clinically accessible biomarkers (level B) and research-grade biomarkers (level C). The CI for the AUROC is calculated for each model. Furthermore, the CI for the specificity at different levels of sensitivity is displayed as a darker blue CI-area. AUROC, area under the receiver operating characteristic curve; CI, confidence interval.

Prognostic performance presented as ROC curves are illustrated in figure 25. For the level A model, based on standard ICU observation variables only (no added biomarkers), prognostic performance remained below 90% during the three days after ICU admission. For the level B model, which also included clinically accessible biomarkers, the prognostic performance improved significantly from day 1 to day 3 (from 81.8% to 94.1%, $p<0.01$). For the level C model, with addition of research-grade biomarkers included, performance was excellent from day 1 through day 3.

Noteworthy is that level C at all time points, and level B at 72h, had a sensitivity above 60% while retaining 100% specificity. When this was further investigated by using a threshold of 100% specificity (in the training set) to predict the outcome in patients within the test set, two models generated false-positive predictions (predicted poor outcome, reported good outcome). The majority of the models had a high rate of false-negative predictions (predicted good outcome, reported poor outcome), but that rate remained under 25% when using research-grade biomarkers (level C). Details on prognostic performance for all models are presented in table 14.

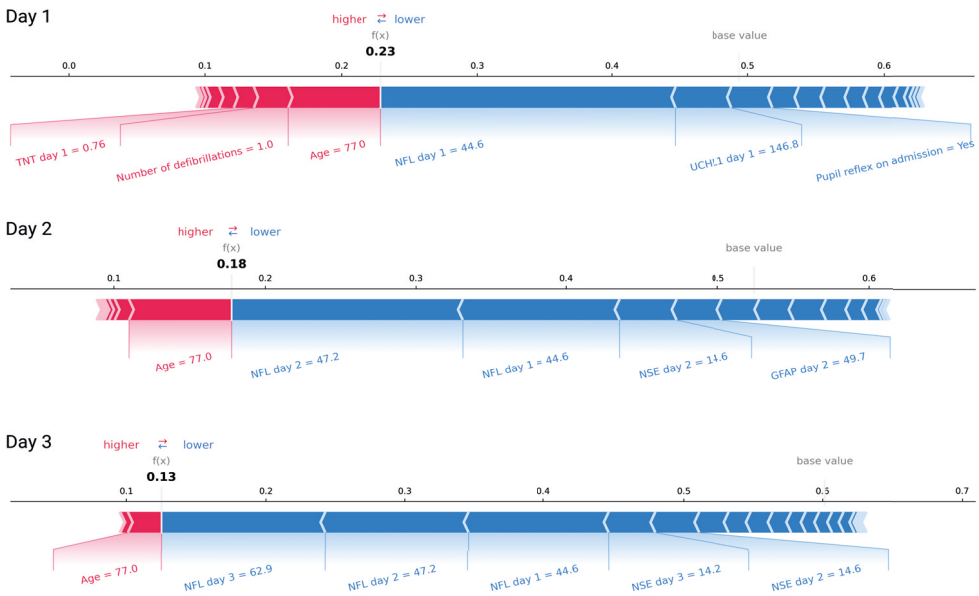


Figure 26. The Shapley additive explanations algorithm (SHAP) used to explain how patient-specific predictions were generated. The patient in this example was predicted to have a risk of a poor outcome of 23% on day 1, 18% on day 2 and 13% on day 3 (using the level C model). On all 3 days, the patient's age was the factor that contributed most to increasing the risk, whereas the modest levels of NFL and NSE contributed most to decrease the level of overall risk. TNT, Troponin-T (ng/L); NFL, Neurofilament light (ng/L); UCHL1, Ubiquitin carboxy-terminal hydrolase L1 (ng/L); NSE, Neuron-specific enolase (ng/mL); GFAP, Glial fibrillary acidic protein (ng/L).

We applied the SHAP algorithm to all models to explain how each prediction had been generated. One example is illustrated in figure 26, where predictions of the ANN model that uses research-grade biomarkers (level C) are explained using the SHAP algorithm on day 1, 2, and 3, respectively. The patient's age, TNT and number of defibrillations adds to an increase in risk of a poor outcome, while the low levels of biomarkers and pupil reflex on admission decrease the risk.

The SHAP algorithm was also used to rank all variables for each model. This is illustrated in figure 27, which demonstrates all three levels of variable importance for day 2. The 10 most important variables change as a consequence when clinically accessible and research-grade biomarkers are added to the models. Age and the administrated dose of adrenaline are the two most important variables in level A. The importance of these two variables is reduced in levels B and C, and in level C the dose of adrenalin is not included among the top 10 variables, while age remains the third most important variable. Similar figures for day 1 and 3 can be found in the supplementary material in paper III.

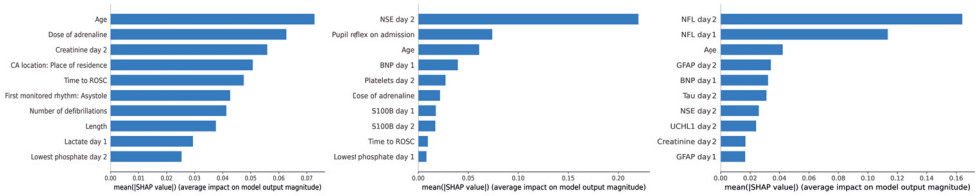


Figure 27. SHAP variable importance on day 2. The global importance of each variable in each model illustrated with the SHAP variable importance. The most important variable has the highest mean of absolute SHAP values. The left panel shows level A, the middle panel shows level B (adding clinically accessible biomarkers) and the rigth panel shows level C (adding research-grade biomarkers). Similar figures for day 1 and 3 can be found in the supplementary material in paper III. CA, cardiac arrest; ROSC, return of spontaneous circulation; NSE, Neuron-specific enolase; BNP, Brain natriuretic protein; S100B, S100 calcium-binding protein B; NFL, Neurofilament light; GFAP, Glial fibrillary acidic protein; UCHL1, Ubiquitin carboxy-terminal hydrolase L1.

Paper IV

Early prediction of functional outcome after OHCA, using two different machine learning models and comparison of internal and external validation

In paper IV we included consecutively collected patients from both INTCAR 1.0 and 2.0 and created a merged cohort (INTCAR1+2) with 5994 adult comatose patients who survived cardiac arrest of any cause and admitted to hospital. After excluding patients with arrest other than out-of-hospital and those with missing data on outcome, 4431 patients were eligible for final analysis as displayed in figure 28 (left side of figure). We selected 29 early variables available on hospital admission as predictors for the development and training of our ANN and XGBoost-algorithms, and one outcome variable (functional status at hospital discharge) as the output variable. To externally test and validate our model's performances we used the TTM trial registry and included patients as presented in figure 28 (right side of figure). Some of the variables in the TTM trial were aligned in order to equate them to the corresponding 29 variables in INTCAR1+2 and thereby enable the algorithms to perform the training and testing processes. Such adjustments included for instance re-coding of variables to make them comparable for the prediction models.

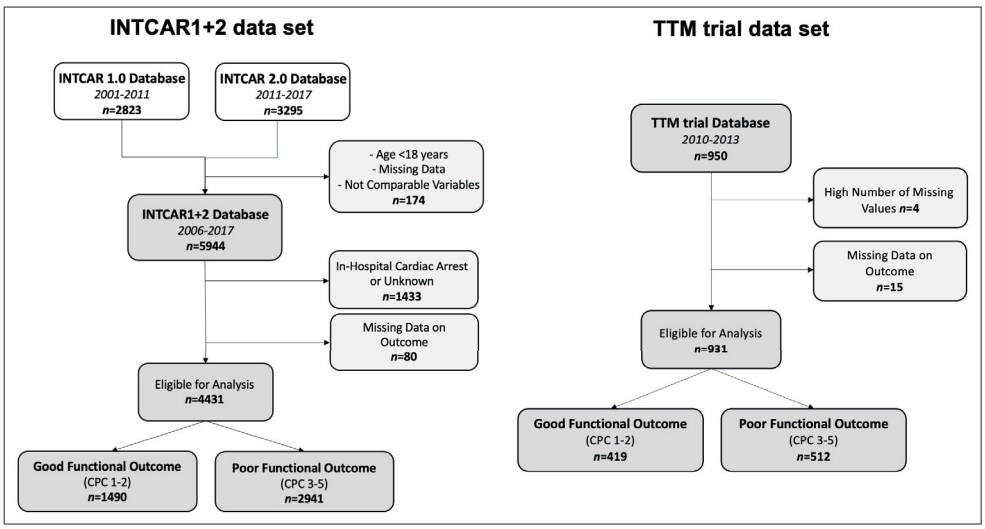


Figure 28. Flow charts displaying exclusion criteria before analyses of the INTCAR1+2 (left) and TTM trial (right) databases. INTCAR, International cardiac arrest registry; TTM, targeted temperature management; CPC, cerebral performance category.

The two cohorts (INTCAR1+2 and TTM) were comparable regarding baseline characteristics and their association with outcome. Older age was associated with poor outcome (INTCAR1+2: 65 vs. 58 years, p<0.001 and TTM: 68 vs. 61 years, p<0.001) as well as female sex (INTCAR1+2: 35,7% vs. 23.1%, p<0.001 and TTM: 20,5% vs. 16,5.1%, p=0.004). Regarding prehospital circumstances and arrest characteristics a good outcome was significantly associated with witnessed arrest, first rhythm shockable, defibrillation performed, shorter time to CPR by emergency medical services and shorter time to ROSC in both INTCAR1+2 and TTM registry. Previously healthy patients were more prone to have a good outcome at hospital discharge (INTCAR1+2: 28,0% vs. 14.4%, p<0.001 and TTM: 43,1% vs. 25.3%, p<0.001).

Using 29 early variables available on hospital admission in the development registry (INTCAR1+2), both ANN and XGBoost models predicted outcome with similar performance with an area under the receiver operating characteristics curve (AUROC) of 0.86 (95% CI 0.85-0.87) (p=0.64) as seen in figure 29.

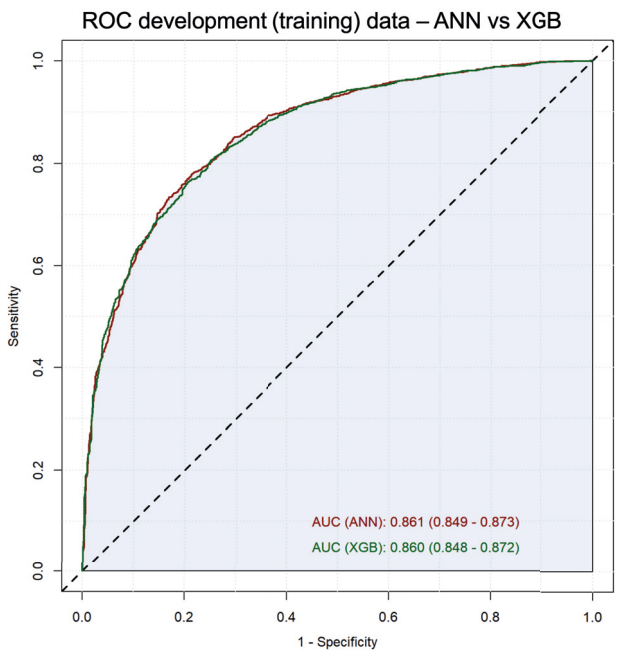


Figure 29. Comparison of predictive performance between ANN and XGBoost models during development. The predictive performance in the development/training set is expressed as the AUC in ROC curves in an ANN and a XGBoost model, using 29 early variables available on hospital admission. Both model was development on a training set (INTCAR1+2) to predict outcome on hospital discharge. The red curve represents ANN and the green curve represent XGBoost. Both models performed well with no significant difference in acurracy (*p*=0.64). ANN, artificial neural network; XGB, eXtreme gradient boost; AUC, area under the curve; ROC, receiver operating characteristics.

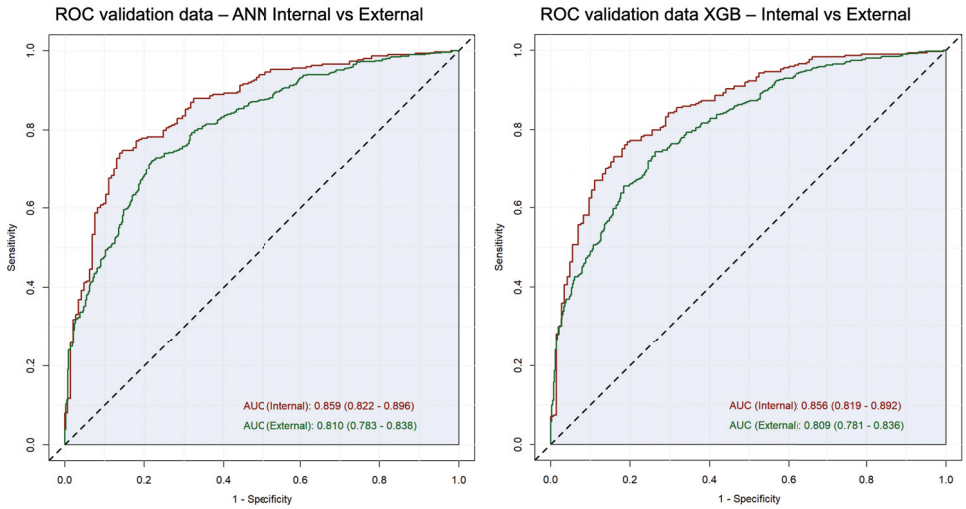


Figure 30. Comparison of predictive performance between internal and external validation using ANN and XGBoost models. The prediction performance is expressed as the AUC in ROC curves using an ANN model (left figure) and a XGBoost model (right figure). The red curves represent the results of internal validation on a test set (10% of randomly chosen and earlier removed from the training set (INTCAR1+2) on which the the two algorithms developed their prediction model). The green curves repesent the results of external validation on a test set (the TTM trial registry). In both models the internal validation significantly outperformed the external validiation in acurracy (*p*=0.04). ANN, artificial neural network; AUC, area under the curve; ROC, receiver operating characteristics

The ANN model yielded an AUROC of 0.86 (95% CI 0.82-0.90) in the internal validation (the same data set), and an AUROC of 0.81 (95% CI 0.78-0.84) in the external validation as seen in figure 30 (left side of figure) The XGBoost model yielded an AUROC of 0.86 (95% CI 0.82-0.89) in the internal validation (the same data set), and an AUROC of 0.81 (95% CI 0.78-0.84) in the external validation as seen in figure 30 (right side of figure). In both models, the internal validation significantly outperformed external validation (p=0.04).

We also performed an inter-model analysis between ANN and XGBoost to compare performance of internal and external validation with no significant difference between the models (p=0.57 and =0.71, respectively) which is displayed in figure 31.

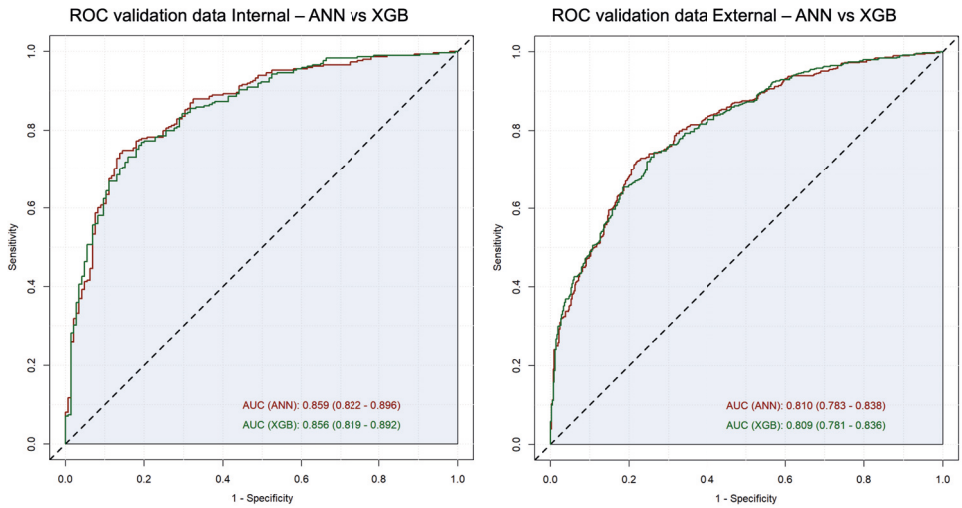


Figure 31. Inter-model analysis between ANN and XGBoost. The prediction performance is expressed as the AUC in ROC curves using an ANN and a XGBoost model. In the left figure, internal validation performance of ANN was compared to internal performance of XGBoost, with no significant difference in acurracy (p=0.57). In the right figure, external validation performance of ANN was compared to external performance of XGBoost, with no significant difference in acurracy (p=0.71). ANN, artificial neural network; AUC, area under the curve; ROC, receiver operating characteristics

By applying Shapley additive explanations (SHAP) algorithm to our models, the predictions were visualised and better explained. An increasing SHAP-value >0.0 indicated a positive impact on model outcome and vice versa. The higher SHAP-value, the more risk was added by each variable for a poor outcome. Both ANN and XGBoost identically ranked longer time to ROSC, older age, PEA/asystole as first monitored rhythm, and absence of spontaneous breathing on hospital admission as highly associated with an increased risk of a poor outcome as illustrated in figure 32.

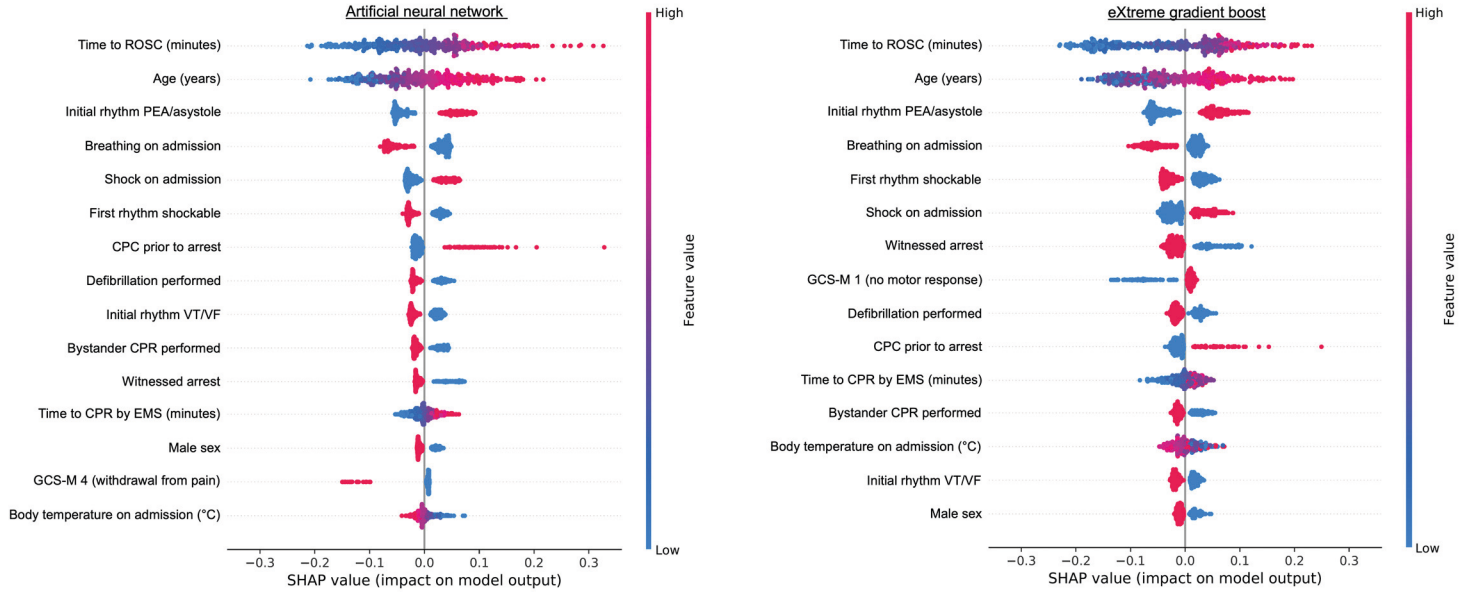


Figure 32. Shapley additive explanations (SHAP) algorithms for ANN and XGB models. The SHAP-algorithms are used to explain how each patient-specific clinical variable (predictor) impact on the model's output (risk of a poor outcome). The left figure is the SHAP-algorithm for ANN and the right figure is the SHAP-algorithm for XGB. Veritcally, the 15 most important predictors in each model are ranked after their individual feature importance from 'high' to 'low'. Horisontally, the predictor's impact on model outcome (SHAP-value) are presented. Each red or blue dot represents a SHAP-value for a patient/prediction performed by the algorithm. Vaiables with positive values (for binary variables) or increasing values (for continuous variables) are marked in red and neagtive values or decreasing values are marked in blue. SHAP-values >0.0 indicate an increased risk of a poor outcome and SHAP-values <0.0 indicate a decreased risk. The four top predictors were identical in both ANN and XGB models. According to the SHAP-algorithms, longer time to ROSC, older age, PEA/asystole as first monitored rhythm and absence of spontaneous breathing on hospital admission, were all associated with an increased risk of a poor outcome. ROSC, Return of spontaneuos circulation; PEA, pulseless electric activity; CPC, cerebral performance category; VT, ventricular tachycardia; VF, ventricular fibrillation; CPR, cardiopulmonary resuscitation; EMS, emergency medical services; GCS-M, Glasgow coma scale motor score

Discussion

Paper I

Analysing functional outcome with varying levels of TTM after OHCA in patients selected from the INTCAR 2.0 registry

This was a large retrospective, observational registry study where we investigated whether the results from the TTM trial could be demonstrated in OHCA patients included in a real life data set (the INTCAR 2.0 registry) containing cardiac arrest data where baseline variables allow for adjusted analyses. Our analyses showed no statistically significant difference in functional outcome at hospital discharge between patients treated with TTM-low (32-34°C) or TTM-high (35-37°C) in either unadjusted or adjusted analyses. When adjusting for confounding factors, the multivariate analysis indicated a non-significant trend towards better functional outcome with TTM-low. This was, however, associated with more hemodynamic instability and discontinuation of TTM therapy.

Although the crude numbers for good functional outcome between the TTM-groups were strikingly similar, the multivariable analysis revealed a tendency towards a more favourable outcome in TTM-low after adjustment for potential confounding factors previously known to be associated with outcome.[204-206] Similar concerns were raised in prior observational studies.[207,208] Although complex mediation analysis of data from 45 935 patients in a study from Bradley et al.[208] suggested inconsistency regarding the role of target temperature in theses outcomes, the lack of randomisation and high potential for bias and confounding suggests great caution when interpreting these results.[209] Similarly, our results must be interpreted with caution, and potential benefit of TTM-low may be worth exploring in further randomised clinical trials.

Our treatment groups differed significantly in baseline characteristics such as age, gender, comorbidities, arrest characteristics, prehospital circumstances, cardiac interventions and shock on admission, all variables significantly associated with outcome after cardiac arrest. These differences may reflect geographic, demographic and policy-related or patient-selection factors specific to treating physicians. In Sweden, the mean age at arrest is higher, male patients suffering cardiac arrest outnumber female patients, shockable rhythms are more common and

the frequency of bystander-CPR is much higher compared to the United States.[210,211] The marked difference in baseline variables between the United States and Europe might indicate the presence of other unidentified and unmeasured factors that differ, resulting in considerable residual confounding.

One strategy to compensate for the differences in background characteristics between intervention groups, would be to use propensity score analysis using sub-class matching. However, the probability of receiving any of the two targeted temperatures was not equal due to country-specific treatment regime (treatment with 33°C more common in the US and 36°C more common in Europe). This means that positivity assumption in our study was not fulfilled, which is mandatory for a propensity score analysis in order to remove treatment bias. Therefore, propensity score analysis was not performed in this study.

Limitations

This was a retrospective study of prospectively collected registry data and the sample size was determined by convenience. No audit or formal quality control was performed, making erroneous data and misinterpreted entries in the INTCAR2 database possible. The generalizability of our findings may be limited, as our results reflect standards in highly specialised OHCA-centres using TTM. Hospital characteristics are associated with OHCA outcome, favouring centres with 24-hour cardiac interventional services.[29,30] Recent studies have shown that the variation in outcome after cardiac arrest may be influenced by variations in withdrawal of life sustaining therapy (WLST) strategies and in-hospital management differences.[88,212]

We did not report data on long-term outcome due to the large amounts of missing values for long-term follow-up. Hospital discharge is not an ideal outcome assessment time point, since functional outcome may evolve after cardiac arrest, and time of discharge varied considerably.[127] The TTM trial, however, showed that the difference in neurological function between hospital survival and 180-day survival was limited.[66] Other studies have also indicated that CPC at hospital discharge is a useful surrogate measure of long-term survival.[135]

The sample-size differed between TTM-low and TTM-high with more patients in the low temperature group. The majority of INTCAR 2.0-patients were registered in the United States where treatment at 33°C was more common in the participating sites. The reverse situation was present for patients included in Sweden (figure 18) where temperature control at 36°C has become standard care after the TTM trial. This difference in treatment strategies in different countries might represent a bias when analysing data from an international multicentre registry. Therefore, patients registered before 2013 were excluded to minimise any treatment bias following the publication of the TTM trial published this year. During the five-year inclusion period, changes may have occurred in cardiac arrest care, including standardised intensive care bundles and more early cardiac intervention. Advanced prehospital

care has also evolved and both availability of public defibrillators and layperson awareness of cardiac arrest and bystander-CPR may have increased. In addition, fewer patients presented with shockable rhythms.[13] Finally, the lack of international standardised processes for prognostication and WLST in cardiac arrest patients may have influenced outcome in these patients.

Strengths

We used a large multinational perspective, a prospective registry, well established cardiac arrest centres, well defined covariables important for adjustment of treatment effects and consecutively entered patients which may better reflect real-world practices than clinical trials do.

While the overall mortality from cardiac arrest remains high, the prognosis for unconscious OHCA patients with initial shockable rhythms and ROSC admitted to the ICU are improving, as more than half will survive with a good functional outcome.[68] Controlling body temperature is a potential treatment that may prevent secondary brain damage but the precise mechanisms are still unknown. Optimal post-resuscitation care remains controversial, including which temperature to target, how long to deliver temperature control, the optimal way of rewarming and whether different target temperatures are appropriate for different patient.[213-216] Overall quality of evidence for this therapy is low or very low, and further studies are necessary to determine benefits and risks related to temperature management.

This study confirmed the results from the first TTM trial but also emphasised the need for validated cardiac arrest-specific severity scoring system to facilitate analyses and comparison between groups with differences in baselines characteristics.

Paper II-IV

Prediction of functional outcome after OHCA using variables accessible already on hospital admission, or cumulatively added information from the first three days after ICU admission – development and validation of machine learning models

These three studies differ somewhat in aim, variable selection and outcome measure (in timing) for various reasons related to either research question (hypothesis) or registry aspects. However, they all have the ambition to present and evaluate an alternative strategy assess cerebral functional outcome in OHCA patients. This strategy includes the use of supervised machine learning algorithms for analyses of OHCA registries, an advanced computer science to enhance challenging clinical real-life situations in the area of modern post-resuscitation care. The studies were designed as post hoc analyses of an international randomised multicentre trial (the TTM trial) and of two international cardiac arrest registries (INTCAR 1.0 and 2.0). This section merges the overall points from the detailed discussions in studies II-IV included in this thesis.

Study II

The aim of study II was to focus on prediction of long-term (6 months follow-up) functional outcome related to neurological status including survival, when using variables available already on admission to hospital. We used patients included in the TTM trial and selected 54 variables with information from patient's background, arrest characteristics and prehospital circumstances. A supervised machine learning model called ANN was developed to identify variables associated with a poor long-term functional outcome through repeated training on the TTM trial registry. The model was then internally validated in terms of performance to discriminate, and presented as an AUROC reflecting the accuracy when assessing the test set (randomly chosen from the development registry). To investigate our model's potential we compared its' performance with another prediction model based on logistic regression ('the TTM score') on the same cohort. Our ANN model outperformed the 'TTM score' in accuracy and performed better compared to most proposed models in the field.[43,150,152,217] Even when the ANN-model used only three to five variables, it showed good predictive performance when internally evaluated on the test set. This approach admittingly decreased the prediction capability compared to the full model, but it was evident that variables from the prehospital setting plus patient age carried the most discriminatory information. This raises questions on how many variables we actually need to include when developing models. In paper II and paper III, we created models which used all accessible information, but also models who used only a limited number of variables. Different

models have different benefits as well as limitations regarding simplicity and accuracy.
Further analysis revealed that prehospital variables (age, time to ROSC and first monitored rhythm) carried the overall most predictive information among all variables, and this has also been confirmed in previous studies. Despite this, early variables are still not included as predictors in the current international prognostication algorithms. Robust and straight forward prediction scores used as a practical decision tool to support clinical assessments would probably improve the overall cardiac arrest care by directing very advanced and potentially high-risk invasive treatment to those patients who may benefit from it. Such scores would hopefully also increase the ability to provide reliable prognostic information to next-of-kin, earlier than the observation time of at least 72h, which is the current recommendation for neurological prognostication after cardiac arrest[56,85,218].

Our models did not show any significant difference in the intervention effect of 33°C or 36°C regarding outcome when dividing the TTM trial population into five different risk classes for a poor outcome. The intervention effect was thus uniform across the risk classes, which strengthens the main conclusion of the trial, but also suggest a possible model for detection of subgroup effect in future clinical trials. We believe that this study was an important step towards improved outcome prediction in comatose patients surviving cardiac arrest with a good functional outcome. There are some obvious medical and ethical implications as well as resource aspects that may benefit from the progression of future reliable cardiac arrest-specific severity scores for early outcome prediction. Future studies should investigate if outcome prediction performance increases significantly by adding additional data and clinical variables such as early electroencephalography, neuroimaging and biomarkers. Finally, to be able to detect subgroups in an OHCA population with an increased risk of a poor outcome or subgroups that may benefit from a specific intervention or need extensive rehabilitation, further studies on larger data sets are necessary to demonstrate significant associations.

Study III

In the post-hoc study of paper III we found that machine learning models using clinical variables paired with biomarkers of brain injury show promising performance in predicting long-term neurological outcome in comatose OHCA patients. Study III had the same outcome measures as study II, but focused instead on the daily prognostication when adding cumulative information (as in clinical reality) until day 3 (72h). According to current international prognostication algorithms for comatose post-resuscitation cardiac arrest patients, should be observed in an intensive care setting for at least 72h post-ROSC prior to the initiation of the prognostication process.[56,111] Study III emphasised on how

important research-grade biomarkers such as NFL are, and the amount of predictive information this variable carry compared to other biomarkers. NFL showed an excellent performance (AUROC of around 95%) already on the first day (24 h) of intensive care. The prognostic performance remained at the same level after day 2 (48h) and day 3 (72h), respectively. These findings confirm previous research in this area. When clinically accessible biomarkers such as NSE were added, there was a significantly improved prognostic performance from day 1 to day3. When adding clinically accessible biomarker (level B) on day 3, the overall performance was similar to when using research-grade biomarkers (level C) on all 3 days of post-resuscitation care. Analyses of the ROC curves from study III shows that our models had comparable performance to the ERC/ESICM guidelines, as presented in a previous study by Moseby-Knappe et al.[219] This appears a bit surprising, considering that our models did not have access to crucial prognostic information such as neurophysiological tests or neurological imaging. These tests were performed only on indication and constituted a minority of the analysed patients from the TTM trial, and thereby not optimal for machine learning models due to missing values. When investigating this matter even further by using a threshold based on 100 % specificity on the training data (cross-validations), to predict the outcome in the test sets, the model with research-grade biomarkers (level C) on day 1 had only one false-positive prediction. The insufficient amounts of investigations performed during to the post-resuscitation care process, may on the other hand have been compensated by other variables such as 'patient's background', 'arrest characteristics' and 'prehospital circumstances', all well known to carry crucial predictive information but still not considered in the current recommended post-resuscitation guidelines.[42,111] In study III as well as in study II, we primarily used all accessible variables and secondarily created a simplified model. These two models had different aims; the simplified model is easy to use bedside but is not as accurate as the model with more variables, which is better suited for use as an integrated part of electronic medical records. In study III, we used a variable selector to find a reasonable number of variables for each model, in some cases reducing the number of variables by 90%. Which method to choose depends on the situation, the prognostic strength of the biomarkers and the aim of the prediction.

To our knowledge, this is the first prediction study using cumulative data in the three days of ICU admission and the first to combine the predictive capability of different groups of biomarkers by adding them to clinical variables. The cumulative approach is a natural step after testing biomarkers individually to understand how they are ranked and interact over time. From an ICU perspective, one of the strengths of this study was that we modified the study population and excluded patients who were either deceased or woke up during the first 72h after ICU admission. This strategy distilled the dataset to those patients that were at risk of a poor prognosis at each time point. Without doing so, the prediction models would be falsely enhanced as we would be predicting patients who had already woken up.

The black-box nature off machine learning was addressed in both study III and IV with the SHAP algorithm. SHAP can help explain how a prediction model works and mitigate some of the concerns about "black-box" modeling. It allows for each prediction to be individually explained, i.e. explains how each variable either increase or decrease the risk of a predicted outcome. It is important to note that the SHAP algorithm, powerful as it is, does not improve the model performance, but rather explains the reason behind each prediction. This can nevertheless aid the understanding of relations and complex dynamics between variables for the individual patient and as a group, and probably facilitate the use of future machine learning predictions models in both outcome-studies and clinical practice. Studies II and III in particular illustrate how machine learning could aid various aspects of intensive care medicine in the future.

Study IV

In this study, we developed and validated two machine learning algorithms to predict outcome in OHCA patients included in two separate registers - a merged cohort from INTCAR 1.0 and 2.0, and the cohort from the TTM1 trial. We compared an ANN to an XGBoost model and investigated their performance to predict the functional outcome at hospital discharge. In this study we focused on early prediction on admission to hospital as well as in study II, but due to the lack of data on long-term follow up in the INTCAR registries, we chose functional status on admission to hospital as the outcome measure. After development, both models were validated internally, on a randomly assigned test set of the same cohort (INTCAR1+2), and then externally on the TTM cohort. Both models performed equally well during training, and with no significant difference between AUROCs. They also performed equally well in the internal validation. However, when the models were tested in an external data set, they continued to perform well but with lower accuracy than in the development registry. Machine learning models, and other prediction models, tend to perform worse when externally tested and this is therefore somewhat expected. Two registries from different time periods will undoubtedly involve variation in case mix and the defined variables collected will slightly differ.

Study IV used two different cohorts of OHCA patients for analyses. The INTCAR1+2 with consecutively entered patients which may better reflect real-world practices compared to clinical trials with strict inclusion criteria. The TTM registry, on the other hand, contains a well-defined cohort of OHCA patients. The TTM trial was an international multicentre randomised controlled trial with predefined protocol-based criteria for inclusion and treatment with minimal data loss on the follow-up which had strict rules for multimodal neurological prognostication and WLST.

Study IV aimed to be a step forward towards an improved and more reliable prediction of functional outcome in comatose patients surviving cardiac arrest as external validation of predictive algorithms is essential to prove generalisability of the models. The quality of data is crucial for algorithm training and to develop reliable prediction models. Analyses of variable importance are essential to detect variables that carry much predictive information to simplify algorithms and test if models with fewer variables available have similar performance. However, machine learning models are still only suitable for group-level analyses, and not reliable enough for prediction of individual patients in an unselected OHCA population and therefore not yet suitable for clinical implementation. Future larger studies with more data points collected in high-resolution registries are crucial for the development of reliable machine learning models to be used in a clinical practice.

Limitations and study strengths were similar in all three studies regarding registry aspects. Both the TTM trial registry and the INTCAR-registries contain a well-defined cohort of comatose OHCA patients who regained ROSC, but nevertheless they differ in characteristics which of course must be considered when interpreting the results. All studies are post hoc analyses of the TTM trial, which enrolled only patients admitted to the hospital or ICU after OHCA of presumed cardiac cause. The TTM trial also had several exclusion criteria (such as limitations in therapy, including 'known illness making survival to 180 days unlikely' or 'do-not-resuscitate' orders, or). This makes the TTM cohort influenced by a selection bias that must be considered when interpreting the results. In study IV the TTM registry constituted the cohort used for external validation of our machine learning models whereas development and internal validation was performed on a merged data set from the INTCAR 1.0 and 2.0 registries. Early versions of INTCAR contain less high-resolution data and unfortunately lack clinical information from the post-resuscitation care in the ICU including standard blood tests, biomarkers, neurophysiological examinations and neuroimaging which are collected in later updated versions of OHCA registries. However, earlier studies (including study II and IV) have shown that information regarding patients' background, prehospital circumstances and arrest characteristics are closely related to later cerebral functional outcome.

It is essential to acknowledge the uncertainty present when using datasets that are the size of the TTM-trial dataset. The performance of models using such datasets can be too dependent on the train/test split and vulnerable to outliers. For example, the model performance was presumably better at the time of hospital admission (study II) than after 24h without the use of biomarkers (study III). The difference is noteworthy and must be kept in perspective when discussing this approach to cardiac arrest prognostication. Furthermore, the CPC-score was used in these studies has a certain inter-reviewer variability, which also can lead to considerable uncertainty.[134] All studies in this thesis used a dichotomised CPC score as the

outcome of the prediction models. Simplified as it is, this score gives information about the neurological outcome. Optimally the model would instead predict the specific CPC or mRS in each patient. However, considering the size of the TTM trial dataset, an attempt to predict a specific outcome score would probably result in prediction uncertainties. Larger and more high-resolution data from the 1900 patients enrolled in the TTM2 trial will probably increase the utility of multiclass predictions and allow valuable predictions at different time points: CPC and mRS score at ICU discharge, hospital discharge, after 6 months and after 24 months.

Artificial intelligence is associated with a lot of promises and expectations, but also with challenges and apprehension related to the difficulties to integrate them into existing processes and product. The implementation of machine learning models in a clinical context with critically ill patients will at this point probably raise more questions than it answers. Algorithms developed to predict outcome in individual OHCA patients in a real life clinical context, are still to be considered somewhat experimental and thereby probably not accepted by the majority of medical communities. Digitalisation is advancing and temptation to automate processes is constantly looming, not least in medical contexts. The principle of 'whatever can be automated will be automated' is continuously being integrated in modern society and is also evident in various medical systems. Machine learning is a form of automation or can at least be part of such processes. The benefits of establishing large high-resolution clinical registers for high quality clinical studies are becoming increasingly obvious and there is a genuine interest in adopting more effective methods to improve understanding of the complexity in such registries. It is crucial to define what type of problems can be formulated as machine learning problems, and how these new models can be implemented into current systems including risk prediction and to further improve our understanding of its pros and cons.

An interesting future use of ANN algorithms would be the possibility to reliably assess individual risk of a poor outcome in OHCA patients which could have clinical implications for early allocation to specific interventions (tailored therapy) and later in the clinical course to inform prognosis and continued life support. An area of more immediate use is the possibility to assign a predicted severity of illness on admission to help stratify analyses of clinical trials. In recent years, machine learning has been used increasingly in various studies and proved to be a promising method for data analyses. Machine learning has advantages compared to traditional regression models, i.e. the ability to detect correlations between independent variables in large complex data sets and to find trends or patterns in subsets of data. Finally, thorough ethical considerations will be crucial to ensure meaningful integration and unbiased use of machine learning. This applies for future clinical studies on cardiac patients as well as for developing clinical decision tools to reliable predict functional outcome and to tailor therapy in patients treated with post-resuscitation care in the ICU.

Hazards of adjustment in studies on temperature interventions

The publication of the Target Temperature Management after Out-of-hospital cardiac arrest (TTM) trial[66] has had significant effects on the utilization of targeted temperature management (TTM) for cardiac arrest.[220] Many of subsequent publications have shown a decrease in the use of TTM, despite continued strong recommendations in guidelines[56,111] and some indicate a tendency towards worse outcomes and increased mortality.[208,220] A variety of different factors might explain why before and after studies have tended towards increased mortality. One obvious reason could be that TTM to 33ºC is beneficial and that there exist clinically important differences between 33ºC and 36ºC, not detected in the TTM trial. Admittingly the confidence limits of the primary outcome included both possible clinically significant benefit and harm of 33ºC.[66] Another explanation could be that the results of the TTM trial were miss-interpreted and have resulted in a laissez faire attitude to temperature control; this could in turn be associated with less time bedside and less rigorous general intensive care. Differences could in some cases also be due to a change in case-mix over time. It is well known that the incidence of ST-elevation myocardial infarctions is decreasing, and it is reasonable to assume that elderly patients with low ejection fraction and cardiomyopathy represent an increasing proportion of cardiac arrest patients.[13] These patients have a higher baseline risk for mortality and might skew towards worse results over time.

To adjust for changes in case-mix over time, adjustment for comorbidity and clinical factors is warranted and propensity score analyses are sometimes considered. Propensity scores are appealing but could in some cases be inappropriate. In the case of cardiac arrest, many centres changed their TTM strategy from one day to another, leaving no room for overlap. This lack of overlap invalidates one of prerequisites for a propensity score analysis; it must be possible for each observation to receive either of the studied interventions. When intensive care units (ICUs) have moved completely to another target temperature this is no longer the case. In addition, any temperature effect will be a surrogate for all changes in care between the two time-periods. When centres continue to use a variety of temperatures, a propensity score might however be a useful tool.

Severity scoring systems are commonly used to predict mortality in ICU patients and some of them, including the Acute Physiology and Chronic Health Evaluation (APACHE) III and IV, use temperature as a variable when calculating illness severity. Several observational studies have indicated the association between lower admission temperature and worse outcome, both in cardiac arrest and in sepsis.[221,222] However, there are hazards when adjusting for severity of illness and comorbidities in trials on temperature interventions. One worth highlighting is related to the Simplified Acute Physiology Score (SAPS) 3. The SAPS 3 study group conducted

the SAPS 3 study in 2003 and published its findings starting from 2005.[77,223] The score has become a widely used system and uses temperature as one of the covariates to estimate ICU mortality, which has implications for statistical adjustments. According to SAPS 3 patients with a body temperature below 35ºC are assigned an additional score of (+7) at admission to ICU. Many centres initiate active cooling in the emergency department and permissive cooling is often started during transportation to hospital or when in the angiography suite, lowering the body temperature. Below, we will show how adjustment for SAPS 3 score might imply a non-existent temperature effect when adjusting for the estimated mortality rate in cardiac arrest patients treated in the ICU.

Methods

A simulation data set with 10.000 cases was created. We used variables presented in a recent article describing TTM-use in a national registry (age, sex, witnessed arrest, bystander-CPR, comorbidity-score and SAPS 3 score), and assigned random values based on the means and standard deviations.[220] These variables were used to predict outcome (death). We created a logistic regression model using coefficient estimates from the TTM trial which are generally similar to what has been found in other trials. For SAPS 3 we used a ß-estimate of 0.05 (the corresponding odds ratio (OR) being 1.05)[83] as shown in table 15. The model intercept was chosen to achieve an overall mortality rate of approximately 49% (as in the TTM trial).

A temperature variable (TTM33 or TTM36) was added at random with no relation to the outcome variable. Logistic regression was performed to predict mortality. As temperature was a totally random variable the OR approximates 1 (not significant). The SAPS 3 scores were adjusted by adding a score of (+7) to a varying proportion on random cases in the TTM33 group. This corresponds to the clinical treatment of patients who are managed at 33ºC, where cooling is often initiated prior to ICU arrival and SAPS 3 temperature is captured within the first hour of ICU admission. In the TTM trial 54% of patients in the 33ºC-group had a temperature below 35ºC 1 hour after randomization. In the 36ºC-group, 37% had a core temperature below 35ºC at this time point. Thus, 17% of patients in the 33ºC group would have a higher predicted mortality using the SAPS 3 algorithm due to targeted and induced cooling in itself.

Since the dataset used was simulated and contained no real-life patient information this study was not sent for ethics approval. R was used for all statistical analysis (R Core Team, 2013).[224]

Table 15. Estimated mortality prediction of a poor outcome with variables presented in a recent article describing TTM-use in a national registry. The table lists variables previously shown to influence outcome after out-of-hospital cardiac arrest with their coefficient and odds ratio of a poor outcome. The logistic regression model used coefficient estimates from the TTM trial which are generally similar to what has been found in other trials. For SAPS 3 we used a ß-estimate of 0.05 (the corresponding odds ratio (OR) being 1.05). The model intercept was chosen to achieve an overall mortality rate of 49% (as in the TTM trial). SAPS 3, Simplified Acute Physiology Score 3 score; CPR, cardiopulmonary resuscitation.

	Coefficient	Odds Ratio (OR)
Intercept	-5.68	-
SAPS 3 score	0.05	1.05
>4 comorbidities	0.55	1.73
1-3 comorbidities	0.22	1.25
Bystander-CPR	-0.51	0.60
Age	0.05	1.05
Male sex	-0.43	0.65
Witnessed arrest	-0.50	0.61

Results

In table 16 we show the effect of the SAPS 3 risk-scoring. The scores in the 33ºC-group increase gradually with an increased proportion of simulated cardiac arrest patients targeted to TTM33 having reached a body temperature below 35°C on admission to ICU, compared to those targeted to TTM36 where the SAPS 3 scores remain the same. As a consequence, the OR for a poor outcome when receiving TTM36 increases with the number of patients having been cooled below 35ºC at time of admission to ICU. In the current example a nominally significant effect (p<0.05) of a lower target temperature could be shown by entering a lower temperature on ICU admission for 15% in patients managed at the lower of two temperatures. Figure 33 illustrates outcome in relation to SAPS 3 with three proportions of patients in TTM33 having reached below 35°C at the time of SAPS 3 registration. The proportion of patients in the lower temperature group that needs to cross the ”SAPS 3-35ºC-threshold” will vary depending on covariate estimates.

Table 16. SAPS 3 score in relation to the proportion of simulated patients having reached a targeted temperature below 35ºC by SAPS 3 registration. The table shows how the SAPS 3 scores in the 33ºC Group increase with an increased proportion of simulated cardiac arrest patients targeted to TTM33 having reached a body temperature below 35°C on admission to ICU, compared to those targeted to TTM36 where the SAPS 3 scores remain the same. As a consequence, the odds ratio for a poor outcome when receiving TTM36 increases spuriously with the number of patients having been cooled below 35ºC at time of admission to ICU. Results were obtained from a simulated model adjusting for known predictors of outcome. The model intercept was chosen to achieve an overall mortality rate in both temperature groups of approximately 49% (as in the TTM trial). ICU, intensive care unit, SAPS 3, Simplified Acute Physiology Score 3 score; TTM36, targeted temperature management at 36ºC.

	SAPS 3 score 33°C group (mean)	SAPS 3 score 36°C group (mean)	Odds Ratio of a poor outcome with TTM36
No temperature difference	71.33	71.33	0.97
20% cooled below 35°C	72.69	71.33	1.03
40% cooled below 35°C	74.05	71.33	1.10
60% cooled below 35°C	75.41	71.33	1.18
All cooled below 35°C	78.14	71.33	1.37

Discussion

Our findings show that the temperature component used to calculate SAPS 3 score during the first hour following admission to intensive care greatly influence the predicted hospital mortality rate in cardiac arrest patients. There are several pitfalls when analysing observational outcomes due to temperature interventions, including propensity scoring and adjustment for baseline risk. Patients treated with a target temperature of 33ºC will consequently have higher predicted mortality rate if their induced lower body temperature is used to calculate a SAPS 3 score compared to patients treated at 36ºC as illustrated in figure 33. This relation will therefore be beneficial to the 33ºC-group if the temperature component of SAPS 3 is not excluded. This phenomenon is also evident in analyses of other interventions (e.g. blood pressure intervened by vasopressors or Glasgow Coma Score intervened by sedation) where a variable with direct effect on the calculated risk score is not excluded from the model.

Validated yet simple early prognostic scoring models are particularly sought for in medical conditions with high mortality such as cardiac arrest. Earlier scores used to predict in-hospital mortality (SAPS II and APACHE II)[225,226] or severity (SOFA)[227] of patients admitted to ICU were calculated from data collected during the first 24h of ICU admission. The SAPS 3 score is calculated from clinical and laboratory data available immediately at the time of ICU admission.[83] Albeit the suggestion from The American Heart Association (AHA) that the SAPS 3 score could effectively predict hospital mortality in the population of resuscitated post-cardiac arrest patients[126] several studies have failed to prove any convincing accuracy of outcome prediction or illness severity scoring capability in this specific population.[83,84]

Since the TTM trial in 2013, many studies analysing a change in cardiac arrest outcome following the global modification of clinical practice and a shift in TTM from 33ºC to 36ºC have been published.[69,207,208,215,220] Some of these studies reported a tendency to an increase in both in-hospital- and long-term (6 months) mortality. 4However, some of these studies have not adjusted for the temperature component of the SAPS 3 score used for risk classification and comparison between temperature groups with a subsequent hazard to correctly interpret the results.

Limitations

All analyses were performed on simulated data-sets which means that the real effect can first be evaluated after re-analyses of already published studies or by examining the effect prospectively. It is also noteworthy that the proportion of change in temperature and the choice of covariate coefficients obviously affect the results.

Figure 33. SAPS 3 score predicted probability of death from a simulated model with all other variables set to mean values. The figure exemplifies the SAPS 3 score predicted probability of death when 0%, 40% and 100% of simulated cardiac arrest patients targeted to TTM33 reach a body temperature below 35°C before admission to ICU. This simulation indicates that patients treated with TTM36 will have a worse predicted outcome compared to those treated with TTM33 when using SAPS 3 for severity scoring without adjusting for temperature. The red curve represent the predicted mortality in both temperature groups when 0% are cooled. In an adjusted model, the 33ºC-group has a lower predicted probability of death when 40% (green curve) and 100% (blue curve) are cooled due to higher SPAS 3 scores. ICU, intensive care unit; SAPS 3, Simplified Acute Physiology Score 3 score; TTM36, targeted temperature management at 36ºC; TTM33, targeted temperature management at 33ºC

Conclusion

Observed changes in outcomes and use of temperature management after cardiac arrest are intriguing and complex. When adjusting for SAPS 3 the temperature component of the score can result in spurious findings in outcome studies on target temperature management which need to be taken into account when analysing data and interpreting the results. A solution would be to create TTM-adjusted severity scoring models for cardiac arrest patients. The observations from this project were published in a condensed letter to the Editor in *Resuscitation* in February 2021 (paper V).

Conclusions

- No significant difference in functional outcome at hospital discharge was found in OHCA patients selected from the INTCAR 2.0-registry, when treated with lower- vs higher targeted temperature. This supports the findings from the TTM trial.
- Our ANN predicted long-term functional outcome in OHCA patients well and outperformed a conventional model based on logistic regression using information available at time of hospital admission. Factors related to the pre-hospital setting carried the most predictive information. No specific risk group was identified to benefit from TTM of either 33°C nor 36°C. ANN may stratify a heterogenous trial population in risk classes and help determine intervention effect across subgroups.
- When adding cumulative information collected during the initial three days of post-resuscitation care, the ANNs provided good to excellent prognostic accuracy in predicting functional outcome in comatose patients post-OHCA. Clinically accessible biomarkers such as NSE and research-grade biomarkers such as NFL increased the prognostic performance.
- Both ANN and XGBoost models developed on a clinical OHCA cohort predicted functional outcome at hospital discharge equally and accurately. The models were validated internally, as well as externally on data from the TTM trial registry. Both models continued to perform well, indicating generalisability, but with lower discrimination than in the development cohort.
- SAPS 3 is not an optimal severity scoring model when used in outcome studies on patients undergoing temperature intervention. This needs to be taken into account when analysing data and interpreting the results. A solution would be to create TTM-adjusted severity scoring models for cardiac arrest patients.
- TTM-adjusted severity scoring models would probably improve the assessment of mortality in cardiac patients undergoing temperature intervention.

Future perspectives

Larger trials with more data points collected in high-resolution registries are crucial for the development of reliable machine learning models to be used in a clinical context. The Targeted Temperature Management 2 (TTM2) trial is a recently finished not yet published trial with 1900 adult OHCA patients randomised to a targeted temperature of 33°C or to normothermia with early treatment of fever (≥37,8°C).[73] This trial will hopefully result in a better understanding of the role for TTM in clinical practice but also in improved and clinically valuable prediction models, which may enable outcome prognostication focused at different time points during the post-cardiac arrest process, i.e. CPC and mRS score at ICU discharge, hospital discharge, after 6 months and after 24 months.

The rapid and prodigious increase in computational power in recent decades, have led to an impressive development of artificial intelligence techniques, to be used as an alternative to conventional statistical methods when analysing large and complex high-resolution registries. These supervised machine learning models will continue to develop and progress in their ability to capture non-linear feature correlations in big-data collections. This will not only improve their accuracy to discriminate between dependencies, but also their generalisability to perform when exposed to new data.

As the models continue to improve in performance by training on big data sets with clinical information from patient's background and comorbidities, cardiac arrest characteristics and prehospital circumstances, findings from the primary assessment in the ER and finally information with high prognostic value (including biomarkers of neurological injury, neurophysiology and neuroimaging) cumulatively collected during post-resuscitation care, a future use could be the further development of clinically reliably multimodal prognostication algorithms for prediction of individual long-term outcome in cardiac arrest patients. Such algorithms should constantly be exposed to real-time clinical data and continuously predict functional outcome to guide clinical decision and tailor further therapy for the individual cardiac arrest patient.

Summary in Swedish

Populärvetenskaplig sammanfattning

Primär hjärtsjukdom som leder till plötsligt hjärtstopp är trots den senaste tidens positiva trend till förbättrad överlevnad både nationellt och internationellt, ett fortsatt mycket farligt tillstånd med hög dödlighet. I samband med hjärtstoppet upphör blodcirkulationen till kroppens samtliga organ och om ingenting görs är döden ett oundvikligt faktum. Hjärtstopp är globalt sett en mycket vanlig orsak till inläggning på en intensivvårdsavdelning för de patienter som kan återupplivas och därmed återfår spontan hjärtaktivitet och återställd blodcirkulation till hjärnan.

Mycket har hänt inom hjärtstoppsjukvården, såväl inom den medicinska vården som i ett större samhällsperspektiv. Tillgången på hjärtstartare i publika områden har väsentligen ökat, liksom lekmäns kunskaper och utbildning i hjärtlungräddning. Ambulanssjukvården har utvecklats och har numera bättre möjligheter än någonsin att snabbt kunna ta över en förhoppningsvis redan påbörjad basal hjärtlungräddning och därmed bidra med ytterligare avancerad hjärtlungräddning med bland annat hjärtstartare, utrustning för automatiska hjärtkompressioner samt möjlighet att ge olika akuta hjärtläkemedel på väg in till akutmottagningen. Inom sjukvården har tillgången på avancerad hjärtsjukvård ökat, däribland akut ballongvidgning av hjärtats kranskärl, och för de patienter som återupplivas erbjuds modern intensivvård med protokollstyrd hjärtstoppsvård, inklusive behandling med temperaturkontroll under avancerad övervakning. De yttre förutsättningarna för en framgångsrik behandling och ökad överlevnad har aldrig varit bättre än nu. Men trots alla dessa insatser och positiva framsteg är överlevnaden efter hjärtstopp fortsatt låg. Fortfarande avlider varannan av de patienter som vårdas på en intensivvårdsavdelning efter hjärtstopp. Majoriteten av dessa dödsfall beror dock på att den pågående intensivvården avslutas i förtid vilket i sin tur baseras på en bedömning att patienten ådragit sig alltför omfattande hjärnskador till följd av den syrebrist som uppstått i hjärnan. Denna hjärnskada uppstår primärt som en direkt konsekvens av syrebristen i samband med cirkulationsstillståndet, men även till följd av olika sekundära skademekanismer och inflammationsprocesser som uppkommer efter återställd blodcirkulation till kroppens vävnader och organ. Att fortsätta, alternativt avsluta intensivvården och därmed de livsuppehållande åtgärderna för en hjärtstoppspatient, är en svår men ytterst viktig uppgift för den medicinska professionen. En uppgift som innefattar komplexa medicinska

bedömningar men även mycket svåra etiska överväganden med direkt livsavgörande konsekvenser för såväl patienten som för dennes anhöriga.

Beslutet kring det fortsatta omhändertagandet grundar sig på en prognosbedömning av patientens sannolikhet för en god neurologisk återhämtning. Detta innebär med andra ord, att hos de patienter som är fortsatt medvetslösa till följd av hjärtstillestånd, försöka förutspå sannolikheten för ett bra funktionellt utfall. Ett utfall som naturligtvis innebär överlevnad, men ännu viktigare återhämtning till en god och acceptabel livssituation ur patientens perspektiv. Ett scenario där sjukvården genom avancerade och ytterst resurskrävande insatser räddar liv, men till priset av svåra bestående hjärnskador och en väsentligt lägre funktionsgrad och livskvalitet än den som patienten uppvisade innan insjuknandet, är varken i patientens bästa intresse eller medicinskt och etiskt försvarbart.

Det finns internationella rekommendationer för hur en prognosbedömning efter hjärtstopp bör gå till och dessa standardiserade rekommendationer uppdateras regelbundet allteftersom kunskapsläget förändras. Enligt den nu gällande och nyligen uppdaterade prognostiseringsalgoritmen, bör förutom den kontinuerliga kliniska undersökningen, även stegvisa kombinationer av olika diagnostiska metoder genomföras inför en samlad prognosbedömning (multimodal prognostisering). Den multimodala prognostiseringen bör ske tidigast 72 timmar efter hjärtstoppet och återspeglas schematiskt i figur 6. De undersökningar som genomförs i samband prognostiseringen speglar olika delar av skademekanismerna sekundärt till syrebristen, t.ex. svullnad av hjärnvävnaden på röntgen, påvisandet samt koncentrationstrenden av en hjärnskademarkör uppmätt i blodet samt olika metoder för att värdera och testa hjärnans elektriska impulser. Genom att kombinera flera undersökningsmetoder, var och en utvecklade till att undersöka och värdera utfall på gruppnivå, vill man skapa en så väl underbyggd vetenskaplig grund som möjligt inför bedömningen av den individuella patientens förutsättningar för en god funktionell återhämtning. En bedömning där risken i princip ska vara obefintlig för att på felaktiga grunder avsluta intensivvården hos en patient som trots allt har förutsättningar för en god återhämtning.

Anmärkningsvärt är att all tillgänglig information gällande patienternas bakgrund, ålder, tidigare/nuvarande sjukdomar, omständigheter kring hjärtstoppet och det prehospitala scenariot, samt den information som erhålls i samband med det primära omhändertagandet på akutmottagning (då patientens sedermera läggs in på sjukhus) överhuvudtaget inte används på ett strukturerat sätt i den senare prognostiseringsprocessen på intensiven. Trots att man i flera tidigare studier kunnat visa på att dessa variabler har stor prognostisk valör för det funktionella utfallet efter hjärtstopp, så används de inte alls som grund för prognosbedömningen enligt de nu gällande internationella riktlinjerna.

Syftet med **Studie I** var att undersöka om olika nivåer av målstyrd temperaturbehandling (TTM) som del av den protokollstyrda intensivvården efter hjärtstopp,

kunde associeras med patienternas funktionella utfall vid utskrivning från sjukhus. Genom att analysera samlad data från ett stort internationellt hjärtstoppsregister (INTCAR) kunde vi se att effekten av själva temperaturbehandlingen inte hade någon signifikant koppling till hur det senare gick för patienterna. Resultatet från denna studie motsvara även tidigare genomförda studier inom området. Vi noterade att patienter som drabbas av hjärtstopp är väldigt olika (heterogena) vilket komplicerar jämförelsen dem emellan då man gör vetenskapliga studier. Detta hade delvis kunnat kompenseras genom indelning av patienterna i olika riskskattningsgrupper för enklare jämförelse. En stor fördel med det neutrala resultatet i studie I var att vi kunde använda detta patientregister för att även genomföra studier med andra frågeställningar utan att ta hänsyn till temperaturbehandlingens eventuella effekt på det funktionella utfallet.

I **Studie II** undersökte vi om det redan tidigt i förloppet efter hjärtstopp kunde gå att förutspå långtidsföljderna för patienterna genom att enbart värdera de variabler som finns tillgängligt redan när patienten läggs in på sjukhus. Sådana variabler utgörs bland annat av allmän bakgrundsinformation, prehospitala faktorer, omständigheter kring själva hjärtstoppet samt resultatet från den initiala läkar-undersökningen på akutmottagningen. Vi valde att använda oss av en så kallad övervakad maskininlärningsmodell (en sorts artificiell intelligens) vid namn ANN (artificiellt neuralt nätverk) i syfte att analysera data från ett detaljerat och välkontrollerat hjärtstoppsregister (TTM registret). Vi jämförde sedan vår egen modells förmåga att förutspå långtidsutfallet, med en tidigare genomförd studie och en modell på samma registerdata men som istället bygger på traditionell statistik. Vårt ANN presterade bäst och förutsåg det funktionella utfallet med god precision och dessutom enbart med informationen som fanns tillgänglig vid tidpunkten för inläggning på sjukhus. Vi noterade även att de prehospitala variablerna var de som bidrog allra mest med prediktiv information.

I **Studie III** undersökte vi om ett ANN i kombination med information från olika blodprover kunde förbättra riskbedömning ytterligare. I analysen inkluderades alla de blodprover, inklusive specifika hjärnskademarkörer, som analyserades dagligen under tre dygn för de patienter som inte vaknat upp efter hjärtstoppet, som inte redan avlidit, och som dessutom inkluderats i TTM studien. Även om vi i denna studie inte hade tillgång till data från röntgenundersökningar av hjärnan eller från undersökningar av hjärnans elektriska impulser, så var resultaten ändå så pass bra att vissa av våra modeller var fullt jämförbara med nu gällande prognostiserings-modeller.

I **Studie IV** undersökte vi två olika typer av maskininlärningsmodeller (ANN och XGBoost) och jämförde deras förmåga att förutspå hjärtstoppspatienters funktionella utfall redan vid utskrivning från sjukhus och enbart genom att analysera variabler redan tillgängliga vid inläggning på sjukhus. Genom att utveckla och träna modellerna på ett data register och sedan testa deras förmåga att förutspå utfallet hos hjärtstoppspatienter från ett separat data register (så kallad extern validering),

fick vi en bättre uppfattning om modellernas generaliserbarhet, dvs hur effektiva de är på att analysera och hitta mönster mellan olika variabler i helt ny data. Vår analys visade att de båda machine learning modellerna presterade precis lika bra under själva träningsfasen som vid intern validering (testad på samma register som träningen skett på) när det gällde att förutsäga funktionellt utfall. Dock noterades att båda modeller presterade något sämre när de testades på ett helt okänd data (extern validering) vilket också är förväntat, då variabler såväl som patientsammansättning (case-mix) skiljer sig åt mellan olika register.

Medvetslösa hjärtstoppspatienter som vårdas på intensiven är en heterogen patient-grupp, vilket innebär stora utmaningar och svårigheter att objektivt utvärdera och pålitligt prognostisera det funktionella långtidsutfallet på individnivå. Även på gruppnivå innebär dessa patienter, särskilt de som genomgår någon form av aktiv temperaturbehandling, svårigheter då ordinarie poängskalor för riskskattning inom intensivvård (till exempel SAPS 3) inte är fullt ut anpassade för att värdera patienter med aktivt sänkt kroppstemperatur som en behandlingsstrategi. Denna problematiken rör företrädesvis analys och tolkning av vissa hjärtstoppstudier vilket sammanfattas i **Publikation V**.

Ett problem med maskininlärningsmodeller är att de är mycket komplicerade och kräver såväl stor datorkraft som expertkunnande inom programmering. Detta borgar för ett fortsatt spännande samarbete mellan de medicinska- och datortekniska professionerna framöver. De olika modellerna har redan visat på lovande förmågor och intressanta resultat inom flera medicinska områden. Detta gäller inte minst analyser av stora komplexa dataregister med hjärtstoppspatienter, vilket även vår senaste forskning har kunnat presentera. Med all sannolikhet kommer denna maskininlärningsteknik fortsätta utvecklas och förbättras med en förhoppning om att i framtiden kunna bidra till ännu säkrare och effektivare prognostiserings-algoritmer som kan underlätta bedömningen av såväl hjärtstoppspatienter som andra patientkategorier inom intensivvården.

Acknowledgements

Niklas Nielsen - my main supervisor, colleague, scientific mentor and role model as researcher, doctor and intensivist. But most of all, a dear friend. I admire your true sense of science in general, and your inspiring spirit of enjoying the truly good things in life in particular. Thank you for your never-ending patience, availability, helpfulness, guidance and perspectives. But most of all – your amazing sense of humour! We finally made it!!

Martin Annborn, Attila Frigyesi and **Sten Walther** – my co-supervisors for invaluable input, support in writing, good advice and for sharing your great knowledge in science, statistics, severity scoring models and intensive care medicine.

Ola Björnsson, for all your excellent work in training and testing our networks but most of all for explaining all incomprehensible aspects of machine learning and ANNs. Thank you for straightening out the question marks and especially for delivering calculations in the middle of the night!

Peder Andersson, a dear colleague and co-author who possesses the unusual and very impressive quality to master computer science as well as intensive care. It has been a privilege to share your knowledge and skills.

Josef Dankiewicz, a true Master of R! Thank you for all help and support with statistics, study design and for your excellent skills in managing cardiac arrest registers.

Josefine Wahlström, my sidekick and main co-author on the first paper. You did a wonderful job!

All national and international co-authors, for making the papers included in this thesis beyond great. It has been such a privilege to write, communicate and learn from you!

Colleagues and friends in the department of Anaesthesia and Intensive Care, Helsingborg Hospital, thank you for all the good times and valuable support during the finishing of this thesis. I will be back for clinical duty any day now!

Ulrika Pahlm, my boss for encouragement and for providing me with necessary research time and financial means despite an ongoing pandemic.

Colleagues and friends on the Ambulance Helicopter and AnOpIVA, Gällivare Hospital, for wonderful memories, exciting experiences and all the good laughs during my northern years. I truly miss working with you!

Ulla-Britt Karlsson, for all the incredibly valuable help and support with research administration and management of financial support from the Gorthon foundation.

Simon Heissler, for magnificent support in delivering computer programs, bookings and for fixing all administrative issues that you may encounter as a PhD-student.

Araz Rawshani and **Johan Herlitz,** for interesting discussions and valuable input in machine learning models and the management of patient registers.

Gisela Lilja and **Erik Blennow Nordström,** for valuable contribution and help regarding the assessment of functional outcome after cardiac arrest.

Florian Ebner and **Sofia Backman,** for letting me use some of your very illustrative figures.

Carolina and Andreas Wickander, for all your support with English grammar, and for correcting some of my dubious 'Google-translate' constructed sentences with endless clauses.

Noah and Axel, my two clever, thoughtful and extremely annoying boys who bring me endless joy and grey hairs. You are, without any competition, my two greatest achievements in life. Always do your best and choose your own paths!

Amelie - my wife and other half in life. Thank you for your love, unfailing support and for putting up with (in your own words) "the most boring man in the whole world", while finishing this thesis. It will finally be the 8th of May any day now…

My dear family and close friends – you know who you are! Thank you for your love, companionship, support and faith in my various projects.

In memory of my mother Lena who left this world far too soon.

Jesper

Financial support

All this work, ending up in a long-awaited doctoral dissertation, would not have been possible without generous research grants and financial support from the Stig och Ragna Gorthon Foundation, the Thelma Zoegas Foundation, VO FoUU Skåne Sjukhus Nordväst, Södra Sjukvårdsregionen, the Thorsten Birger Segerfalk Foundation, the Swedish Society of Medicine and the European Regional Development Fund. Thank you so much for this great opportunity!

References

1. Lagen om kriterier för bestämmande av människans död (1987:269). Socialstyrelsens föreskrifter och allmänna råd (SOSFS 2005:10).
2. Wit AL, Janse MJ. Reperfusion arrhythmias and sudden cardiac death: a century of progress toward an understanding of the mechanisms. Circ Res 2001;89:741-3.
3. Sunde K, Kramer-Johansen J, Pytte M, Steen PA. Predicting survival with good neurologic recovery at hospital admission after successful resuscitation of out-of-hospital cardiac arrest: the OHCA score. Eur Heart J 2007;28:773; author reply -4.
4. Bulut S, Aengevaeren WR, Luijten HJ, Verheugt FW. Successful out-of-hospital cardiopulmonary resuscitation: what is the optimal in-hospital treatment strategy? Resuscitation 2000;47:155-61.
5. Huikuri HV, Castellanos A, Myerburg RJ. Sudden death due to cardiac arrhythmias. N Engl J Med 2001;345:1473-82.
6. Pell JP, Sirel JM, Marsden AK, Ford I, Walker NL, Cobbe SM. Presentation, management, and outcome of out of hospital cardiopulmonary arrest: comparison by underlying aetiology. Heart 2003;89:839-42.
7. Dicker B, Davey P, Smith T, Beck B. Incidence and outcomes of out-of-hospital cardiac arrest: A New Zealand perspective. Emerg Med Australas 2018;30:662-71.
8. Chugh SS, Reinier K, Teodorescu C, et al. Epidemiology of sudden cardiac death: clinical and research implications. Prog Cardiovasc Dis 2008;51:213-28.
9. Marijon E, Uy-Evanado A, Dumas F, et al. Warning Symptoms Are Associated With Survival From Sudden Cardiac Arrest. Ann Intern Med 2016;164:23-9.
10. Patil KD, Halperin HR, Becker LB. Cardiac arrest: resuscitation and reperfusion. Circ Res 2015;116:2041-9.
11. Nichol G, Thomas E, Callaway CW, et al. Regional variation in out-of-hospital cardiac arrest incidence and outcome. JAMA 2008;300:1423-31.

12. Benjamin EJ, Blaha MJ, Chiuve SE, et al. Heart Disease and Stroke Statistics-2017 Update: A Report From the American Heart Association. Circulation 2017;135:e146-e603.
13. Yeh RW, Sidney S, Chandra M, Sorel M, Selby JV, Go AS. Population trends in the incidence and outcomes of acute myocardial infarction. N Engl J Med 2010;362:2155-65.
14. Rabjohns J, Quan T, Boniface K, Pourmand A. Pseudo-pulseless electrical activity in the emergency department, an evidence based approach. Am J Emerg Med 2020;38:371-5.
15. Paradis NA, Martin GB, Goetting MG, Rivers EP, Feingold M, Nowak RM. Aortic pressure during human cardiac arrest. Identification of pseudo-electromechanical dissociation. Chest 1992;101:123-8.
16. Goto Y, Maeda T, Nakatsu-Goto Y. Prognostic implications of conversion from nonshockable to shockable rhythms in out-of-hospital cardiac arrest. Crit Care 2014;18:528.
17. Holmberg M, Holmberg S, Herlitz J. Incidence, duration and survival of ventricular fibrillation in out-of-hospital cardiac arrest patients in sweden. Resuscitation 2000;44:7-17.
18. Wissenberg M, Lippert FK, Folke F, et al. Association of national initiatives to improve cardiac arrest management with rates of bystander intervention and patient survival after out-of-hospital cardiac arrest. JAMA 2013;310:1377-84.
19. Chan PS, Krumholz HM, Nichol G, Nallamothu BK, American Heart Association National Registry of Cardiopulmonary Resuscitation I. Delayed time to defibrillation after in-hospital cardiac arrest. N Engl J Med 2008;358:9-17.
20. Girotra S, Nallamothu BK, Spertus JA, et al. Trends in survival after in-hospital cardiac arrest. N Engl J Med 2012;367:1912-20.
21. Andersen LW, Lind PC, Vammen L, Hoybye M, Holmberg MJ, Granfeldt A. Adult post-cardiac arrest interventions: An overview of randomized clinical trials. Resuscitation 2020;147:1-11.
22. Herlitz J, Bang A, Ekstrom L, et al. A comparison between patients suffering in-hospital and out-of-hospital cardiac arrest in terms of treatment and outcome. J Intern Med 2000;248:53-60.
23. Perkins GD, Brace SJ, Smythe M, Ong G, Gates S. Out-of-hospital cardiac arrest: recent advances in resuscitation and effects on outcome. Heart 2012;98:529-35.

24. Hansen CM, Kragholm K, Granger CB, et al. The role of bystanders, first responders, and emergency medical service providers in timely defibrillation and related outcomes after out-of-hospital cardiac arrest: Results from a statewide registry. Resuscitation 2015;96:303-9.
25. Gates S, Smith JL, Ong GJ, Brace SJ, Perkins GD. Effectiveness of the LUCAS device for mechanical chest compression after cardiac arrest: systematic review of experimental, observational and animal studies. Heart 2012;98:908-13.
26. Yan S, Gan Y, Jiang N, et al. The global survival rate among adult out-of-hospital cardiac arrest patients who received cardiopulmonary resuscitation: a systematic review and meta-analysis. Crit Care 2020;24:61.
27. Rajan S, Wissenberg M, Folke F, et al. Association of Bystander Cardiopulmonary Resuscitation and Survival According to Ambulance Response Times After Out-of-Hospital Cardiac Arrest. Circulation 2016;134:2095-104.
28. Hasselqvist-Ax I, Riva G, Herlitz J, et al. Early cardiopulmonary resuscitation in out-of-hospital cardiac arrest. N Engl J Med 2015;372:2307-15.
29. Stub D, Smith K, Bray JE, Bernard S, Duffy SJ, Kaye DM. Hospital characteristics are associated with patient outcomes following out-of-hospital cardiac arrest. Heart 2011;97:1489-94.
30. Schober A, Sterz F, Laggner AN, et al. Admission of out-of-hospital cardiac arrest victims to a high volume cardiac arrest center is linked to improved outcome. Resuscitation 2016;106:42-8.
31. Thomas H, Diamond J, Vieco A, et al. Global Atlas of Cardiovascular Disease 2000-2016: The Path to Prevention and Control. Glob Heart 2018;13:143-63.
32. Berdowski J, Berg RA, Tijssen JG, Koster RW. Global incidences of out-of-hospital cardiac arrest and survival rates: Systematic review of 67 prospective studies. Resuscitation 2010;81:1479-87.
33. Kiguchi T, Okubo M, Nishiyama C, et al. Out-of-hospital cardiac arrest across the World: First report from the International Liaison Committee on Resuscitation (ILCOR). Resuscitation 2020;152:39-49.
34. Grasner JT, Wnent J, Herlitz J, et al. Survival after out-of-hospital cardiac arrest in Europe - Results of the EuReCa TWO study. Resuscitation 2020;148:218-26.
35. Rawshani AH, J. et al. Annual report of cardiac arrest from the Swedish National Quality Registry for Cardiopulmonary Resuscitation. 2020.

36. Taglieri N, Saia F, Bacchi Reggiani ML, et al. Prognostic significance of shockable and non-shockable cardiac arrest in ST-segment elevation myocardial infarction patients undergoing primary angioplasty. Resuscitation 2018;123:8-14.
37. Kouwenhoven WB, Jude JR, Knickerbocker GG. Closed-chest cardiac massage. JAMA 1960;173:1064-7.
38. Cardiopulmonary resuscitation. JAMA 1966;198:372-9.
39. Pantridge JF, Geddes JS. A mobile intensive-care unit in the management of myocardial infarction. Lancet 1967;2:271-3.
40. Cummins RO, Ornato JP, Thies WH, Pepe PE. Improving survival from sudden cardiac arrest: the "chain of survival" concept. A statement for health professionals from the Advanced Cardiac Life Support Subcommittee and the Emergency Cardiac Care Committee, American Heart Association. Circulation 1991;83:1832-47.
41. Nolan J, Soar J, Eikeland H. The chain of survival. Resuscitation 2006;71:270-1.
42. Soar JB, B. W.; Carli, P.; Couper, K.; Deakin, C. D.; Djärv T.; Lott, C.; Olasveengen, T.; Paal, P.; Pellis, T.; Perkins, G. D.; Sandroni, C.; Nolan, J. P. European Resuscitation Council Guidelines 2021: Adult advanced life support. Resuscitation 2021;161:115-51.
43. Martinell L, Nielsen N, Herlitz J, et al. Early predictors of poor outcome after out-of-hospital cardiac arrest. Crit Care 2017;21:96.
44. Wibrandt I, Norsted K, Schmidt H, Schierbeck J. Predictors for outcome among cardiac arrest patients: the importance of initial cardiac arrest rhythm versus time to return of spontaneous circulation, a retrospective cohort study. BMC Emerg Med 2015;15:3.
45. Hazinski MF, Nolan JP, Aickin R, et al. Part 1: Executive Summary: 2015 International Consensus on Cardiopulmonary Resuscitation and Emergency Cardiovascular Care Science With Treatment Recommendations. Circulation 2015;132:S2-39.
46. Perkins GD, Ji C, Deakin CD, et al. A Randomized Trial of Epinephrine in Out-of-Hospital Cardiac Arrest. N Engl J Med 2018;379:711-21.
47. Neumar RW, Nolan JP, Adrie C, et al. Post–Cardiac Arrest Syndrome. Circulation 2008;118:2452-83.
48. Stub D, Bernard S, Duffy SJ, Kaye DM. Post cardiac arrest syndrome: a review of therapeutic strategies. Circulation 2011;123:1428-35.
49. Laver S, Farrow C, Turner D, Nolan J. Mode of death after admission to an intensive care unit following cardiac arrest. Intensive Care Med 2004;30:2126-8.

50. Laurent I, Monchi M, Chiche JD, et al. Reversible myocardial dysfunction in survivors of out-of-hospital cardiac arrest. J Am Coll Cardiol 2002;40:2110-6.
51. Sunde K, Pytte M, Jacobsen D, et al. Implementation of a standardised treatment protocol for post resuscitation care after out-of-hospital cardiac arrest. Resuscitation 2007;73:29-39.
52. Adrie C, Adib-Conquy M, Laurent I, et al. Successful cardiopulmonary resuscitation after cardiac arrest as a "sepsis-like" syndrome. Circulation 2002;106:562-8.
53. Sasson C, Rogers MA, Dahl J, Kellermann AL. Predictors of survival from out-of-hospital cardiac arrest: a systematic review and meta-analysis. Circ Cardiovasc Qual Outcomes 2010;3:63-81.
54. Coles JP, Fryer TD, Coleman MR, et al. Hyperventilation following head injury: effect on ischemic burden and cerebral oxidative metabolism. Crit Care Med 2007;35:568-78.
55. Nolan JP, Neumar RW, Adrie C, et al. Post-cardiac arrest syndrome: epidemiology, pathophysiology, treatment, and prognostication. A Scientific Statement from the International Liaison Committee on Resuscitation; the American Heart Association Emergency Cardiovascular Care Committee; the Council on Cardiovascular Surgery and Anesthesia; the Council on Cardiopulmonary, Perioperative, and Critical Care; the Council on Clinical Cardiology; the Council on Stroke. Resuscitation 2008;79:350-79.
56. Nolan JPS, C.; Böttiger, B. W.; Cariou, A.; Cronberg, T.; Friberg, H.; Genbrugge, C.; Haywood, K.; Lilja, G.; Moulaert, V. R. M.; Nikolaou, N.; Olasveengen, T. M.; Skrifvars, M. B.; Taccone, F.; Soar, J. . European Resuscitation Council and European Society of Intensive Care Medicine Guidelines 2021: Post-resuscitation care. Resuscitation 2021;161:220-69.
57. Sterz F, Safar P, Tisherman S, Radovsky A, Kuboyama K, Oku K. Mild hypothermic cardiopulmonary resuscitation improves outcome after prolonged cardiac arrest in dogs. Crit Care Med 1991;19:379-89.
58. Coimbra C, Wieloch T. Moderate hypothermia mitigates neuronal damage in the rat brain when initiated several hours following transient cerebral ischemia. Acta Neuropathol 1994;87:325-31.
59. Bro-Jeppesen J, Hassager C, Wanscher M, et al. Post-hypothermia fever is associated with increased mortality after out-of-hospital cardiac arrest. Resuscitation 2013;84:1734-40.
60. Zeiner A, Holzer M, Sterz F, et al. Hyperthermia after cardiac arrest is associated with an unfavorable neurologic outcome. Arch Intern Med 2001;161:2007-12.

61. Gilbert M, Busund R, Skagseth A, Nilsen PA, Solbo JP. Resuscitation from accidental hypothermia of 13.7 degrees C with circulatory arrest. Lancet 2000;355:375-6.
62. Hughes A, Riou P, Day C. Full neurological recovery from profound (18.0 degrees C) acute accidental hypothermia: successful resuscitation using active invasive rewarming techniques. Emerg Med J 2007;24:511-2.
63. Hypothermia after Cardiac Arrest Study G. Mild therapeutic hypothermia to improve the neurologic outcome after cardiac arrest. N Engl J Med 2002;346:549-56.
64. Bernard SA, Gray TW, Buist MD, et al. Treatment of comatose survivors of out-of-hospital cardiac arrest with induced hypothermia. N Engl J Med 2002;346:557-63.
65. Nielsen N, Friberg H, Gluud C, Herlitz J, Wetterslev J. Hypothermia after cardiac arrest should be further evaluated--a systematic review of randomised trials with meta-analysis and trial sequential analysis. Int J Cardiol 2011;151:333-41.
66. Nielsen N, Wetterslev J, Cronberg T, et al. Targeted temperature management at 33 degrees C versus 36 degrees C after cardiac arrest. N Engl J Med 2013;369:2197-206.
67. Annborn M, Bro-Jeppesen J, Nielsen N, et al. The association of targeted temperature management at 33 and 36 degrees C with outcome in patients with moderate shock on admission after out-of-hospital cardiac arrest: a post hoc analysis of the Target Temperature Management trial. Intensive Care Med 2014;40:1210-9.
68. Cronberg T, Lilja G, Horn J, et al. Neurologic Function and Health-Related Quality of Life in Patients Following Targeted Temperature Management at 33 degrees C vs 36 degrees C After Out-of-Hospital Cardiac Arrest: A Randomized Clinical Trial. JAMA Neurol 2015;72:634-41.
69. Deye N, Vincent F, Michel P, et al. Changes in cardiac arrest patients' temperature management after the 2013 "TTM" trial: results from an international survey. Ann Intensive Care 2016;6:4.
70. Frydland M, Kjaergaard J, Erlinge D, et al. Target temperature management of 33 degrees C and 36 degrees C in patients with out-of-hospital cardiac arrest with initial non-shockable rhythm - a TTM sub-study. Resuscitation 2015;89:142-8.
71. Winther-Jensen M, Kjaergaard J, Wanscher M, et al. No difference in mortality between men and women after out-of-hospital cardiac arrest. Resuscitation 2015;96:78-84.

72. Winther-Jensen M, Pellis T, Kuiper M, et al. Mortality and neurological outcome in the elderly after target temperature management for out-of-hospital cardiac arrest. Resuscitation 2015;91:92-8.
73. Dankiewicz J, Cronberg T, Lilja G, et al. Targeted hypothermia versus targeted Normothermia after out-of-hospital cardiac arrest (TTM2): A randomized clinical trial-Rationale and design. Am Heart J 2019;217:23-31.
74. Siegel T, Adamski J, Nowakowski P, Onichimowski D, Weigl W. Prospective assessment of standardized mortality ratio (SMR) as a measure of quality of care in intensive care unit--a single-centre study. Anaesthesiol Intensive Ther 2015;47:328-32.
75. Vincent JL, Moreno R. Clinical review: scoring systems in the critically ill. Crit Care 2010;14:207.
76. Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. Intensive Care Med 1996;22:707-10.
77. Moreno RP, Metnitz PG, Almeida E, et al. SAPS 3--From evaluation of the patient to evaluation of the intensive care unit. Part 2: Development of a prognostic model for hospital mortality at ICU admission. Intensive Care Med 2005;31:1345-55.
78. Rydenfelt K, Engerstrom L, Walther S, Sjoberg F, Stromberg U, Samuelsson C. In-hospital vs. 30-day mortality in the critically ill - a 2-year Swedish intensive care cohort analysis. Acta Anaesthesiol Scand 2015;59:846-58.
79. Riskjusteringsmodeller inom svensk intensivvård. 2020. at https://www.icuregswe.org/globalassets/riktlinjer/riskjustering_16.0.pdf.)
80. Engerstrom L, Kramer AA, Nolin T, et al. Comparing Time-Fixed Mortality Prediction Models and Their Effect on ICU Performance Metrics Using the Simplified Acute Physiology Score 3. Crit Care Med 2016;44:e1038-e44.
81. Bouch DC, Thompson JP. Severity scoring systems in the critically ill. Continuing Education in Anaesthesia Critical Care & Pain 2008;8:181-5.
82. Skrifvars MB, Varghese B, Parr MJ. Survival and outcome prediction using the Apache III and the out-of-hospital cardiac arrest (OHCA) score in patients treated in the intensive care unit (ICU) following out-of-hospital, in-hospital or ICU cardiac arrest. Resuscitation 2012;83:728-33.
83. Salciccioli JD, Cristia C, Chase M, et al. Performance of SAPS II and SAPS III scores in post-cardiac arrest. Minerva Anestesiol 2012;78:1341-7.
84. Bisbal M, Jouve E, Papazian L, et al. Effectiveness of SAPS III to predict hospital mortality for post-cardiac arrest patients. Resuscitation 2014;85:939-44.

85. Sandroni C, D'Arrigo S, Nolan JP. Prognostication after cardiac arrest. Crit Care 2018;22:150.
86. Wijdicks EF, Hijdra A, Young GB, Bassetti CL, Wiebe S, Quality Standards Subcommittee of the American Academy of N. Practice parameter: prediction of outcome in comatose survivors after cardiopulmonary resuscitation (an evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology 2006;67:203-10.
87. Lybeck A, Cronberg T, Aneman A, et al. Time to awakening after cardiac arrest and the association with target temperature management. Resuscitation 2018;126:166-71.
88. May TL, Ruthazer R, Riker RR, et al. Early withdrawal of life support after resuscitation from cardiac arrest is common and may result in additional deaths. Resuscitation 2019.
89. Sandroni C, D'Arrigo S, Cacciola S, et al. Prediction of poor neurological outcome in comatose survivors of cardiac arrest: a systematic review. Intensive Care Med 2020;46:1803-51.
90. Dragancea I, Horn J, Kuiper M, et al. Neurological prognostication after cardiac arrest and targeted temperature management 33 degrees C versus 36 degrees C: Results from a randomised controlled clinical trial. Resuscitation 2015;93:164-70.
91. Lybeck A, Friberg H, Aneman A, et al. Prognostic significance of clinical seizures after cardiac arrest and target temperature management. Resuscitation 2017;114:146-51.
92. Reynolds AS, Claassen J. Treatment of Seizures and Postanoxic Status Epilepticus. Semin Neurol 2017;37:33-9.
93. Bouwes A, van Poppelen D, Koelman JH, et al. Acute posthypoxic myoclonus after cardiopulmonary resuscitation. BMC Neurol 2012;12:63.
94. Rossetti AO, Oddo M, Logroscino G, Kaplan PW. Prognostication after cardiac arrest and hypothermia: a prospective study. Ann Neurol 2010;67:301-7.
95. Seder DB, Sunde K, Rubertsson S, et al. Neurologic outcomes and postresuscitation care of patients with myoclonus following cardiac arrest. Crit Care Med 2015;43:965-72.
96. Elmer J, Rittenberger JC, Faro J, et al. Clinically distinct electroencephalographic phenotypes of early myoclonus after cardiac arrest. Ann Neurol 2016;80:175-84.
97. Lance JW, Adams RD. The syndrome of intention or action myoclonus as a sequel to hypoxic encephalopathy. Brain 1963;86:111-36.

98. Annborn M, Nilsson F, Dankiewicz J, et al. The Combination of Biomarkers for Prognostication of Long-Term Outcome in Patients Treated with Mild Hypothermia After Out-of-Hospital Cardiac Arrest-A Pilot Study. Ther Hypothermia Temp Manag 2016;6:85-90.
99. Annborn M, Dankiewicz J, Erlinge D, et al. Procalcitonin after cardiac arrest - an indicator of severity of illness, ischemia-reperfusion injury and outcome. Resuscitation 2013;84:782-7.
100. Haque A, Ray SK, Cox A, Banik NL. Neuron specific enolase: a promising therapeutic target in acute spinal cord injury. Metab Brain Dis 2016;31:487-95.
101. Zetterberg H, Blennow K. Fluid biomarkers for mild traumatic brain injury and related conditions. Nat Rev Neurol 2016;12:563-74.
102. Stammet P, Collignon O, Hassager C, et al. Neuron-Specific Enolase as a Predictor of Death or Poor Neurological Outcome After Out-of-Hospital Cardiac Arrest and Targeted Temperature Management at 33 degrees C and 36 degrees C. J Am Coll Cardiol 2015;65:2104-14.
103. Gaetani L, Blennow K, Calabresi P, Di Filippo M, Parnetti L, Zetterberg H. Neurofilament light chain as a biomarker in neurological disorders. J Neurol Neurosurg Psychiatry 2019;90:870-81.
104. Moseby-Knappe M, Mattsson N, Nielsen N, et al. Serum Neurofilament Light Chain for Prognosis of Outcome After Cardiac Arrest. JAMA Neurol 2019;76:64-71.
105. Wihersaari L, Ashton NJ, Reinikainen M, et al. Neurofilament light as an outcome predictor after cardiac arrest: a post hoc analysis of the COMACARE trial. Intensive Care Med 2021;47:39-48.
106. Donato R, Sorci G, Riuzzi F, et al. S100B's double life: intracellular regulator and extracellular signal. Biochim Biophys Acta 2009;1793:1008-22.
107. Stammet P, Dankiewicz J, Nielsen N, et al. Protein S100 as outcome predictor after out-of-hospital cardiac arrest and targeted temperature management at 33 degrees C and 36 degrees C. Crit Care 2017;21:153.
108. Jang JH, Park WB, Lim YS, et al. Combination of S100B and procalcitonin improves prognostic performance compared to either alone in patients with cardiac arrest: A prospective observational study. Medicine (Baltimore) 2019;98:e14496.
109. Duez CHV, Grejs AM, Jeppesen AN, et al. Neuron-specific enolase and S-100b in prolonged targeted temperature management after cardiac arrest: A randomised study. Resuscitation 2018;122:79-86.
110. Choi S, Park K, Ryu S, et al. Use of S-100B, NSE, CRP and ESR to predict neurological outcomes in patients with return of spontaneous circulation and treated with hypothermia. Emerg Med J 2016;33:690-5.

111. Panchal AR, Bartos JA, Cabanas JG, et al. Part 3: Adult Basic and Advanced Life Support: 2020 American Heart Association Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. Circulation 2020;142:S366-S468.
112. Mattsson N, Zetterberg H, Nielsen N, et al. Serum tau and neurological outcome in cardiac arrest. Ann Neurol 2017;82:665-75.
113. Bazarian JJ, Biberthaler P, Welch RD, et al. Serum GFAP and UCH-L1 for prediction of absence of intracranial injuries on head CT (ALERT-TBI): a multicentre observational study. Lancet Neurol 2018;17:782-9.
114. Fink EL, Berger RP, Clark RS, et al. Exploratory study of serum ubiquitin carboxyl-terminal esterase L1 and glial fibrillary acidic protein for outcome prognostication after pediatric cardiac arrest. Resuscitation 2016;101:65-70.
115. Larsson IM, Wallin E, Kristofferzon ML, Niessner M, Zetterberg H, Rubertsson S. Post-cardiac arrest serum levels of glial fibrillary acidic protein for predicting neurological outcome. Resuscitation 2014;85:1654-61.
116. Helwig K, Seeger F, Holschermann H, et al. Elevated Serum Glial Fibrillary Acidic Protein (GFAP) is Associated with Poor Functional Outcome After Cardiopulmonary Resuscitation. Neurocrit Care 2017;27:68-74.
117. Friberg H, Cronberg T, Dunser MW, Duranteau J, Horn J, Oddo M. Survey on current practices for neurological prognostication after cardiac arrest. Resuscitation 2015;90:158-62.
118. Westhall E, Rossetti AO, van Rootselaar AF, et al. Standardized EEG interpretation accurately predicts prognosis after cardiac arrest. Neurology 2016;86:1482-90.
119. Westhall E, Rosen I, Rossetti AO, et al. Interrater variability of EEG interpretation in comatose cardiac arrest patients. Clin Neurophysiol 2015;126:2397-404.
120. Hirsch LJ, LaRoche SM, Gaspard N, et al. American Clinical Neurophysiology Society's Standardized Critical Care EEG Terminology: 2012 version. J Clin Neurophysiol 2013;30:1-27.
121. Attia ZI, Noseworthy PA, Lopez-Jimenez F, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. Lancet 2019;394:861-7.
122. van Putten M, Jansen C, Tjepkema-Cloostermans MC, et al. Postmortem histopathology of electroencephalography and evoked potentials in postanoxic coma. Resuscitation 2019;134:26-32.
123. Wu O, Sorensen AG, Benner T, Singhal AB, Furie KL, Greer DM. Comatose patients with cardiac arrest: predicting clinical outcome with diffusion-weighted MR imaging. Radiology 2009;252:173-81.

124. Gutierrez LG, Rovira A, Portela LA, Leite Cda C, Lucato LT. CT and MR in non-neonatal hypoxic-ischemic encephalopathy: radiological findings with pathophysiological correlations. Neuroradiology 2010;52:949-76.
125. Tong JT, Eyngorn I, Mlynash M, Albers GW, Hirsch KG. Functional Neurologic Outcomes Change Over the First 6 Months After Cardiac Arrest. Crit Care Med 2016;44:e1202-e7.
126. Becker LB, Aufderheide TP, Geocadin RG, et al. Primary outcomes for resuscitation science studies: a consensus statement from the American Heart Association. Circulation 2011;124:2158-77.
127. Arrich J, Zeiner A, Sterz F, et al. Factors associated with a change in functional outcome between one month and six months after cardiac arrest: a retrospective cohort study. Resuscitation 2009;80:876-80.
128. Peskine A, Cariou A, Hajage D, et al. Long-Term Disabilities of Survivors of Out-of-Hospital Cardiac Arrest: The Hanox Study. Chest 2021;159:699-711.
129. Brain Resuscitation Clinical Trial ISG. Randomized clinical study of thiopental loading in comatose survivors of cardiac arrest. N Engl J Med 1986;314:397-403.
130. Blondin NA, Greer DM. Neurologic prognosis in cardiac arrest patients treated with therapeutic hypothermia. Neurologist 2011;17:241-8.
131. Perkins GD, Jacobs IG, Nadkarni VM, et al. Cardiac arrest and cardiopulmonary resuscitation outcome reports: update of the Utstein Resuscitation Registry Templates for Out-of-Hospital Cardiac Arrest: a statement for healthcare professionals from a task force of the International Liaison Committee on Resuscitation (American Heart Association, European Resuscitation Council, Australian and New Zealand Council on Resuscitation, Heart and Stroke Foundation of Canada, InterAmerican Heart Foundation, Resuscitation Council of Southern Africa, Resuscitation Council of Asia); and the American Heart Association Emergency Cardiovascular Care Committee and the Council on Cardiopulmonary, Critical Care, Perioperative and Resuscitation. Circulation 2015;132:1286-300.
132. Raina KD, Callaway C, Rittenberger JC, Holm MB. Neurological and functional status following cardiac arrest: method and tool utility. Resuscitation 2008;79:249-56.
133. Rittenberger JC, Raina K, Holm MB, Kim YJ, Callaway CW. Association between Cerebral Performance Category, Modified Rankin Scale, and discharge disposition after cardiac arrest. Resuscitation 2011;82:1036-40.
134. Ajam K, Gold LS, Beck SS, Damon S, Phelps R, Rea TD. Reliability of the Cerebral Performance Category to classify neurological status among survivors of ventricular fibrillation arrest: a cohort study. Scand J Trauma Resusc Emerg Med 2011;19:38.

135. Phelps R, Dumas F, Maynard C, Silver J, Rea T. Cerebral Performance Category and long-term prognosis following out-of-hospital cardiac arrest. Crit Care Med 2013;41:1252-7.
136. Moulaert VR, Verbunt JA, van Heugten CM, Wade DT. Cognitive impairments in survivors of out-of-hospital cardiac arrest: a systematic review. Resuscitation 2009;80:297-305.
137. Lilja G, Nielsen N, Friberg H, et al. Cognitive function in survivors of out-of-hospital cardiac arrest after target temperature management at 33 degrees C versus 36 degrees C. Circulation 2015;131:1340-9.
138. Juan E, De Lucia M, Beaud V, et al. How Do You Feel? Subjective Perception of Recovery as a Reliable Surrogate of Cognitive and Functional Outcome in Cardiac Arrest Survivors. Crit Care Med 2018;46:e286-e93.
139. Sawyer KN, Camp-Rogers TR, Kotini-Shah P, et al. Sudden Cardiac Arrest Survivorship: A Scientific Statement From the American Heart Association. Circulation 2020;141:e654-e85.
140. Smith K, Andrew E, Lijovic M, Nehme Z, Bernard S. Quality of life and functional outcomes 12 months after out-of-hospital cardiac arrest. Circulation 2015;131:174-81.
141. Lilja G, Nilsson G, Nielsen N, et al. Anxiety and depression among out-of-hospital cardiac arrest survivors. Resuscitation 2015;97:68-75.
142. Wilder Schaaf KP, Artman LK, Peberdy MA, et al. Anxiety, depression, and PTSD following cardiac arrest: a systematic review of the literature. Resuscitation 2013;84:873-7.
143. Moulaert VR, Wachelder EM, Verbunt JA, Wade DT, van Heugten CM. Determinants of quality of life in survivors of cardiac arrest. J Rehabil Med 2010;42:553-8.
144. Wachelder EM, Moulaert VR, van Heugten C, Verbunt JA, Bekkers SC, Wade DT. Life after survival: long-term daily functioning and quality of life after an out-of-hospital cardiac arrest. Resuscitation 2009;80:517-22.
145. Lilja G, Nielsen N, Bro-Jeppesen J, et al. Return to Work and Participation in Society After Out-of-Hospital Cardiac Arrest. Circ Cardiovasc Qual Outcomes 2018;11:e003566.
146. Haywood K, Whitehead L, Nadkarni VM, et al. COSCA (Core Outcome Set for Cardiac Arrest) in Adults: An Advisory Statement From the International Liaison Committee on Resuscitation. Resuscitation 2018;127:147-63.
147. Tasker RC, Menon DK. Critical Care and the Brain. JAMA 2016;315:749-50.
148. Harhash AA, May TL, Hsu CH, et al. Risk Stratification Among Survivors of Cardiac Arrest Considered for Coronary Angiography. J Am Coll Cardiol 2021;77:360-71.

149. Coppler PJ, Callaway CW, Guyette FX, Baldwin M, Elmer J. Early risk stratification after resuscitation from cardiac arrest. J Am Coll Emerg Physicians Open 2020;1:922-31.
150. Maupain C, Bougouin W, Lamhaut L, et al. The CAHP (Cardiac Arrest Hospital Prognosis) score: a tool for risk stratification after out-of-hospital cardiac arrest. Eur Heart J 2016;37:3222-8.
151. Grasner JT, Meybohm P, Lefering R, et al. ROSC after cardiac arrest--the RACA score to predict outcome after out-of-hospital cardiac arrest. Eur Heart J 2011;32:1649-56.
152. Adrie C, Cariou A, Mourvillier B, et al. Predicting survival with good neurological recovery at hospital admission after successful resuscitation of out-of-hospital cardiac arrest: the OHCA score. Eur Heart J 2006;27:2840-5.
153. Pareek N, Kordis P, Beckley-Hoelscher N, et al. A practical risk score for early prediction of neurological outcome after out-of-hospital cardiac arrest: MIRACLE2. Eur Heart J 2020;41:4508-17.
154. Eertmans W, Tran TMP, Genbrugge C, et al. A prediction model for good neurological outcome in successfully resuscitated out-of-hospital cardiac arrest patients. Scand J Trauma Resusc Emerg Med 2018;26:93.
155. Deans KJ, Minneci PC, Danner RL, Eichacker PQ, Natanson C. Practice misalignments in randomized controlled trials: Identification, impact, and potential solutions. Anesth Analg 2010;111:444-50.
156. Johnsson J, Bjornsson O, Andersson P, et al. Artificial neural networks improve early outcome prediction and risk classification in out-of-hospital cardiac arrest patients admitted to intensive care. Crit Care 2020;24:474.
157. Holmgren G, Andersson P, Jakobsson A, Frigyesi A. Artificial neural networks improve and simplify intensive care mortality prognostication: a national cohort study of 217,289 first-time intensive care unit admissions. J Intensive Care 2019;7:44.
158. Adamson AS, Welch HG. Machine Learning and the Cancer-Diagnosis Problem - No Gold Standard. N Engl J Med 2019;381:2285-7.
159. Chen TG, C. XGBoost: A Scalable Tree Boosting System. 2016.
160. Raj R, Luostarinen T, Pursiainen E, et al. Machine learning-based dynamic mortality prediction after traumatic brain injury. Sci Rep 2019;9:17672.
161. Kang DY, Cho KJ, Kwon O, et al. Artificial intelligence algorithm to predict the need for critical care in prehospital emergency medical services. Scand J Trauma Resusc Emerg Med 2020;28:17.
162. Jang DH, Kim J, Jo YH, et al. Developing neural network models for early detection of cardiac arrest in emergency department. Am J Emerg Med 2019.

163. Blomberg SN, Folke F, Ersboll AK, et al. Machine learning as a supportive tool to recognize cardiac arrest in emergency calls. Resuscitation 2019;138:322-9.
164. Seki T, Tamura T, Suzuki M, Group S-KS. Outcome prediction of out-of-hospital cardiac arrest with presumed cardiac aetiology using an advanced machine learning technique. Resuscitation 2019;141:128-35.
165. Kwon JM, Jeon KH, Kim HM, et al. Deep-learning-based out-of-hospital cardiac arrest prognostic system to predict clinical outcomes. Resuscitation 2019;139:84-91.
166. Gillon R. Medical ethics: four principles plus attention to scope. BMJ 1994;309:184-8.
167. Rey A, Rossetti AO, Miroz JP, Eckert P, Oddo M. Late Awakening in Survivors of Postanoxic Coma: Early Neurophysiologic Predictors and Association With ICU and Long-Term Neurologic Recovery. Crit Care Med 2019;47:85-92.
168. Mulder M, Gibbs HG, Smith SW, et al. Awakening and withdrawal of life-sustaining treatment in cardiac arrest survivors treated with therapeutic hypothermia*. Crit Care Med 2014;42:2493-9.
169. Gold B, Puertas L, Davis SP, et al. Awakening after cardiac arrest and post resuscitation hypothermia: are we pulling the plug too early? Resuscitation 2014;85:211-4.
170. Perkins GD, Jacobs IG, Nadkarni VM, et al. Cardiac Arrest and Cardiopulmonary Resuscitation Outcome Reports: Update of the Utstein Resuscitation Registry Templates for Out-of-Hospital Cardiac Arrest: A Statement for Healthcare Professionals From a Task Force of the International Liaison Committee on Resuscitation (American Heart Association, European Resuscitation Council, Australian and New Zealand Council on Resuscitation, Heart and Stroke Foundation of Canada, InterAmerican Heart Foundation, Resuscitation Council of Southern Africa, Resuscitation Council of Asia); and the American Heart Association Emergency Cardiovascular Care Committee and the Council on Cardiopulmonary, Critical Care, Perioperative and Resuscitation. Resuscitation 2015;96:328-40.
171. Cummins RO, Chamberlain D, Hazinski MF, et al. Recommended guidelines for reviewing, reporting, and conducting research on in-hospital resuscitation: the in-hospital 'Utstein style'. A statement for healthcare professionals from the American Heart Association, the European Resuscitation Council, the Heart and Stroke Foundation of Canada, the Australian Resuscitation Council, and the Resuscitation Councils of Southern Africa. Resuscitation 1997;34:151-83.

172. Zaritsky A, Nadkarni V, Hazinski MF, et al. Recommended guidelines for uniform reporting of pediatric advanced life support: the Pediatric Utstein Style. A statement for healthcare professionals from a task force of the American Academy of Pediatrics, the American Heart Association, and the European Resuscitation Council. Resuscitation 1995;30:95-115.
173. Dyson K, Brown SP, May S, et al. International variation in survival after out-of-hospital cardiac arrest: A validation study of the Utstein template. Resuscitation 2019;138:168-81.
174. Nielsen N, Hovdenes J, Nilsson F, et al. Outcome, timing and adverse events in therapeutic hypothermia after out-of-hospital cardiac arrest. Acta Anaesthesiol Scand 2009;53:926-34.
175. Jacobs I, Nadkarni V, Bahr J, et al. Cardiac arrest and cardiopulmonary resuscitation outcome reports: update and simplification of the Utstein templates for resuscitation registries. A statement for healthcare professionals from a task force of the international liaison committee on resuscitation (American Heart Association, European Resuscitation Council, Australian Resuscitation Council, New Zealand Resuscitation Council, Heart and Stroke Foundation of Canada, InterAmerican Heart Foundation, Resuscitation Council of Southern Africa). Resuscitation 2004;63:233-49.
176. Langhelle A, Nolan J, Herlitz J, et al. Recommended guidelines for reviewing, reporting, and conducting research on post-resuscitation care: the Utstein style. Resuscitation 2005;66:271-83.
177. Nielsen N, Wetterslev J, al-Subaie N, et al. Target Temperature Management after out-of-hospital cardiac arrest--a randomized, parallel-group, assessor-blinded clinical trial--rationale and design. Am Heart J 2012;163:541-8.
178. Kjaergaard J, Nielsen N, Winther-Jensen M, et al. Impact of time to return of spontaneous circulation on neuroprotective effect of targeted temperature management at 33 or 36 degrees in comatose survivors of out-of hospital cardiac arrest. Resuscitation 2015;96:310-6.
179. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. BMC Med Res Methodol 2017;17:162.
180. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837-45.
181. Sun X, Xu W. Fast Implementation of DeLong's Algorithm for Comparing the Areas Under Correlated Receiver Operating Characteristic Curves. IEEE Signal Processing Letters 2014;21:1389-93.

182. Safari S, Baratloo A, Elfil M, Negida A. Evidence Based Emergency Medicine; Part 5 Receiver Operating Curve and Area under the Curve. Emerg (Tehran) 2016;4:111-3.
183. Fawcett T. An introduction to ROC analysis. Pattern Recognition Letters 2006;27:861-74.
184. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol 2010;5:1315-6.
185. Van Calster B, McLernon DJ, van Smeden M, et al. Calibration: the Achilles heel of predictive analytics. BMC Medicine 2019;17:230.
186. Finazzi S, Poole D, Luciani D, Cogo PE, Bertolini G. Calibration belt for quality-of-care assessment based on dichotomous outcomes. PloS one 2011;6:e16110-e.
187. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. J Clin Epidemiol 2016;74:167-76.
188. Bergstra JY, D.; Cox, D. Making a science of model search: hyperparameter optimization in hundreds of dimensions for vision architectures.
Proceedings of the 30th International Conference on International Conference on Machine Learning 2013;28:115-23.
189. Lundberg SML, S. A. A unified approach to interpreting model predictions. Proceedings of the 31st International Conference on Neural Information. Processing Systems; Long Beach, California. USA: Curran Associates Inc.; 2017. p.4768-77 2017.
190. Lundberg SM, Nair B, Vavilala MS, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat Biomed Eng 2018;2:749-60.
191. Slack D, et al. Fooling lime and shap: Adversarial attacks on post hoc explanation methods. Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society. 2020.
192. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing,; 2020.
193. Python Core Team. Python: A dynamic, open source programming language. Python version 3.7 ed: Python Software Foundation; 2020.
194. Abadi M, Barham P, Chen JM, et al. TensorFlow: A system for large-scale machine learning. Proceedings of Osdi'16: 12th Usenix Symposium on Operating Systems Design and Implementation 2016:265-83.
195. tableone: Create 'Table 1' to Describe Baseline Characteristics with or without Propensity Score Weights. 2020. at https://CRAN.R-project.org/package=tableone.)

196. forestplot: Advanced Forest Plot Using 'grid' Graphics. 2020. at https://CRAN.R-project.org/package=forestplot.)
197. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011;12:77.
198. López-Ratón M, Rodríguez-Álvarez MX, Suárez CC, Sampedro FG. OptimalCutpoints : An R Package for Selecting Optimal Cutpoints in Diagnostic Tests. Journal of statistical software 2014;61:1-36.
199. BorutaShap 1.0.14. 2020. (Accessed Oct 3, 2020, at https://pypi.org/project/BorutaShap/.)
200. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, California, USA: Curran Associates Inc.; 2017:4768–77.
201. Tantau T. Graph Drawing in TikZ. 2013; Berlin, Heidelberg: Springer Berlin Heidelberg. p. 517-28.
202. https://xgboost.readthedocs.io/en/latest/python/python_intro.html.
203. Martinell L, Nielsen N, Herlitz J, et al. Early predictors of poor outcome after out-of-hospital cardiac arrest. Critical care (London, England) 2017;21:96-.
204. Karlsson V, Dankiewicz J, Nielsen N, et al. Association of gender to outcome after out-of-hospital cardiac arrest--a report from the International Cardiac Arrest Registry. Crit Care 2015;19:182.
205. Geri G, Dumas F, Bougouin W, et al. Immediate Percutaneous Coronary Intervention Is Associated With Improved Short- and Long-Term Survival After Out-of-Hospital Cardiac Arrest. Circ Cardiovasc Interv 2015;8.
206. Terman SW, Shields TA, Hume B, Silbergleit R. The influence of age and chronic medical conditions on neurological outcomes in out of hospital cardiac arrest. Resuscitation 2015;89:169-76.
207. Salter R, Bailey M, Bellomo R, et al. Changes in Temperature Management of Cardiac Arrest Patients Following Publication of the Target Temperature Management Trial. Crit Care Med 2018;46:1722-30.
208. Bradley SM, Liu W, McNally B, et al. Temporal trends in the use of therapeutic hypothermia for out-of-hospital cardiac arrest. JAMA Network Open 2018;1:e184511.
209. Lee H, Herbert RD, McAuley JH. Mediation Analysis. JAMA 2019;321:697-8.
210. Herlitz J. Svenska hjärt- lungräddningsregistret Årsrapport 20172017.

211. Kim LK, Looser P, Swaminathan RV, et al. Sex-Based Disparities in Incidence, Treatment, and Outcomes of Cardiac Arrest in the United States, 2003-2012. J Am Heart Assoc 2016;5.
212. May TL, Lary CW, Riker RR, et al. Variability in functional outcome and treatment practices by treatment center after out-of-hospital cardiac arrest: analysis of International Cardiac Arrest Registry. Intensive Care Med 2019.
213. Kapinos G, Becker LB. The American Academy of Neurology affirms the revival of cooling for the revived. Neurology 2017;88:2076-7.
214. Nichol G, Polderman KH, Friberg H, Kurz M, Kapinos G. Perspectives on Temperature Management. Ther Hypothermia Temp Manag 2018.
215. Bray JE, Stub D, Bloom JE, et al. Changing target temperature from 33 degrees C to 36 degrees C in the ICU management of out-of-hospital cardiac arrest: A before and after study. Resuscitation 2017;113:39-43.
216. Polderman KH, Varon J. We should not abandon therapeutic cooling after cardiac arrest. Crit Care 2014;18:130.
217. Aschauer S, Dorffner G, Sterz F, Erdogmus A, Laggner A. A prediction tool for initial out-of-hospital cardiac arrest survivors. Resuscitation 2014;85:1225-31.
218. Oddo M, Bracard S, Cariou A, et al. Update in Neurocritical Care: a summary of the 2018 Paris international conference of the French Society of Intensive Care. Ann Intensive Care 2019;9:47.
219. Moseby-Knappe M, Westhall E, Backman S, et al. Performance of a guideline-recommended algorithm for prognostication of poor neurological outcome after cardiac arrest. Intensive Care Med 2020.
220. Abazi L, Awad A, Nordberg P, et al. Long-term survival in out-of-hospital cardiac arrest patients treated with targeted temperature control at 33 degrees C or 36 degrees C: A national registry study. Resuscitation 2019;143:142-7.
221. Kushimoto S, Abe T, Ogura H, et al. Impact of Body Temperature Abnormalities on the Implementation of Sepsis Bundles and Outcomes in Patients With Severe Sepsis: A Retrospective Sub-Analysis of the Focused Outcome Research on Emergency Care for Acute Respiratory Distress Syndrome, Sepsis and Trauma Study. Crit Care Med 2019;47:691-9.
222. Nolan JP, Laver SR, Welch CA, Harrison DA, Gupta V, Rowan K. Outcome following admission to UK intensive care units after cardiac arrest: a secondary analysis of the ICNARC Case Mix Programme Database. Anaesthesia 2007;62:1207-16.
223. Metnitz PG, Moreno RP, Almeida E, et al. SAPS 3--From evaluation of the patient to evaluation of the intensive care unit. Part 1: Objectives, methods and cohort description. Intensive Care Med 2005;31:1336-44.

224. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, 2020. at https://www.R-project.org.)
225. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med 1985;13:818-29.
226. Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. JAMA 1993;270:2957-63.
227. Vincent JL, de Mendonca A, Cantraine F, et al. Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study. Working group on "sepsis-related problems" of the European Society of Intensive Care Medicine. Crit Care Med 1998;26:1793-800.