Observational Data to Improve Clinical Decision Making

in Acute Care

Benjamin Y. Gravesteijn

Observational Data to Improve Clinical Decision Making

in Acute Care

This thesis was printed on FSC certified paper

Publication of this thesis was financially supported by the Department of Public Health, Erasmus MC and the Erasmus University

Illustrations: Jesse Ceelen
Design: Benjamin Gravesteijn
Printed by: Proefschriftenmaken.nl

ISBN: 978-94-6423-379-7

Observational Data to Improve Clinical Decision Making in Acute Care

Het Verbeteren van Klinische Beslisvorming met Observationele Data in de Acute Geneeskunde

Proefschrift

ter verkrijging van de graad van doctor aan de Erasmus Universiteit Rotterdam op gezag van de rector magnificus

Prof.dr. L.A. Bredenoord

en volgens besluit van het College voor Promoties. De openbare verdediging zal plaatsvinden op

Woensdag 9 maart 2022 om 15:30

door

Benjamin Yaël Gravesteijn
geboren te Capelle aan den IJssel

Erasmus University Rotterdam

Promotiecommissie

Promotoren:	Prof.dr. E. W. Steyerberg
	Prof.dr. H. F. Lingsma
Overige leden:	Prof.dr. S. le Cessie
	Prof.dr. M. Sabbe
	Prof.dr. M. G. M. Hunink

The premise of Big Data in acute medicine is to make medicine more efficient and effective. However, the translation of large observational data to knowledge is difficult. This thesis explores and discusses the three main types of research questions which can be asked from large observational data:

1. What is current clinical practice?
2. What is best practice?
3. What patients need to be prioritised?

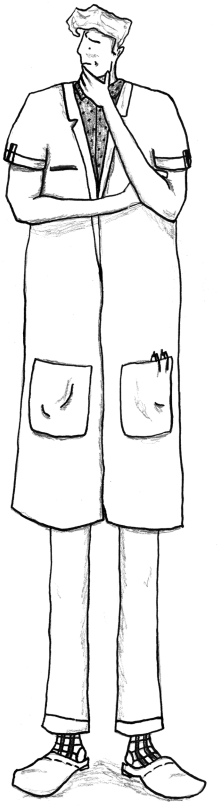
This thesis will focus on traumatic brain injury and in-hospital cardiac arrest.

Big Data in acute geneeskunde belooft de zorg efficiënter en effectiever te maken. De weg van grote observationele data tot kennis is echter lang en gecompliceerd. Deze thesis onderzoekt en behandelt de drie voornaamste type vragen die we kunnen stellen aan grote observationele data:

1. *Hoe ziet de huidige klinische praktijk eruit?*
2. *Welke klinische praktijk zorgt voor de beste uitkomsten?*
3. *Welke patiënten moeten geprioriteerd worden?*

In deze thesis zullen patiënten met traumatisch hersenletsel en hartstilstanden in het ziekenhuis bestudeerd worden.

Contents



1

Introduction

Practicing medicine starts by taking a patient's history, or anamnesis. Anamnesis (αναμνεσις) translates to recollecting, and has been described by Socrates as the ability to recognise something as it is, based on previously acquired knowledge [1]. During anamnesis, a doctor tries to discover what the cause of the symptoms of the patient is by recognising a pattern. One of the first crossroads in the process of anamnesis is to recognise whether the patient suffers from a disease that is chronic, or acute.

A patient with an acute disease often needs immediate attention, to start adequate therapy as fast as possible. Decisions on the short-term largely impact the survival of patients. However, since not all necessary information might be available yet, decisions have to be made under uncertainty. For example, when a trauma patient arrives at the emergency room, the decision to stabilise the cervical spine might prevent long-lasting paralysis of the upper limbs. However, the doctor can only be certain that a cervical spine is actually fractured after having seen an X-ray or CT scan. In the primary assessment, during which these images are not yet avail-

able, which patients actually require stabilisation to prevent paralysis?

Two acute diseases are studied in this thesis: traumatic brain injury, and in-hospital cardiac arrest. Although both diseases are acute, they affect different organ systems, and require completely different interventions. However, studying both diseases illustrates what challenges can be expected when performing research to improve clinical decision making in acute care.

1.1 | Traumatic brain injury

The global burden of traumatic brain injury (TBI) is high [2, 3]. Although the rates vary between countries, TBI is estimated to be responsible for around 300 hospital admissions and 12 deaths per 100,000 persons per year in Europe [4].

TBI is defined as an impairment of neurocognitive functioning, caused by an external force [5]. Commonly observed impairments in neurocognitive functioning are amnesia, deprived level of consciousness, and dizziness. In survivors, long-term sequelae such as psychiatric, emotional, cognitive, and physical disabilities disrupt lives of patients and their relatives [2].

Although the primary brain injury is defined by the trauma itself, secondary brain injury – especially due to hypoxia and hypotension – must be prevented [6–8]. This should preferably happen by intervening as early as possible. For example by securing the airway: inserting a tube into the tracheas of patients with a depressed level of consciousness (intubation) [9–11] secures the exchange of carbon dioxide and oxygen [12]. An often used rule of thumb is to intubate patients with a Glasgow Coma Scale

(GCS) below eight [12, 13] (figure 1.1), but the evidence underpinning this recommendation is thin. Moreover, intubating the patient on-scene might be preferred over securing the airway at arrival in the hospital, according to an Australian RCT [14]. This effect has not yet been validated to the European setting, where a higher density of hospitals provides a different context. Due to this different context, it might be hypothesised that the positive effect of prehospital intubation might be lower because the travel distance to the nearest hospital is shorter.

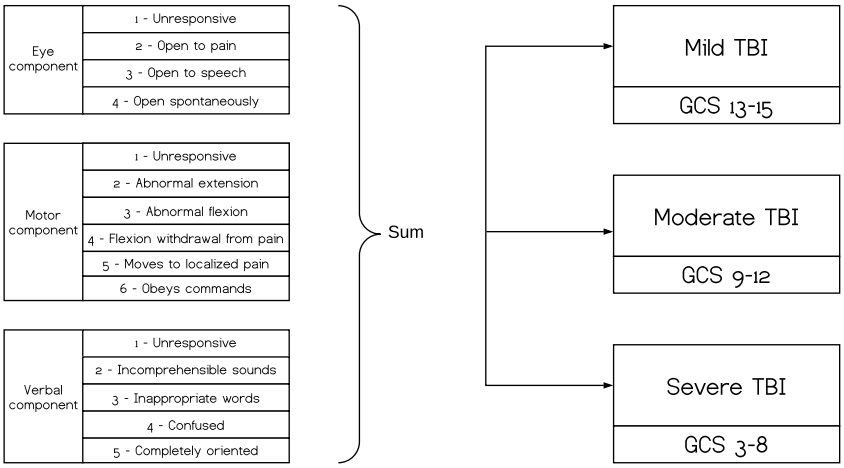


Figure 1.1: *Classification of clinical severity of TBI using the Glasgow Coma Scale.*

Moreover, TBI patients often require treatments which are only available in specialised neurosurgical centres, e.g. intracranial pressure monitoring, and decompressive craniectomy [15]. There is evidence that admission in such centres improves outcome of these patients [16–18]. However, it is not well known if patients need to be transported directly to these centres, or whether it is safe to stabilise them in non-neurosurgical hospitals

first.

Due to the lack of evidence underpinning many of the interventions for TBI, the self-reported adherence to the available international guidelines is low [19, 20]. It is also known that there is a large variation in outcome following treatment for TBI: the risk of unfavourable outcome is described to differ up to 3.3-fold, depending on where the patient is treated [21]. Due to these observations, it is likely that there is also variation in the way interventions are being applied in different areas. However, the extent to which these vary has not yet been established.

There is a limited number of treatments proven effective to improve outcome of TBI patients. One of the reasons is that TBI is a heterogeneous disease [22]. At this moment, TBI is most commonly classified into 3 groups (mild, moderate, severe) based on the GCS [13] (figure 1.1), but this simple classification is often criticised [22]. TBI is more heterogeneous, with various intracranial abnormalities (e.g. epidural hematoma, contusions, axonal injury), physiological phenotypes (e.g. high intracranial pressure, dysfunctional auto-regulation), or epidemiological entities with various physiological or psychological reserve (e.g. high- or low educated, elderly or children). Treatments are developed and tested in a broad range of TBI patients. The only specific characteristic they often do take into account, is injury severity measured by GCS [23]. Better characterisation, classification, and prediction models are needed to better understand which patients require an intervention [2].

1.2 | In-hospital cardiac arrest

Cardiac arrest, cardiopulmonary arrest, or circulatory arrest is the loss of effective blood circulation. Potential causes are for example myocardial infarction, shock, or a large pulmonary embolus [24]. Cardiac arrest inevitably leads to death if cardiopulmonary resuscitation (CPR) is not started.

Cardiac arrest is usually classified into either out-of-hospital cardiac arrest (OHCA) and in-hospital cardiac arrest (IHCA). OHCA is described to occur around 19–104 times per 100,000 people per year and results in 10% short-term survival [25]. The incidence of IHCA is 1–6 events per 1000 hospital admissions [26–28] and results in 15% short-term survival [29]. Because the outcome after IHCA is poorly described, the ROUTiNE study has been designed and executed to fill this knowledge gap. This thesis therefore also focuses on IHCA.

A possible advantage for patients suffering IHCA versus OHCA is that hospitals are equipped with teams, who could employ advanced life support using highly specialised equipment. In the last two decades, physicians have increasingly used veno-arterial extracorporeal membrane oxygenation (VA-ECMO, figure 1.2) during CPR [30, 31]. Using ECMO during CPR is often referred to as extracorporeal cardiopulmonary resuscitation (ECPR). By taking over cardiac and respiratory function, VA-ECMO ensures oxygenation, ventilation, and circulation, paving the way for a potential recovery [32]. Although evidence from randomised controlled trials is lacking [33], observational studies have repeatedly shown better survival after ECPR compared to conventional CPR [34, 35]. Nevertheless, ECPR is costly and labour intensive, and its cost-effectiveness has not yet been established.

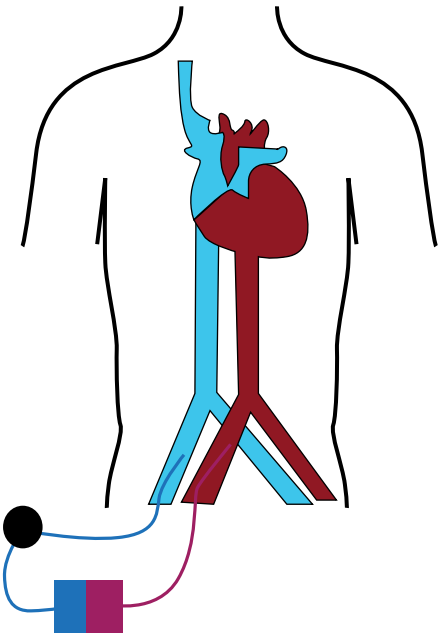


Figure 1.2: *Schematic representation of VA-ECMO.*

1.3 | Methods for improving clinical decision making

To improve clinical decision making, we need scientific evidence. If a quantitative approach is taken, three types of research questions can be distinguished (ordered by increasing methodological complexity) [36]. First, we can simply describe relationships between parameters: a descriptive question. Second, we can try to combine multiple parameters to predict an outcome: a predictive question. Third, we can try to ascertain whether a

relation we find in the data is potentially a causal relation (e.g.: the effect of a treatment on outcome): a causal question. These types of questions can be considered the starting point for many research endeavours.

1.3.1 | Study design

If the researcher wants to produce new evidence to answer a research question, a new study might be designed and executed. The research question should guide the design of the study, because some types of designs are better suited to answer specific types of research questions. Medical studies can broadly be classified into interventional and non-interventional studies, often called observational studies (figure 1.3). Whereas interventional studies are often more appropriate to answer causal questions, observational studies might often be more appropriate to answer predictive questions (if they are larger, and have a more varied patient population) [37, 38].

Figure 1.3: *Proposed overview of quantitative study approaches for answering research questions, either by producing or aggregating evidence. Moreover, the ability of study designs to detect causal effects is displayed.*

Interventional studies, as the name suggests, intervene in clinical practice. The effect attributable to an intervention (e.g. a drug, a procedure) on a relevant patient outcome is often the main interest of such a study. Interventional studies can be subdivided into randomised or non-randomised controlled trials, according to the method of allocation of the intervention. In a randomised controlled trial (RCT), patients are randomly assigned to either the interventional or the control arm, and their outcomes are compared. Because the two groups are on expectation comparable, except for the intervention they received, these studies can demonstrate causality: we can observe the so called counterfactual ("what would be the outcome in patients if they had not received intervention A, but intervention B"). There are a large number of variation on the traditional RCT, which fall outside the scope of this thesis, for example cluster RCTs [39], step-wedged RCTs [40], adaptive trial designs such as the Multi-Arm Multi-stage trial [41], play-the-winner randomisation [42], or non-parallel designs such as cross-over trials [43]. Finally, non-randomised controlled trials allocate interventions based on a rule, for example based on an age cut-off.

In contrast to interventional studies, observational studies simply allow clinical practice to be performed as usual. Observational studies can be classified into prospective and retrospective: a prospective observational study selects a predefined period, and records all data of interest as clinical practice unfolds. A retrospective observational study starts after a period of clinical practice, and evaluates what has happened during that period by evaluating patient records. Proving causality in observational study designs is complicated due to a variety of biases, but primarily confounding bias (by indication). This bias arises for example for life-saving interventions. The sicker the patient, the more likely the patient receives this intervention. If then outcome of receivers of this life-saving interven-

tion would be compared to non-receivers, the results would indicate that the intervention is associated with worse outcome. Fortunately, there are multiple ways to address these biases, which mostly rely on adjusting for a particular set of other observed variables. If we have observed a variable which is randomly allocated between patients (for example: the hospital in which they are treated), we can use instrumental variable analysis to exploit this naturally occurring randomisation [44, 45]. If there is a clear cut-off in the allocation of treatment based on a characteristic, for example age, we can use regression discontinuity to assess the effectiveness of the intervention; around the man-made cut-off (which is arbitrary for nature), treatment allocation is nearing random allocation [46]. More generally, we can use visual representations of the assumed causal models which gave rise to the relationships in the data. These visual representations (directed acyclic graphs [47]) guide what variables to control for to reduce bias, when trying to prove causality [48] (figure 1.4).

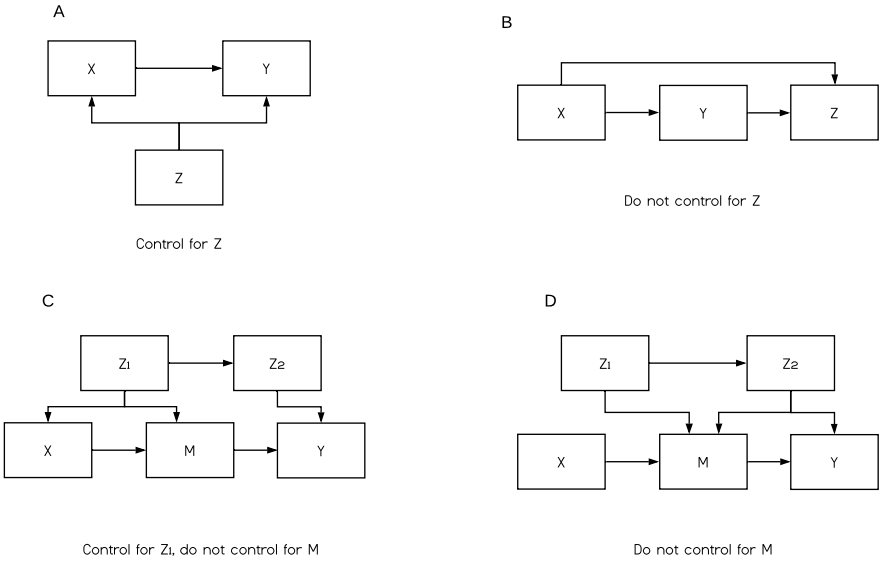


Figure 1.4: *Examples of directed acyclic graphs (DAGs), and the consequence for the analysis to evaluate the effect of X on Y. In panel A, Z is a confounder, and needs to be controlled for; In panel B, Z is a collider, so controlling for Z will result in the path $X \rightarrow Z \rightarrow Y$ to be opened; In panel C, Controlling for $Z1$ closes the path $X \rightarrow Z1 \rightarrow Z2 \rightarrow Y$, and controlling for M closes the path $X \rightarrow M \rightarrow Y$ which is the effect of interest. In panel D, $Z1$ and $Z2$ both collide on M, so this pathway is closed unless M is controlled for, which also should not be controlled for since it closes the path of interest $X \rightarrow M \rightarrow Y$.*

In order to answer the questions of this thesis, we will exploit large observational data. We aim to produce new evidence with large observational data, as well as aggregate evidence from these type of data. Advantages of observational data include:

1. that it is easier to arrive at a large sample, improving precision;
2. good accessibility since based on routine care;
3. direct translation to the real-world situation since the data comprises of current practice; and
4. ethically and financially, observational studies are often more feasible than interventional studies.

Because of these reasons and more, people often call the current period in time the uprising of "Big Data" [49]. However, the boundaries of what can be learned from these type of data should be evaluated.

1.3.2 | Analytical methods

To analyse the data produced by these studies, epidemiologists traditionally use regression-based techniques. These regression models can be used to study the inter-relatedness of the variables of interest. The researcher first assumes what outcome is most relevant to their research question (the y-variable, outcome, or dependent variable). Next, the researcher assumes what variables influence this outcome (the x-variables, predictors, or independent variables). Finally, the model is "fitted" to the data, and the relationships between the x-variables and the y-variable is estimated. These models are general and flexible, and can be used to answer almost every

quantitative research question, of course with the appropriate data. More specifically, regression can be used for descriptive, causal, and predictive research questions, but their interpretation depends on the study design, and included parameters.

A more modern approach to analyse data is by using machine learning algorithms. These algorithms vary widely in mathematical structure, but can broadly be classified into supervised and unsupervised algorithms (table 1.1). Supervised machine learning algorithms are similar to regression techniques: the researcher assumes the relevant x- and y-variables, and the model is again fitted (= trained) to the data. When using unsupervised machine learning algorithms, the researcher does not have to assume what variables are exposures, treatments, predictors, or outcomes. Rather, the algorithm evaluates inter-relatedness of any included variables.

Table 1.1: *Relation between definitions of machine learning algorithm and regression techniques.*

Technique	Conceptual aim	Examples
Machine learning		
Supervised learning	Y ∼X	Bayesian network, neural network
Unsupervised learning	X	K-means clustering, PCA
Regression	Y ∼X	Logistic/linear/cox-regression

However, it can be argued that above mentioned classification is too simplistic, and divides techniques based on arbitrary definitions. More broadly speaking, these techniques are all algorithms which adjust their parameters to optimally fit data. The more parameters, the more flexible algorithms are, and more complex relationships might be caught by the algorithm. If the data are large enough and consist of complex relationships, a higher performance is expected for an algorithm with more parameters.

A proposal for a representation of this continuum, and some exemplary algorithms, are shown in figure 1.5. An important question, however, is whether this flexibility is required or even helpful when applied to clinical data to answer predictive questions.

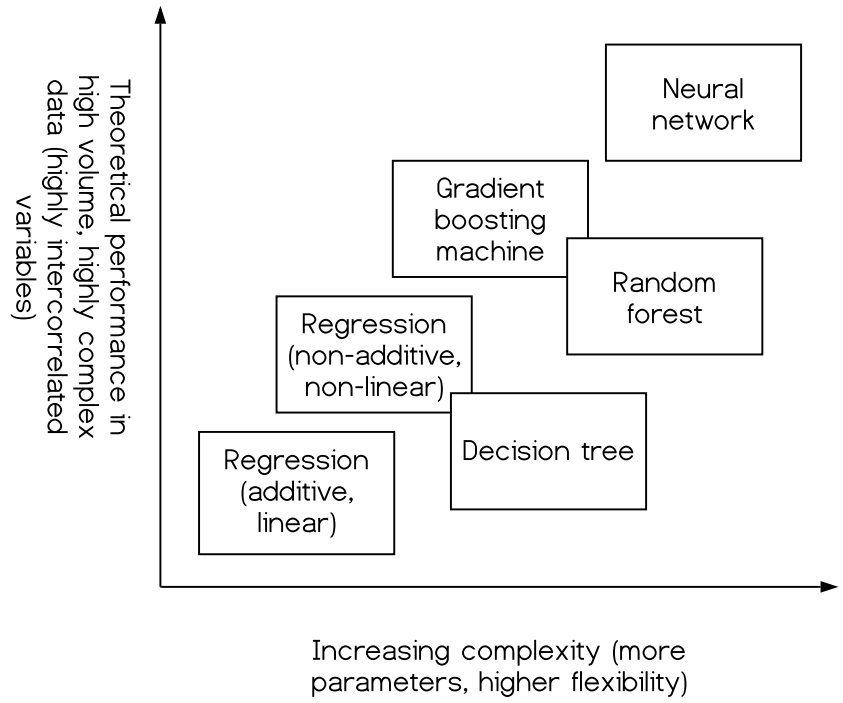


Figure 1.5: *Potential performance (in terms of explained variance) of a variety of algorithms, given the data is large and complex enough.*

Neither regression or machine learning algorithms can process data points with missing data, a common problem in medical research. Instead, they both simply delete patients with any missing values. Deleting these patients decreases sample size and renders the analysed data set more se-

lected: patients with no missing data might be different from patients with some missing values. A sophisticated and broadly advocated method to deal with missing data is multiple imputation. There are considerations in multiple dimensions which need to be taken into account when performing imputation, such as the effective sample size of the data set, the actual assumed imputation model, missing data mechanism etc. Clear methodological guidance is lacking how to deal with missing data in predictive research.

1.3.3 | Aggregating evidence

In addition to performing a study and analysing its data, researchers can aggregate existing evidence to produce new or more precise evidence (figure 1.3). In this thesis, two different ways are used to aggregate already existing data.

On the one hand, one might use meta-analysis to aggregate estimates of multiple studies. A more precise estimate based on data from all these studies might then be more conclusive, or definitive. However, the literature should not be too heterogeneous, because pooling should be reasonable: either the characteristics of the design and execution of studies (methodological heterogeneity) or the characteristics associated with the participants, treatments, or outcome, (clinical heterogeneity) should not be too different between studies [50].

On the other, one might use a decision model to aggregate evidence from various sources of evidence [51]. Decision models are especially useful to evaluate the impact of decision on the long-term [52]. Therefore, they are often used in cost-effectiveness analyses. Another important reason for using decision models, is the premise to estimate the (causal) effect

of decisions when randomised data is not available [53].

Compared to regression models or machine learning algorithms, decision models arrive at their estimate in a different way. An illustrative example would be the calculation of travel time between Amsterdam and Rotterdam. The decision model approach would be to take the distance between the two cities, and the average allowed maximum speed limit of every road. These "parameters" can be used to calculate the travel time. The alternative, more directly empirical approach would be to measure the time between arrival and destination of a sample of travellers and take the average. Theoretically, both would arrive at the same estimate. For complex multidimensional decisions, only decision modelling is feasible.

In this thesis, all of the above mentioned methods will be applied to improve clinical decision making in acute medicine, and advantages and disadvantages of these methods will be addressed.

1.4 | Aims

The overall aim is to contribute to more efficient and effective clinical decision making for traumatic brain injury (TBI) and in-hospital cardiac arrest (IHCA), based on empirical evidence. For this aim, I use large observational data sets. Three specific questions will be addressed in this thesis:

1. Variation in care:
    - what is the variation in (prehospital) interventions for TBI, and
    - what is the expected long-term outcome after IHCA, and how much variation exists?

    – a descriptive question.

2. What is the best practice, or what is the (cost-)effectiveness of currently performed interventions for TBI and IHCA? — a causal question.

3. How can we better characterise and predict outcome in TBI? — a predictive question, focusing only on TBI.

These questions have a logical sequence. The premise of studying current practice, is that areas which need improvement can be identified. If those areas are identified, the consecutive question that arises must be what the best practice is. When it is known that substantial variation exists between centres, for example in the use of ECPR, one needs to assess what practice has the best results. The best practice, when identified, must then be implemented generally. And finally, patients at high risk of poor outcome need to be identified, for example to inform triage decisions or to give relatives relevant information about outcome.

1.5 | Data used in this thesis

The data we will use in this thesis, is the CENTER-TBI [54, 55] and IMPACT data set [56] for TBI, and the ROUTiNE data set [57] for cardiac arrest. For characteristics of these databases, see table 1.2. Moreover, I will use already published data for systematic reviews and decision models.

1.6 | Outline of this thesis

As this thesis answers three main questions, it is divided into three parts.

Table 1.2: *Characteristics of the datasets included in this thesis.*

	CENTER-TBI	IMPACT-II	ROUTiNE
Study type	Prospective	Prospective(4)/RCT(11)	Prospective
Included patients	TBI	Moderate-severe TBI	IHCA
N patients	4509	11022	701
N centers	53	355	14
N countries	18	-	1
Inclusion period	2014 - 2018	1984 - 2004	2017 - 2018

The first part, *current practice*, begins with two studies in the field of TBI. **Chapters 2 and 3** cover the variation in prehospital treatments for TBI throughout Europe. **Chapter 2** aims to provide a general overview, where **chapter 3** zooms in on intubation as one of the most vital prehospital interventions in moderate to severe TBI. **Part I** continues with studies on in-hospital cardiac arrest, where in **chapters 4 and 5** the outcome after IHCA is described. **Chapter 4** describes the overall outcome of these patients, and **chapter 6** describes the outcome for IHCA patients who underwent ECPR. Finally, in **chapter 6**, I assess variation in interventions, next to variation in outcome in IHCA patients in the Netherlands.

The next part, *best practice*, again begins with two studies in the field of TBI. **Chapters 7 and 8** assess the benefit of two interventions in TBI: In **chapter 7** I study intubation, and in **chapter 8** I study direct transfer to a specialized center. This part ends with a study in IHCA: in **chapter 9**, I assess the cost-effectiveness of ECPR in IHCA patients.

The last part, *identifying patients at risk*, focuses mainly on TBI. In the first chapter, **chapter 10**, I focus on grouping TBI patients based on clinical characteristics. **Chapter 11** evaluates what kind of algorithms can best be

used to predict outcome in these patients. Moreover, I assess one way to improve prognostication in TBI, using biomarkers (**chapter 12**). Finally, methodological guidance to deal with missing data in predictive research is given in **chapter 13**.

References

1. Allen, R. E. Anamnesis in Plato's "Meno and Phaedo". *The review of metaphysics,* 165–174 (1959).
2. Maas, A. I. R. *et al.* Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. *The Lancet Neurology* **4422.** ISSN: 14744422 (2017).
3. Feigin, V. L. *et al.* Global, regional, and national burden of neurological disorders, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Neurology* **18,** 459–480. ISSN: 14744465 (2019).
4. Majdan, M. *et al.* Epidemiology of traumatic brain injuries in Europe: a cross-sectional analysis. *The Lancet Public Health* **1,** e76–e83. ISSN: 2468-2667 (2016).
5. Menon, D. K., Schwab, K., Wright, D. W. & Maas, A. I. Position Statement: Definition of Traumatic Brain Injury. *Archives of Physical Medicine and Rehabilitation* **91,** 1637–1640 (2010).
6. Advanced trauma life support (ATLS): The ninth edition (2013).
7. Green, R. S., Butler, M. B. & Erdogan, M. Increased mortality in trauma patients who develop postintubation hypotension. *Journal of Trauma and Acute Care Surgery* **83,** 569–574 (2017).
8. Manley, G. *et al.* Hypotension, Hypoxia, and Head Injury. *JAMA surgery* **136,** 1118–1123 (2001).

9. Moulton, C, Pennycook, A & Makower, R. Relation between Glasgow coma scale and the gag reflex. *British Medical Journal* **303,** 1240–1241. ISSN: 09598146 (1991).

10. Boidin, M. P. Airway patency in the unconscious patient. *British journal of anaesthesia* **57,** 306–310. ISSN: 0007-0912 (Print) (1985).

11. Lockey, D. J., Coats, T. & Parr, M. J. Aspiration in severe trauma: A prospective study. *Anaesthesia* **54,** 1097–1098. ISSN: 00032409 (1999).

12. Badjatia, N. *et al.* Guidelines for Prehospital Management of Traumatic Brain Injury 2nd Edition. *Prehospital Emergency Care* **12,** S1–S52. ISSN: 1090-3127 (2008).

13. Teasdale, G & Jennett, B. Assessment of coma and impaired consciousness. A practical scale. *The Lancet* **2,** 81–84. ISSN: 0140-6736 (Print) (1974).

14. Bernard, S. A. *et al.* Prehospital Rapid Sequence Intubation Improves Functional Outcome for Patients With Severe Traumatic Brain Injury. *Annals of Surgery* **252,** 959–965. ISSN: 0003-4932 (2010).

15. Stocchetti, N. & Maas, A. I. Traumatic Intracranial Hypertension. *New England Journal of Medicine* **370,** 2121–2130. ISSN: 0028-4793 (2014).

16. McConnell, K. J., Newgard, C. D., Mullins, R. J., Arthur, M. & Hedges, J. R. Mortality benefit of transfer to level I versus level II trauma centers for head-injured patients. *Health services research* **40,** 435–458 (2005).

17. Mendeloff, J. M. & Cayten, C. G. Trauma Systems and Public Policy. *Annual Review of Public Health* **12,** 401–424. ISSN: 0163-7525 (1991).

18. DuBose, J. J. *et al.* Effect of trauma center designation on outcome in patients with severe traumatic brain injury. *Archives of surgery* **143,** 1213–1217 (2008).

19. Cnossen, M. C. *et al.* Adherence to Guidelines in Adult Patients with Traumatic Brain Injury: A Living Systematic Review. *Journal of Neurotrauma* **14.** ISSN: 0897-7151 (2016).

20. Cnossen, M. C. *et al.* Variation in structure and process of care in traumatic brain injury: Provider profiles of European Neurotrauma Centers participating in the CENTER-TBI study. *PLoS ONE* **11,** 1–21. ISSN: 19326203 (2016).

21. Lingsma, H. F. *et al.* Large Between-Center Differences in Outcome After Moderate and Severe Traumatic Brain Injury in the International Mission on Prognosis and Clinical Trial Design in Traumatic Brain Injury (IMPACT) Study. *Neurosurgery* **68,** 601–608. ISSN: 0148-396X (2011).

22. Saatman, K. E. *et al.* Classification of traumatic brain injury for targeted therapies. *Journal of neurotrauma* **25,** 719–738 (2008).

23. Maas, A. I. *et al.* Re-orientation of clinical research in traumatic brain injury: report of an international workshop on comparative effectiveness research. *Journal of neurotrauma* **29,** 32–46 (2012).

24. Bergum, D, Nordseth, T, Mjolstad, O. C., Skogvoll, E & Haugen, B. O. Causes of in-hospital cardiac arrest - Incidences and rate of recognition. *Resuscitation* **87,** 63–68 (2015).

25. Gräsner, J.-T. *et al.* EuReCa ONE—27 Nations, ONE Europe, ONE Registry: A prospective one month analysis of out-of-hospital cardiac arrest outcomes in 27 countries in Europe. *Resuscitation* **105,** 188–195 (2016).

26. Skogvoll, E, Isern, E, Sangolt, G. K. & Gisvold, S. E. In-hospital cardiopulmonary resuscitation. 5 years' incidence and survival according to the Utstein template. *Acta anaesthesiologica Scandinavica* **43,** 177–84. ISSN: 0001-5172 (1999).

27. Hodgetts, T. J. *et al.* Incidence, location and reasons for avoidable in-hospital cardiac arrest in a district general hospital. *Resuscitation* **44,** 115–123 (2002).

28. Sandroni, C, Nolan, J, Cavallaro, F & Antonelli, M. In-hospital cardiac arrest: Incidence, prognosis and possible measures to improve survival. *Intensive Care Medicine* **33,** 237–245 (2007).

29. Zhu, A. & Zhang, J. Meta-analysis of outcomes of the 2005 and 2010 cardiopulmonary resuscitation guidelines for adults with in-hospital cardiac arrest. *The American journal of emergency medicine* **34,** 1133–1139 (2016).

30. Karagiannidis, C. *et al.* Extracorporeal membrane oxygenation: evolving epidemiology and mortality. *Intensive Care Medicine* **42,** 889–896. ISSN: 0342-4642 (2016).

31. Extracorporeal life support organisation. *ECLS registry report* tech. rep. (2019).

32. Massetti, M *et al.* Back from irreversibility: Extracorporeal life support for prolonged cardiac arrest. *Annals of Thoracic Surgery* **79,** 178–183. ISSN: 0003-4975 (2005).

33. Tramm, R. *et al.* Extracorporeal membrane oxygenation for critically ill adults. *Cochrane Database of Systematic Reviews* (2015).

34. Chen, Y. S. *et al.* Comparison of outcome of extracorporeal cardiopulmonary resuscitation for out-of-hospital and in-hospital cardiac arrest. *Circulation* **128** (2013).

35. Holmberg, M. J. *et al.* Extracorporeal cardiopulmonary resuscitation for cardiac arrest: A systematic review. *Resuscitation* **131,** 91–100. ISSN: 1873-1570 (2018).

36. Shmueli, G. To explain or to predict? *Statistical Science* **25,** 289–310. ISSN: 08834237. arXiv: 1101.0891 (2010).

37. Steyerberg, E. W. *Clinical Prediction Models* ISBN: 978-3-030-16398-3. http://link.springer.com/10.1007/978-3-030-16399-0 (Springer International Publishing, Cham, 2019).

38. Harrell, F. E. *Regression Modeling Strategies* ISBN: 978-1-4419-2918-1 (Springer New York, New York, NY, 2001).

39. Medical Research Council. *Cluster randomised trials: methodological and ethical considerations* (2002).

40. Hemming, K., Haines, T. P., Chilton, P. J., Girling, A. J. & Lilford, R. J. The stepped wedge cluster randomised trial: Rationale, design, analysis, and reporting. *British Medical Journal (Online)* **350.** ISSN: 17561833 (2015).

41. Royston, P., Parmar, M. K. & Qian, W. Novel designs for multi-arm clinical trials with survival outcomes with an application in ovarian cancer. *Statistics in Medicine* **22,** 2239–2256. ISSN: 02776715 (2003).

42. Rosenberger, W. F. Randomized play-the-winner clinical trials: Review and recommendations. *Controlled Clinical Trials* **20,** 328–342. ISSN: 01972456 (1999).

43. Louis, T. A., Lavori, P. W., Bailar, J. C. & Polansky, M. Crossover and Self-Controlled Designs in Clinical Research. *New England Journal of Medicine* **310,** 24–31. ISSN: 0028-4793 (1984).

44. Swanson, S. A. & Hernán, M. A. Think globally, act globally: an epidemiologist's perspective on instrumental variable estimation. *Statistical science: a review journal of the Institute of Mathematical Statistics* **29,** 371 (2014).

45. Lousdal, M. L. An introduction to instrumental variable assumptions, validation and estimation. *Emerging themes in epidemiology* **15,** 1. ISSN: 1742-7622 (2018).

46. Van Leeuwen, N. *et al.* Regression Discontinuity Design. *Epidemiology* **27,** 503–511 (2016).

47. VanderWeele, T. J., Hernán, M. A. & Robins, J. M. Causal directed acyclic graphs and the direction of unmeasured confounding bias. *Epidemiology* **19,** 720–728. ISSN: 10443983 (2008).

48. Pearl, J. & Mackenzie, D. *The book of why: the new science of cause and effect* ISBN: 0465097618 (Basic Books, 2018).

49. Andreu-Perez, J., Poon, C. C., Merrifield, R. D., Wong, S. T. & Yang, G.-Z. Big data for health. *Institute of Electrical and Electronics Engineers: journal of biomedical and health informatics* **19,** 1193–1208 (2015).

50. Ryan, R. Heterogeneity and subgroup analyses in Cochrane Consumers and Communication Review Group reviews: Planning the analysis at protocol stage Heterogeneity (2014).

51. Hunink, M. M. *et al. Decision making in health and medicine: integrating evidence and values* (Cambridge University Press, 2014).

52. Sonnenberg, F. A. & Beck, J. R. Markov Models in Medical Decision Making. *Medical Decision Making* **13,** 322–338. ISSN: 0272-989X (1993).

53. Murray, E. J., Robins, J. M., Seage, G. R., Freedberg, K. A. & Hernán, M. A. A comparison of agent-based models and the parametric g-formula for causal inference. *American journal of epidemiology* **186,** 131–142 (2017).

54. Maas, A. I. *et al.* Collaborative European neurotrauma effectiveness research in traumatic brain injury (CENTER-TBI): A prospective longitudinal observational study. *Neurosurgery* **76,** 67–80. ISSN: 15244040 (2015).

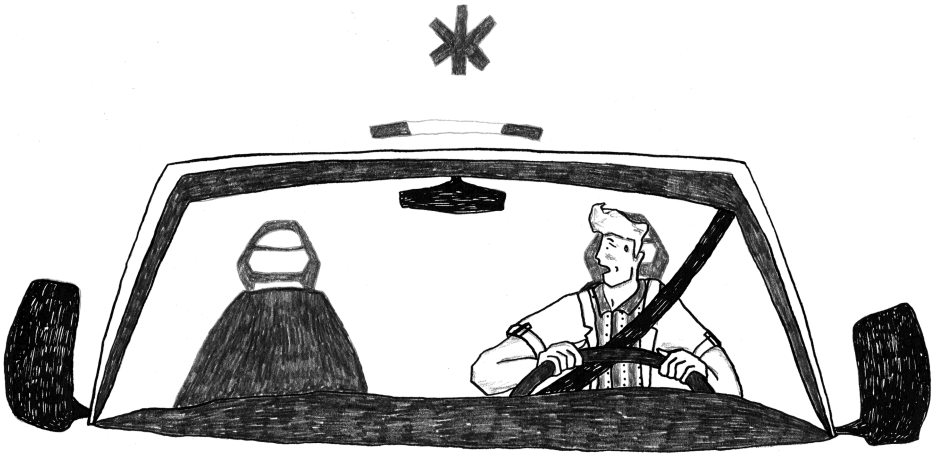
55. Steyerberg, E. W. *et al.* Case-mix, care pathways, and outcomes in patients with traumatic brain injury in CENTER-TBI: a European prospective, multicentre, longitudinal, cohort study. *The Lancet Neurology* **18,** 923–934 (2019).

56. Marmarou, A. *et al.* IMPACT Database of Traumatic Brain Injury: Design And Description. *Journal of Neurotrauma* **24,** 239–250. ISSN: 0897-7151 (2007).

57. Schluep, M. M. *et al.* Long-term survival and health-related quality of life after in-hospital cardiac arrest. *Resuscitation* (2021).

Part I

Current practice



2

Prehospital management of traumatic brain injury across Europe: a CENTER-TBI study

Benjamin Y Gravesteijn*, Charlie A Sewalt*, Nino Stocchetti, Giuseppe Citerio, Ari Ercole, Hester F Lingsma, Nicole von Steinbüchel, Ewout W Steyerberg, Lindsay Wilson, Andrew I R Maas, David K Menon, Fiona E Lecky, and CENTER-TBI collaborators. **both authors contributed equally*

Prehospital Emergency Care

2.1 | Abstract

Prehospital care for traumatic brain injury (TBI) is important to prevent secondary brain injury. We aim to compare prehospital care systems within Europe and investigate the association of system characteristics with the stability of patients at hospital arrival. We studied TBI patients who were transported to CENTER-TBI centres, a pan-European, prospective TBI cohort study, by emergency medical services between 2014 and 2017.

The association of demographic factors, injury severity, situational factors, and interventions associated with on-scene time was assessed using linear regression. We used mixed effects models to investigate the case mix adjusted variation between countries in prehospital times and interventions. The case mix adjusted impact of on-scene time and interventions on hypoxia (oxygen saturation <90%) and hypotension (systolic blood pressure <100mmHg) at hospital arrival was analysed with logistic regression.

Among 3878 patients, the greatest driver of longer on-scene time was intubation (+8.3 min, 95% CI: 5.6 - 11.1). Secondary referral was associated with shorter on-scene time (-5.0 min 95% CI: -6.2 - -3.8). Between countries, there was a large variation in response (range: 12 - 25 min), on-scene (range: 16 - 36 min) and travel time (range: 15 - 32 min) and in prehospital interventions. These variations were not explained by patient factors such as conscious level or severity of injury (expected OR between countries: 1.8 for intubation, 1.8 for IV fluids, 2.0 for helicopter). On-scene time was not associated with the regional EMS policy (p=0.58). Hypotension and/or hypoxia were seen in 180 (6%) and 97 (3%) patients in the overall cohort and in 13% and 7% of patients with severe TBI (GCS < 8). The largest association with secondary insults at hospital arrival was with major extracranial injury: the OR was 3.6 (95% CI: 2.6 - 5.0) for hypotension and 4.4 (95% CI:

2.9 - 6.7) for hypoxia.

To conclude, hypoxia and hypotension continue to occur in patients who suffer a TBI, and remain relatively common in severe TBI. Substantial variation in prehospital care exists for patients after TBI in Europe, which is only partially explained by patient factors.

2.2 | Introduction

Traumatic brain injury (TBI) remains an important cause of death and disability globally [1]. Although rates vary between countries, TBI is estimated to be responsible for around 300 hospital admissions and 12 deaths per 100,000 persons per year in Europe [2].

After the initial TBI, secondary insults, such as hypotension, hypoxia and intracranial hypertension may worsen the brain damage [3, 4]. Prehospital care for TBI focuses on preventing secondary brain injury by on-scene stabilisation and rapid transportation to an appropriate hospital. There is no universally accepted and implemented international guideline aimed at avoiding secondary injury in the prehospital environment. While national guidelines do exist, these vary substantially. Moreover, the extent to which they are adopted and implemented is unclear, since real-life data on international variations in prehospital care are limited. Provider profiling of study centres in the CENTER-TBI study [5–8], a large prospective observational cohort study of TBI across Europe and Israel, highlighted substantial reported variation in advanced life support capability of prehospital staff, degree of preference for stabilising on scene versus immediate transport, and in preferred destination from scene (specialist centre versus nearest hospital) [8]. However, these reported preferences were based on clinicians' reports of local protocols rather than objective patient data.

Objective assessment of such data is important. There is a trade-off between prehospital stabilization and prompt transportation to hospital. Stabilizing the patient in the prehospital environment with complex interventions can cause an important time delay reaching the hospital and starting appropriate diagnostic and tailored treatments. This delay could worsen outcome [9]. Conversely other studies suggest that stabilizing pa-

tients on-scene for transportation to more distant specialist centres could improve outcomes [10–13]. The decision between prehospital stabilization and immediate transport is made on-scene by prehospital staff based on clinical parameters, injury characteristics, skill levels available and the local policy.

The current study aimed to compare prehospital management of patients with TBI across Europe, and to investigate the association of prehospital care system characteristics with stability of patients at Emergency Department (ED) arrival.

2.3 | Methods

This study is reported according to the STROBE reporting guidelines [14]. Ethical approval was obtained from all local institutional revision boards, according to various national standards[1].

2.3.1 | Study design

CENTER-TBI is a multicentre, longitudinal, prospective, observational study in 18 countries across Europe which enrolled patients between December 2014 and December 2017 [5]. The core cohort includes patients presenting within 24 hours of injury, with a clinical diagnosis of TBI and an indication for computed tomography [6]. Analyses in this manuscript were undertaken on the CENTER-TBI dataset (version 2.0), and accessed using a bespoke data management tool, Neurobot[2].

[1] `https://www.center-tbi.eu/project/ethical-approval`

[2] Details available on the SciCrunch Resource Identification Portal, using the Research Resource Identifier RRID $SCR_0$17004.

Prehospital data were collected by physicians and researchers at participating study centres. Unfortunately, no data was available on prehospital physiology. Response time was defined as time between injury and arrival of first EMS crew. On scene time was defined as time between first EMS crew arrival until the conveying crew left the injury scene. Travel time was the time between patient leaving the scene and arrival at first hospital [15]. Major extracranial injury (MEI) was defined as any injury in all areas except head with an Abbreviated Injury Score (AIS) above 3.

2.3.2 | Patient selection

Patients with TBI who were transported by ambulance or helicopter to participating hospitals (n = 56), either directly or by secondary transfer, were included. For the centre-level analysis, secondary transfer patients were excluded.

2.3.3 | Statistical analysis

We have first compared baseline characteristics between patients that were immediately transported or that were stabilised on scene. This distinction was based on an a-priori defined cut-off of 20 minutes on scene. Continuous variables were described by the median and interquartile range (IQR). Categorical variables were described by the number of patients and the corresponding percentage.

Second, we have assessed the drivers of on-scene time, as a continuous variable, using linear regression. The included predictors were demographic factors (age, sex), severity (GCS, pupil reactivity, major extracranial injury), situational factors (travel time – as proxy to travel distance,

physician at scene, road traffic incident, high energy trauma), and interventions (intubation, IV fluids, CPR, ventilation). Within this analysis, we also assessed the adjusted between-country variation in prehospital times and prehospital interventions with mixed effects modelling. A random intercept for centres was applied to correct for between centre differences. To assess the effect of between-centre differences, the partial R^2 for the random intercept was calculated by comparing the R^2 of the model with and without random intercept.

Third, we have assessed the adjusted impact of on-scene times and prehospital interventions (intubation, ventilation, IV fluids, secondary referral) on hypoxia (Saturation <90%) and hypotension (Systolic Blood Pressure <100mmHg) at arrival with logistic regression. We adjusted for the following patient characteristics: age, GCS, pupil reactivity, major extracranial injury [16]. We also measured the influence of these surrogate prehospital endpoints on functional outcome using ordinal logistic regression, which was adjusted for the aforementioned patient characteristics and utilised the imputed optimised 6-month Extended Glasgow Outcome Scale (GOS-E [6]) as the dependent variable. We allowed for a non-linear effect of systolic blood pressure and saturation with restricted cubic splines (3 degrees of freedom).

Fourth, we have illustrated the unadjusted and adjusted variation between countries in prehospital times and rates of prehospital interventions (prehospital intubation, IV fluids, helicopter usage) across Europe. Bar charts depict unadjusted variation whilst the aforementioned mixed effects model enabled illustration of adjusted variation. Values of the random intercept for country were visually depicted on a map of Europe. Furthermore, the variation was adjusted for the core variables of the prediction model developed in the International Mission for Prognosis and Analysis

of Clinical Trials in TBI (IMPACT) study (age, number of reactive pupils, and Glasgow Coma Score at baseline) [16], and the CENTER-TBI stratum (ER/Admission/ICU) in which the patient was enrolled. Also, the median odds ratio (OR) was calculated, which quantifies the expected OR - of interventions performed or times taken - when two randomly picked countries are compared [17].

Additionally, the adjusted on-scene times were compared across centres which had indicated that they have a policy of immediate transportation, or a policy of stabilizing on scene based on the Provider Profiling questionnaires [8]. Therefore mixed effects models were applied, with on-scene time as dependent variable, on-scene policy as independent variable and country as random intercept. The on-scene times were adjusted for GCS, travel time to study centre, intubation, pupils and sex.

The effects of continuous predictors were presented as the odds ratio for comparing the 75th and the 25th percentile of the specific variable. This was calculated by multiplying the regression coefficient and standard error by the width of the interquartile range of that variable.

We performed the multiple imputation method to impute the covariates for all regression analyses using the MICE package in R. The following covariates were included in the imputation model: age, pupil reactivity, GCS, MEI, sex, prehospital intubation, IV fluids, CPR, ventilation, secondary referral and helicopter usage. The percentage of missing data can be found in table 2.1. These results were compared with complete case analysis as a sensitivity analysis. The results of the complete case analysis of each analysis are shown in the supplemental material, available online.

Table 2.1: *Descriptive analysis of patients who received a short on-scene time (< 20 min), or long on-scene time (> 20 min).*

		On-scene time			
	Overall	"Short", <20min, n=1744	"Long", >20min, n=2118	p	Missing %
Age (median [IQR])	51 [31, 67]	52 [31, 67]	50 [31, 67]	0.518	0.0
Male (%)	2647 (68.3)	1125 (64.5)	1511 (71.3)	<0.001	0.0
MEI (%)	670 (17.3)	209 (12.0)	456 (21.5)	<0.001	0.0
Cause (%)				0.081	10
RTI	1589 (45.6)	699 (44.8)	883 (46.3)		
Fall	1657 (47.5)	756 (48.4)	895 (46.9)		
Violence	191 (5.5)	92 (5.9)	97 (5.1)		
Intentional self-harm	48 (1.4)	14 (0.9)	34 (1.8)		
Type (%)				0.020	1
Closed	3702 (96.5)	1683 (97.4)	2004 (95.7)		
Blast	5 (0.1)	3 (0.2)	2 (0.1)		
Crush	91 (2.4)	27 (1.6)	63 (3.0)		
Penetrating	39 (1.0)	15 (0.9)	24 (1.1)		
Rural area (%)	742 (19.9)	235 (14.0)	502 (24.6)	<0.001	4
Place (%)				0.001	2
Street	2070 (54.6)	985 (57.5)	1077 (52.2)		
Home	941 (24.8)	381 (22.2)	557 (27.0)		
Work/school	240 (6.3)	94 (5.5)	146 (7.1)		
Sport	236 (6.2)	106 (6.2)	129 (6.3)		
Military	2 (0.1)	0 (0.0)	2 (0.1)		
Public location	303 (8.0)	148 (8.6)	152 (7.4)		
Highest trained bystander (%)				<0.001	0.5
None	33 (0.9)	5 (0.3)	27 (1.3)		
Bystander	23 (0.6)	17 (1.0)	6 (0.3)		
Paramedic	1173 (30.4)	664 (38.3)	503 (23.9)		
Nurse	658 (17.1)	400 (23.1)	258 (12.3)		
Physician	1044 (27.1)	456 (26.3)	583 (27.7)		
Medical rescue team	926 (24.0)	193 (11.1)	729 (34.6)		
Secondary referral (%)	594 (15.3)	352 (20.2)	241 (11.4)	<0.001	0.0
Arrival Method (%)				<0.001	0.0
Ambulance	3141 (81.0)	1585 (90.9)	1547 (73.0)		
Helicopter	483 (12.5)	97 (5.6)	381 (18.0)		
Mobile medical team	254 (6.5)	62 (3.6)	190 (9.0)		
GCS motor, baseline (median [IQR])	6 [4, 6]	6 [6, 6]	6 [2, 6]	<0.001	2
GCS, baseline(median [IQR])	14 [8, 15]	15 [13, 15]	13 [6, 15]	<0.001	4
Pupils, baseline (%)				<0.001	5
Two reactive	3273 (88.7)	1545 (92.7)	1717 (85.4)		
One reactive	150 (4.1)	53 (3.2)	96 (4.8)		
None reactive	269 (7.3)	69 (4.1)	197 (9.8)		
CPR (%)	51 (1.3)	10 (0.6)	40 (1.9)	0.001	0.0
IV Fluids (%)	1469 (37.9)	442 (25.3)	1019 (48.1)	<0.001	0.0
Intubation (%)	885 (23.7)	123 (7.4)	754 (36.7)	<0.001	4
Supplemental oxygen (%)	1612 (46.3)	485 (31.8)	1118 (57.5)	<0.001	10
Ventilation (%)	815 (22.0)	114 (6.9)	693 (34.1)	<0.001	4
On-scene time (median [IQR])	22 [15, 32]	14 [10, 17]	30 [25, 40]	<0.001	0.4
Arrival time (median [IQR])	17 [10, 30]	16 [10, 30]	18 [10, 30]	0.276	41
Travel time (median [IQR])	18 [11, 28]	15 [10, 23]	20 [12, 32]	<0.001	42
Prehospital time (median [IQR])	62 [44, 90]	45 [32, 60]	80 [61, 109]	<0.001	3

All analyses were performed using R[3]. The code applied can be found online[4].

2.4 | Results

We included 3878 patients from 56 centres in 17 European Countries from a total of 4509 patients enrolled into the core CENTER-TBI study. Patients who had self-presented to hospital without EMS activation (n = 616) or where prehospital details were missing or misreported (one country systematically misreported times for the 15 patients they included), were excluded (Figure S1 available online).

2.4.1 | On-scene time

The median on-scene time was 22 (IQR: 15 - 32) minutes, with 1744 (45) patients having an on-scene time of less than 20 minutes, and 2118 (55) more than 20 minutes (Table 1). Patients with TBI and longer on-scene times were more severely injured (GCS, pupil reactivity, MEI) and had more complex prehospital interventions (CPR, IV fluids, intubation and ventilation). The two characteristics with the largest association with longer on-scene time were prehospital tracheal intubation (+8.3 min, 95 CI: 5.6 - 11.1), and secondary referral (-5.0 min, 95 CI: -6.2 - -3.8). Other characteristics with smaller (though statistically significant) associations with longer on-scene times were travel time to the hospital (on average +0.6 min, 95 CI: 0.34 – 0.90), having a physician present at scene (+2.1 min , 95 CI: 1.1 – 3.2),

[3]R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

[4]`https://github.com/bgravesteijn/Code_Core_prehospital`

administration of IV fluids (+1.5 min, 95 CI: 0.5 – 2.4), initiation of ventilatory support (+3.1 min, 95 CI: 0.4 – 5.7), and male gender (+1.4 min, 95 CI: 0.6 - 2.3) (Figure 2.1; Table 1, S1 available online). The full model explained 36% of the variation in on-scene time (R^2). Of that variation explained, 42% was due to between centre differences.

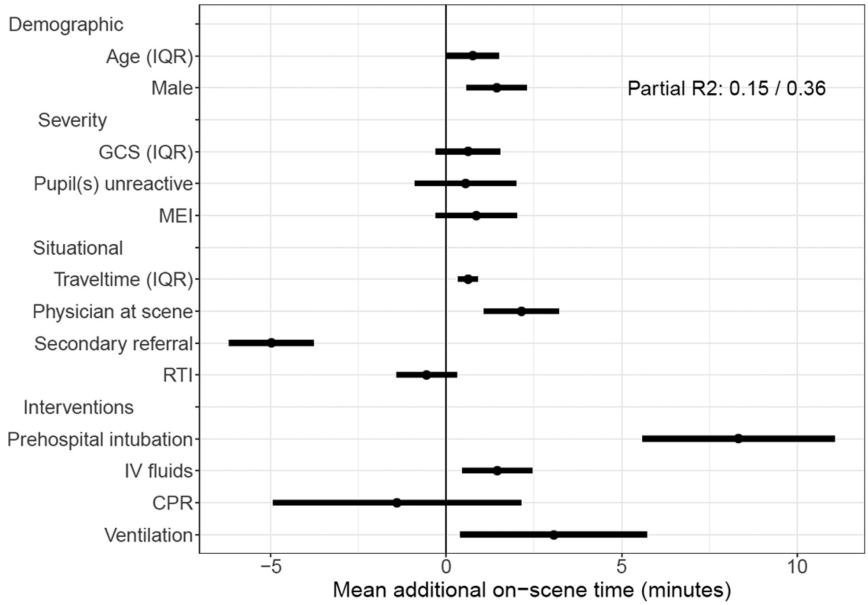


Figure 2.1: *A forest plot showing the associations of demographic factors, injury severity, situational factors, and given interventions with on on-scene time. The associations are the result of a multivariable analysis, which includes all variables in the model. The estimates can be interpreted as follows: this factor increases or decreases the on-scene time by x minutes, independent of the other factors displayed. This is the result of a multivariable mixed effects linear regression model with a random intercept for centre conditional on country. The coefficients (and 95% confidence intervals) of the model are displayed. The partial R^2 displayed is the percentage of the full model attributable to between country differences. RTI: Road traffic incident; MEI: major extracranial injury; GCS: Glasgow Coma Scale; IQR: interquartile range; CPR: cardiopulmonary resuscitation; IV: intravenous.*

2.4.2 | Predictors of hypotension and hypoxia

In total, 159 (5%) of the patients arrived at the ED with hypotension, 76 (2%) with hypoxia, and 21 (1%) with both (table 2.2). The proportions of hypoxia and hypotension were higher in severe TBI patients (defined as a GCS < 8), 90 (11%) arrived with hypotension, 38 (5%) with hypoxia, and 17 (2%) with both (table 2). Moreover, of the patients who were intubated on-scene, 92 (12%) had hypotension, 31 (4%) had hypoxia, and 14 (2%) had both.

Table 2.2: *The number and percentage of patients with hypotension or hypoxia at arrival at the ED.*

	N	Hypotension+Hypoxia	Hypotension	Hypoxia	Neither
Overall	3348	21 (1%)	159 (5%)	76 (2%)	3092 (92%)
Intubated	759	14 (2%)	92 (12%)	31 (4%)	622 (82%)
Not intubated	2485	6 (0%)	62 (2%)	42 (2%)	2375 (96%)
Primary referral	2871	20 (1%)	140 (5%)	67 (2%)	2644 (92%)
Secondary referral	477	1 (0%)	19 (4%)	9 (2%)	448 (94%)
GCS >12	2096	4 (0%)	45 (2%)	26 (1%)	2021 (96%)
GCS 9-12	318	0 (0%)	17 (5%)	9 (3%)	292 (92%)
GCS <9	842	17 (2%)	90 (11%)	38 (5%)	697 (83%)

The largest association with secondary insults on arrival was with major extracranial injury: the OR was 3.6 (95% CI: 2.6 – 5.0) for hypotension and 4.4 (95% CI: 2.9 – 6.7) for hypoxia. Other patient factors were also independently associated with arrival secondary insults including a higher GCS at scene, which was associated with less hypotension (OR 0.7, 95% CI: 0.5 - 0.9) and hypoxia (OR 0.6, 95%CI 0.4 - 0.8) on arrival; the presence of on scene unilaterally or bilaterally non-reactive pupils(s) predicted arrival hypoxia (OR: 1.9, 95% CI: 1.1 – 3.1). In terms of interventions, the requirement for IV fluids was associated with hypotension at arrival (OR 1.8, 95% CI: 1.3 – 2.5), while prehospital time (average OR 1.1 (1.01 - 1.20)) predicted hypoxia at arrival (Figure 2.2; Table 2.2 S1 available online). The complete case analysis showed the same direction and range of effects (Figure 4 S1 available online). The case mix adjusted variation by country in rates of arrival hypoxia and hypotension was small with a median OR of 1.11 and 1.05 respectively (Figure 6 S1 available online).

The adjusted association of these surrogate endpoints with functional outcome was significant (Figure 5 S1 available online): for saturation, lower values were associated with worse GOSE scores, plateauing at a saturation above 95%. For systolic blood pressure, lower (<100 mmHg) as well as higher (>180 mmHg) values were associated with worse functional outcome.

2.4.3 | National variation

There was large variation between prehospital times across European countries (unadjusted analyses, Figure 2.3). The shortest prehospital times for primary referrals were seen in Sweden (49 [IQR: 39-64] minutes) and Serbia (44 [IQR: 28 - 85] minutes) whereas the longest prehospital times were seen

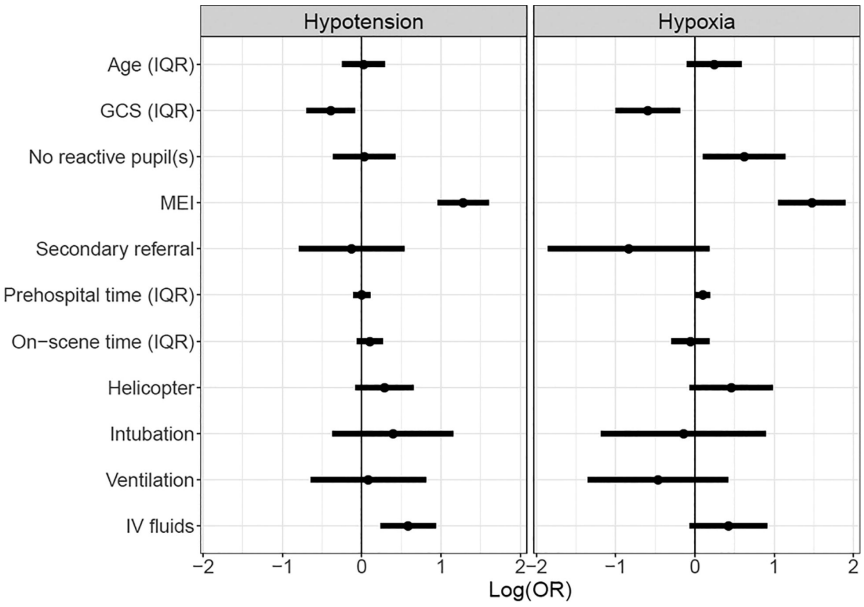


Figure 2.2: *The association of demographic factors, injury severity, situational factors, and interventions given on hypotension (systolic blood pressure < 100mmHg) or hypoxia (oxygen saturation < 90%) at arrival at the emergency department. The effects are based on a logistic multivariable regression model, which includes all variables shown in the plot.*

in the United Kingdom (96 [IQR: 72 – 127] minutes) and France (101 [IQR: 74 – 146] minutes). Secondary referral extended the time until arrival at the study hospital to a greater degree (to hours rather than minutes). In Sweden, the time to arrival at the study hospital for secondary referrals was the longest (446 [IQR: 340 – 560] minutes). There was also large between-country variation in therapies the patients were provided with: intubation rates varied from 10% to 88%, iv fluid administration from 22% to 67%, and use of helicopters from 0% to 31%.

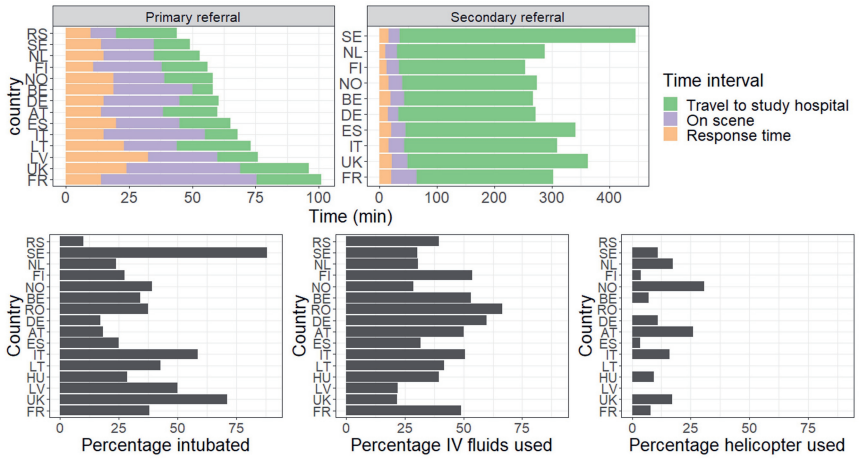


Figure 2.3: *Bar charts showing the time spent in different prehospital phases per country (upper row), and the percentage of prehospital interventions (second row) used. In the upper row, only bars based on more than 10 patients are displayed.*

After adjusting for case mix, the variation in prehospital times and interventions within Europe remained substantial (Figure 2.4). The range of response times adjusted for injury severity was 12-25 minutes; the range of on-scene times was 16-36 minutes; and the range of travel times was 15-32 minutes. The range of response times adjusted for injury severity and prehospital interventions was 9-31 minutes; the range of on-scene times was 15-34 minutes; and the range of travel times was 14-32 minutes. The median odds ratio, expected when two randomly picked countries are compared, was 1.8 for prehospital intubation, 1.8 for IV fluids and 2.0 for helicopter. If prehospital times were also adjusted

Figure 2.4: *The adjusted variation in prehospital time (upper row), and use of key prehospital interventions (bottom row) across Europe. Every map shows the deviation per country from the overall average. In the upper row, the mean of the median time per country is shown. Moreover, secondarily referred patients are excluded from the analysis of travel times, because the time until arrival in the secondary hospital is unknown. The estimates of the random intercepts for each country are displayed. These are adjusted for the IMPACT core variables (age, pupils, and GCS), the CENTER-TBI stratum in which the patient was included, and the random variation at the centre level.*

For the interventions that individual patients received, the model fit improved significantly (likelihood ratio tests, p<0.001). However, the values of the random intercepts (which represent the average difference to the European average) did not differ from the models that only adjust for injury severity (figure S7 available online).

The unadjusted difference between the on-scene times of centres was not significantly different for patients from study hospitals reporting their EMS having a policy of stabilising on scene versus a policy of immediate transport (p=0.49) (18). After adjustment, the two centres reporting to have only a policy of immediate transport as part of provider profiling had on average the shortest average on-scene times (Figure 2.5). However, the overall difference in on-scene times between hospitals that reported the two different prehospital EMS policies was not significant (p=0.58).

2.5 | Discussion

To our knowledge this is the most comprehensive analysis comparing prehospital care for patients after TBI across Europe. Our multicentre, multinational, prospective cohort study suggests large variations across European countries in the prehospital care provided to patients who suffer a TBI, largely unexplained by patient characteristics. Despite the common availability of national guidelines for prehospital care, patients after TBI continue to present at the ER with hypotension and hypoxia, although these are less common than in the past (6% and 3% of cases, respectively). These physiological insults are most common in severe TBI, where they occur in 13% and 7% of cases, respectively. The main determinant of such physiological instability on arrival at hospital were major extracranial in-

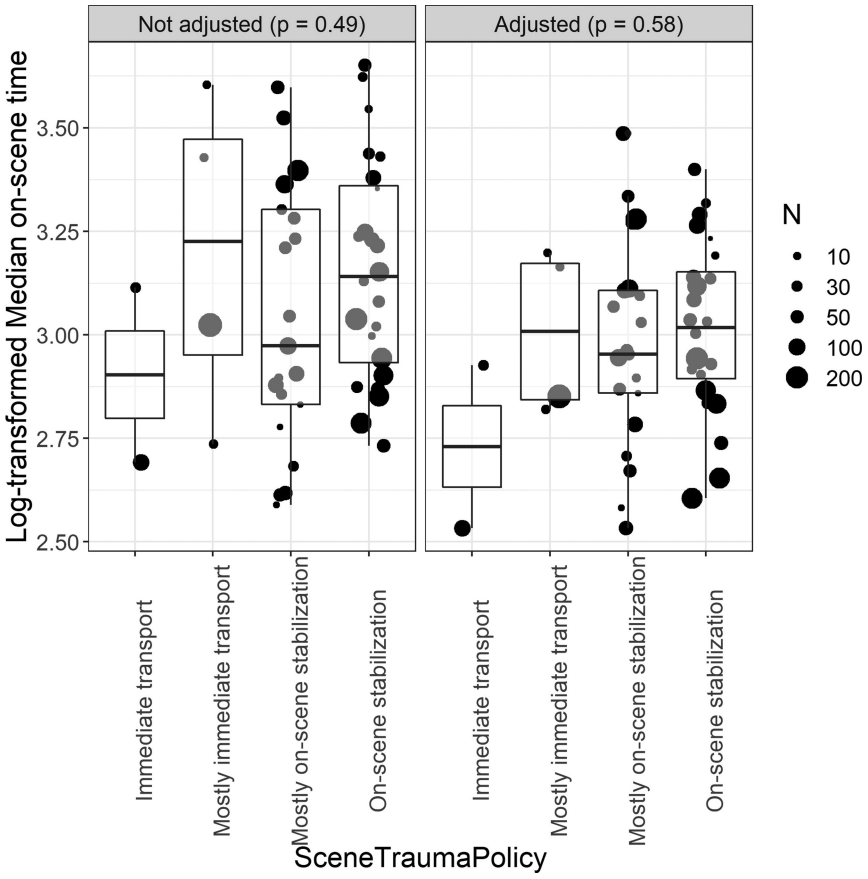


Figure 2.5: *The unadjusted and adjusted log transformed median on-scene times. The bubbles represent the random intercept value for the model predicting on-scene time with centre as random intercept. The right panel shows log transformed median on-scene times adjusted for GCS, traveltime, intubation, pupils, and sex (which were identified drivers of on-scene time).*

juries. We found that the main determinants of longer on-scenes time were interventional and situational rather than patient-related, for example on-

scene intubation and primary referral to the study centre.

However, we also determined that variation across Europe in prehospital times and interventions was only partly concordant with the prehospital policy (immediate transport or stabilise on scene) reported by clinicians in the CENTER-TBI provider profiling exercise [8]. We discovered that the probability of a patient with TBI being intubated at the injury scene, receiving IV fluids, or being transported by helicopter, was highly dependent on the country where the patient suffered the injury.

Not only did we see variation in prehospital interventions, but also in prehospital times. For on-scene times, this can partially be explained by the variation in provided interventions: for example, we found that prehospital intubation increased the on-scene time by 10 minutes, similar to an American retrospective study [18]. Other interventions (IV-fluids, mechanical ventilation) also slightly increased the on-scene times. It is likely that the association of prolonged on-scene time and greater intervention may have been, in part, due to greater injury severity, requiring more on-scene stabilization before transfer. Although this explanation might be true for variations observed concerning the patient-level, the explanation for country-level variation in hospital times requires a different explanation: the diverse geographical landscapes of Europe, and the large between-centre variation in the size and type of population of hospital catchment areas are more likely to drive the variation in prehospital times. Unsurprisingly, the use of helicopters was most prevalent in Norway which has large areas with low population density. Interestingly though, the longest total prehospital times (even after adjustment for patient and some situational factors) occurred in France and the United Kingdom. Potential explanations vary: France had the highest case-mix adjusted rates of prehospital intubation concordant with their surveyed response of stabilizing patients

on scene; while the United Kingdom had the highest travel times from scene to hospital, perhaps reflecting traffic congestion and/or recent centralization of major trauma care to just 30 out of over 200 hospitals (8 of which participated in CENTER-TBI).

Despite large variation in performed interventions and prehospital times were observed, the rates of hypoxia and hypotension at arrival at the Emergency Department were lower than those in historical TBI studies: for example, even in the group of severe patients, only 11% had hypotension at arrival, compared to 35% in a large historical study [3, 19]. In part, these lower rates may be explained by differences in case selection or definitions. While we only report documented hypoxia, the Traumatic Coma Data Bank also inferred hypoxia if there was clinically reported cyanosis or apnea. For example, we included intoxicated GCS < 9 patients in CENTER-TBI, similar to the study by Miller et al [20], who found a similar incidence of hypotension. Historically, TBI patients not in coma were generally not thought to have sustained a significant injury and imaging by CT scan was rarely conducted if intoxication was thought to be the root cause of a low GCS. Therefore, these patients were not included in historical TBI studies. The lower rates of hypoxia and hypotension at arrival can be explained by a higher inclusion rate of mild TBI patients with less severe extracranial injury than in previous studies. Our study reflects modern Emergency Medicine practice, which is to image all severities of TBI. However, there remains the possibility that prehospital care has simply improved over the last decades - in particular the almost universal use of supplemental oxygen, increased use of tracheal intubation, and the common use of prehospital IV fluids, - may have markedly reduced the incidence of hypoxia and hypotension. However, there continues to be room for improvement - both physiological insults still occur at high rates, particularly in patients after

severe TBI.

A limitation of this international, multicentre trial is the proportion of missing data. This is unfortunately unavoidable in such a logistically challenging study. Since complete case analysis is both inefficient, and potentially biased, we imputed the data [21]: both single imputation for the on-scene time, as well as multiple imputation for the main analyses were used. The single imputation was reliable, but not perfect: 60% of the variation could be explained by the model. For the analysis with multiple imputed datasets, similar results were observed as the complete case analysis. This supports the validity of the selected imputation method.

Another limitation is that some prehospital physiological parameters (oxygen saturation and blood pressure) were not entered into the database. We used hypotension and hypoxia at arrival at the Emergency Department as a proxy for secondary insult. However, interventions such as intubation may have restored normal oxygen levels for some patients who were hypoxic at scene. There were some situational factors such as difficult extrication from the scene due to entrapment or stairs that may be valid factors for prolonging on scene times – and vary by country – that we could not account for using the data.

Finally, we acknowledge the fact that the centres that contributed patients to CENTER-TBI are a selected population of centres: these centres were mostly the equivalent of North American level 1 trauma centres [8]. Our conclusions are based on extrapolation of the preferences and policies of these specialised centres towards the entire country.

Nevertheless, the prospective nature of the study, the large number of centres and countries, and the size of the CENTER-TBI cohort do provide high external validity. Additionally, the data are acquired as "real-world" data, with lenient exclusion criteria. Therefore, we believe our results are

applicable to the majority of settings.

We suggest that the large variation in administered prehospital interventions can be explained by two factors. First, the most relevant guidelines for prehospital management of TBI are national guidelines, which vary substantially across countries [8]. However, even within countries, local policies vary according to the Provider Profiling questionnaires [22]. Moreover, these local policies might not be concordant with practice, as research suggests that the adherence to guidelines is low [23]. However, it is also possible that the prehospital guidelines are not (or not perceived as being) relevant to clinical practice in these contexts, and/or may be difficult to implement [7]. Understanding and reconciling this discordance is essential if we are to provide a better evidence base for clinical practice in these contexts and ensure its appropriate adoption.

Second, the resources for prehospital care vary substantially across Europe. Even for prehospital intubation, for which the benefit - for severe TBI - has been shown in a randomized controlled trial [24], large variation was observed irrespective of patient factors [25]: the practice variation is therefore likely to be also attributable (in part) to variation in resources. In many countries the academic basis for prehospital care is now only becoming a routine part of training for paramedics and other practitioners, whereas it has been established for Hospital based Emergency Medicine for at least 20 years. Some elements of prehospital care – such as helicopters - are costly, so research should also take account of cost-effectiveness. We need to identify prehospital interventions with proven clinical and cost effectiveness, prioritize their integration into guidelines then monitor adherence and impact on outcomes.

2.5.1 | Conclusion

Across Europe, there are large variations in prehospital interventions for patients after TBI and in the associated on scene times. This variation is only partially explained by patient factors. Additional drivers of variation are likely to include EMS resource and organizational differences, and a low evidence base. While hypoxia and hypotension are less common than observed in past studies, they continue to occur in a substantial minority of patients after TBI, are particularly frequent following severe TBI or extracranial injury, and are associated with substantially worse outcomes. These data make a strong case for further research to facilitate the development and implementation of guidelines that support best practice in the prehospital care of patients with TBI.

2.6 | Supplementary Material

Supplementary materials can be found in the online version of this article: `https://tinyurl.com/mxfvk4ve`

References

1. Feigin, V. L. *et al.* Global, regional, and national burden of neurological disorders, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Neurology* **18,** 459–480. ISSN: 14744465 (2019).
2. Majdan, M. *et al.* Epidemiology of traumatic brain injuries in Europe: a cross-sectional analysis. *The Lancet Public Health* **1,** e76–e83. ISSN: 2468-2667 (2016).

3. Chesnut, R. M. *et al.* The role of secondary brain injury in determining outcome from severe head injury. *Journal of Trauma - Injury, Infection and Critical Care* **34,** 216. ISSN: 15298809 (1993).

4. Davis, D. *et al.* The Impact of Hypoxia and Hyperventilation on Outcome after Paramedic Rapid Sequence Intubation of Severely Head-injured Patients. *The Journal of Trauma: Injury, Infection, and Critical Care* **57,** 1–10. ISSN: 0022-5282 (2004).

5. Maas, A. I. *et al.* Collaborative European neurotrauma effectiveness research in traumatic brain injury (CENTER-TBI): A prospective longitudinal observational study. *Neurosurgery* **76,** 67–80. ISSN: 15244040 (2015).

6. Steyerberg, E. W. *et al.* Case-mix, care pathways, and outcomes in patients with traumatic brain injury in CENTER-TBI: a European prospective, multicentre, longitudinal, cohort study. *The Lancet Neurology* **18,** 923–934 (2019).

7. Cnossen, M. C. *et al.* Adherence to Guidelines in Adult Patients with Traumatic Brain Injury: A Living Systematic Review. *Journal of Neurotrauma* **14.** ISSN: 0897-7151 (2016).

8. Cnossen, M. C. *et al.* Prehospital trauma care among 68 European neurotrauma centers: Results of the CENTER-TBI Provider Profiling Questionnaires. *Journal of neurotrauma* **36,** 176–181 (2019).

9. Raj, R. *et al.* Factors correlating with delayed trauma center admission following traumatic brain injury. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine* **21,** 67. ISSN: 1757-7241 (2013).

10. Lodwick, G. & Edwards, L. Paediatric retrieval services: is it better to 'stay and play' or 'scoop and run'? *British Journal of Hospital Medicine* **78,** 118–118. ISSN: 1750-8460 (2017).

11. Nirula, R., Maier, R., Moore, E., Sperry, J. & Gentilello, L. Scoop and run to the trauma center or stay and play at the local hospital: hospital transfer's effect on mortality. *The Journal of trauma* **69,** 595–9; discussion 599–601. ISSN: 1529-8809 (2010).

12. Smith, R. M. & Conn, A. K. Prehospital care, Scoop and run or stay and play? *Injury* **40,** S23–S26. ISSN: 0020-1383 (2009).

13. King, S. Stay & play vs. scoop & run. *Journal of Emergency Medical Services* **28,** 14. ISSN: 0197-2510 (2003).

14. Von Elm, E. *et al.* Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *British Medical Journal* **335,** 806–8. ISSN: 1756-1833 (2007).

15. Ringdal, K. G. *et al.* The Utstein template for uniform reporting of data following major trauma: a joint revision by SCANTEM, TARN, DGU-TR and RITG. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine* **16,** 7. ISSN: 1757-7241 (2008).

16. Steyerberg, E. W. *et al.* Predicting Outcome after Traumatic Brain Injury: Development and International Validation of Prognostic Scores Based on Admission Characteristics. *PLoS Medicine* **5** (ed Singer, M.) e165. ISSN: 1549-1676 (2008).

17. Merlo, J. *et al.* A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *Journal of Epidemiol Community Health* **60,** 290–297 (2006).

18. Cudnik, M. T., Newgard, C. D., Wang, H., Bangs, C. & Herrington, R. Endotracheal Intubation Increases Out-of-Hospital Time in Trauma Patients. *Prehospital Emergency Care* **11,** 224–229. ISSN: 1090-3127 (2007).

19. Chesnut, R. M. *et al.* Early and late systemic hypotension as a frequent and fundamental source of cerebral ischemia following severe brain injury in the Traumatic Coma Data Bank. *Acta neurochirurgica. Supplementum* **59,** 121–5 (1993).

20. Miller, J. D., Sweet, R. C., Narayan, R. & Becker, D. P. Early Insults to the Injured Brain. *JAMA: The Journal of the American Medical Association* **240,** 439. ISSN: 0098-7484 (1978).

21. Little, R. J. A. & Rubin, D. B. Statistical Analysis with Missing Data. *Journal of Educational Statistics* **16,** 150. ISSN: 03629791 (1991).

22. Cnossen, M. C. *et al.* Variation in structure and process of care in traumatic brain injury: Provider profiles of European Neurotrauma Centers participating in the CENTER-TBI study. *PLoS ONE* **11,** 1–21. ISSN: 19326203 (2016).

23. Ebben, R. H. *et al.* Adherence to guidelines and protocols in the prehospital and emergency care setting: a systematic review. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine* **21,** 9. ISSN: 1757-7241 (2013).

24. Bernard, S. A. *et al.* Prehospital Rapid Sequence Intubation Improves Functional Outcome for Patients With Severe Traumatic Brain Injury. *Annals of Surgery* **252,** 959–965. ISSN: 0003-4932 (2010).

25. Gravesteijn, B. Y. *et al.* Variation in the practice of tracheal intubation in Europe after traumatic brain injury: a prospective cohort study. *Anaesthesia,* anae.14838. ISSN: 0003-2409 (2019).

3

Intubation practice in traumatic brain injury in Europe: a prospective cohort study

Benjamin Y Gravesteijn, Charlie A Sewalt, Ari Ercole, Fiona E Lecky, David K Menon, Ewout W Steyerberg, Andrew I R Maas, Hester F Lingsma, Markus Klimek, and the CENTER-TBI collaborators

Anaesthesia

3.1 | Abstract

Traumatic brain injury patients frequently undergo tracheal intubation. We aimed to assess current intubation practice in Europe and identify variation in practice.

We analysed data from patients with traumatic brain injury included in the prospective cohort study collaborative European neurotrauma effectiveness research in traumatic brain injury (CENTER-TBI) in 45 centres in 16 European countries. We included patients who were transported to hospital by emergency medical services. We used mixed-effects multinomial regression to quantify the effects on pre-hospital or in-hospital tracheal intubation of the following: patient characteristics, injury characteristics, centre, and trauma system characteristics.

A total of 3843 patients were included. Of these, 1322 (34%) had their tracheas intubated, 839 (22%) pre-hospital and 483 (13%) in-hospital. The fit of the model with only patient characteristics predicting intubation was good (Nagelkerke R^2 64%). The probability of tracheal intubation increased with the following: younger age, lower pre-hospital or emergency department GCS, higher abbreviated injury scale scores (head and neck, thorax and chest, face or abdomen abbreviated injury score), and one or more unreactive pupils. The adjusted median odds ratio for intubation between two randomly chosen centres was 3.1 (95% CI 2.1 – 4.3) for pre-hospital intubation, and 2.7 (95% CI 1.9 - 3.5) for in-hospital intubation. Furthermore, the presence of an anaesthetist was independently associated with more pre-hospital intubation (OR 2.9, 95% CI 1.3 - 6.6), in contrast to the presence of ambulance personnel who are allowed to intubate (OR 0.5, 95% CI 0.3 – 0.8).

In conclusion, patient and injury characteristics are key drivers of tra-

cheal intubation. Between-centre differences were also substantial. Further studies are needed to improve the evidence base supporting recommendations for tracheal intubation.

3.2 | Introduction

The burden of traumatic brain injury (TBI) is high; it is a leading cause of injury-related death and disability [1]. Although the rates vary between countries, TBI is estimated to be responsible for around 300 hospital admissions and 12 deaths per 100,000 persons per year in Europe [2]. Although the primary brain injury is defined by the trauma itself, secondary brain injury – especially due to hypoxia and hypotension – must be prevented [3–5]. Secondary insults might be prevented by securing the airway, by intubating the tracheas of patients with a depressed level of consciousness, compromised airway reflexes and induced central respiratory depression [6–9] to protect the airway and sustain normoxia and normocapnia [10, 11].

There are also potential risks of intubation. Injudicious use of anaesthetic agents required for intubation and positive pressure ventilation can cause hypotension, particularly in hypovolaemic trauma patients [12]. On the other hand, inadequate depth of anaesthesia during laryngoscopy may precipitate hypertension and lead to surges in blood pressure and/or intracranial pressure (ICP) [13]. Moreover, failure to rapidly control the airway may lead to hypoxia or hypercapnia. These insults (hypotension, intracranial hypertension and hypoxia) may all cause harm [4, 14–17].

There are few data available regarding which patients should have their airways secured. Although a GCS $\leq$ 8 is generally considered as the threshold for mandatory tracheal intubation [11, 18, 19], there is little evidence to support this recommendation. Traumatic brain injury intubation guidelines are based primarily on level-3 evidence [11]. The only exception is a randomised controlled trial recommending pre-hospital intubation in TBI patients with a GCS $\leq$ 9 [20]. Rates of adherence to guidelines for pre-hospital intubation are around 80%, with a wide range of 44–92% re-

ported in the literature [21, 22]. This lack of evidence and low adherence to guidelines could possibly result in differences in local intubation protocols or preferences.

We aimed to gain insights into the current practice of tracheal intubation after TBI across Europe by conducting this prospective cohort study, and to quantify the effects of: patient and trauma factors, centre, and trauma system characteristics on intubation practice.

3.3 | Methods

This study conforms with the STROBE reporting guidelines [23]. Data from the collaborative European neurotrauma effectiveness research in TBI (CENTER-TBI) were used [24]. In brief, CENTER-TBI was a prospective cohort study comprising 4509 patients with TBI of all severities. Traumatic brain injury patients presenting within 24 h after injury to one of the 61 participating study sites in Europe (mainly level-1 trauma centres), or referred from another hospital to the participating study site within 24 h, were eligible for this study. We collected data from 2014 until 2018. More details, including details concerning ethics approval, have previously been reported [24, 25].

For this analysis, we did not include patients who self-presented to the study site, because pre-hospital intubation can only be considered by medical services. We also did not include patients presenting to hospitals that included less than 20 patients, to allow for reliable statistical analysis. Although an intensive phase of data cleaning had already been completed, the CENTER-TBI database continues to be improved whenever data entry errors are found. Data for the CENTER-TBI study were collected through

the Quesgen e-CRF (Quesgen Systems Inc, Burlingame, CA, USA), hosted on the INCF platform and extracted via the INCF Neurobot tool (INCF, Stockholm, Sweden). We used Version 1.1 of the database for this analysis.

We defined in-hospital intubation by the variables that described whether a patient had their trachea intubated in the referring hospital (if they were referred), or in the study hospital. Pre-hospital intubation was defined by the variable that described whether a patient received pre-hospital intubation. All other patients were considered as having not had their tracheas intubated.

Since we were interested in the effect of baseline characteristics on both in-hospital and pre-hospital intubation, we mostly considered predictors that could influence both. However, readily available vital signs such as oxygen saturation or respiratory rate in the pre-hospital setting were not taken into account because they were not registered in the study. Instead, the baseline patient and trauma characteristics which were considered for the models included: age, the thorax, abdominal, facial and head and neck anatomical subscales abbreviated injury scale (AIS) of the injury severity score (ISS), the highest pre-hospital or emergency department (ED) GCS, and pre-hospital pupil reactivity.

Every participating study centre completed provider profiling questionnaires to gain insight into general operational structures and treatment policies for trauma patients. Details and the design of the questionnaires have previously been described [26–28]. For this study, we used questions that addressed the trauma system or policies regarding intubation. These included whether the physician on the pre-hospital care team was an anaesthetist, whether the ambulance personnel were trained to intubate without drugs and whether the policy on scene was best described as "stay-and-play" (giving treatment for stabilisation before transportation) or

"scoop-and-run" (transport the patient as quickly as possible to the hospital).

The data analysis plan was approved by the management committee of the CENTER-TBI study before commencement. Firstly, we compared patient and trauma characteristics of patients whose tracheas were intubated in the pre-hospital setting, in the in-hospital setting and patients whose tracheas were not intubated. Categorical variables were compared using Chi-square tests, or Fisher's exact test where appropriate. We tested continuous variables with one-way ANOVA or Kruskall–Wallis tests. The correlation between incidence of pre-hospital intubation and in-hospital intubation per centre was calculated with the Spearman's correlation coefficient.

For the models predicting intubation, we imputed missing data with a multiple imputation method (five datasets), using the MICE package [29], assuming data to be missing at random. The imputation model included relevant predictors and the outcome (intubation). After imputation, patients with missing outcome (intubation) were not included ('imputation then deletion') [30].

We used multinomial regression models to study associations with pre-hospital and in-hospital intubation. Candidate variables were selected based on the descriptive analysis ($p < 0.05$) and clinical knowledge, and were then included in the model. We did not categorise continuous variables.

Subsequently, the models, including patient and trauma characteristics, were extended with random intercepts for centre, conditional on country, to estimate the difference in probability of intubation between centres. Finally, we added the relevant trauma system characteristics from the provider profiling questionnaires to the model.

The different models were compared using the Nagelkerke R^2 as a measure for explained variance. The mean log-likelihood of the fitted models was compared with the log-likelihood of the null model [31]. To quantify the between-centre and between-country differences in intubation, we calculated the median odds ratio [32]. The median odds ratio can be interpreted as the odds ratio for intubation in two randomly selected centres or countries, comparing the high risk with the low-risk group. The estimates and standard errors of the random intercepts and variance of the random intercepts were pooled using Rubin's rules [33].

Two sensitivity analyses were performed. First, we performed a complete case sensitivity analysis, not including patients with some missing value in any of the predictors or outcome. The results were compared with the analysis on the imputed dataset, to observe whether imputation changed the effect estimates. Second, a sensitivity analysis was performed by not including the patients who underwent in-hospital tracheal intubation in a referring hospital. This was done to observe whether the two in-hospital intubated groups were comparable.

We performed the analyses using R (R Foundation for Statistical Computing, Vienna, Austria). For the multinomial model, the 'multinom' function from the 'nnet' package was used. The mixed-effects multinomial regression was performed using the PROC GLIMMIX function in SAS[1] [34]. The code can be found online[2].

[1] SAS Institute Inc. SAS, Cary, NC, USA

[2] `https://github.com/ErasmusCMB/CENTER-TBI/blob/master/final_script_pv_intub.Rmd`

3.4 | Results

After excluding patients who did not arrive by medical services (n = 487), patients from centres with fewer than 20 patients (n = 176) and patients from whom no information on intubation was present (n = 3), we included 3843 patients from 45 centres in the analysis (figure 3.1). The median number of patients was 62 per centre, and 115 per country (figure 3.2).

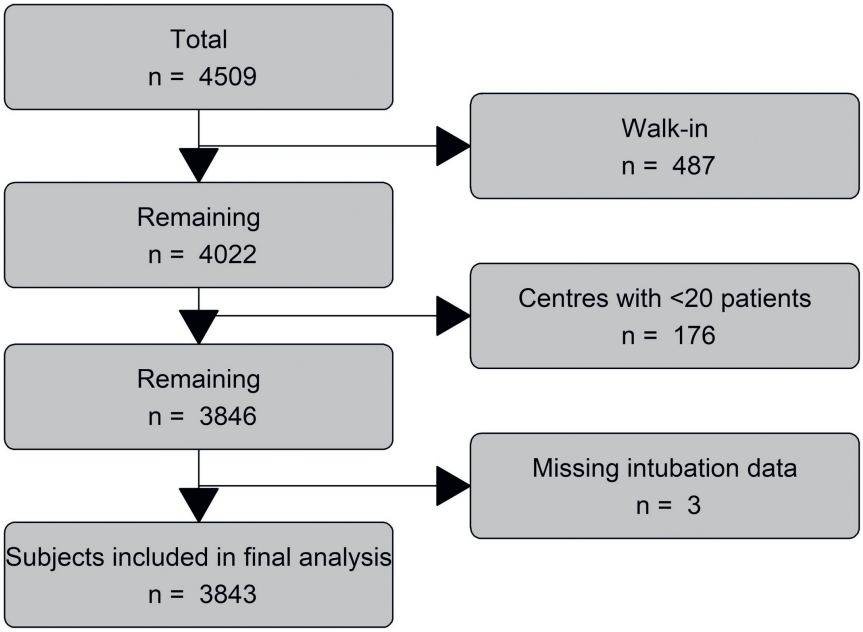


Figure 3.1: *Flow chart of patients included in this analysis.*

In total, 839 (22%) had their tracheas intubated in the pre-hospital setting, while 483 (13%) had their tracheas intubated in hospital, of which 194 (40%) were performed in the referring hospital. The observed pre-hospital intubation rates differed from 0% to 60% between centres, and from 2%

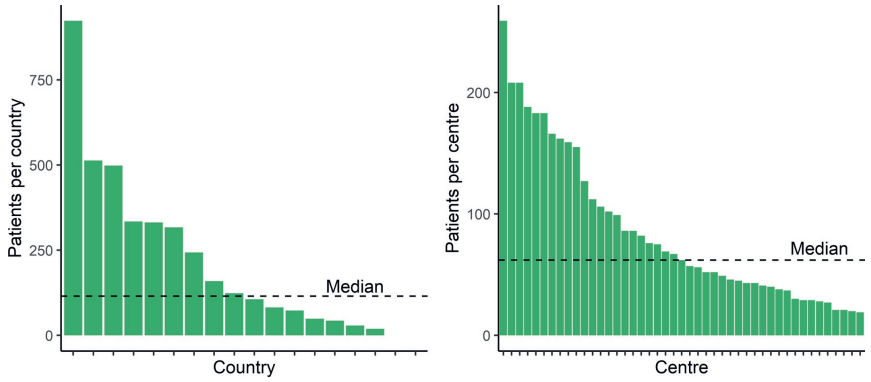


Figure 3.2: *Number of observations per participating country and centre. The median is displayed (115 per country, 62 per centre).*

to 56% between countries. In-hospital intubation rates differed from 0% to 73% between centres, and from 1% to 41% between countries (figure 3.3). Centres who performed more pre-hospital intubation did not perform more or less in-hospital intubation (rho = 0.05, p = 0.73).

Patients whose tracheas were intubated had lower pre-hospital motor GCS, median (IQR [range]) 3 (1–5 [1 – 6]) and higher ISS than patients whose tracheas were not intubated. The pre-hospital intubation group most often had one or two non-reactive pupils (187, 29%), followed by the in-hospital intubation group. Patients in the pre-hospital intubation group were 7.0 (95% CI 5.1 – 8.2) years younger than the other groups. Road traffic incident was the cause of injury in the majority of the pre-hospital intubation group (458, 56%), whereas falls were more common in the other groups; 195 (43%) in the in-hospital intubation group and 1195 (48%) in the group who were not intubated. The pre-hospital time was 0.3 h longer in the pre-hospital intubation group (95% CI 0.2 – 0.3 h). The travel time, however, was similar in all groups; the median was 0.3 (0.2 – 0.5 [0 – 1.37]),

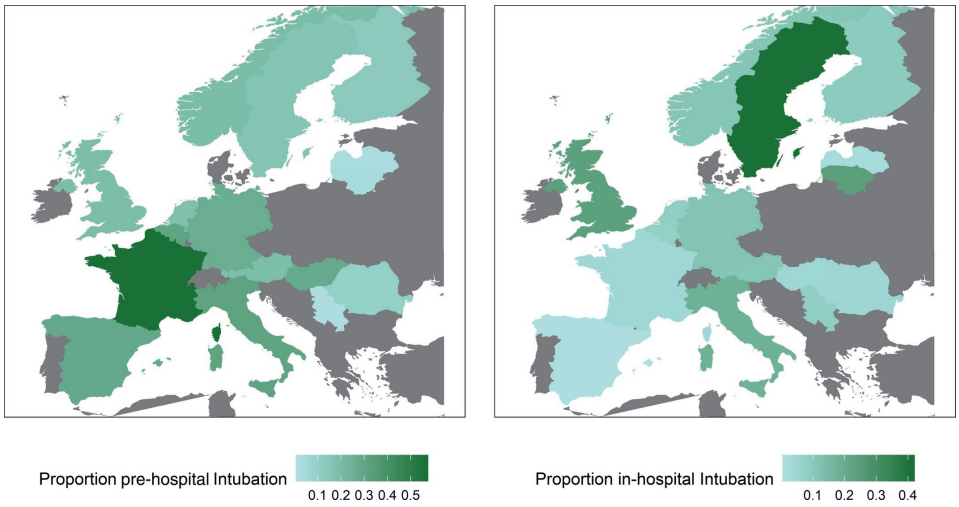


Figure 3.3: *Proportion of pre-hospital and in-hospital patients who had their tracheas intubated across Europe.*

0.2 (0.2 – 0.4 [0 – 1.37]) and 0.3 (0.2 – 0.4 [0 – 1.37]) in the pre-hospital, in-hospital and no intubation groups, respectively. The highest proportion of missing values was seen for the pupil assessments (43% pre-hospital, 50% in-hospital) and the travel time (50%) (Table 3.1).

Table 3.1: *Baseline characteristics. Values are number (proportion) or median (IQR [range])*

	Pre-hospital intubation n = 839	In-hospital intubation n = 483	Not intubated n = 2521	Missing data
Trachea intubated in referring hospital	–	194 (40.2%)	–	–
Age, years	44 (25–60 [3–92])	52 (31–68 [0–95])	53 (33–68 [1–94])	0%
Male	616 (73.4%)	353 (73.1%)	1639 (65.0%)	0%
BMI	24.7 (22.6–27.7 [14–52])	24.7 (22.6–27.6 [15–42])	24.8 (22.3–27.6 [13–57])	32%
Total injury severity score	35 (25–50 [1–75])	29 (25–41 [1–75])	13 (8–18 [1–75])	1%
Head/neck AIS	3 (0–4 [0–6])	2 (0–5 [0–6])	1 (0–3 [0–6])	0%
Thorax/chest AIS	2 (0–3.5 [0–5])	0 (0–3 [0–5])	0 (0–0 [0–5])	0%
Face AIS	0 (0–3 [0–6])	0 (0–2 [0–5])	0 (0–1 [0–5])	0%
Abdomen/pelvis AIS	0 (0–0 [0–5])	0 (0–0 [0–6])	0 (0–0 [0–5])	0%
Cause of injury				
RTI	458 (55.9%)	178 (39.1%)	919 (37.1%)	2%
Fall	265 (32.4%)	195 (42.9%)	1195 (48.2%)	
Other	54 (6.6%)	45 (9.9%)	203 (8.2%)	
Violence	42 (5.1%)	37 (8.1%)	161 (6.5%)	
Highest pre-hospital mGCS	3 (1–5 [1–6])	5 (3–6 [1–6])	6 (6–6 [1–6])	34%
mGCS at arrival at the first ED	1 (1–1 [1–6])	5 (1–6 [1–6])	6 (6–6 [1–6])	22%
Most predictive GCS	4 (3–8 [3–15])	8 (5–13 [3–15])	15 (14–15 [3–15])	4%
GCS >12	74 (9.7%)	117 (25.2%)	2236 (90.2%)	4%
Pupil(s) unreactive pre-hospital	187 (28.9%)	28 (11%)	28 (2.1%)	43%
Pupils unreactive in-hospital	227 (40.4%)	52 (23.2%)	37 (3.3%)	50%
Pre-hospital time, h	1.3 (1.0–1.7 [0–5])	1.0 (0.7–1.3 [0–4])	1.0 (0.7–1.4 [0–5])	8%
Travel time, min	0.3 (0.2–0.5 [0–1])	0.2 (0.2–0.4 [0–1])	0.3 (0.2–0.4 [0–1])	50%

Every variable differed significantly between the groups, $p < 0.001$. BMI: body mass index, AIS: abbreviated injury scale, RTI: road traffic incident, mGCS: motor component of the Glasgow Coma Score, ED: emergency department.

Consecutively, we fitted the model with seven predictors of intubation (Fig. 3.5). The strongest predictor was GCS (OR 0.57, 95% CI 0.55 – 0.59 per point increase in GCS for pre-hospital intubation, and OR 0.64, 95% CI 0.62 – 0.67 for in-hospital intubation). The model with GCS only already had a good fit on the data, the Nagelkerke R^2 was 60%. Pre-hospital unreactive pupil(s) increased the odds of pre-hospital intubation (OR 3.0, 95% CI 1.5 – 6.0), but not for in-hospital intubation (OR 1.0, 95% CI 0.5 – 2.0). Higher AIS increased the odds for intubation, the strongest predictors of the AIS were thorax and chest AIS (OR 1.5, 95% CI 1.4 – 1.6 per point increase for pre-hospital intubation, and OR 1.3, 95% CI 1.1 – 1.4 for in-hospital intubation) and face AIS (OR 1.3, 95% CI 1.2 – 1.5 per point increase for pre-hospital intubation, and OR 1.3, 95% CI 1.2 – 1.4 for in-hospital intubation). Finally, age lowered the odds of pre-hospital intubation (OR 0.98, 95% CI 0.98 – 0.99 per decade), but not of in-hospital intubation (OR 0.99, 95% CI 0.99 – 1.00). These predictors, other than GCS, increased the fit of the model to 64% (Table S1 available online). A complete case analysis of this model showed the same magnitude and direction of the associations (Table S2 available online). Similarly, a sensitivity analysis not including patients whose tracheas were intubated in a referring hospital showed the same magnitude and directions of the associations (Table S3 available online).

The fit of the model increased to 71% with the inclusion of country and centre, indicating substantial practice variation. The median odds ratio between two randomly chosen centres was 3.1 (95% CI 2.1–4.3) for pre-hospital intubation and 2.7 (95% CI 1.9–3.5) for in-hospital intubation (Table S1 available online). The predicted probability for an average patient to undergo pre-hospital intubation was highest in the south and west of Europe, and the probability of undergoing in-hospital intubation was higher in northern Europe (figure 3.4).

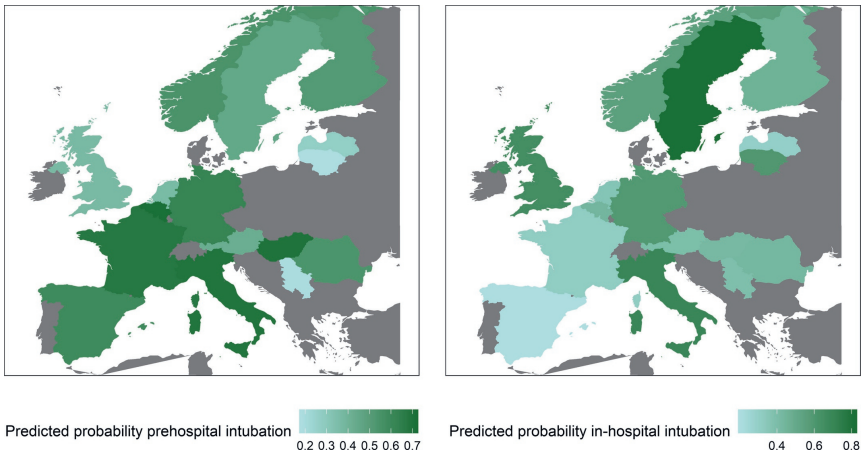


Figure 3.4: *Intubation practice variation. The left panel shows the predicted probabilities of pre-hospital tracheal intubation for the average patient in each country, and the right panel shows the same result for in-hospital intubation.*

The variation attributable to centre was partly explained by trauma system characteristics. In particular, trauma system characteristics were strongly associated with pre-hospital intubation: the odds of pre-hospital intubation were larger (OR 2.9, 95% CI 1.3 – 6.6) when the physician on the pre-hospital care team was an anaesthetist, smaller (OR 0.5, 95% CI 0.3 – 0.8) when the ambulance personnel were allowed to intubate without drugs and smaller still (OR 0.1, 95% CI 0.0 – 0.4) when the main policy was scoop-and-run, instead of stay-and-play (figure 3.5 and Table S1 available online).

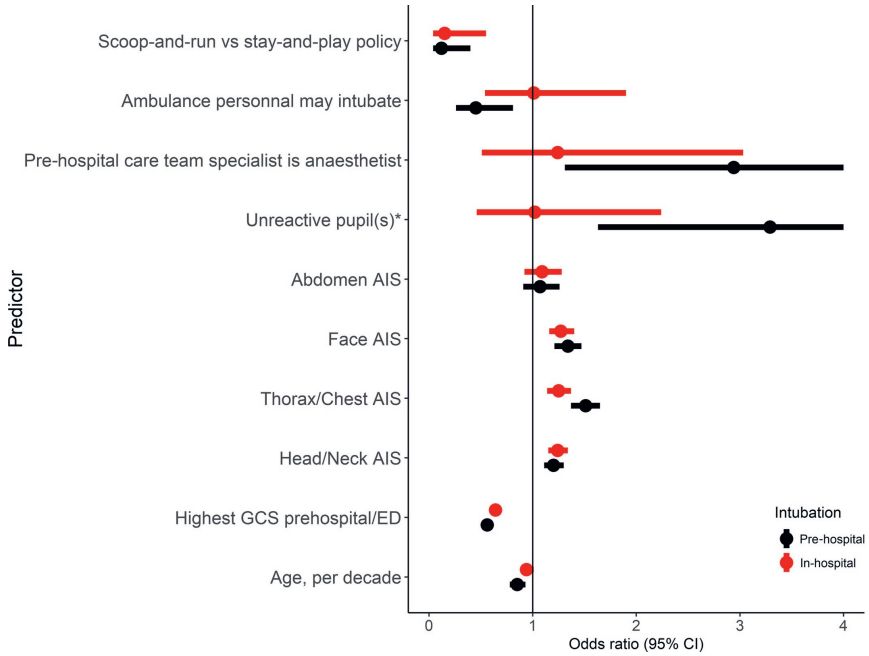


Figure 3.5: *The adjusted effect of the individual predictors on intubation. The results of the full model, including random intercept for centre conditional on country is presented. * pre-hospital assessment. AIS, abbreviated injury score, GCS, Glasgow coma score, ED, emergency department.*

3.5 | Discussion

This study provides insights into current intubation practice for TBI patients in Europe. We found that the main driver of intubation was the GCS. However, other patient and trauma characteristics were also important regarding the decision to intubate, such as unreactive pupils, face injury and thorax and chest injury. In addition, this study describes significant variations in tracheal intubation practice between centres and countries in Eu-

rope, the effect of centre on the odds of intubation was similar to the effect of unreactive pupils. This large variation could be partially explained by trauma system characteristics.

The finding that other patient characteristics besides GCS played a role in the decision to intubate contrasts with current guidelines. Currently, international guidelines include only GCS as an objective clinical parameter with a specific threshold for intubation [11]. Therefore, it is a self-fulfilling prophecy that this patient characteristic should explain the majority of the variation. However, the substantial added effects of additional clinical parameters in our models indicates that, in practice, the decision is also based on other factors. An illustrative example is the absence of pupillary reflexes, which indicate compromised brainstem function and therefore potentially jeopardised airway reflexes. Another example is the severity of facial injury, which could be suggestive of airway obstruction. Further research should focus on whether they could be included as indications for intubation.

Another finding of this study is the large regional variations in frequency of tracheal intubation. We found that these differences might be caused by regional differences in the composition of pre-hospital care teams, and their experience of intubation.

In particular, this study found that the availability of pre-hospital personnel who are skilled in pre-hospital intubation without drugs, actually lowered the chances of intubation. In the CENTER-TBI study, 43% of the centres indicated that personnel on the ambulance were able to intubate on scene without drugs [28]. These trauma systems might consist of more intensively trained ambulance personnel, often operating without the assistance of a physician. However, since they are not allowed to perform tracheal intubation with drugs, they can only do so on moribund patients

(GCS of 3). Since the majority of moderate to severe TBI patients still have (partially) intact motor GCS responses in the pre-hospital setting [35], they would not be eligible for intubation in these trauma systems, explaining the lower overall intubation rates.

On the other end of the spectrum, we found that involvement of anaesthetists, with extensive training in intubation, increased the probability of intubation. Experience decreases the risk of harmful intubation, especially in non-elective settings [36, 37]. However, it is undesirable that the indication for intubation is the presence of specific professionals, instead of patient and trauma characteristics.

Our study confirms that, in the TBI field, paucity of evidence often results in low adherence to guidelines [21]. This was confirmed by observing large variations between countries and centres. Since this variation was corrected for patient and trauma characteristics, it is more likely the result of guidelines based on low-quality evidence. In general, it is uncertain if not adhering to guidelines with low quality of evidence represents bad clinical practice. This is similar as the effectiveness of parachutes in preventing death after jumping from an airplane: the absence of evidence does not imply that current practice is problematic [38]. For intubation, however, it has been suggested that low adherence rates with guidelines do affect the outcome of patients [22].

The variation in intubation practice does offer an eloquent solution, since it enables us to identify best clinical practice by comparing regions [24, 39]. This will possibly improve the evidence base regarding intubation, and eventually improve adherence. Moreover, more personalised identification of TBI patients requiring tracheal intubation could be investigated using this method.

Missing data, especially from the pre-hospital scene, was a substan-

tial problem in our study. We dealt with this by focusing on the well-documented factors, and otherwise using multiple imputation, a method proven to give valid estimates under the missing-at-random assumption [33]. It is in the nature of this logistically challenging study that non-observation of data can probably be attributed to random non-administration of data. This mechanism at least does not result in a missing not-at-random pattern. Since we found substantial correlation between variables and sufficient observed auxiliary variables, imputation is likely to be successful. Additionally, it is reassuring that the complete case analysis of the main model showed similar magnitude and direction of the coefficients.

Furthermore, there may have been unmeasured policy characteristics that explain variations in the incidence of tracheal intubation. Even though the thorough development of the questionnaires attempted to ensure the completeness of the topics, they still lacked some specific questions of interest for this analysis. For example, we were not able to assess the following: whether the physician was in favour of intubation when neurological deterioration was anticipated (based on clinical insight); whether the physician was in favour of intubation in patients with mild TBI, or in cases of mild TBI; or whether intubation occurred to facilitate safe treatment and transfer after TBI in cases of severe agitation, even though the airway may have been uncompromised.

Finally, not all data which we would have wanted for this analysis were registered in the CENTER-TBI database. First of all, it was not possible to distinguish whether patients had their tracheas intubated using rapid sequence induction (RSI) of anaesthesia, or without drugs. Since RSI was the preferred method for intubation in trauma patients who were not moribund, patients who underwent RSI are likely to be different from patients who underwent tracheal intubation without drugs. By not distinguishing

between the two, we might have missed some subtle differences in variation. Secondly, we did not document the pre-hospital respiratory rate and oxygen saturation. These are likely to have influenced the decision to intubate, and therefore could have been included as a predictor in the models. Future studies should focus specifically on these aspects to provide additional insights.

However, our study was based on a large sample size and with few exclusion criteria in the analysis. This suggests a high degree of generalisability of our findings.

Although the GCS is the main driver of tracheal intubation, other patient and trauma characteristics, such as injury severity and neurological impairment, play a role in the decision as well. Furthermore, unexplained differences are substantial between countries and between centres. It remains unclear which patients benefit most from tracheal intubation, and further studies are needed to improve the evidence base in TBI patients.

3.6 | Supplementary Material

Supplementary materials can be found in the online version of this article: `https://tinyurl.com/bwn3rcfs`

References

1. Maas, A. I. R. *et al.* Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. *The Lancet Neurology* **4422.** ISSN: 14744422 (2017).

2. Majdan, M. *et al.* Epidemiology of traumatic brain injuries in Europe: a cross-sectional analysis. *The Lancet Public Health* **1,** e76–e83. ISSN: 2468-2667 (2016).

3. Advanced trauma life support (ATLS): The ninth edition (2013).

4. Green, R. S., Butler, M. B. & Erdogan, M. Increased mortality in trauma patients who develop postintubation hypotension. *Journal of Trauma and Acute Care Surgery* **83,** 569–574 (2017).

5. Manley, G. *et al.* Hypotension, Hypoxia, and Head Injury. *JAMA surgery* **136,** 1118–1123 (2001).

6. Boidin, M. P. Airway patency in the unconscious patient. *British journal of anaesthesia* **57,** 306–310. ISSN: 0007-0912 (Print) (1985).

7. Moulton, C, Pennycook, A & Makower, R. Relation between Glasgow coma scale and the gag reflex. *British Medical Journal* **303,** 1240–1241. ISSN: 09598146 (1991).

8. Lockey, D. J., Coats, T. & Parr, M. J. Aspiration in severe trauma: A prospective study. *Anaesthesia* **54,** 1097–1098. ISSN: 00032409 (1999).

9. Atkinson, J. L. The neglected prehospital phase of head injury: apnea and catecholamine surge. *Mayo Clinic proceedings* **75,** 37–47. ISSN: 0025-6196 (Print) (2000).

10. Carney, N. *et al.* Guidelines for the management of severe traumatic brain injury. *Neurosurgery* **80,** 6–15 (2017).

11. Badjatia, N. *et al.* Guidelines for Prehospital Management of Traumatic Brain Injury 2nd Edition. *Prehospital Emergency Care* **12,** S1–S52. ISSN: 1090-3127 (2008).

12. Shafi, S. & Gentilello, L. Pre-hospital endotracheal intubation and positive pressure ventilation is associated with hypotension and decreased survival in hypovolemic trauma patients: an analysis of the National Trauma Data Bank. *Journal of Trauma and Acute Care Surgery* **59,** 1140–1147 (2005).

13. Burney, R. G. & Winn, R. Increased cerbrospinal fluid pressure during laryngoscopy and intubation for induction of anesthesia. *Anesthesia and analgesia* **54,** 687–90. ISSN: 0003-2999 (1975).

14. Davis, D. *et al.* The Impact of Hypoxia and Hyperventilation on Outcome after Paramedic Rapid Sequence Intubation of Severely Head-injured Patients. *The Journal of Trauma: Injury, Infection, and Critical Care* **57,** 1–10. ISSN: 0022-5282 (2004).

15. Davis, D. P. *et al.* Both hypoxemia and extreme hyperoxemia may be detrimental in patients with severe traumatic brain injury. *Journal of neurotrauma* **26,** 2217–2223. ISSN: 0897-7151 (2009).

16. Marmarou, A. *et al.* IMPACT Database of Traumatic Brain Injury: Design And Description. *Journal of Neurotrauma* **24,** 239–250. ISSN: 0897-7151 (2007).

17. Stocchetti, N. & Maas, A. I. Traumatic Intracranial Hypertension. *New England Journal of Medicine* **370,** 2121–2130. ISSN: 0028-4793 (2014).

18. Dinsmore, J. Traumatic brain injury: an evidence-based review of management. *Continuing Education in Anaesthesia, Critical Care & Pain* **13,** 189–195 (2013).

19. Hoffmann, M. *et al.* The impact of prehospital intubation with and without sedation on outcome in trauma patients with a GCS of 8 or less. *Journal of neurosurgical anesthesiology* **29,** 161–167 (2017).

20. Bernard, S. A. *et al.* Prehospital Rapid Sequence Intubation Improves Functional Outcome for Patients With Severe Traumatic Brain Injury. *Annals of Surgery* **252,** 959–965. ISSN: 0003-4932 (2010).

21. Cnossen, M. C. *et al.* Adherence to Guidelines in Adult Patients with Traumatic Brain Injury: A Living Systematic Review. *Journal of Neurotrauma* **14.** ISSN: 0897-7151 (2016).

22. Franschman, G *et al.* Prehospital endotracheal intubation in patients with severe traumatic brain injury: Guidelines versus reality. *Resuscitation* **80,** 1147–1151. ISSN: 0300-9572 (2009).

23. Von Elm, E. *et al.* Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *British Medical Journal* **335,** 806–8. ISSN: 1756-1833 (2007).

24. Maas, A. I. *et al.* Collaborative European neurotrauma effectiveness research in traumatic brain injury (CENTER-TBI): A prospective longitudinal observational study. *Neurosurgery* **76,** 67–80. ISSN: 15244040 (2015).

25. Steyerberg, E. W. *et al.* Case-mix, care pathways, and outcomes in patients with traumatic brain injury in CENTER-TBI: a European prospective, multicentre, longitudinal, cohort study. *The Lancet Neurology* **18,** 923–934 (2019).

26. Cnossen, M. C. *et al.* Variation in structure and process of care in traumatic brain injury: Provider profiles of European Neurotrauma Centers participating in the CENTER-TBI study. *PLoS ONE* **11,** 1–21. ISSN: 19326203 (2016).

27. Foks, K. A. *et al.* Management of Mild Traumatic Brain Injury at the Emergency Department and Hospital Admission in Europe: A Survey of 71 Neurotrauma Centers Participating in the CENTER-TBI Study. *Journal of Neurotrauma* **34,** 2529–2535. ISSN: 0897-7151 (2017).

28. Cnossen, M. C. *et al.* Prehospital trauma care among 68 European neurotrauma centers: Results of the CENTER-TBI Provider Profiling Questionnaires. *Journal of neurotrauma* **36,** 176–181 (2019).

29. Van Buuren, S. *Flexible imputation of missing data* 415. ISBN: 9781138588318 (CRC Press, 2018).

30. Von Hippel, P. T. Regression with Missing Ys: An Improved Strategy for Analyzing Multiply Imputed Data. *Sociological Methodology* **37,** 83–117. ISSN: 0081-1750 (2007).

31. Nagelkerke, N. J. D. *A Note on a General Definition of the Coefficient of Determination* tech. rep. 3 (1991), 691–692.

32. Merlo, J. *et al.* A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *Journal of Epidemiol Community Health* **60,** 290–297 (2006).

33. Little, R. J. A. & Rubin, D. B. Statistical Analysis with Missing Data. *Journal of Educational Statistics* **16,** 150. ISSN: 03629791 (1991).

34. Lesaffre, E., Steyerberg, E. W., Lingsma, H. F. & Li, B. Logistic random effects regression models: a comparison of statistical packages for binary and ordinal outcomes (2011).

35. Mauritz, W. *et al.* Glasgow Coma Scale Motor Score and Pupillary Reaction To Predict Six-Month Mortality in Patients with Traumatic Brain Injury: Comparison of Field and Admission Assessment. *Journal of Neurotrauma.* ISSN: 0897-7151 (2014).

36. Bossers, S. M. *et al.* Experience in Prehospital Endotracheal Intubation Significantly Influences Mortality of Patients with Severe Traumatic Brain Injury: A Systematic Review and Meta-Analysis. *PLoS One* **10,** e0141034. ISSN: 1932-6203 (2015).

37. Buis, M. L., Maissan, I. M., Hoeks, S. E., Klimek, M. & Stolker, R. J. Defining the learning curve for endotracheal intubation using direct laryngoscopy: A systematic review. *Resuscitation* **99,** 63–71 (2016).

38. Smith, G. C. S. & Pell, J. P. Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. *British Medical Journal* **327,** 1459 (2003).

39. Of Medicine, I. *Initial National Priorities for Comparative Effectiveness Research* ISBN: 978-0-309-13836-9 (National Academies Press, Washington, D.C., 2009).

4

One-year survival after in-hospital cardiac arrest: A systematic review and meta-analysis

Marc Schluep, Benjamin Y Gravesteijn, Robert Jan Stolker, Henrik Endeman, Sanne E Hoeks

Resuscitation

4.1 | Abstract

4.1.1 | Introduction

In-hospital cardiac arrest is a major adverse event with an incidence of 1–6/1000 admissions. It has been poorly researched and data on survival is limited. The outcome of interest in IHCA research is predominantly survival to discharge, however recent guidelines warrant for more long-term outcomes. In this systematic review we sought to quantitatively summarize one-year survival after in-hospital cardiac arrest.

4.1.2 | Methods

For this systematic review and meta-analysis we performed a systematic search of all published data on one-year survival after IHCA up to March 9th, 2018. Results of the meta-analyses are presented as pooled proportions with corresponding 95% prediction intervals (95% PI). Between-study heterogeneity was assessed using I^2 statistic and the DerSimonian–Laird estimator for τ^2. Subgroup analyses were performed for cardiac and non-cardiac patients.

4.1.3 | Results

We included 40 studies in our systematic review and meta-analysis. The pooled one-year survival after in-hospital cardiac arrest was 13.4% (95%PI: 5.6 – 28.8%, I^2=100%). In the subgroup of patients with a cardiac admission characteristic one-year survival was 39.3% (16.1% – 68.6%), whereas in the subgroup of non-cardiac patients one-year survival was 10.7% (4.4% –

23.6%). These data cover the period 1985 - 2018 and show a modest change in survival over that period (10-year OR: 1.70, 95% CI: 1.04 – 2.76).

4.1.4 | Conclusion

One-year survival after in-hospital cardiac arrest is poor. Survival is higher in patients admitted to cardiac wards. The time trend between 1985–2018 has shown a modest improvement in one-year survival rates. Research into IHCA population characteristics might elicit the issue of heterogeneity and stagnated survival over the past decades.

4.2 | Introduction

Cardiac arrest, cardiopulmonary arrest, or circulatory arrest is the loss of mechanical heart function and effective blood circulation. If not treated by cardiopulmonary resuscitation (CPR) it inevitably means the end of life. However, if treated, circulation can be restored. Cardiac arrest is usually divided into two categories: out-of-hospital cardiac arrest (OHCA) and in-hospital cardiac arrest (IHCA). The latter is poorly researched; data on incidence and survival of IHCA are limited. Current literature describes an incidence of 1–6 events per 1000 hospital admissions [1–4].

The outcome of interest in IHCA research is predominantly survival to discharge. A recent meta-analysis shows a pooled survival rate at discharge of 15.0% (95% CI, 12.0 – 18.0%) with little change over time [5], while an analysis in the UK over the same period of time shows a significant increase in hospital survival after IHCA (9.0% in 2004 to 12.2% in 2014) [6]. Survival to discharge is an important outcome measure, however little is known about the long-term outcomes of patients discharged from the hospital. Recent guidelines warrant for more research into long-term outcomes and associated factors [7]. As patient-centred outcomes are increasingly important to biomedical and clinical research, long-term survival could be regarded as such and could serve as important information in clinical decision-making. This systematic review aims to quantitatively summarise one-year survival after in-hospital cardiac arrest.

4.3 | Methods

4.3.1 | Search strategy and study selection

This systematic review and meta-analysis was reported following the PRISMA and MOOSE guidelines for reporting of systematic reviews and meta-analyses of observational studies [8, 9]. The protocol was registered with PROSPERO[1]. We performed a systematic search of published data on one-year survival of IHCA using Embase, Medline Ovid, Cochrane Central, Web of Science, PubMed recent and Google scholar from their inception through March 9th, 2018. The search strategy is shown in supplemental Table 1. We set no limitations on type of study or language. Mendeley (2017 Mendeley Ltd.) was used for the selection of relevant articles. Study selection was performed in a 2-staged process. Two authors (MS and BG) independently screened titles and abstracts (stage 1), and full-text papers for inclusion (stage 2). Disagreements were resolved with discussion and involvement of a third author (SH). Pre-defined inclusion criteria were: 1) In-hospital cardiac arrest, using conventional CPR (CCPR); 2) One year survival reported; 3) Adult patients; 4) Clinical study. Cardiac arrest definitions per article are provided in supplemental Table 2 available online. Conventional CPR is defined as chest compressions with or without use of compression devices, as opposed to extracorporeal CPR via cardiopulmonary bypass. Studies were excluded if they did not fit inclusion criteria, if they were only published as abstract or written in a language none of the reviewers was proficient in.

[1] 2017:CRD42017072037

4.3.2 | Data extraction and quality assessment

Data extraction from selected studies was performed independently by two investigators (MS and BG) using a standardized form. To describe the study design, we extracted the sample size of patients who underwent CCPR, the country of origin, the investigated period, the definition of the study population, whether the study was retrospective or prospective, how the investigators attained their data, which comparisons were made, how they defined one year survival and which patients were excluded from the cohort. Patient populations were checked for overlap to prevent patients from appearing multiple times in our analysis. If this was the case the study with the smallest sample size was excluded. The characteristics of the study population included were: age, gender, prevalence of cardiac patients, percentage of witnessed arrests or monitored patients and prevalence of ventricle fibrillation or ventricle tachycardia as initial rhythm. A common denominator for comorbidity or severity of disease was sought. If age was defined in strata or ranges a weighed mean was calculated without standard deviation. Finally, one-year survival of patients who underwent CCPR in hospital was extracted. Survival was defined as the survival of one single CPR attempt. Authors were contacted for the exact survival rate when the one-year survival was not directly available from the manuscript. We specifically looked at conventional CPR, and excluded extracorporeal CPR. When a study included both, only the conventional CPR group was extracted.

The quality of the studies was evaluated using the method of Hayden et al. for the evaluation of the quality of prognosis studies in systematic reviews [10]. Known prognostic factors such as initial rhythm and witnessed arrest were assessed. Two authors individually assessed all six items and

discrepancies were resolved by a third researcher (SH).

4.3.3 | Statistical analysis

One-year survival data were pooled across studies using the inverse variance method. A random-effects model was used to estimate the pooled one-year survival probability after IHCA as considerable heterogeneity was expected. A random-effects meta-analysis model assumes the observed estimates can vary across studies because of real differences in each study as well as sampling variability (chance). Results of the meta-analyses are presented as pooled proportions with corresponding 95% confidence intervals (CI). Between-study heterogeneity was assessed using I^2 statistic and the DerSimonian–Laird estimator for τ^2. Furthermore in order to address heterogeneity between studies better, a 95% prediction interval was reported [11, 12].

A sensitivity analysis was performed to assess the presence or absence of heterogeneity. Subgroup analyses were performed for cardiac and other patients. Cardiac, or a cardiac admission characteristic, was defined as a study in which all patients came from cardio (-thoracic) units, or were predominantly admitted to the hospital for cardiac disease or cardiac surgery. The non-cardiac subgroup consisted of studies that included patients who were not specifically admitted for cardiologic or cardiac surgical reasons (i.e. general nursing wards, but also critical care units). Other subgroup analyses were done for study quality, geographical distribution (i.e. continents) and initial arrest rhythm. Furthermore, a random intercept meta-regression analysis (binomial-normal model) with corresponding bubble plot was carried out to assess the influence of study period on one-year survival. This model is appropriate for probability meta regression, since

it avoids the bias that occur when a normal-normal model would be used for logit transformed proportion [13, 14]. Studies were allocated in time using the median of the period the study covered. After careful evaluation of all articles a post-hoc analysis of cognitive outcome was done with use of a random effects model to analyse available data on the fraction of one-year survivors with a cerebral performance category score (CPC) of 1 or 2. Secondly a post-hoc analysis was performed for survival to discharge.

All data was extracted into Microsoft Excel and then statistically analysed by importing the data in R[2]. The packages used for the analysis were 'meta' and 'metafor', of which we used the 'metaprop',' forrest' and 'rma.glmm functions.

4.4 | Results

4.4.1 | Search results and characteristics of included studies

Our search strategy retrieved 7331 records, of which 4999 remained after duplicates were removed. The parallel exclusion of studies based on title and abstract resulted in 239 full text articles eligible for detailed assessment. Finally, we included 39 studies in our systematic review and meta-analysis [15–54]. Full details of study selection are summarised in figure 4.1.

Characteristics of the included studies and study populations are given in table 4.1 and table 4.2. All studies were performed between 1985 and

[2]R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

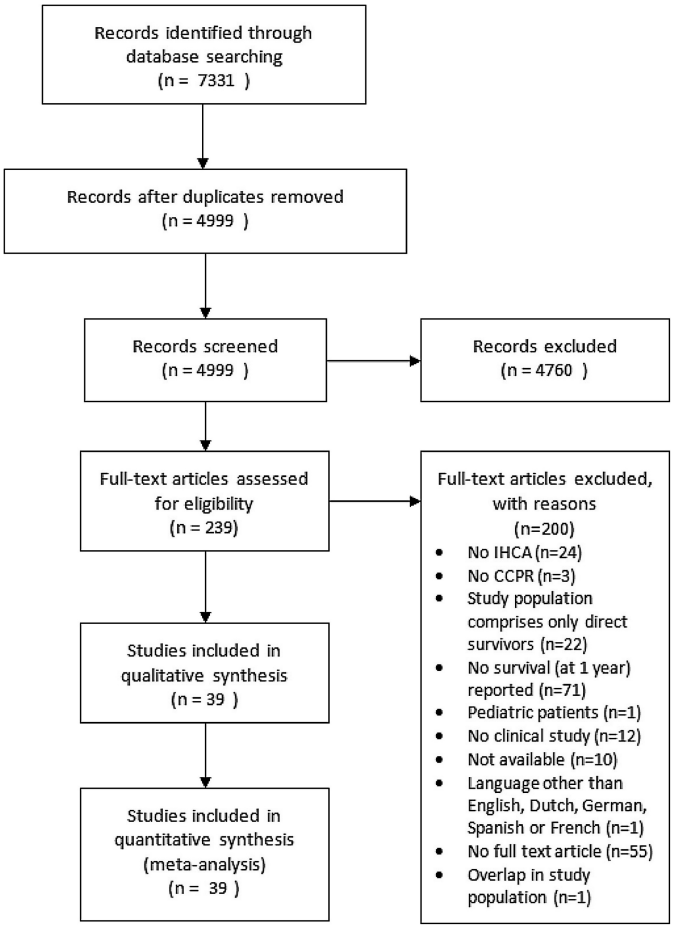


Figure 4.1: *PRISMA Flow Diagram of search strategy and included studies.*

2015, predominantly in North America and Western Europe. Data was available on age in 35 (89.7%) studies, on gender in 33 (84.6%), on the proportion of cardiac patients in 14 (35.9%) studies and on shockable rhythm in 27 (69.2%) of the included studies. Of the included studies 18 (46.1%)

described level of patient monitoring at the time of arrest (e.g. critical care units). Number of inclusions ranged from 25 to 471,962 patients and mean age of the study population ranged from 54 to 86 years.

Table 4.1: *General characteristics of included studies (n=39). Study design: PC=prospective cohort, RC=retrospective cohort.*

Author	Year	N	Country	Design	Excluded	Outcome
Al-Alwan [15]	2014	471962	USA	RC	Patients who were just intubated	1 year survival post-CA
Berger [16]	1994	255	USA	PC	–	1 year survival post-discharge
Beuret [27]	1992	181	Switzerland	RC	Respiratory arrests, seizures	1 year survival post-CA
Bloom [38]	2007	732	USA	RC	–	1 year survival post-CA
Blumenstein [49]	2016	272	Germany	RC	–	1 year survival post-CA
Chen [50]	2014	5151	Australia	RC	–	1 year survival from discharge
Colmenero [51]	2004	89	Spain	PC	Perioperative cardiac arrests	1 year survival post-CA
Dimopoulou [52]	2001	29	Greece	PC	IABP, massive bleeding <2 h post-op	1 year survival from discharge
Ezquerra [53]	2009	90	Spain	RC	DNAR	1 year survival post-CA
Feingold [54]	2016	1262	USA	PC	–	1 year survival from discharge
Fredriksson [17]	2006	833	Sweden	PC	–	1 year survival from discharge
Gomes [18]	2005	452	Brazil	PC	–	1 year survival post-CA
Heller [19]	1995	308	Australia	PC	–	1 year survival post-CA
Herlitz [20]	2000	216	Sweden	PC	–	1 year survival from discharge
Hessulf [46]	2018	18069	Sweden	RC	–	1 year survival post-CA
Huang [21]	2002	103	Taiwan	PC	<17 years	1 year survival from discharge
Joshi [22]	2015	260	India	PC	–	1 year survival post-CA
Karetzky [23]	1995	668	USA	RC	-	1 year survival post-CA
Kutsogiannis [24]	2011	517	Canada	PC	Secondary arrests	1 year survival post-CA
Lees [25]	2012	99	UK	PC	–	1 year survival post-CA
Lin [26]	2010	63	Taiwan	PC	CPR <10 min, non-witnessed and no ROSC	1 year survival post-CA
Menon [28]	2014	413403	USA	RC	<65 years	1 year survival post-CA
Möhnle [29]	2012	189	Germany	RC	–	1 year survival post-CA
Moretti [30]	2007	156	Brazil	PC	<20 years, futile CPR	1 year survival post-CA
O'Sullivan [31]	2016	63	Ireland	RC	DNAR order	1 year survival post-CA
Paniagua [32]	2002	474	USA	RC	<80 years	1 year survival post-discharge
Rankin [33]	1998	133	New-Zealand	PC	–	1 year survival post-CA
Rudiger [34]	2004	25	Switzerland	PC	ICU patients	1 year survival post-CA
Saklayen [35]	1995	340	USA	RC	–	1 year survival post-CA
Shin [36]	2013	321	Korea	RC	Patients with bad prognosis	1 year survival post-CA
Skrifvars [37]	2003	204	Finland	PC	–	1 year survival post-CA
Skrifvars [39]	2005	183	Finland	PC	–	1 year survival post-CA
Stapleton [40]	2014	358682	USA	RC	–	1 year survival post-CA
Thompson [48]	2018	45567	USA	PC	Patients without documented initial rhythm	1 year survival post-CA
Tunstall-Pedoe [41]	1992	2838	UK	PC	False alarms, recurrences within 24 h	1 year survival post-CA
Vakil [42]	2016	182	USA	RC	<18 years	1 year survival post-CA
Varon [43]	1998	83	USA	RC	Respiratory arrests, patients in shock	1 year survival from discharge
Wong [44]	2015	33731	USA	RC	–	1 year survival from discharge
Yi [45]	2006	214	South-Korea	RC	Non-neurosurgical patients	1 year survival post-CA

Table 4.2: *Patient characteristics of included studies (n = 39). * = Intubated versus non-intubated; ** = Mean (range); *** = Median with/without IQR; † = With versus without cardiac life support training groups (the survival is the overall survival).*

First author	Mean age (±SD)	% male	% cardiac patients	% monitored/witnessed	% VF/VT	% CPC 1 or 2 at 1 year
Al-Alwan* [15]	73.3 (±11.9) vs 75.0 (±11.4)	52.3 vs 51.6	N/A	N/A	N/A	N/A
Berger [16]	67.4	N/A	N/A	N/A	25.0	N/A
Beuret [27]	61.5 (17.0-89.0)**	69.0	N/A	34.0	39.0	N/A
Bloom [38]	59.0	N/A	N/A	N/A	N/A	N/A
Blumenstein [49]	75.3 (67.4 – 79.1)***	61.4	100	100	2.9	N/A
Chen [50]	68.2 (±16.9)	61.2	N/A	N/A	N/A	N/A
Colmenero [51]	68.0 (56-74.5)**	57.3	N/A	N/A	34.8	100
Dimopoulou [52]	61.0 (±11.0)	87.5	100	N/A	44.0	N/A
Ezquerra [53]	73.1 (±12.3)	68.9	N/A	N/A	22.2	93.0
Feingold [54]	61.1 (±14.3)	50.8	N/A	N/A	N/A	N/A
Fredriksson [17]	69.4	63.0	66.0	N/A	48.6	N/A
Gomes [18]	54.1	54.9	N/A	76.8	39.0	N/A
Heller [19]	60.4	63.0	N/A	N/A	N/A	N/A
Herlitz [20]	68.0***	62.0	N/A	N/A	N/A	95.0
Hessulf [46]	75***	71.0	29.0	50.0	32.0	N/A
Huang [21]	66.8	71.0	17.0	N/A	14.0	N/A
Joshi [22]	N/A	N/A	31.2	91.0	21.9	96.0
Karetzky [23]	59.2	48.2	N/A	65.7	15.7	N/A
Kutsogiannis [24]	66.5 (±14.9)	62.3	60.6	100	33.7	N/A
Lees [25]	N/A	N/A	100	100	26.8	N/A
Lin [26]	60.6 (±12.7)	65.1	47.6	N/A	41.3	91.0
Menon [28]	78.3 vs 77.4	50.5 vs 50.7	N/A	N/A	N/A	N/A
Möhnle [29]	65.2 (±16.1)	69.8	N/A	21.7	32.3	N/A
Morettì† [30]	64.4 (±17.2) vs 63.6 (±15.8)	58.6 vs 55.2	N/A	90.3 vs 74.6	32.7 vs 22.1	N/A
O'Sullivan [31]	74.3***	63.4	44.4	87.3	30.2	81.0
Paniagua Paniagua [32]	86.0 (±4.8)	42.0	N/A	N/A	N/A	N/A
Rankin [33]	N/A	N/A	N/A	47.4	32.3	100
Rudiger [34]	72.8	72.0	N/A	N/A	28.0	N/A
Saklayen [35]	66.9	N/A	N/A	57.0	18.0	N/A
Shin [36]	61.6 (±14.2)	62.6	49.5	100	22.7	N/A
Skrifvars [37]	68.0 (±15.8)	59.3	N/A	72.1	28.0	N/A
Skrifvars [39]	73 (64.0 – 78.0)**	60.0	N/A	75.4	33.3	N/A
Stapleton [40]	78.9 (±7.2)	50.3	N/A	N/A	N/A	N/A
Thompson [48]	77.2 (±7.4)	55.5	26.7	25.3	20.3	N/A
Tunstall-Pedoe [41]	N/A	64.2	N/A	N/A	N/A	N/A
Vakil [42]	68.0 (±8.0)	98.0	100	N/A	71.4	N/A
Varon [43]	56.2	49.3	N/A	N/A	N/A	N/A
Wong [55]	>65.0	53.9	16.7	N/A	N/A	N/A
Yi [45]	54.0 (±9.4)	65.5	19.2	100	29.0	N/A

4.4.2 | Quality assessment

The quality assessment of the included studies is given in supplemental table 3 available online. The study population was adequately defined and described in 26 (66.6%) studies. The study attrition, referring to the manner in which patients were recruited for inclusion, was of good quality in 28 (71.8%) studies. Prognostic factors were adequately measured in 21 (53.8%) studies. The means of outcome measurement were not or inadequately described in 16 (41.0%) studies, and were sufficiently described and measured in 12 (30.8%) studies.

4.4.3 | Outcome

The meta-analysis of all studies showed a pooled one-year survival of 13.4% (95% PI: 5.6% – 28.8%) summarized in figure 4.3. Statistical heterogeneity was high: I^2=100%, τ^2=0.22, p<0.01. Subgroup analysis of cardiac patients revealed a one-year survival of 39.3% (95% PI: 16.1% – 68.6%; I^2=85.0%), while repeating this analysis in studies of the non-cardiac subgroup analysis resulted in a one year survival of 10.7% (95% PI: 4.4% – 23.6%; I^2=100%) Survival plots for cardiac and non-cardiac patients are available in supplemental Fig. 1, Fig. 2 available online. As displayed in figure 4.4 survival to discharge was available in 35 studies. Pooled survival to discharge was 17.6% (95%PI: 13.1 – 22.7%, I^2=99%). All survival statistics are summarised in table 4.3.

Table 4.3: *Summary of outcomes from the performed meta-analyses. All survival rates are presented with a 95% prediction interval (95% PI). Non-cardiac was defined as studies not included in the cardiac subgroup analysis.*

Survival rates (%, 95% PI)	Survival to discharge	I^2, τ^2, p-value	One-year survival	I^2, τ^2, p-value
Overall	17.6 (13.1 – 22.7)	99%, 0.03, <0.01	13.4 (5.6 - 28.8)	100%, 0.22, <0.01
Cardiac	49.7 (3.8 - 96.2)	88%, 0.44, <0.01	39.3 (16.1 - 68.6)	85.0%, 0.16, <0.01
Non-cardiac	15.9 (12.0 - 20.7)	99%, 0.02, <0.01	10.7 (4.4 – 23.6)	100%, 0.21, <0.01

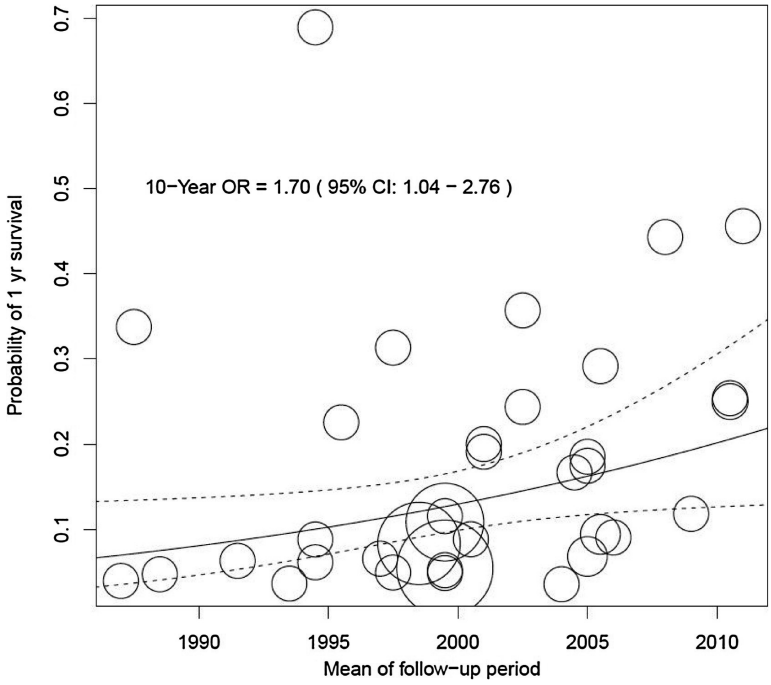


Figure 4.2: *Bubble-plot for meta-regression analysis of the influence of study period on one-year survival (OR=1.70, 95% CI: 1.04 - 2.76 per ten year increase).*

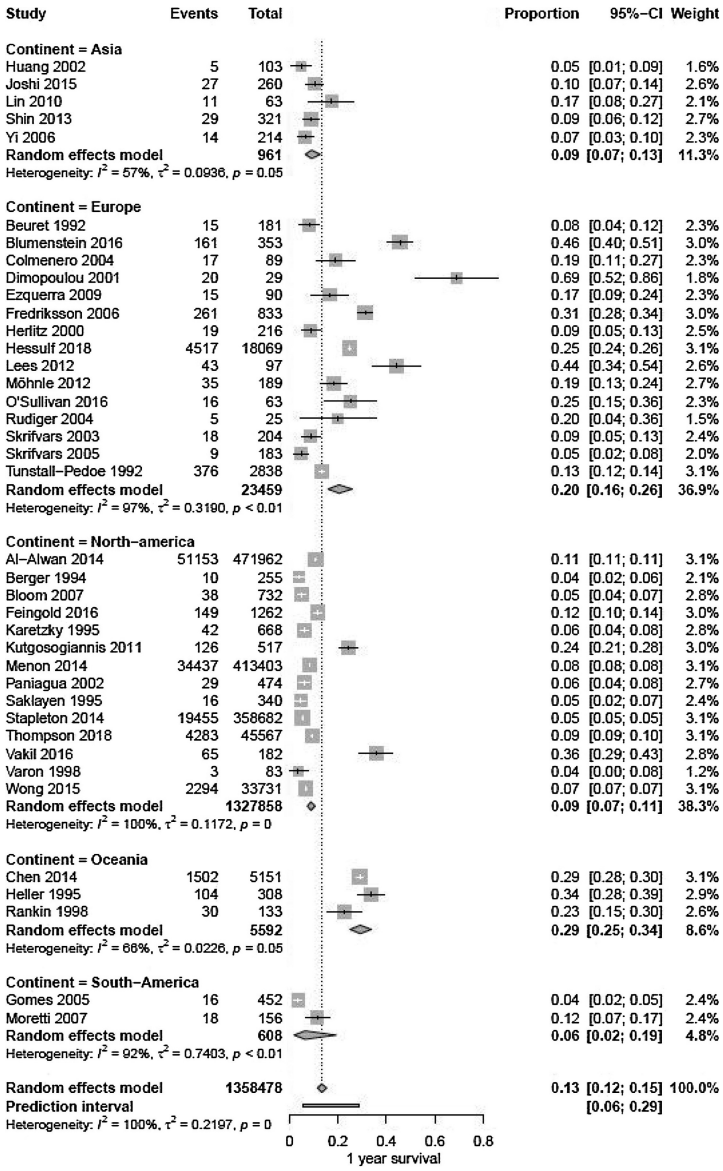


Figure 4.3: *Pooled one-year survival rate after in-hospital cardiac arrest.*

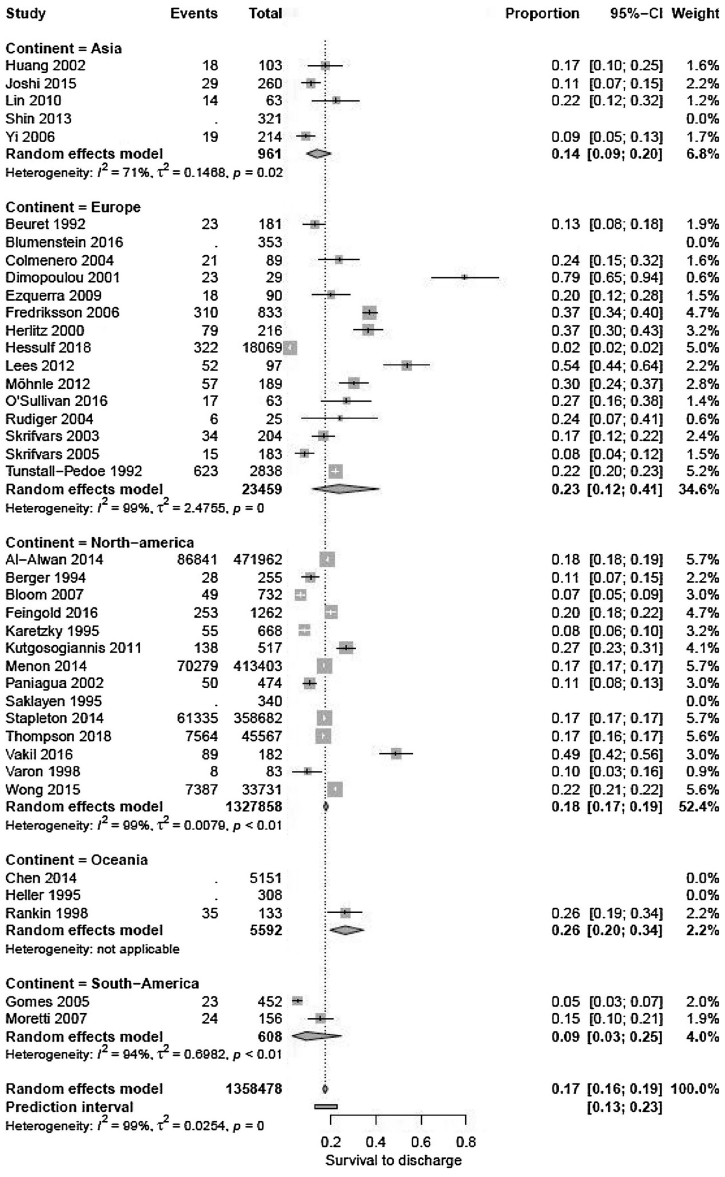


Figure 4.4: *Pooled survival to discharge rate after in-hospital cardiac arrest for studies that reported this outcome measure.*

Finally, when analysing the temporal trend of one year survival, a significant and modestly positive trend was observed (OR=1.70 per 10-year period, 95% CI: 1.04 - 2.76), as shown in figure 4.2. Seven studies reported CPC scores for one-year survivors. A pooled estimate shows 92.0% (95% CI: 85.0% - 96%) of patients alive at one year after cardiac arrest have a CPC score of 1 or 2 (figure 4.5). Pooled estimates stratified by study quality, geographical distribution or initial arrest rhythm did not produce any significant differences in effect estimates or heterogeneity. We were not able to identify a common denominator of co-morbidity or severity of disease to perform analyses on.

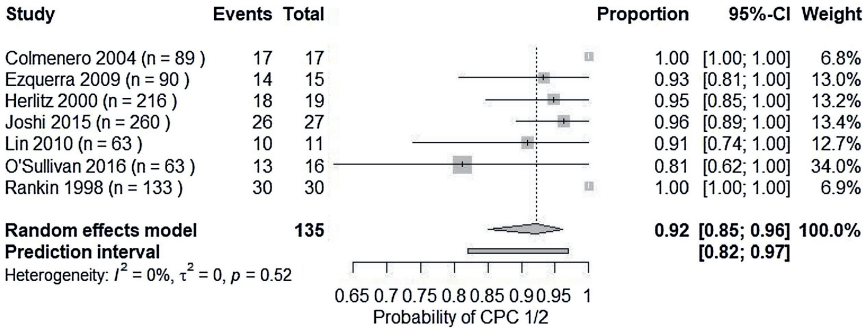


Figure 4.5: *Pooled fraction of survivors at 1 year with a cerebral performance category of 1 or 2.*

4.5 | Discussion

In this systematic review one-year survival after in-hospital cardiac arrest is 13.4% (95% PI: 5.6% - 28.8%). When viewed separately one-year survival in cardiac versus non-cardiac patients is 39.3% and 10.7% respectively. As far as we have found these data represent the first documentation of a sys-

tematic overview on one-year survival after IHCA through most recent publications and covers the period 1985 - 2018.

One-year survival of 13.4% after IHCA is poor. When compared to survival to discharge this implies a large portion of patients discharged alive survive the following year [5, 6]. The low survival rate is probably attributable to the presence of underlying disease. Comorbid disease has been demonstrated to worsen survival. This is most evident for severe COPD, cirrhotic liver disease, chronic kidney disease and heart failure and is supported by recent evidence that links co-morbidity and age to 30-day survival [56]. Although we did not have sufficient data for a subgroup analysis, some of the studies we have included suggest a similar relationship between co-morbidity and long-term survival [40, 56].

We found significant heterogeneity in outcomes across the studies. These differences may be related to the variability in study populations, treatment strategies and/or international differences in life expectancy [57]. With regard to differences in study population, subgroup analyses showed a survival of 39.3% in patients who are admitted to hospital for cardiac disease or cardiac surgery. In these patients survival is higher than for patients admitted for other reasons and part of the heterogeneity can be explained by this subgroup analysis. The higher survival rates are related to the presence of monitored wards, a higher incidence of shockable rhythm (also demonstrated in this review) and presumably a higher incidence of reversible causes (e.g. tamponnade, coronary occlusion) [58]. This supports the hypothesis of earlier recognition and intervention, as well as higher baseline survival in cardiac patients compared to other patients after cardiac arrest. To further explain heterogeneity we have performed several subgroup analyses with the available information, but did not find any sufficient answer.

The heterogeneity of data can to greater extent be attributed to the epidemiological nature of the populations, rather than being selected or randomised groups. We believe that pooling of data was reasonable for outcome measures for different reasons. First (I) this approach is pragmatic and clinically relevant; (II) we took measures to reduce potential clinical heterogeneity by performing subgroup analyses on the basis of clinical criteria (i.e. cardiac versus non-cardiac patients); (III) by contrast with comparative meta-analyses in which the presence of statistical heterogeneity might limit conclusions about effect size or exposure, pooling of data is an accepted method in single-group meta-analyses done for epidemiological purposes and (IV) pooling the data was necessary to appraise the available data on one-year survival in a comprehensive manner that could help inform the clinical context and related clinical decision making [59]. Although generalisability is limited due to a large diversity in study populations, pooling due of data provides a clinically relevant estimate for one-year survival after IHCA. In reporting survival rates we used the prediction interval, rather than the confidence interval. This provides an estimate of what survival rates can be expected in future studies. As to be expected with large heterogeneity in outcomes the prediction intervals we found were very broad and make prognostication difficult.

We compared one-year survival to survival to discharge from a recent meta-analysis (i.e. 15.0% 95% CI: 12.0% - 18.0%) and to survival to discharge from this meta-analysis (i.e. 17.6%, 95% CI 13.1% - 22.7%) [5]. It suggests death after IHCA occurs mainly during hospital admission rather than after discharge. Furthermore, when pooled survival for in-hospital cardiac arrest patients is compared to one-year survival after out-of-hospital cardiac arrest survival it is nearly identical: 13.4% for IHCA versus 12.0% for OHCA [55, 60]. These data give rise to new questions

regarding the aetiology of IHCA in non-cardiac patients and factors that influence survival. It could be argued that factors concerning pre-existing health status have added value in predicting one-year survival after in-hospital cardiac arrest. A positive finding came from our analysis for cognitive performance showed CPC scores were 1 or 2 in 92.0% (95% CI: 85.0% - 96.0%) of one-year survivors. This however pertains to performance and not necessarily to quality of life.

Certain limitations should be taken into account. Most studies have reported one-year survival from the moment of cardiac arrest, with a few reporting survival from the moment of discharge. We have considered this difference to be negligible to the interpretation of our outcome because survival is measured at least one year from the occurrence of cardiac arrest. Secondly we need to consider the heterogeneity of outcomes, as population-level data was not available for many of the included studies and therefore only stratification for cardiac and non-cardiac patients rather than for co-morbidity or age was possible. No difference could be analysed between monitored or non-monitored wards or initial arrest rhythms, as sufficient data was not available. Although some subgroup analyses were attempted no clear explanation for this heterogeneity could be pinpointed. Lastly health care developments and changes in public health will have influenced incidence and outcome of IHCA. The meta-regression we have performed shows a trend in one-year survival that shows a slight improvement when viewed on a basis of 10-year intervals. One could state that survival improves over time, however this trend is only modestly positive and we hope this effect will become more evident in the future. Whether patient case mix has significantly altered, treatment strategies are insufficient or it is a combination of factors remains uncertain.

In the future heterogeneity in structure and processes of care should

be explored. This variation in practice also adds to the heterogeneity in outcome. We do believe that careful assessment of quality of care should be performed, taking into account statistical uncertainty and case-mix. Being able to explain differences in outcome through quality of care could enable improving overall quality of care by identifying the most effective policy [61]. Secondly subgroup analyses can be performed if predefined subgroups are available. These subgroups need to be defined by known predictors and need to be comparable between studies [62]. We would recommend the implementation of nationwide registries and the use of standardised sets for reporting populations and outcomes, in this case the Utstein criteria and Core Outcome Set for Cardiac Arrest (COSCA) [6, 63, 64]. This will help improve comparability and enhance future implementation research [65].

This meta-analysis contains important information pertaining to all patients worldwide. In-hospital cardiac arrest is a global health issue, which concerns all patients and health care workers. Before making decisions about cardiopulmonary resuscitation and treatment restrictions, physicians must communicate accurate expectations of outcome to patients and families. However, one important caveat when reviewing these survival data is that its applicability to individual patients is limited. Although data on long-term outcome can inform patients on medical decisions about CPR, these data represent survival spread over a large population rather than predicting the trajectory for any individual patient. Overall we can conclude that one-year survival is poor in patients admitted to hospital for non-cardiac disease. Specific patient-level prognostication may probably require more knowledge about age, co-morbidity and intercurrent disease.

4.5.1 | Conclusion

In conclusion, our systematic review showed a one-year survival of 13.4% in IHCA patients. The time trend between 1985–2018 has shown a modest improvement in one-year survival rates. Future research is needed, specifically into the subject of prognostic factors for long-term qualitative outcome. Furthermore description of IHCA populations might elicit the issue of stagnated survival over the past decades. Moreover, more studies are published randomizing extracorporeal CPR versus conventional CPR, which in the future could be a more common method of resuscitation [66]. We feel multicentre prospective research in a known source population is needed to improve current knowledge on this subject.

4.6 | Supplementary Material

Supplementary materials can be found in the online version of this article: `https://tinyurl.com/teb6bbhm`

References

1. Skogvoll, E, Isern, E, Sangolt, G. K. & Gisvold, S. E. In-hospital cardiopulmonary resuscitation. 5 years' incidence and survival according to the Utstein template. *Acta anaesthesiologica Scandinavica* **43,** 177–84. ISSN: 0001-5172 (1999).
2. Hodgetts, T. J. *et al.* Incidence, location and reasons for avoidable in-hospital cardiac arrest in a district general hospital. *Resuscitation* **44,** 115–123 (2002).
3. Sandroni, C, Nolan, J, Cavallaro, F & Antonelli, M. In-hospital cardiac arrest: Incidence, prognosis and possible measures to improve survival. *Intensive Care Medicine* **33,** 237–245 (2007).

4. Fennessy, G., Hilton, A., Radford, S., Bellomo, R. & Jones, D. The epidemiology of in-hospital cardiac arrests in Australia and New Zealand. *Internal medicine journal* **46,** 1172–1181 (2016).

5. Zhu, A. & Zhang, J. Meta-analysis of outcomes of the 2005 and 2010 cardiopulmonary resuscitation guidelines for adults with in-hospital cardiac arrest. *The American journal of emergency medicine* **34,** 1133–1139 (2016).

6. Nolan, J. P. *et al.* Increasing survival after admission to UK critical care units following cardiopulmonary resuscitation. *Critical Care* **20,** 1–10 (2016).

7. Bossaert, L. L. *et al.* European Resuscitation Council Guidelines for Resuscitation 2015: Section 11. The ethics of resuscitation and end-of-life decisions. *Resuscitation* **95,** 302–311 (2015).

8. Stroup, D. F. *et al.* Meta-analysis of observational studies in epidemiology: a proposal for reporting. *Journal of the American Medical Association* **283,** 2008–2012 (2000).

9. Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G. & Group, P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS medicine* **6,** e1000097 (2009).

10. Hayden, J. A., Côté, P. & Bombardier, C. Evaluation of the quality of prognosis studies in systematic reviews. *Annals of internal medicine* **144,** 427–437 (2006).

11. IntHout, J., Ioannidis, J. P., Rovers, M. M. & Goeman, J. J. Plea for routinely presenting prediction intervals in meta-analysis. *British Medical Journal open* **6,** e010247 (2016).

12. Borenstein, M., Higgins, J. P., Hedges, L. V. & Rothstein, H. R. Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. *Research synthesis methods* **8,** 5–18 (2017).

13. Stijnen, T., Hamza, T. H. & Özdemir, P. Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Statistics in medicine* **29,** 3046–3067 (2010).

14. Zürcher, K. *et al.* Outcomes of HIV-positive patients lost to follow-up in African treatment programmes. *Tropical medicine & international health* **22,** 375–387 (2017).

15. Al-Alwan, A., Ehlenbach, W. J., Menon, P. R., Young, M. P. & Stapleton, R. D. Cardiopulmonary resuscitation among mechanically ventilated patients. *Intensive care medicine* **40,** 556–563 (2014).

16. Berger, R. & Kelley, M. Survival after in-hospital cardiopulmonary arrest of noncritically ill patients: a prospective study. *Chest* **106,** 872–879 (1994).

17. Fredriksson, M., Aune, S., Thorén, A.-B. & Herlitz, J. In-hospital cardiac arrest—an Utstein style report of seven years experience from the Sahlgrenska University Hospital. *Resuscitation* **68,** 351–358 (2006).

18. Gomes, A. *et al.* Prognostic factors of survival in post-cardiopulmonary-cerebral resuscitation in general hospital. *Arquivos brasileiros de cardiologia* **85,** 262–271 (2005).

19. Heller, R. F., Steele, P. L., Fisher, J. D., Alexander, H. M. & Dobson, A. J. Success of cardiopulmonary resuscitation after heart attack in hospital and outside hospital. *British Medical Journal* **311,** 1332–1336 (1995).

20. Herlitz, J., Andréasson, A.-C., Bång, A., Aune, S. & Lindqvist, J. Long-term prognosis among survivors after in-hospital cardiac arrest. *Resuscitation* **45,** 167–171 (2000).

21. Huang, C.-H. *et al.* Factors influencing the outcomes after in-hospital resuscitation in Taiwan. *Resuscitation* **53,** 265–270 (2002).

22. Joshi, M. A prospective study to determine the circumstances, incidence and outcome of cardiopulmonary resuscitation in a referral hospital in India, in relation to various factors. *Indian journal of anaesthesia* **59,** 31 (2015).

23. Karetzky, M., Zubair, M & Parikh, J. Cardiopulmonary resuscitation in intensive care unit and non—intensive care unit patients: immediate and long-term survival. *Archives of internal medicine* **155,** 1277–1280 (1995).

24. Kutsogiannis, D. J., Bagshaw, S. M., Laing, B. & Brindley, P. G. Predictors of survival after cardiac or respiratory arrest in critical care units. *Canadian Medical Association Journal* **183,** 1589–1595 (2011).

25. Lees, N. J., Powell, S. J. & Mackay, J. H. Six-year prospective audit of 'scoop and run'for chest-reopening after cardiac arrest in a cardiac surgical ward setting. *Interactive cardiovascular and thoracic surgery* **15,** 816–823 (2012).

26. Lin, J.-W. *et al.* Comparing the survival between extracorporeal rescue and conventional resuscitation in adult in-hospital cardiac arrests: propensity analysis of three-year data. *Resuscitation* **81,** 796–803 (2010).

27. Beuret, P. *et al.* Cardiac arrest: prognostic factors and outcome at one year. *Resuscitation* **25,** 171–179 (1993).

28. Menon, P. R., Ehlenbach, W. J., Ford, D. W. & Stapleton, R. D. Multiple in-hospital resuscitation efforts in the elderly. *Critical care medicine* **42** (2014).

29. Möhnle, P. *et al.* Survival after cardiac arrest and changing task profile of the cardiac arrest team in a tertiary care center. *The Scientific World Journal* **2012** (2012).

30. Moretti, M. A. *et al.* Advanced cardiac life support training improves long-term survival from in-hospital cardiac arrest. *Resuscitation* **72,** 458–465 (2007).

31. O'Sullivan, E & Deasy, C. In-hospital cardiac arrest at Cork University Hospital. *Irish Medical Journal* **109,** 335–8 (2016).

32. Paniagua, D. *et al.* Outcome and cost-effectiveness of cardiopulmonary resuscitation after in-hospital cardiac arrest in octogenarians. *Cardiology* **97,** 6–11 (2002).

33. Rankin, A. P. N. The in-hospital Utstein style: use in reporting outcome from cardiac arrest in Middlemore Hospital 1995–1996. *Resuscitation* **36,** 91–94 (1998).

34. Rudiger, A., Tobler, D. & Estlinbaum, W. Frequency and outcome of in-hospital rescuscitation outside the ICU-setting. *Swiss medical weekly* **134,** 59 (2004).

35. Saklayen, M., Liss, H. & Markert, R. In-hospital cardiopulmonary resuscitation. Survival in 1 hospital and literature review. *Medicine* **74,** 163–175 (1995).

36. Shin, T. G. *et al.* Two-year survival and neurological outcome of in-hospital cardiac arrest patients rescued by extracorporeal cardiopulmonary resuscitation. *International journal of cardiology* **168,** 3424–3430 (2013).

37. Skrifvars, M. *et al.* Evaluation of the in-hospital Utstein template in cardiopulmonary resuscitation in secondary hospitals. *Resuscitation* **56,** 275–282 (2003).

38. Bloom, H. L. *et al.* Long-term survival after successful inhospital cardiac arrest resuscitation. *American heart journal* **153,** 831–836 (2007).

39. Skrifvars, M., Saarinen, K, Ikola, K & Kuisma, M. Improved survival after in-hospital cardiac arrest outside critical care areas. *Acta anaesthesiologica scandinavica* **49,** 1534–1539 (2005).

40. Stapleton, R. D., Ehlenbach, W. J., Deyo, R. A. & Curtis, J. R. Long-term outcomes after in-hospital CPR in older adults with chronic illness. *Chest* **146,** 1214–1225 (2014).

41. Tunstall-Pedoe, H. *et al.* Survey of 3765 cardiopulmonary resuscitations in British hospitals (the BRESUS Study): methods and overall results. *British Medical Journal* **304,** 1347–1351 (1992).

42. Vakil, K. *et al.* Long-Term Outcomes of Patients Who Had Cardiac Arrest After Cardiac Operations. *The Annals of Thoracic Surgery* **102,** 512–517 (2016).

43. Varon, J., Walsh, G. L., Marik, P. E. & Fromm, R. E. Should a cancer patient be resuscitated following an in-hospital cardiac arrest? *Resuscitation* **36,** 165–168 (1998).

44. Wong, S. P., Kreuter, W., Curtis, J. R., Hall, Y. N. & O'Hare, A. M. Trends in in-hospital cardiopulmonary resuscitation and survival in adults receiving maintenance dialysis. *JAMA internal medicine* **175,** 1028–1035 (2015).

45. Yi, H.-J. *et al.* Factors Associated with Survival and Neurological Outcome after Cardiopulmonary Resuscitation of Neurosurgical Intensive Care Unit Patients. *Neurosurgery* **59,** 838–846 (2006).

46. Hessulf, F. *et al.* Factors of importance to 30-day survival after in hospital cardiac arrest in Sweden – A population-based registry study of 15, 000 cases. *Resuscitation* **96,** 102 (2015).

47. Memar, M. *et al.* Long-term mortality and morbidity among 30-day survivors after in-hospital cardiac arrests - a Swedish cohort study. *Resuscitation* **124,** 76–79 (2018).

48. Thompson, L. E. *et al.* Trends in long-term survival after in-hospital cardiac arrest: insights from get with the guidelines-resuscitation®. *Circulation* **132,** A18923–A18923 (2015).

49. Blumenstein, J. *et al.* Extracorporeal life support in cardiovascular patients with observed refractory in-hospital cardiac arrest is associated with favourable short and long-term outcomes: A propensity-matched analysis. *European Heart Journal: Acute Cardiovascular Care* **5,** 13–22 (2016).

50. Chen, J. *et al.* The impact of implementing a rapid response system: A comparison of cardiopulmonary arrests and mortality among four teaching hospitals in Australia. *Resuscitation* **85,** 1275–1281 (2014).

51. Ruiz, M. C. *et al.* Outcome after cardiorespiratory arrest in a referral hospital reported in Utstein style. *Medicina Intensiva* **28,** 49–56 (2004).

52. Dimopoulou, I., Anthi, A., Michalis, A. & Tzelepis, G. E. Functional status and quality of life in long-term survivors of cardiac arrest after cardiac surgery. *Critical Care Medicine* **29,** 1408–1411 (2001).

53. García, A. E., Fernández, I. S. & Pesquera, M. P. Evaluación de la efectividad de un sistema de alarma cardiaca intrahospitalaria. *Enfermería Intensiva* **20,** 58–68 (2009).

54. Feingold, P. L. *et al.* Long-term survival following in-hospital cardiac arrest: A matched cohort study. *Resuscitation* **99,** 72–78 (2016).

55. Wong, M. K. *et al.* Trends in Short- and Long-Term Survival Among Out-of-Hospital Cardiac Arrest Patients Alive at Hospital Arrival. *Circulation* **130,** 1883–1890 (2014).

56. Piscator, E., Hedberg, P., Göransson, K. & Djärv, T. Survival after in-hospital cardiac arrest is highly associated with the Age-combined Charlson Co-morbidity Index in a cohort study from a two-site Swedish University hospital. *Resuscitation* **99,** 79–83 (2016).

57. Wang, H. *et al.* Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet* **388,** 1459–1544 (2016).

58. Bapat, V., Allen, D., Young, C., Roxburgh, J. & Ibrahim, M. Survival and Quality of Life After Cardiac Surgery Complicated by Prolonged Intensive Care. *Journal of Cardiac Surgery* **20,** 212–217 (2005).

59. Damuth, E., Mitchell, J. A., Bartock, J. L., Roberts, B. W. & Trzeciak, S. Long-term survival of critically ill patients treated with prolonged mechanical ventilation: a systematic review and meta-analysis. *The Lancet Respiratory Medicine* **3,** 544–553 (2015).

60. Meaney, P. A. *et al.* Rhythms and outcomes of adult in-hospital cardiac arrest∗. *Critical Care Medicine* **38,** 101–108 (2010).

61. Lingsma, H. *Measuring quality of care : methods and applications to acute neurological diseases* ISBN: 9789077283110 (Erasmus University Rotterdam, 2010).

62. Assmann, S. F., Pocock, S. J., Enos, L. E. & Kasten, L. E. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *The Lancet* **355,** 1064–1069 (2000).

63. Haywood, K. *et al.* COSCA (Core Outcome Set for Cardiac Arrest) in Adults: An Advisory Statement From the International Liaison Committee on Resuscitation. *Circulation* **137** (2018).

64. Perkins, G. D. *et al.* Cardiac Arrest and Cardiopulmonary Resuscitation Outcome Reports: Update of the Utstein Resuscitation Registry Templates for Out-of-Hospital Cardiac Arrest. *Resuscitation* **96,** 328–340 (2015).

65. Regenbogen, S. E. & Dimick, J. B. Using Clinical Registries to Enhance Implementation Research. *Journal of the American Medical Association, Surgery* **153,** 366. ISSN: 2168-6254 (2018).

66. Ahn, C. *et al.* Efficacy of extracorporeal cardiopulmonary resuscitation compared to conventional cardiopulmonary resuscitation for adult cardiac arrest patients: a systematic review and meta-analysis. *Scientific Reports* **6,** 34208. ISSN: 2045-2322 (2016).

5

Neurological outcome after extracorporeal cardiopulmonary resuscitation for in-hospital cardiac arrest: a systematic review and meta-analysis

Benjamin Y Gravesteijn, Marc Schluep, Maksud Disli, Prakriti Garkhail, Dinis Dos Reis Miranda, Robert Jan Stolker, Henrik Endeman, Sanne E Hoeks

Critical Care

5.1 | Abstract

5.1.1 | Background

In-hospital cardiac arrest (IHCA) is a major adverse event with a high mortality rate if not treated appropriately. Extracorporeal cardiopulmonary resuscitation (ECPR), as adjunct to conventional cardiopulmonary resuscitation (CCPR), is a promising technique for IHCA treatment. Evidence pertaining to neurological outcomes after ECPR is still scarce.

5.1.2 | Methods

We performed a comprehensive systematic search of all studies up to December 20, 2019. Our primary outcome was neurological outcome after ECPR at any moment after hospital discharge, defined by the Cerebral Performance Category (CPC) score. A score of 1 or 2 was defined as favourable outcome. Our secondary outcome was post-discharge mortality. A fixed-effects meta-analysis was performed.

5.1.3 | Results

Our search yielded 1215 results, of which 19 studies were included in this systematic review. The average survival rate was 30% (95% CI 28 – 33%, I^2 = 0%, p = 0.24). In the surviving patients, the pooled percentage of favourable neurological outcome was 84% (95% CI 80 – 88%, I^2 = 24%, p = 0.90).

5.1.4 | Conclusion

ECPR as treatment for in-hospital cardiac arrest is associated with a large proportion of patients with good neurological outcome. The large proportion of favourable outcome could potentially be explained by the selection of patients for treatment using ECPR. Moreover, survival is higher than described in the conventional CPR literature. As indications for ECPR might extend to older or more fragile patient populations in the future, research should focus on increasing survival, while maintaining optimal neurological outcome.

5.2 | Introduction

In-hospital cardiac arrest (IHCA) is a serious adverse event in hospitalized patients that inevitably leads to death if not treated appropriately. It is associated with low survival rates at discharge and at 1-year follow-up (13%, 95% prediction interval 6–29%) [1, 2]. The use of extracorporeal membrane oxygenation (ECMO) in addition to chest compressions for cardiopulmonary resuscitation may improve survival after IHCA [3]. Recent guidelines state the use of ECMO for CPR (ECPR) as potentially beneficial for specific patient populations [4]. However, they also stress the lack of evidence for this novel technique [5]. To our knowledge, there is no large-scale evidence pertaining to neurologic outcomes after ECPR for IHCA [6, 7].

Survivors of cardiac arrest also suffer from neurological sequelae, which have been described as the post-cardiac arrest syndrome [8]. An important measure for neurological outcome is the aforementioned CPC. Although the CPC scoring suffers from limited discriminatory capacity and has a potential ceiling effect and possible overestimation of function, it is to date the most used outcome measure [9]. The neurological outcome of 1-year survivors after conventional CPR (CCPR) tends to be high: 92% of patients score a cerebral performance category (CPC) of 1 or 2 (95% prediction interval 82–97%) [2]. Another important neurological score is the Glasgow Outcome Scale (GOS). This outcome scale was developed for scoring outcome after acquired brain injury, but also is used to assess functional outcome after cardiac arrest [10, 11].

ECPR facilitates return of circulation, albeit artificial. However, it is much more uncertain whether this recovery of circulation translates into survival, or acceptable neurological outcome. Furthermore, the association

between neurologic outcomes and prognostic factors should be elicited, the effect of time to ECMO on outcome in particular [12]. This systematic review aims to summarise the evidence on neurologic outcomes after hospital discharge of patients treated with ECPR for in-hospital cardiac arrest.

5.3 | Methods

5.3.1 | Literature search and selection criteria

This systematic review and meta-analysis is reported following the PRISMA and MOOSE guidelines for reporting of systematic reviews and meta-analyses of observational studies [13, 14]. For this systematic review, we performed a systematic search of all published data on post-discharge neurological outcome after IHCA treated by ECPR up to December 20, 2019. We used the search engines PubMed, EMBASE, Medline Ovid, Web of Science and Cochrane Central. Our searches contained the following keywords: in-hospital cardiac arrest, ECMO, neurological outcome, brain injury and neurological outcome. The exact search strategies are included in Additional file 1: Appendix 1 available online.

Our inclusion criteria were as follows: (1) use of ECPR for in-hospital cardiac arrest, (2) adult patients, (3) reporting of neurological outcome (CPC or GOS), (4) clinical studies, and (5) written in English, German, French or Dutch. We included studies that reported outcome upon or after discharge from hospital. Studies were excluded if they did not fit inclusion criteria or if they were only published as abstract.

After the initial screening, the remaining articles were assessed by reading the full text. Studies often reported characteristics and outcomes of in-hospital and out-of-hospital cardiac arrest simultaneously. The authors

of articles in which data for the IHCA cohort was not reported separately were contacted. Data extraction from selected studies was performed independently by two investigators (MD, PG) using a standardized form. Subsequently, the discrepancies were resolved by discussion with the other authors (BG, MS, SH).

5.3.2 | Definitions

The primary outcome was defined as favourable neurological outcome post-discharge from hospital using CPC or GOS score. A measurement was considered post-discharge, when the outcome was reported at discharge or later. For a description of the CPC and the GOS score, see Table 5.1 and Additional file 2: Appendix 2 available online. A CPC score of 1 or 2 or a GOS score of 4 or 5 was defined as favourable outcome. The secondary outcome was post-discharge survival. If a study reported survival and neurological outcome at different follow-up moments, we ensured extracting the data for the same follow-up moment per study. Additionally, out of interest in the time to ECMO cannulation on the effect of ECPR, we extracted the average time to ECMO per study. Only the effect of the average time to ECMO cannulation on the primary outcome (favourable outcome) was investigated.

5.3.3 | Quality assessment

The quality of the included studies was evaluated using the method of Hayden et al. for prognosis studies in systematic reviews [34]. The quality assessment is based on six categories: (1) study population: whether the study correctly defines and describes the study population; (2) study attri-

tion: whether the study was able to obtain a complete follow-up; (3) prognostic factor measurement: whether the study reports the most important prognostic characteristics; (4) outcome measurement: whether the neurological outcome was measured in a valid and robust way; (5) confounding measurement: whether the authors explored what factors influenced neurological outcome; and (5) account and analysis: whether the study reports a correct methodology of statistical analysis. Up to 2 points can be scored in each category. Therefore, the maximum score was 12 points, indicating high quality.

5.3.4 | Statistical analysis

For the analysis of the primary outcome, a fixed-effects model was used, because little heterogeneity was observed. Results of the meta-analyses are presented as pooled proportions with corresponding 95% confidence intervals (CI). Between-study heterogeneity was assessed using I^2 statistic and the DerSimonian–Laird estimator for τ^2. Moreover, heterogeneity was analysed by assessing statistical significance based on Cochran's Q statistic.

Furthermore, because of specific interest in the relationship between time to ECMO and outcome in these patients, a meta-regression analysis was performed. A random intercept meta-regression analysis (binomial-normal model) was used with favourable outcome as outcome. This model is appropriate for meta-regression of probabilities, since it avoids the bias that occurs when a normal-normal model would be used for logit-transformed probabilities [35].

Finally, we considered multiple follow-up moments for our primary and secondary outcome. Therefore, a sensitivity analysis was performed

for the studies that used the most frequently reported follow-up moment (i.e. at discharge).

5.4 | Results

5.4.1 | Included articles

Our search yielded 1215 results. Subsequently, 1130 articles were excluded by screening of title and abstract (2 because of a language different than Dutch, English, French or German). Full-text screening resulted in inclusion of 28 articles, of which 9 did not report characteristics and outcome of the IHCA cohort separately. For these articles, authors were contacted to provide this data for the IHCA cohort. None replied after multiple attempts; therefore, these studies were excluded. Finally, 19 articles were included [15–33] (figure 5.1).

The sample size ranged between 10 and 200 patients. The mean age ranged between 18 and 86. All studies were observational studies, of which 10 (53%) were retrospective (table 5.1). All studies mentioned contra-indications. The most frequently reported contra-indications were CPR duration (58%), advanced age (58%), terminal cancer (84%), previous severe or irreversible brain damage (63%) and uncontrollable bleeding (63%). These are summarized in table 5.2.

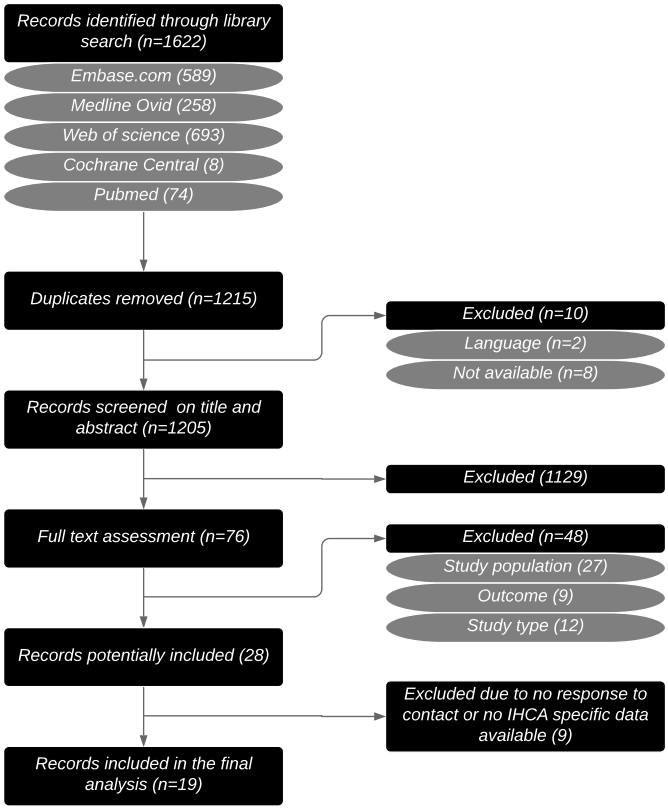


Figure 5.1: *Flowchart showing the process of inclusion of studies. The search strategy was performed on 20 December 2019.*

Table 5.1: *Overview and characteristics of the included studies*

Author	Year	Study type	Country	ECPR age median	CA to ECMO time (range)	Time of CPC assessment*
Avalli [15]	2012	Retrospective	Italy	67 (61–73)	55 (40–70)	6 months
Bednarczyk [16]	2014	Retrospective	Canada	–	49 ± 21	Discharge
Blumenstein [17]	2016	Retrospective	Germany	72 (55–72.9)	33.0 (19.0–47.0)	Discharge (30d)
Chen [18]	2008	Prospective	Taiwan	61.5 (18–74)	52.8 ± 37.2	Discharge
Dennis [19]	2017	Retrospective	Australia	–	40 (30–55)	Discharge
Ellouze [20]	2018	Retrospective	France	–	60 (45–89)	6 months
Fagnoul [21]	2013	Prospective	Belgium	–	55 (42–59.5)	Discharge from ICU
Jung [22]	2016	Retrospective	Germany	66 (56–78)	–	Discharge (30d)
Lazzeri [23]	2013	Prospective	Italy	54.8 ± 9 years (24–74)	51.9 ± 24.8	Discharge
Lee [24]	2016	Prospective	S. Korea	–	–	Discharge
Lin [25]	2010	Prospective	Taiwan	62.3 (21–73)	40 (16–150)	Discharge
Liu [26]	2011	Prospective	Taiwan	53 (50–69)	–	Discharge
Mazzeffi [27]	2016	Prospective	USA	57 ± 15 (34–86)	31 (15–52)	Discharge
Peigh [28]	2015	Retrospective	USA	46 ± 12	54 ± 30	4-6w after discharge
Pozzi [29]	2019	Prospective	France	46.2 ± 13.5 (18–76)	46.9 ± 19.0	Discharge
Shin [30]	2013	Retrospective	S. Korea	59.9 ± 15.3	38.8	6 months
Spangenberg [31]	2016	Retrospective	Germany	–	42.9 ± 28.6	Discharge
Stub [32]	2015	Prospective	Australia	–	56 (40–85)	Discharge
Wang [33]	2014	Retrospective	Taiwan	55.7 ± 15.1	40 (15–162)	Discharge

**Neurological outcome and mortality was extracted at this follow-up moment*

Table 5.2: *Reported contra-indications for ECPR per study.*

Author	CPR duration (min)	Non-witnessed arrest	Severe comorbidity	Age (years)	Terminal cancer	Advanced heart failure
Avalli [15]	<30			>75	x	X
Bednarczyk [16]	<15	x			x	
Blumenstein [17]		x			x	X
Chen [18]	<10	x		>75	x	
Dennis [19]					x	
Ellouze [20]	<30	x		>75	x	
Fagnoul [21]	<10	x	x	>65	x	X
Lazzeri [23]	<30			>75	x	
Liu [26]	<10				x	
Jung [22]				>74	x	
Lee [24]					x	
Lin [25]				x	x	X
Mazzeffi [27]	<10					
Peigh [28]				>70	x	X
Pozzi [29]	<20	x			x	X
Shin [30]				>80	x	
Spangenberg [31]	<20					
Stub [32]				>65	x	
Wang [33]	<10	x	x	>80	x	

**Some studies selected pre-specified groups based on cardiac arrest aetiology.*
Liu: acute myocardial infarction; Mazeffi: post-cardiac surgery; Chen/Spangenberg: cardiac origin; Stub: cardiac origin with ventricular fibrillation; Jung: cardiac origin or pulmonary embolism

Author	Post-hoc*	Pre-existing brain damage	Liver cirrhosis	Renal failure	Uncontrollable sepsis	Uncontrollable bleeding
Avalli				x	x	x
Bednarczyk		x			x	x
Blumenstein	x	x			x	x
Chen		x				
Dennis		x				
Ellouze		x	x		x	x
Fagnoul		x			x	x
Lazzeri	x	x				
Liu		x				
Jung						
Lee						
Lin		x			x	x
Mazzeffi	x	x		x	x	x
Peigh		x				
Pozzi		x				
Shin		x				
Spangenberg	x					
Stub		x	x			
Wang		x				

**Some studies selected pre-specified groups based on cardiac arrest aetiology.*
Liu: acute myocardial infarction; Mazzeffi: post-cardiac surgery; Chen/Spangenberg: cardiac origin; Stub: cardiac origin with ventricular fibrillation; Jung: cardiac origin or pulmonary embolism

Author	Irreversible (multi) organ failure	Arrest of septic origin	Coagulation disorder	BMI >40	Weight <30 kg	Aortic dissection
Avalli	x					x
Bednarczyk	x	x	x	x		x
Blumenstein			x			x
Chen						
Dennis						
Ellouze						
Fagnoul			x		x	
Lazzeri	x	x				
Liu						
Jung						
Lee						
Lin						
Mazzeffi						
Peigh						
Pozzi						
Shin	x	x				
Spangenberg						
Stub						
Wang	x					

**Some studies selected pre-specified groups based on cardiac arrest aetiology.*
Liu: acute myocardial infarction; Mazeffi: post-cardiac surgery; Chen/Spangenberg: cardiac origin; Stub: cardiac origin with ventricular fibrillation; Jung: cardiac origin or pulmonary embolism

Author	Extensive peripheral artery disease	Bedridden, care-dependant	At the discretion of the CPR team
Avalli	x		
Bednardyzk	x		
Blumenstein			
Chen			x
Dennis			x
Ellouze			
Fagnoul			
Lazzeri			
Liu			x
Jung			x
Lee			
Lin			
Mazzeffi			x
Peigh			
Pozzi			
Shin			
Spangenberg			
Stub			X
Wang	x		

**Some studies selected pre-specified groups based on cardiac arrest aetiology.*
Liu: acute myocardial infarction; Mazeffi: post-cardiac surgery; Chen/Spangenberg: cardiac origin; Stub: cardiac origin with ventricular fibrillation; Jung: cardiac origin or pulmonary embolism

Fifteen (79%) of the included studies had a score $\geq$ 9 (out of 12) in the Hayden method for quality assessment (table 5.3). Thirteen studies (68%) did not sufficiently adjust for confounding bias, while 18 studies (95%) reported important prognostic characteristics. Overall, high quality was observed for study participation (13 studies, 68%, received maximum scores), study attrition (14 studies, 74%, received maximum scores), outcome measurement (14 studies, 74%, received maximum scores) and analysis (17 studies, 89%, received maximum scores).

Table 5.3: *Risk of bias assessment, using the method of Hayden et al. for prognosis studies in systematic reviews. The colours correspond to the points given per attribute.*

Author	Participation	Attrition	Prognostic factors	Outcome	Confounding	Analysis	Total
Avalli et al.	+	±	+	+	-	+	9
Bednarczyk (	+	+	+	+	-	±	9
Blumenstein	+	+	+	±	+	+	11
Chen et al.	+	±	+	+	+	+	11
Dennis et al.	+	+	+	+	-	+	10
Ellouze et al.	+	+	+	+	-	+	10
Fagnoul et a	+	+	+	+	-	+	9
Jung et al.	±	+	+	+	-	+	9
Lazzeri et al.	±	±	+	+	-	+	8
Lee et al.	±	+	+	±	-	+	8
Liu et al	±	+	+	±	-	+	8
Mazzeffi et a	±	+	+	+	-	+	9
Lin et al.	+	+	+	+	+	+	12
Peigh et al.	+	+	+	±	±	±	9
Pozzi et al.	+	+	+	+	-	+	10
Shin et al.	+	+	+	+	+	+	12
Spangenber	+	+	±	±	-	+	8
Stub et al.	+	±	+	+	-	+	9
Wang et al.	+	±	+	+	±	+	10

Hayden scoring system. + is 2 points, ± is 1 point, and - is 0 points

Study participatioin comprises whether the study correctly defines and describes the study population

Study attrition comprises whether the study was able to obtain a complete follow-up

Prognostic factor management comprises whether the study reports the most important prognostic characteristics, as described in the characteristics table (e.g. Utstein-style reporting)

Outcome measurement comprises whether the neurological oucome was measured in a confidential manner

Confounding measurement and account comprises whether the authors explored what influenced neurological outcome/survival data

Analysis comprises whether the study reports a correct methodology of statictical analysis

None of the included articles expressed neurological outcome in GOS. Six studies showed that all survivors were classified as CPC 1–2 [20, 23, 25, 29, 32, 33]. The largest study reported 52 patients with CPC 1–2 (84%) versus 10 patients with CPC 3–4 (16%) [26]. There was variation in the timing of assessment of outcome: 15 studies (79%) reported CPC and mortality at discharge, 2 (11%) studies reported CPC and mortality at 6 months, 1 (5%) study reported CPC and mortality at 4–6 weeks after discharge and 1 (5%) study reported CPC and mortality at discharge from ICU.

5.4.2 | Meta-analysis

The average post-discharge survival rate (i.e. discharge until 6 months) was 30% (95% CI 28–33%). Heterogeneity was low: I^2 = 0%, p = 0.24. At the same follow-up moment in these survivors, the pooled proportion of favourable outcome was 84% (95% CI 80–88%). The heterogeneity was again low: I^2 = 24%, p = 0.90 (figures 5.2 and 5.3).

As previously described, there was a variation in timing of assessment of outcome. In the 15 studies (79%) which reported survival to discharge, the pooled survival rate was 30% (95% CI 0.27–0.34%), with low heterogeneity (I^2 = 0%, p = 0.15). In these survivors, the pooled proportion of favourable neurological outcome was 83% (95% CI 78–87%), with again low heterogeneity (I^2 = 0%, p = 0.93).

5.4.3 | Meta-regression

A total of 16 studies (84%) reported an average time to ECMO (time to cannulation/time to start ECMO), and the reported range was large (31–60 min). However, the OR per 10 min for favourable outcome was 1.29 (95%

CI 0.73–2.29): favourable outcome was not explained by the average time to ECMO per study.

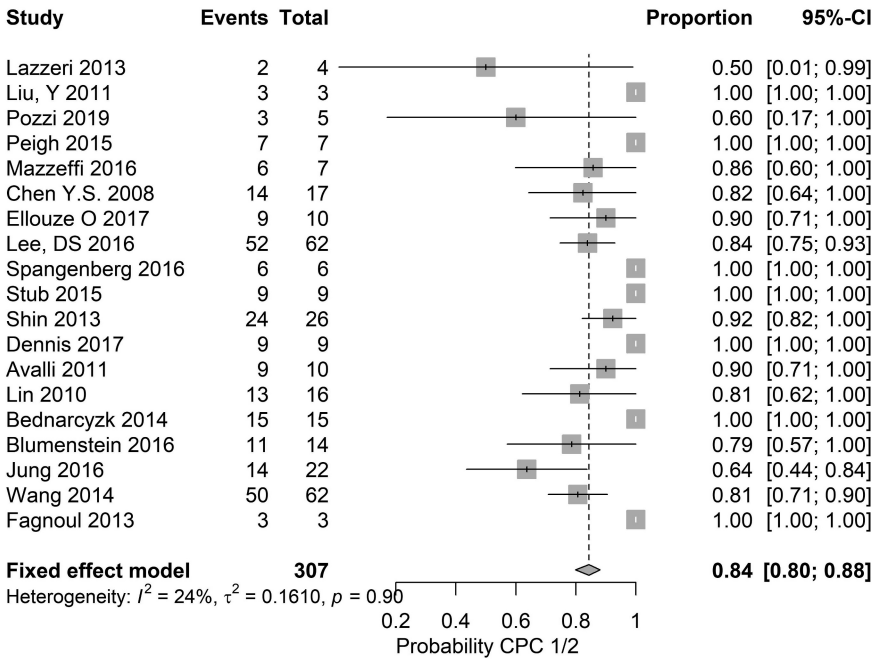


Figure 5.2: *Forest plot showing the results for the primary outcome of this study, neurological outcome.*

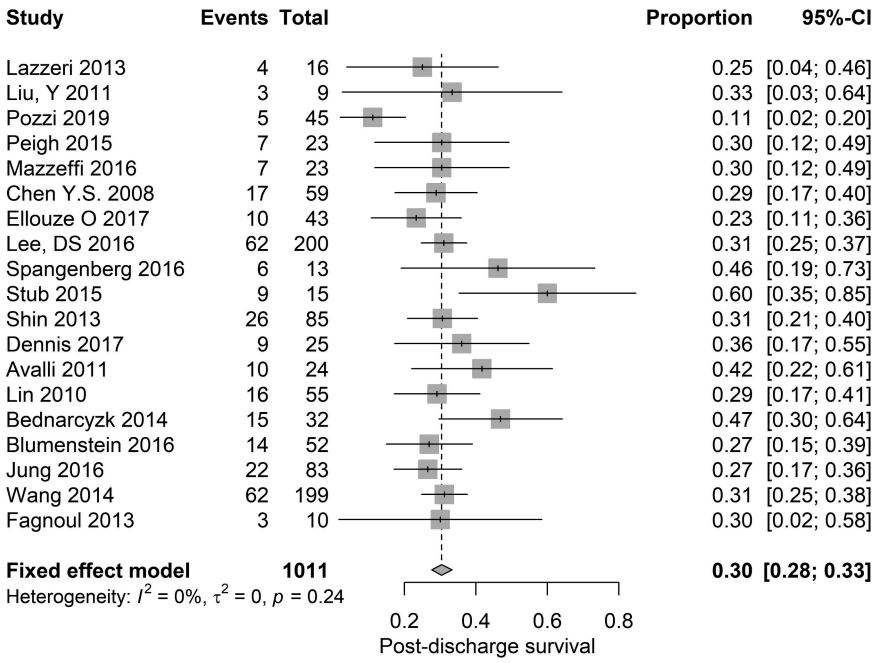


Figure 5.3: *Forest plot showing the results for the secondary outcome of this study, post-discharge survival.*

5.5 | Discussion

Our primary goal was to provide a comprehensive overview of current literature pertaining to neurological outcome after ECPR for in-hospital cardiac arrest. In post-discharge survivors, we found a high proportion of patients with a CPC 1–2 (84% [95% CI 80–88%]), which is lower than described for 1-year survivors CCPR (92% [95% PI 82–97%] [2]). Post-discharge survival was higher than reported for the general IHCA populations (30% [95% CI 28–33%] versus 17% [95% PI 13–23%] [1, 2]). We found little heterogeneity in outcome between studies.

Although neurological outcome is good, it remains inconclusive whether neurological outcome of patients receiving ECPR is better than patients receiving CCPR. We did find a lower percentage of "good" neurological outcome (CPC 1–2) than in a systematic review in a conventional CPR population [2]. However, in this review, CPC score was a secondary outcome. In this review, the proportion outcome assessment was also specifically set for 1 year, rather than after hospital discharge. A systematic review aimed at comparing ECPR and CCPR suggests that the neurological outcome is better in IHCA patients treated with ECPR compared to CCPR [7]. Due to the observational nature of the studies included in these reviews, the selection of patients for ECPR could still lead to better outcomes for this group. For literature pertaining to OHCA, the same caveats apply [36, 37].

Comparing this study to the literature suggests that survival of IHCA patients undergoing ECPR is higher than IHCA populations who receive conventional CPR (chest compressions) [1, 2]. Our estimate of survival is also comparable to the reported survival rate of adult ECPR patients by the ELSO registry [38]. This high survival might be explained by the selection of patients with a high chance of good outcome. The American Heart

Association guidelines state that ECPR should be considered in patients for whom the suspected aetiology of the cardiac arrest is potentially reversible during a limited period of mechanical cardiorespiratory support [39]. In contrast, the European Resuscitation Council simply declares that the technique is a potential rescue therapy in patients where standard advanced life support (ALS) measures are not successful [5]. In practice, however, a much broader range of contra-indications are being used: this study found that the primary reported contra-indications were CPR duration, age, severe comorbidities such as terminal cancer or pre-existing neurological impairments and uncontrolled bleeding. These contra-indications are known to impact prognosis. Excluding these patients from ECPR effectively results in a higher survival compared to patients receiving conventional CCPR. Especially the age criteria are quite stringent and therefore likely affect the apparent survival [40], given the average age of the CPR population [41]. Moreover, the finding that we found substantially less heterogeneity in survival rates between studies than a systematic review of the CCPR literature [1, 2] also supports the hypothesis that this is a selected population. Nevertheless, part of the difference might be explained by the effect of ECPR versus CCPR on outcome [33, 42, 43].

On the other hand, ECPR is only indicated in patients with refractory cardiac arrest. Therefore, patients eligible for ECPR have, by indication, a worse prognosis than patients with conventional CPR as a portion of these patients ROSC after a short resuscitation period [44]. As a result, ECPR patients might not be the patient population with the most favourable outcome.

Evidence in the literature suggests that a longer time to ECMO time is associated with lower benefit of ECPR [45–48]. Bartos et al. suggest the association between time to ECMO and survival is explained by the

metabolic derangements, which develop during prolonged low-flow time, leading to a worse outcome [40]. In our meta-analysis, this association between time to ECMO and survival is not found. However, most of the studies included in our meta-analysis do find a relationship between time to ECMO and survival, when this was investigated [20, 24, 29–33]. Possibly, our results can be explained due to an aggregation effect: our results imply that—because the variation in outcome between studies was small—differences in mean calculated time to ECMO do not explain differences in mean survival between studies. Additionally, our results might be explained by the long time to ECMO in the included studies (> 30 min). Given that the success rate of CPR is very low when the duration is longer than 30 min [49, 50], it might be more relevant to assess the effect of time to ECMO in when the time to ECMO is shorter. Since the effect of timing of ECPR on outcome impacts implementation, more high-quality evidence is needed.

Certain limitations should be taken into account. First, the time of CPC assessment was not the same for all studies. Most studies only scored CPC at the moment of discharge. This was not clearly defined in all studies. Some studies mentioned CPC scores at 6 months; others report a CPC score at discharge. We did show in a sensitivity analysis with the studies that reported data for the same follow-up moment that the estimates were very similar to the main analysis. However, a standardized and comprehensive assessment of neurologic and functional outcomes in cardiac arrest research is needed [9]. In spite of these differences, we encountered homogenous results, which suggest that the time of outcome assessment did not significantly influence the results: the neurological outcome and survival seem to remain constant at different follow-up times. Second, the included studies had two main shortcomings: they were relatively small

(the largest study included 200 patients) and often reported their data non-standardized and non-structured, which complicated the process of data extraction. Remarkably, we observed little heterogeneity between these small studies, which enabled us to perform a fixed-effects meta-analysis. Finally, we were not able to do an individual patient data meta-analysis. Since heterogeneity between studies was found, the effect of prognostic factors on outcome in these patients could not be explored effectively. An individual patient data meta-analysis would enable this [51] and could be of interest for future research.

By showing that treating a selected group of IHCA patients with ECPR can result in a high proportion of good neurological outcome, this study illustrates what next step the field should take. When centres become more experienced, the indications of ECPR will shift towards a less selected, but probably also more fragile patient population: older patients with more comorbidities might be considered eligible for ECPR in the near future. Nevertheless, we should focus on treating these patients while maintaining such a high proportion of favourable neurological outcome.

5.5.1 | Conclusion

ECPR as treatment for in-hospital cardiac arrest is associated with a large proportion of patients with good neurological outcome (CPC 1–2). The large proportion of favourable outcome could potentially be explained by the selection of patients for treatment using ECPR. Nevertheless, both conventional and extracorporeal CPR are associated with low survival rates. The survival after ECPR, however, is higher than described in the conventional CPR literature. As indications for ECPR might extend to older or more fragile patient populations in the future, research should focus on

increasing survival, while maintaining optimal neurological outcome.

5.6 | Supplementary Material

Supplementary materials can be found in the online version of this article: `https://tinyurl.com/dd5nt9cz`

References

1. Zhu, A. & Zhang, J. Meta-analysis of outcomes of the 2005 and 2010 cardiopulmonary resuscitation guidelines for adults with in-hospital cardiac arrest. *The American journal of emergency medicine* **34,** 1133–1139 (2016).
2. Schluep, M., Gravesteijn, B. Y., Stolker, R. J., Endeman, H. & Hoeks, S. E. One-year survival after in-hospital cardiac arrest: A systematic review and meta-analysis. *Resuscitation* **132,** 90–100. ISSN: 1873-1570 (2018).
3. Chen, Y.-S. *et al.* Analysis and results of prolonged resuscitation in cardiac arrest patients rescued by extracorporeal membrane oxygenation. *Journal of the American College of Cardiology* **41,** 197–203 (2003).
4. Conrad, S. A., Rycus, P. T. & Dalton, H. *Extracorporeal life support registry report 2004* tech. rep. 1 (2005), 4–10.
5. Soar, J. *et al.* European Resuscitation Council Guidelines for Resuscitation 2015. *Resuscitation* **95,** 100–147. ISSN: 03009572 (2015).
6. Wang, J., Ma, Q., Zhang, H., Liu, S. & Zheng, Y. Predictors of survival and neurologic outcome for adults with extracorporeal cardiopulmonary resuscitation. *Medicine* **97,** e13257 (2018).
7. Chen, Z. *et al. Clinical Efficacy of Extracorporeal Cardiopulmonary Resuscitation for Adults with Cardiac Arrest: Meta-Analysis with Trial Sequential Analysis* 2019.

8. ACC/AHA *et al.* Post-cardiac arrest syndrome: epidemiology, pathophysiology, treatment, and prognostication. A consensus statement from the International Liaison Committee on Resuscitation (American Heart Association, Australian and New Zealand Council on Resuscitation, *Circulation* **118,** 2452–2483 (2008).

9. Haywood, K. *et al.* COSCA (Core Outcome Set for Cardiac Arrest) in Adults: An Advisory Statement From the International Liaison Committee on Resuscitation. *Circulation* **137** (2018).

10. Jennett, B & Bond, M. Assessment of outcome after severe brain damage. *Lancet* **1,** 480–484. ISSN: 0140-6736 (1975).

11. Sivaraju, A. *et al.* Prognostication of post-cardiac arrest coma: early clinical and electroencephalographic predictors of outcome. *Intensive Care Medicine* **41,** 1264–1272. ISSN: 14321238 (2015).

12. Sandroni, C, Nolan, J, Cavallaro, F & Antonelli, M. In-hospital cardiac arrest: Incidence, prognosis and possible measures to improve survival. *Intensive Care Medicine* **33,** 237–245 (2007).

13. Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G. & Group, P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS medicine* **6,** e1000097 (2009).

14. Stroup, D. F. *et al.* Meta-analysis of observational studies in epidemiology: a proposal for reporting. *Journal of the American Medical Association* **283,** 2008–2012 (2000).

15. Avalli, L. *et al.* Favourable survival of in-hospital compared to out-of-hospital refractory cardiac arrest patients treated with extracorporeal membrane oxygenation: An Italian tertiary care centre experience. *Resuscitation* **83,** 579–583. ISSN: 03009572 (2012).

16. Bednarczyk, J. M. *et al.* Resuscitative extracorporeal membrane oxygenation for in hospital cardiac arrest: A Canadian observational experience. *Resuscitation* **85,** 1713–1719. ISSN: 18731570 (2014).

17. Blumenstein, J. *et al.* Extracorporeal life support in cardiovascular patients with observed refractory in-hospital cardiac arrest is associated with favourable short and long-term outcomes: A propensity-matched analysis. *European heart journal. Acute cardiovascular care* **5,** 13–22. ISSN: 20488734 (2016).

18. Chen, Y.-S. *et al.* Cardiopulmonary resuscitation with assisted extracorporeal life-support versus conventional cardiopulmonary resuscitation in adults with in-hospital cardiac arrest: an observational study and propensity analysis. *The Lancet* **372,** 554–561. ISSN: 01406736 (2008).

19. Dennis, M. *et al.* Extracorporeal cardiopulmonary resuscitation for refractory cardiac arrest: A multicentre experience. *International Journal of Cardiology* **231,** 131–136. ISSN: 18741754 (2017).

20. Ellouze, O. *et al.* Comparable Outcome of Out-of-Hospital Cardiac Arrest and In-Hospital Cardiac Arrest Treated With Extracorporeal Life Support. *Artificial Organs* **42,** 15–21. ISSN: 0160564X (2018).

21. Fagnoul, D *et al.* Extracorporeal life support associated with hypothermia and normoxemia in refractory cardiac arrest. English. *Resuscitation* **84,** 1519–1524. ISSN: 0300-9572 1873-1570 (2013).

22. Jung, C. *et al.* Outcome predictors in cardiopulmonary resuscitation facilitated by extracorporeal membrane oxygenation. *Clinical Research in Cardiology* **105,** 196–205. ISSN: 1861-0684 (2016).

23. Lazzeri, C. *et al.* In-hospital refractory cardiac arrest treated with extracorporeal membrane oxygenation: A tertiary single center experience. *Acute Cardiac Care* **15,** 47–51. ISSN: 1748-2941 (2013).

24. Lee, D. S. *et al.* Survival after extracorporeal cardiopulmonary resuscitation on weekends in comparison with weekdays. *Annals of Thoracic Surgery* **101,** 133–140. ISSN: 15526259 (2016).

25. Lin, J.-W. *et al.* Comparing the survival between extracorporeal rescue and conventional resuscitation in adult in-hospital cardiac arrests: propensity analysis of three-year data. *Resuscitation* **81,** 796–803 (2010).

26. Liu, Y., Cheng, Y. T., Chang, J. C., Chao, S. F. & Chang, B. S. Extracorporeal membrane oxygenation to support prolonged conventional cardiopulmonary resuscitation in adults with cardiac arrest from acute myocardial infarction at a very low-volume centre. *Interactive CardioVascular and Thoracic Surgery* **12,** 389–393. ISSN: 1569-9293 (2011).

27. Mazzeffi, M. A. *et al.* Outcomes of extracorporeal cardiopulmonary resuscitation for refractory cardiac arrest in adult cardiac surgery patients. *Journal of Thoracic and Cardiovascular Surgery* **152,** 1133–1139. ISSN: 1097685X (2016).

28. Peigh, G., Cavarocchi, N. & Hirose, H. Saving life and brain with extracorporeal cardiopulmonary resuscitation: A single-center analysis of in-hospital cardiac arrests. *Journal of Thoracic and Cardiovascular Surgery* **150,** 1344–1349. ISSN: 1097685X (2015).

29. M., P. *et al.* Extracorporeal Life Support for Refractory Cardiac Arrest: A 10-Year Comparative Analysis. English. *Annals of Thoracic Surgery* **107,** 809–816. ISSN: 1552-6259 (2019).

30. Shin, T. G. *et al.* Two-year survival and neurological outcome of in-hospital cardiac arrest patients rescued by extracorporeal cardiopulmonary resuscitation. *International journal of cardiology* **168,** 3424–3430 (2013).

31. Spangenberg, T. *et al.* "Shock and Go?" extracorporeal cardio-pulmonary resuscitation in the golden-hour of ROSC. *Catheterization and Cardiovascular Interventions* **88,** 691–696. ISSN: 15221946 (2016).

32. Stub, D. *et al.* Refractory cardiac arrest treated with mechanical CPR, hypothermia, ECMO and early reperfusion (the CHEER trial). *Resuscitation* **86,** 88–94. ISSN: 18731570 (2015).

33. Wang, C.-H. *et al.* Improved outcome of extracorporeal cardiopulmonary resuscitation for out-of-hospital cardiac arrest – A comparison with that for extracorporeal rescue for in-hospital cardiac arrest. *Resuscitation* **85,** 1219–1224. ISSN: 03009572 (2014).

34. Hayden, J. A., Côté, P. & Bombardier, C. Evaluation of the quality of prognosis studies in systematic reviews. *Annals of internal medicine* **144,** 427–437 (2006).

35. Stijnen, T., Hamza, T. H. & Özdemir, P. Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Statistics in medicine* **29,** 3046–3067 (2010).

36. Holmberg, M. J. *et al.* Extracorporeal cardiopulmonary resuscitation for cardiac arrest: A systematic review. *Resuscitation* **131,** 91–100. ISSN: 1873-1570 (2018).

37. Beyea, M. M. *et al.* Neurologic outcomes after extracorporeal membrane oxygenation assisted CPR for resuscitation of out-of-hospital cardiac arrest patients: A systematic review. *Resuscitation* **130,** 146–158 (2018).

38. Extracorporeal life support organisation. *ECLS registry report* tech. rep. (2019).

39. Link, M. S. *et al.* Part 7: Adult advanced cardiovascular life support: 2015 American Heart Association guidelines update for cardiopulmonary resuscitation and emergency cardiovascular care. *Circulation* **132,** S444–S464. ISSN: 15244539 (2015).

40. Bartos, J. A. *et al.* Improved Survival With Extracorporeal Cardiopulmonary Resuscitation Despite Progressive Metabolic Derangement Associated With Prolonged Resuscitation. *Circulation* **141,** 877–886 (2020).

41. Girotra, S. *et al.* Trends in survival after in-hospital cardiac arrest. *New England Journal of Medicine* **367,** 1912–1920 (2012).

42. Ouweneel, D. M. *et al.* Extracorporeal life support during cardiac arrest and cardiogenic shock: a systematic review and meta-analysis. *Intensive Care Medicine* **42,** 1922–1934. ISSN: 0342-4642 (2016).

43. Tramm, R. *et al.* Extracorporeal membrane oxygenation for critically ill adults. *Cochrane Database of Systematic Reviews* (2015).

44. Gravesteijn, B. Y. *et al.* Cost-effectiveness of extracorporeal cardiopulmonary resuscitation after in-hospital cardiac arrest: A Markov decision model. *Resuscitation* **143,** 150–157. ISSN: 03009572 (2019).

45. Twohig, C. J., Singer, B., Grier, G. & Finney, S. J. A systematic literature review and meta-analysis of the effectiveness of extracorporeal-CPR versus conventional-CPR for adult patients in cardiac arrest. *Journal of the Intensive Care Society* **20,** 347–357 (2019).

46. Otani, T. *et al.* Low-flow time is associated with a favorable neurological outcome in out-of-hospital cardiac arrest patients resuscitated with extracorporeal cardiopulmonary resuscitation. *Journal of Critical Care* **48,** 15–20. ISSN: 15578615 (2018).

47. Ryu, J.-A. *et al.* Neurologic Outcomes in Patients Who Undergo Extracorporeal Cardiopulmonary Resuscitation. *The Annals of Thoracic Surgery* **108,** 749–755 (2019).

48. Yu, H.-Y. *et al.* Effect of interplay between age and low-flow duration on neurologic outcomes of extracorporeal cardiopulmonary resuscitation. *Intensive Care Medicine* **45,** 44–54. ISSN: 0342-4642 (2019).

49. Rohlin, O., Taeri, T., Netzereab, S., Ullemark, E. & Djärv, T. Duration of CPR and impact on 30-day survival after ROSC for in-hospital cardiac arrest—A Swedish cohort study. *Resuscitation* **132,** 1–5. ISSN: 18731570 (2018).

50. Funada, A. *et al.* Duration of cardiopulmonary resuscitation in patients without prehospital return of spontaneous circulation after out-of-hospital cardiac arrest: Results from a severity stratification analysis. *Resuscitation* **124,** 69–75. ISSN: 18731570 (2018).

51. Riley, R. D., Lambert, P. C. & Abo-Zaid, G. Meta-analysis of individual participant data: Rationale, conduct, and reporting. *British Medical Journal (Online)* **340,** 521–525. ISSN: 17561833 (2010).

6

Between-centre differences in care for in-hospital cardiac arrest: a prospective cohort study

Benjamin Y Gravesteijn, Marc Schluep, Hester Lingsma, Robert Jan Stolker, Henrik Endeman, Sanne E Hoeks

Critical Care

6.1 | Abstract

6.1.1 | Background

Survival after in-hospital cardiac arrest is poor universally, but current literature shows substantial heterogeneity in the measure. This study aims to evaluate care for patients suffering in-hospital cardiac arrest (IHCA) in the Netherlands by assessing between-hospital heterogeneity in outcomes, and to explain this heterogeneity stemming from differences in case-mix or differences in quality of care.

6.1.2 | Approach and results

A prospective multicentre study was conducted comprising 14 centres. All IHCA patients were included. The variation in structure and process indicators of quality of care and outcomes (in-hospital mortality and cerebral performance category [CPC] scale) was assessed with mixed effects regression with centre as random intercept. Variation was quantified using the median odds ratio (MOR), representing the expected odds ratio for poor outcome between two randomly picked centres. After excluding centres with less than 10 inclusions (2 centres, n=12), 701 patients were included of whom, 218 (32%) survived to discharge. The unadjusted and case-mix adjusted MOR for mortality was 1.19 and 1.05, respectively. The unadjusted and case-mix adjusted MOR for CPC score was 1.24 and 1.19, respectively. In hospitals where personnel received CPR training twice per year, 183 (64.7%) versus 290 (71.4%) patients died or were in a vegetative state, and 59 (20.8%) versus 68 (16.7%) patients showed full recovery ($p<0.001$).

6.1.3 | Conclusion

In the Netherlands, survival after IHCA is relatively high and between-centre differences in outcomes are small. The existing differences in survival are mainly attributable to differences in case-mix. Variation in neurological outcome is less attributable to case-mix. CPR training could potentially prove beneficiary to improve neurological outcome.

6.2 | Background

In-hospital cardiac arrest (IHCA) is a major adverse event in hospitalized patients. Previous studies have documented the incidence of IHCA between 1-6 events per 1000 hospital admissions [1–3], and both short- and long-term survival after IHCA is poor. A meta-analysis yielded a one-year survival rate of 13.4% but showed substantial heterogeneity between studied cohorts. [4]. A US study also showed heterogeneity in incidence and outcomes after IHCA between centres [5]. This observed heterogeneity may be attributed in part to differences in case-mix, or to differences in improvable facets of care (quality of care) at the provider- and hospital-level. In other fields, such as stroke, targeted quality improvement measures have led to improved outcomes [6]. However, it is not known whether outcomes after IHCA can be improved through a similar focus on quality improvement.

Quality of care can be assessed through structures and processes of care, as well as through patient outcomes [7]. Structure of care indicators pertain to hospital-level factors, which apply to all patients. Notable examples of hospital-level structure of care factors relevant to IHCA are availability of advanced-life-support (ALS) trained personnel, cardiopulmonary resuscitation (CPR) training frequency of personnel, assigned roles of specialists in the cardiopulmonary resuscitation team, and the availability of an intensive care physician. These particular structural indicators have been shown to vary substantially between Dutch hospitals [8]. Secondly, there are process of care indicators, which can vary on the patient-level and can easily be acted upon. A potentially relevant process of care indicator for IHCA is the time until ALS is started, at which point the ALS practitioners can provide additional life-sustaining measures: e.g. endo-

tracheal intubation, administration of epinephrine, and potentially initiate extracorporeal life support [9]. A shorter time between IHCA and these interventions could improve short and long-term outcomes. The registration of a rapid response team warning score (RRS) could be an additional relevant process indicator: these scores (the early warning score, EWS; the modified early warning score, the MEWS; the national early warning score, NEWS) may help in identifying patients at-risk for cardiac arrest, in which case extra precautions could be taken [10]. Finally, outcome metrics such as mortality and cerebral performance category (CPC) score at discharge are relevant patient-level quality indicators [11].

This study aims to assess variation in outcomes between hospitals, and to explain heterogeneity in these outcomes by differences in case-mix or by differences in quality of care stemming from structural and procedural metrics.

6.3 | Methods

6.3.1 | Study population

The Resuscitation Outcomes in the Netherlands (ROUTiNE) study is a multicentre prospective study aiming to assess care and outcome of IHCA patients [12]. All patients in the 14 participating hospitals who received CPR (i.e. chest compressions) for IHCA between January 2017 and May 2018 were included in the study. This study period was predetermined in the study protocol, as reviewed by the Institutional Review Board at Erasmus MC. Data was collected on patient demographics and clinical characteristics related to cardiac arrest and post-CPR treatment, according to Utstein and COSCA templates [13, 14]. For the current hospital-based anal-

ysis hospitals that contributed $\leq$ 10 patients will be excluded, because a reliable measurement of 'standard' care could not be inferred from such a small sample size.

Hospital characteristics and structural indicators were assessed with a structured questionnaire as part of an earlier project completed in February 2018. Details of this questionnaire can be found in a prior publication [8]. In the current study, we compared hospital characteristics from our sample to the other hospitals that participated in this questionnaire.

6.3.2 | Definitions

The patient characteristics that were selected as potential confounders were based on existing literature [2, 15]. These factors consisted of pre-arrest patient characteristics indicative of morbidity and frailty, including: age, the Charlson comorbidity index [16], the pre-arrest modified Rankin scale (MRS), and the pre-arrest cerebral performance category (CPC) (see Supplementary Material 2 available online for a description of the scales).

The time to advanced life support (ALS) and the reporting of a rapid response team score (RRS) were included as process indicators. The time to advanced life support was defined as the time between ascertaining circulatory arrest (and consequently starting BLS) and the moment the ALS team arrived, in minutes. Reporting of RRS was defined as any RRS reported during the 24 hours prior to cardiac arrest. Since processes of care indicators are likely embedded in a complex clinical framework, we assumed the causal models for the data as illustrated in Figure 1 of Supplementary Material 1 available online. As structure of care indicators, we investigated the 24/7 availability of an ALS-certified physician or the 24/7 availability of an intensivist (also ALS certified), and whether the training

frequency of CPR for medical staff was at least twice per year. Finally, as outcome indicators, we considered in-hospital mortality and CPC score at discharge separately. The CPC score was measured and analysed ordinally, ranging from 0 (asymptomatic) to 5 (death).

6.3.3 | Statistical analysis

We performed multiple imputation and imputed five datasets under the assumption of missing at random (MAR) for all missing predictor and outcome data, using the MICE package in R [17, 18]. The outcomes were included in the imputation model. For the descriptive analysis, patients of the following two groups were compared: patients who died in-hospital and patients who survived after discharge from hospital. Continuous variables were compared using Mann-Whitney U tests and categorical variables using χ^2 tests or Fisher's exact test where appropriate. A complete case analysis for the main analyses was performed as sensitivity analysis to assess whether the results are sensitive for imputation.

It is not reliable to crudely compare hospitals on these potential process or outcome indicators of quality of care. Due to small sample sizes within hospitals, there is often random variation (noise) between hospitals. Furthermore, a difference in case-mix results in confounding bias. Random variation and confounding bias unjustifiably contribute to the variation between hospitals, and should be adjusted for [19, 20]: assessment of quality of care should reflect the complexity of hospital care [21].

We first used fixed-effects logistic regression to model in-hospital mortality and a proportional odds logistic regression to model the CPC score. The fixed-effects logistic regression model was subsequently extended with a random intercept for each individual centre in order to assess between-

centre variation in outcomes. Including random intercepts also takes into account random variation between centres due to small sample size [19, 20]. The random intercept values of the unadjusted (without the potential confounders) and the adjusted model were compared to assess what part of the variation was attributable to patient characteristics (age, the Charlson comorbidity index, the pre-arrest MRS, pre-arrest CPC). The variation was further quantified using the median odds ratio (MOR). The MOR is the typical odds ratio between two randomly selected centres, when the centre with higher odds is compared to the centre with the lower odds [22]. Moreover, to assess how much of the variation in outcome could be explained by our predefined case-mix variables, the Nagelkerke R^2 was calculated.

To explore the variation in potential process indicators, mixed effects linear (time to ALS) and logistic regression (registration of RRS) were used. Similar to the variation in outcome, the variation between centres was visually assessed by the comparing the adjusted and non-adjusted random intercept values. The MOR (for registration of RRS only) was also calculated.

Moreover, the rankability of the outcome and process indicators is defined as the variation between hospitals that cannot be attributed to chance. To calculate this measure, the following formula was used:

$$\rho = \tau^2 / (\tau^2 + median(\sigma^2))$$

In this formula, τ^2 indicates the variance of the random intercept of centre. σ^2 is the median variance of the fixed-effects (the coefficients for the centres in the model with centre included as categorical predictor). Finally, ρ is a probability and can be interpreted as the proportion of varia-

tion between hospitals not explained by chance. Therefore, this measure quantifies how reliable it is to rank hospitals by this indicator [19].

Finally, the effect of process and structure of care indicators on outcome was assessed. Only outcomes with variation not attributable to differences in case-mix were selected. For the structure of care processes, the previously mentioned causal model (Figure 1 Supplemental Material 1 available online) was assumed. To specify the variables to correct for in our analysis, we used the back-door-criterion to guide what characteristics to include in our regression model [23]. The back-door criterion is fulfilled when no (causal) paths can be drawn from the exposure of interest to the outcome in the assumed causal model. Using this criterion, we adjusted the effect of time to ALS on functional outcome for timing (weekend versus weekdays, night or evening versus day), whether the arrest was witnessed, and whether an RRS was reported. The effect of the reporting of an RRS on outcome could not be investigated in this study. The reason is that we assume that reporting RRS affects outcome by preventing cardiac arrest. The ROUTiNE study only included patients who experienced cardiac arrest. Therefore, we did not include the relevant control group (patients without cardiac arrest). Finally, the outcome of patients treated in centres with certain structure of care indicators were compared using Fisher's Exact test (while combining score 4 – vegetative state, and 5 – dead), because no confounders were assumed between structure of care and outcome.

All analyses were performed using R[1]. Used packages include the lme4 and ordinal package for the random effects models, and the mice package for the multiple imputation framework. Significance was evaluated an alpha level of 0.05.

[1]R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

6.4 | Results

6.4.1 | Descriptive statistics

The ROUTiNE study included 713 patients from 14 different hospitals. Two hospitals included 10 patients or less, so these patients were excluded (n=12). Therefore, this analysis comprises of 701 patients, included in 12 different hospitals. Of the included patients, 230 (33%) survived to discharge, and 12 (1.7%) patients had missing CPC scores at discharge. The median number of inclusions per hospital was 49 (Figure 6.1). The sample of hospitals is primarily comprised of teaching hospitals (83.3%), and most hospitals are located in urban or metropolitan areas (91.7%). Compared to other hospitals in the Netherlands, the hospitals included in this study were more often trauma centres (66.7% versus 26.3%), performed thoracic surgery (41.7% versus 17.2%), and were more often able to facilitate extracorporeal membrane oxygenation (ECMO) life support (50.0% versus 14.3%, see Table 6.1). [8, 12].

Table 6.1: *Characteristics of the studied hospitals, as part of survey research published earlier [24]. If hospitals had multiple locations, only one location is shown: the highest level of care is reported, and if the facilities are present in one location, it is reported as present.*

Characteristic, n (%)	Total N	Not-included centres (N=58)	Included centres (N=12)
General aspects			
Urban area	70		
Metropolitan		22 (37.9)	5 (41.7)
Urban		18 (31.0)	6 (50.0)
Rural			1 (8.3)
Hospital level	70		
University		6 (10.3)	1 (8.3)
Non-teaching		23 (39.7)	1 (8.3)
Teaching		29 (50.0)	10 (83.3)
Hospital size, n beds	70		
<300		23 (39.7)	2 (16.7)
300-600		25 (43.1)	6 (50)
>600		10 (17.2)	4 (33.3)
Availability of			
Emergency department	70	57 (98.3)	12 (100.0)
Trauma centre	69	15 (26.3)	8 (66.7)
Thoracic surgery	70	10 (17.2)	5 (41.7)
Neurosurgery	70	12 (20.7)	4 (33.3)
Aortic surgery	70	38 (65.5)	12 (100.0)
Cardiac care unit	70	57 (98.3)	12 (100.0)
Rapid Response Team	70	57 (98.3)	12 (100.0)
Rapid response system	70	56 (96.6)	12 (100.0)
Type of rapid resonse system	70		
(M)EWS		54 (93.1)	9 (75.0)
NEWS		1 (1.7)	1 (8.3)
Own modified system		1 (1.7)	2 (16.7)
ICU	70	57 (98.3)	12 (100.0)
Level of ICU*	69		
1		19 (33.3)	1 (8.3)
2		24 (42.1)	4 (33.3)
3		14 (24.6)	7 (58.3)
Intensivist 24/7	69	33 (57.9)	5 (41.7)
ECMO	68	8 (14.3)	6 (50.0)
Both vv and va	14	8 (100.0)	5 (83.3)
Mechanical CPR device	70	26 (44.8)	7 (58.3)

Characteristic, n (%)	Total N	Not-included centres (N=58)	Included centres (N=12)
Practice guidelines/adherence			
Targeted temperature	65		
33 ºC		19 (33.3)	1 (8.3)
Both 33 and 36 ºC		24 (42.1)	4 (33.3)
36 ºC		14 (24.6)	7 (58.3)
Mandatory DNR-counselling upon admission	70	51 (87.9)	10 (83.3)
Advanced life support protocol is ERC 2015	70	57 (98.3)	11 (91.7)
No. of CPR training sessions per year	70		
Twice a year		26 (44.8)	4 (33.3)
Once a year		29 (50.0)	8 (66.7)
Less than once a year		3 (5.1)	0 (0.0)
ERC ALS-certified physician available	70	55 (94.8)	12 (100.0)
ERC ALS-certified physician 24/7 available	70	32 (55.2)	10 (83.3)

* See table 2 supplementary material 1 available online for a detailed description of icu level designation in the netherlands. vv = venous-venous; VA = venous-arterial

We compared patients who survived to hospital discharge with patients who died in hospital. Survivors were younger (a median of 67 [56-73], versus 70 [62-77]), more often had normal neurological function prior to hospital admission (CPC score of 0: 192 [85.3%] versus 334 [74.1%]), and had lower Charlson comorbidity index (median of 1 [0-2], versus 2 [0-3]). Survivors sustained IHCA more often at daytime than non-survivors, and patients at daytime also had better neurological outcomes (Table 2, and Figure 2 Supplementary Material 1 available online). In survivors, cardiac arrest was more often witnessed (212 [92.2%], compared to 339 [72%]), possibly because the location of cardiac arrest was more often at the emergency department (survivors: 26 [11.3%], versus non-survivors: 44 [9.3%]), the intensive care unit (40 [17.4%], 65 [13.8%]), and the operation theatre (19 [8.3%], 13 [2.8%]). Also, the first observed rhythm in survivors was more often shockable (102 [44.3%], versus 82 [17.4%], Table 6.2 and Table 1 Supplementary Material 1 available online). Only 22 (3.1%) patients received extracorporeal membrane oxygenation (ECMO) during CPR (ECPR), of which 7 patients survived to hospital discharge and 1 patient survived,

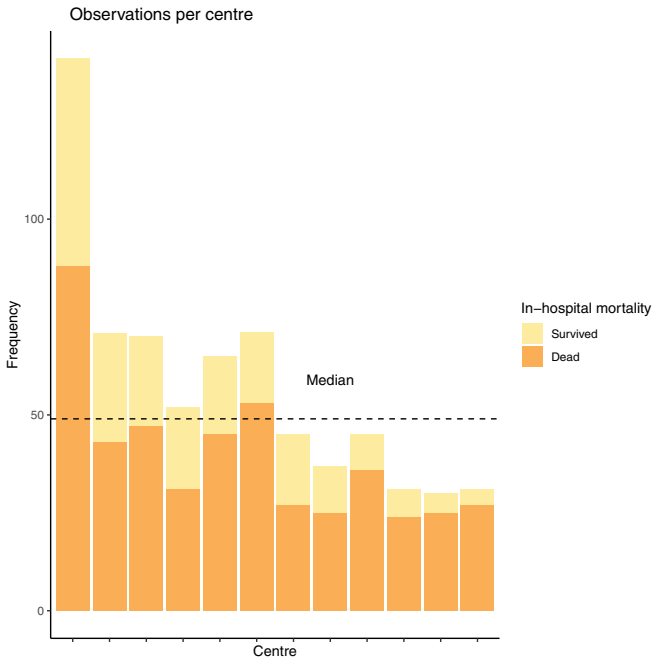


Figure 6.1: *The number of inclusions per participating centre (displayed anonymously) and the primary outcome measure in-hospital mortality per centre.*

but was in a vegetative state (Table 6 Supplementary Material 1 available online).

Table 6.2: *Characteristics of the patients. Twelve patients had missing outcome values and were excluded from this table.*

In-hospital mortality	Observed	Survivors (n = 230)	Non-survivors (n = 471)
Pre-arrest			
Age (median [IQR])	701	67 [56, 73]	70 [62, 77]
Female (%)	701	83 (36.7)	165 (35.0)
Charlson comorbidity score (median [IQR])	701	1 [0, 2]	2 [0, 3]
Pre-arrest CPC (%)	676		
0 – asymptomatic		192 (85.3)	334 (74.1)
1 – Good cerebral performance		26 (11.6)	72 (16.0)
2 – Moderate cerebral disability		3 (1.3)	32 (7.1)
3 – Several cerebral disability		4 (1.8)	12 (2.7)
4 – coma or vegetative state		0 (0.0)	1 (0.2)
Pre-arrest MRS (%)	674		
0 – asymptomatic		100 (45.9)	148 (33.8)
1 – No significant disability		104 (46.6)	155 (34.4)
2 – Slight disability		22 (9.9)	67 (14.9)
3 – moderate disability		19 (8.5)	64 (14.2)
4 – Moderately severe disability		3 (1.3)	14 (3.1)
5 – severe disability		2 (0.9)	3 (0.7)
Any RRS score registered 24 before arrest (%)	701	51 (23.4)	179 (38.0)
Type of RRS score	234		
EWS		20 (36.4)	69 (38.5)
MEWS		13 (23.6)	50 (27.9)
NEWS		3 (5.5)	9 (5.0)
Own system		10 (18.2)	29 (16.2)
Not specified		9 (16.4)	22 (12.3)
Trauma (%)	694	6 (2.6)	14 (3.0)
Sepsis (%)	691	19 (8.3)	65 (14.1)
Reversible diagnosis of arrest (%)	689		
Hypoxia		70 (31.0)	173 (37.4)
Hypovolemia		37 (16.4)	83 (17.9)
Hypothermia		0 (0.0)	0 (0.0)
Hypo-/Hyperkalemia/metabolic		8 (3.7)	22 (4.8)
Tamponade		8 (3.7)	25 (5.4)
Thrombo-embolic		86 (38.1)	145 (31.3)
Toxines		15 (6.6)	11 (2.4)
Tension pneumothorax		2 (0.9)	4 (0.9)

Characteristic	Observed	Survivors (n = 230)	Non-survivors (n = 471)
Hypotension before the arrest (%)*	649		
Yes		32 (15.0)	69 (15.8)
Yes, with vasopressors		8 (3.8)	32 (7.3)
No		173 (81.2)	335 (76.8)
Location (%)	701		
Ward		77 (33.5)	240 (51.0)
Emergency department		26 (11.3)	44 (9.3)
Intensive care unit		40 (17.4)	65 (13.8)
Cardiac care unit		28 (12.2)	54 (11.5)
Interventional radiology theatre		15 (6.5)	25 (5.3)
Operation theatre		19 (8.3)	13 (2.8)
Other		8 (3.5)	5 (1.1)
During arrest			
Shockable rhythm (%)	701	102 (44.3)	82 (17.4)
Witnessed arrest (%)	701	212 (92.2)	339 (72.0)
Time of day	678		
Day ($08:00-16:00$), (%)		68 (30.2)	168 (37.1)
Evening ($16:00-22:00$), (%)		127 (56.4)	208 (45.9)
Night ($22:00-08:00$), (%)		30 (13.3)	77 (17.0)
Time to ALS (median [IQR])	694	2 [0, 3]	2 [1, 4]
CPR duration, ROSC	395	5 [2, 10]	10 [5, 20]
CPR Duration, no ROSC	306	-	30 [21, 50]

* Not defined, subjectively reported by each registrar

6.4.2 | Outcome

All considered pre-arrest patient characteristics were independently associated with in-hospital mortality, except for pre-arrest MRS (OR: 1.10 per unit increase, 95% CI: 0.93 – 1.31). Similar effects were found on the ordinal CPC score (table 6.3). For in-hospital mortality, the explained variance (Nagelkerke R^2) of the model with these predefined predictors was

9.6%. For CPC score, the Nagelkerke R2 was 8.4%. A complete case analysis showed similar results (Table 4, Supplementary Material 1 available online).

Table 6.3: *The results of logistic regression models with outcome as an independent variable, and baseline characteristics as dependent variables. The considered outcomes were in-hospital mortality, and CPC score (worse neurological outcome). An odds ration above one indicates a higher chance of mortality, or a higher chance of a worse CPC score.*

	In-hospital mortality	Worse neurological outcome (CPC)
Charlson comorbidity index	1.17 (1.08 - 1.27)	1.16 (1.07 - 1.26)
MRS score at baseline	1.10 (0.93 - 1.31)	1.11 (0.94 - 1.31)
CPC score at baseline	1.43 (1.04 - 1.95)	1.55 (1.14 - 2.12)
Age, per decade	1.25 (1.10 - 1.41)	1.22 (1.08 - 1.37)

Including a random intercept in these models enables capturing the variation attributable to centre. There was small variation in mortality (median odds ratio [MOR] was 1.19), which decreased by 12% by adjusting for case-mix (adjusted MOR was 1.05). There was moderate variation in CPC score (MOR was 1.24), which decreased 4% by adjusting for case-mix (adjusted MOR was 1.19). This implies that variation in mortality was more dependent on patient characteristics than variation in CPC score (figure 6.2). The rankability, however, of these outcome indicators was low: The variation in mortality and CPC score was for 1.0% and 12% not attributable to chance, respectively (figure 2, supplementary table 1 available online).

6.4.3 | Processes of care

There was little variation in time to ALS across the patient cohort, and this did not change substantially after adjusting for case-mix (Figure 6.3, left panel). The longest times to ALS were observed in two centres, in which the ALS team arrived 1.9 and 1.6 minutes later than average. The

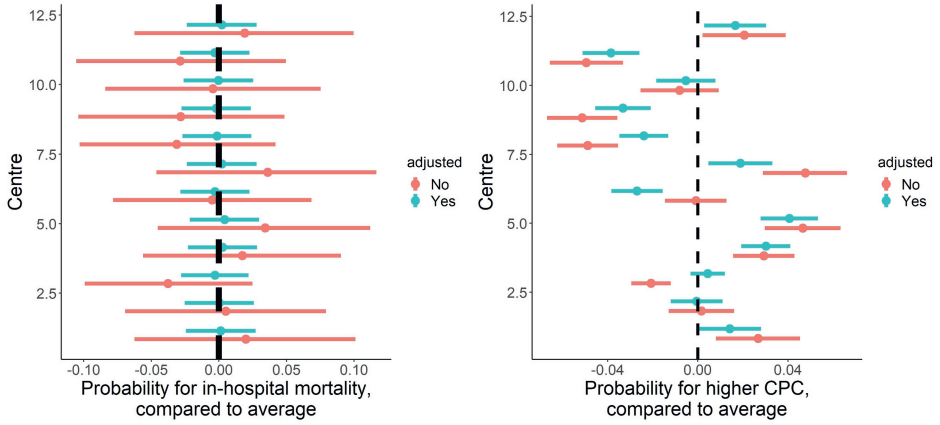


Figure 6.2: *The individual effects of each centre on outcome indicators: mortality on the left, and CPC score on the right. The estimates are the random intercept values of a mixed effects model including the predictors in table 6.3.*

rankability of this indicator was high: 79% of the variation between centres was not attributable to chance (table 2, supplementary material 1 available online). There was no evidence that higher time to ALS increases the odds of a worse CPC score (OR: 0.99, 95% CI: 0.92 – 1.07).

The variation in the reporting of an RRS was large and did not change substantially after adjusting for case-mix (figure 6.3, right panel). The adjusted median odds ratio (MOR) was 2.95. The rankability of this indicator was high: 77% of the variation between centres was not attributable to chance (table 2, supplementary material 1 available online).

6.4.4 | Structure of care

Hospitals which provided CPR training twice a year had a better functional outcome (Figure 6.4, Table 5 and 7 Supplementary Material 1 avail-

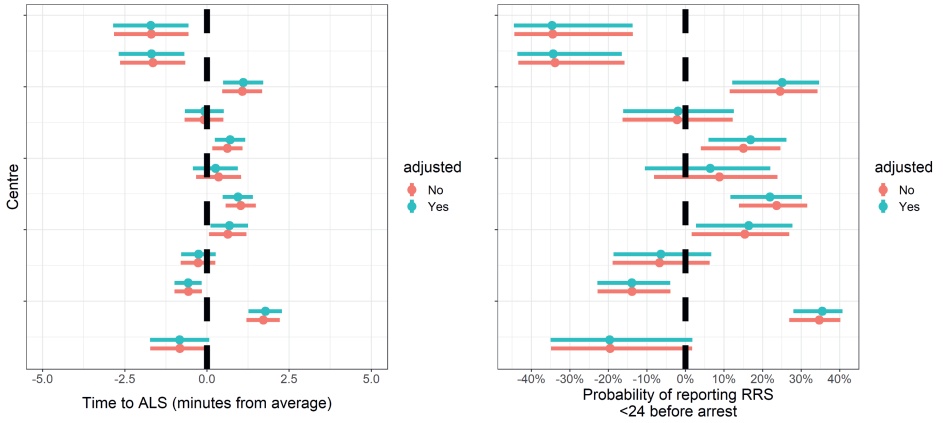


Figure 6.3: *The individual effects of each center on process indicators: time to ALS on the left, and reporting any RRS score <24 hours before arrest on the right. The estimates are the random intercept values of a mixed effects model including the predictors in table 6.3.*

able online): 183 (64.7%) versus 290 (71.4%) patients died or were in a vegetative state, and 59 (20.8%) versus 68 (16.7%) patients showed full recovery ($p < 0.001$). However, patients in hospitals where personnel was trained twice per year were younger (66 [IQR 56-74], versus 71 [IQR 63 – 78]), and had better initial CPC scores (229 [82.4%] had a CPC score of 0, versus 297 [74.6%], Table 8, Supplementary Material 1 available online). The 24/7 availability of an intensivist showed a similar trend towards more favourable CPC scores, but the effect was not significant.

6.5 | Discussion

In this study we first assessed whether there is substantial variation in outcomes between hospitals in the Netherlands after IHCA. We found small to

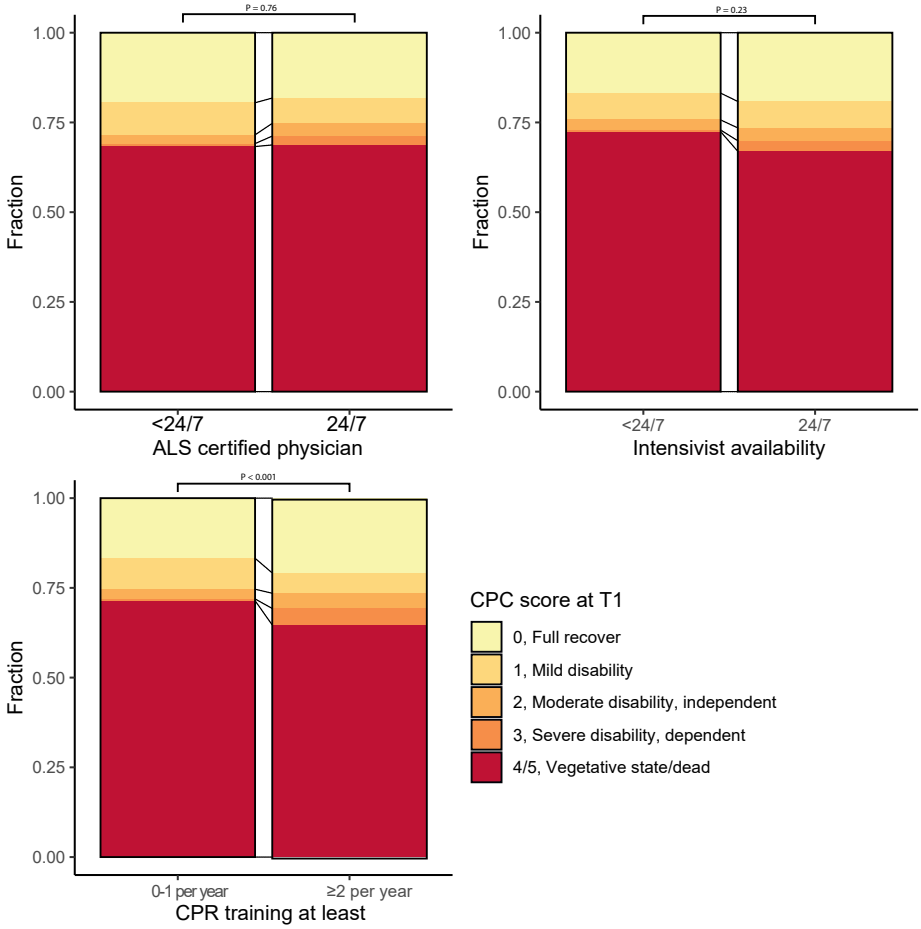


Figure 6.4: *The CPC scores at discharge, stratified per investigated structure of care indicator. The p-value as a result of a Fisher's Exact test are displayed above the barcharts. Only patients with known CPC scores are included. For the absolute numbers, see table 5, supplementary material 1 available online.*

moderate variation in mortality and functional outcomes. Between-centre differences in mortality rates could largely be explained by case-mix, but between-centre differences in CPC scores at discharge persisted after adjustment for case-mix.

To potentially improve functional outcomes, we investigated the reliability and relevance (in terms of association with outcomes) of processes and structure of care indicators. The reliability of the two process indicators was high, but their relevance could not be established with current data. We could not establish this relevance either due to the design of our study, or because our data did not provide evidence against the null hypothesis. In general, quality of care does not often significantly explain variation in outcomes, because treatment effects are generally modest and not all processes of care apply to all patients [25, 26]. However, our data did suggest a positive effect of a structure of care indicator: offering multiple CPR trainings per year to personnel was associated with better functional outcomes of survivors at discharge.

The group of included centres consisted of teaching hospitals with more extensive facilities than the typical Dutch hospital. Within this group of centres, there was little variation in both mortality as well as CPC score. This finding is in contrast with a U.S. study, which described substantial variation in outcome between centres [5]. One explanation is that this study included a much broader range of hospital levels, while our sample mainly includes teaching hospitals.

Nevertheless, the finding that the observed variation in mortality is explained by differences in case-mix can be seen as a strong indication for a cohesive hospital system with uniform adherence to guidelines carried out by highly-trained personnel. We should consider the possibility that that participating hospitals might have performed better, or re-

ported selectively, simply because they were observed within this study (the Hawthorne effect) [27]. Nevertheless, we hypothesize that the homogeneity in quality of care is an important explanation why survival in our population is higher than described in literature [4, 28].

On the contrary, the variation in CPC score could not be entirely explained by differences in case-mix. It can be argued that the explained variance of our models was not high enough. Although the Nagelkerke R^2 is lower than other prognostic studies in cardiac arrest [29, 30], it is known that R^2 measures for categorical outcomes are much lower than those of continuous outcomes [31]. Also, because our aim was to explain (and not to predict) outcomes [32], we think this finding has important implications for cardiac arrest care in the Netherlands: improving care might not improve survival rates, but it might improve functional outcomes. We recommend that other hospital systems identify local processes and structures of care indicators and enact appropriate improvements that could lead to better patient outcomes.

Although the reliability for the investigated processes and structure of care indicators was high, only the relevance for structure of care indicators could be confirmed with the current study. We will here discuss the investigated processes and structure of care indicators, and the implication of our evaluation.

First, we found an indication that CPR training frequency of twice per year might improve functional outcomes. However, patients in centres who trained twice per year were younger and had slightly better pre-existing neurological status, coincidentally. Nevertheless, as only 45% of the Dutch hospitals are described to offer CPR training twice per year [8], increasing adherence to this structure of care indicator could result in improvements in outcome: decreasing intervals between CPR training in-

creases CPR quality in terms of compression depth and rate, and complete chest recoils [33, 34].

Second, our results did not suggest that 24/7 availability of intensivists improves outcomes, in spite of evidence to the contrary [35–37]. We believe that the 24/7 availability of intensivists could indeed improve neurological outcomes, but that our study lacks sufficient power to detect an effect due to the small number of included centres with an intensivist 24/7. With 24/7 intensivist coverage, similar mortality between weekdays and weekends have been reported [38, 39]. If neurological outcomes indeed also improve by such a system, this would add another argument in favour of 24/7 availability of intensivists. It might be hypothesised that we would have found a significant effect if we would have included more hospitals without 24/7 availability of intensivists.

Third, the absolute variation in time to ALS was limited, but consistent and reliable: the rankability was more than the 70% threshold that is suggested as reasonable for quality indicator to be valid [40]. The effect on outcome, however, could not be established: the assumed mechanism through which a lower time to ALS improves functional outcome is by enabling early treatment of reversible causes [41]. We recommend that future studies register whether a reversible cause was present, and whether this was effectively resolved, to better establish the relevance of this process indicator.

Fourth, the reporting of an RRS varied substantially between hospitals, and was again a reliable process indicator. The presumed effect of RRS on outcomes, however, primarily impacts outcomes through preventing cardiac arrest [10]. Therefore, a study which only includes patients with cardiac arrest cannot evaluate the relevance of this indicator. Nevertheless, as other studies have showed evidence for effective prevention of cardiac

arrest [10, 42], our results mainly indicate that the implementation of these scores in clinical practice could be more stringent.

This study is limited because we study a selected group of centres due to logistical reasons. The observed variation in outcome could partly be explained by case-mix in these centres, but perhaps this cannot be generalised to all centres. Fortunately, we collected data about characteristics of these centres and were able to compare our sample's characteristics to those of the universe of hospitals in the Netherlands. Because we are transparent about these differences, the data can be interpreted with more context.

Another limitation of our study is the presence of missing data. We dealt with missing data by using multiple imputations. Using this method we have assumed that the data was missing at random. Unfortunately, there is no empirical way to check this assumption. The fact that a complete case analysis showed same direction and uncertainty of effects is reassuring.

Finally, we only were able to assess the process and structure of care indicators which we collected in this study. Other potential process indicators are the time to defibrillation in patients with IHCA by shockable rhythm, or time to BLS. Both indicators were not (accurately) collected, and therefore could be of interest in future studies. That is, if unexplained differences in outcome are found between centres.

This study introduces metrics for the evaluation and improvement of resuscitation care. Notable strengths of our study include the large sample size and the comprehensive adjustment for both random variation and case-mix. Based on our findings, the following two recommendations for clinical management and research for IHCA can be proposed:

1. we should improve care for IHCA mainly to improve neurological outcomes, i.e. through more frequent CPR training of staff;
2. existing outcome measures of IHCA cannot be reliably used to compare hospitals on quality of care, as opposed to processes and structure of care indicators.

6.5.1 | Conclusion

In our sample of Dutch hospitals, the variation in both mortality and neurological outcome is not substantial after cardiopulmonary resuscitation for in-hospital cardiac arrest. Survival is relatively high and mainly attributable to differences in case-mix, rather than differences in quality of care. The variation in neurological outcome was less attributable to case-mix, suggesting that improvements in care can lead to better neurological outcomes. Multiple CPR trainings per year could be a way forward to improve care for in-hospital cardiac arrest patients. Finally, this study provides an exemplary framework for the evaluation of resuscitative care and the identification of improvable facets of resuscitative care.

6.6 | Supplementary Material

Supplementary materials can be found in the online version of this article: `https://tinyurl.com/2rjfs77a`

References

1. Skogvoll, E, Isern, E, Sangolt, G. K. & Gisvold, S. E. In-hospital cardiopulmonary resuscitation. 5 years' incidence and survival according to the Utstein template. *Acta anaesthesiologica Scandinavica* **43,** 177–84. ISSN: 0001-5172 (1999).
2. Sandroni, C, Nolan, J, Cavallaro, F & Antonelli, M. In-hospital cardiac arrest: Incidence, prognosis and possible measures to improve survival. *Intensive Care Medicine* **33,** 237–245 (2007).
3. Hodgetts, T. J. *et al.* Incidence, location and reasons for avoidable in-hospital cardiac arrest in a district general hospital. *Resuscitation* **44,** 115–123 (2002).
4. Schluep, M., Gravesteijn, B. Y., Stolker, R. J., Endeman, H. & Hoeks, S. E. One-year survival after in-hospital cardiac arrest: A systematic review and meta-analysis. *Resuscitation* **132,** 90–100. ISSN: 1873-1570 (2018).
5. Bradley, S. M. *et al.* Temporal trends and hospital-level variation of inhospital cardiac arrest incidence and outcomes in the Veterans Health Administration. *American heart journal* **193,** 117–123 (2017).
6. Cadilhac, D. A. *et al.* Quality of Acute Care and Long-Term Quality of Life and Survival: The Australian Stroke Clinical Registry. *Stroke.* ISSN: 15244628 (2017).
7. Donabedian, A. The quality of care: how can it be assessed? *Journal of the American Medical Association* **260,** 1743–1748 (1988).
8. Schluep, M., van Limpt, G. J. C., Stolker, R. J., Hoeks, S. E. & Endeman, H. Cardiopulmonary resuscitation practices in the Netherlands: results from a nationwide survey. *BioMed Central: health services research* **19,** 333 (2019).
9. Perkins, N. *et al.* Principled Approaches to Missing Data in Epidemiologic Studies. English. *American Journal of Epidemiology* **187,** 568–575. ISSN: 1476-6256 (2018).

10. Alam, N. *et al.* The impact of the use of the Early Warning Score (EWS) on patient outcomes: a systematic review. *Resuscitation* **85,** 587–594 (2014).

11. Edgren, E. *et al.* Assessment of neurological prognosis in comatose survivors of cardiac arrest. *The Lancet* **343,** 1055–1059 (1994).

12. Schluep, M. M. *et al.* Long-term survival and health-related quality of life after in-hospital cardiac arrest. *Resuscitation* (2021).

13. Perkins, G. D. *et al.* Cardiac Arrest and Cardiopulmonary Resuscitation Outcome Reports: Update of the Utstein Resuscitation Registry Templates for Out-of-Hospital Cardiac Arrest. *Resuscitation* **96,** 328–340 (2015).

14. Haywood, K. *et al.* COSCA (Core Outcome Set for Cardiac Arrest) in adults: an advisory statement from the International Liaison Committee on Resuscitation. *Circulation* **137,** e783–e801 (2018).

15. Fernando, S. M. *et al.* Pre-arrest and intra-arrest prognostic factors associated with survival after in-hospital cardiac arrest: systematic review and meta-analysis. *British Medical Journal* **367** (2019).

16. Charlson, M., Szatrowski, T. P., Peterson, J. & Gold, J. Validation of a combined comorbidity index. *Journal of clinical epidemiology* **47,** 1245–1251 (1994).

17. Rubin, D. B. *Multiple imputation for nonresponse in surveys* (John Wiley & Sons, 2004).

18. Van Buuren, S. *Flexible imputation of missing data* 415. ISBN: 9781138588318 (CRC Press, 2018).

19. Van Dishoeck, A.-M., Lingsma, H. F., Mackenbach, J. P. & Steyerberg, E. W. Random variation and rankability of hospitals using outcome indicators. *British Medical Journal: quality & safety* **20,** 869–874 (2011).

20. Lingsma, H. F. *et al.* Comparing and ranking hospitals based on outcome: Results from The Netherlands Stroke Survey. *Quarterly Journal of Medicine* **103,** 99–108. ISSN: 14602725 (2009).

21. Black, N. Time for a new approach to assessing the quality of hospitals in England. *British Medical Journal* **347,** f4421 (2013).

22. Merlo, J. *et al.* A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *Journal of Epidemiol Community Health* **60,** 290–297 (2006).

23. Pearl, J. & Mackenzie, D. *The book of why: the new science of cause and effect* ISBN: 0465097618 (Basic Books, 2018).

24. Schluep, M., Van Limpt, G. J. C., Stolker, R. J., Hoeks, S. E. & Endeman, H. Cardiopulmonary resuscitation practices in the Netherlands: Results from a nationwide survey. *BioMed Central: Health Services Research* **19.** ISSN: 14726963 (2019).

25. Amini, M. *et al.* Improving quality of stroke care through benchmarking center performance: why focusing on outcomes is not enough. *BioMed Central: health services research* **20,** 1–10 (2020).

26. Lingsma, H. F. *et al.* Variation between hospitals in patient outcome after stroke is only partly explained by differences in quality of care: results from the Netherlands Stroke Survey. *Journal of Neurology, Neurosurgery & Psychiatry* **79,** 888–894 (2008).

27. Franke, R. H. & Kaul, J. D. The Hawthorne experiments: First statistical interpretation. *American sociological review,* 623–643 (1978).

28. Zhu, A. & Zhang, J. Meta-analysis of outcomes of the 2005 and 2010 cardiopulmonary resuscitation guidelines for adults with in-hospital cardiac arrest. *The American journal of emergency medicine* **34,** 1133–1139 (2016).

29. Murray, G. D. *et al.* Multivariable prognostic analysis in traumatic brain injury: results from the IMPACT study. *Journal of neurotrauma* **24,** 329–337 (2007).

30. Barry, T. *et al.* Ten years of cardiac arrest resuscitation in Irish general practice. *Resuscitation* **126,** 43–48 (2018).

31. Smith, T. J. & McKenna, C. M. A comparison of logistic regression pseudo R2 indices. *Multiple Linear Regression Viewpoints* **39,** 17–26 (2013).

32. Shmueli, G. To explain or to predict? *Statistical Science* **25,** 289–310. ISSN: 08834237. arXiv: 1101.0891 (2010).

33. Anderson, R., Sebaldt, A., Lin, Y. & Cheng, A. Optimal training frequency for acquisition and retention of high-quality CPR skills: a randomized trial. *Resuscitation* **135,** 153–161 (2019).

34. Oermann, M. H., Krusmark, M. A., Kardong-Edgren, S., Jastrzembski, T. S. & Gluck, K. A. Training interval in cardiopulmonary resuscitation. *PLoS One* **15,** e0226786 (2020).

35. Blunt, M. C. & Burchett, K. R. Out-of-hours consultant cover and case-mix-adjusted mortality in intensive care. *The Lancet* **356,** 735–736 (2000).

36. Goh, A. Y.-T., Lum, L. C.-S. & Abdel-Latif, M. E.-A. Impact of 24 hour critical care physician staffing on case-mix adjusted mortality in paediatric intensive care. *The Lancet* **357,** 445–446 (2001).

37. Benoit, M. A. *et al.* Postoperative complications and outcomes associated with a transition to 24/7 intensivist management of cardiac surgery patients. *Critical care medicine* **45,** 993–1000 (2017).

38. Arabi, Y., Alshimemeri, A. & Taher, S. Weekend and weeknight admissions have the same outcome of weekday admissions to an intensive care unit with onsite intensivist coverage. *Critical care medicine* **34,** 605–611 (2006).

39. Brunot, V. *et al.* Mortality associated with night and weekend admissions to ICU with on-site intensivist coverage: results of a nine-year cohort study (2006-2014). *PLoS One* **11,** e0168548 (2016).

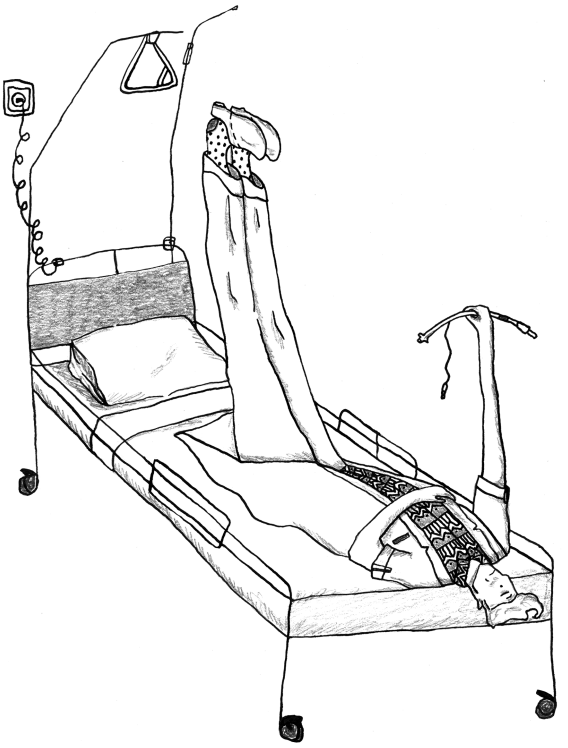
40. Lingsma, H. F. *et al.* Comparing and ranking hospitals based on outcome: results from The Netherlands Stroke Survey. *Quarterly Journal of Medicine: An International Journal of Medicine* **103,** 99–108 (2010).

41. Soar, J. *et al.* European Resuscitation Council Guidelines for Resuscitation 2015. *Resuscitation* **95,** 100–147. ISSN: 03009572 (2015).

42. Moon, A, Cosgrove, J., Lea, D, Fairs, A & Cressey, D. An eight year audit before and after the introduction of modified early warning score (MEWS) charts, of patients admitted to a tertiary referral intensive care unit after CPR. *Resuscitation* **82,** 150–154 (2011).

Part II

Best practice



7

Tracheal intubation in traumatic brain injury: a multicentre prospective observational study.

Benjamin Y Gravesteijn, Charlie A Sewalt, Daan Nieboer,David K Menon, Andrew IR Maas, Fiona E Lecky, Markus Klimek, Hester F Lingsma, and the CENTER-TBI collaborators

British Journal of Anaesthesia

7.1 | Abstract

7.1.1 | Background

We aimed to study the associations between pre- and in-hospital tracheal intubation and outcomes in traumatic brain injury (TBI), and whether the association varied according to injury severity.

7.1.2 | Methods

Data from the international prospective pan-European cohort study, Collaborative European NeuroTrauma Effectiveness Research for TBI (CENTER-TBI), were used (n = 4509). For prehospital intubation, we excluded self-presenters. For in-hospital intubation, patients whose tracheas were intubated on-scene were excluded. The association between intubation and outcome was analysed with ordinal regression with adjustment for the International Mission for Prognosis and Analysis of Clinical Trials in TBI variables and extracranial injury. We assessed whether the effect of intubation varied by injury severity by testing the added value of an interaction term with likelihood ratio tests.

7.1.3 | Results

In the prehospital analysis, 890/3736 (24%) patients had their tracheas intubated at scene. In the in-hospital analysis, 460/2930 (16%) patients had their tracheas intubated in the emergency department. There was no adjusted overall effect on functional outcome of prehospital intubation (odds ratio=1.01; 95% confidence interval, 0.79 - 1.28; $P = 0.96$), and the adjusted overall effect of in-hospital intubation was not significant (odds ratio=0.86;

95% confidence interval, 0.65 - 1.13; $P = 0.28$). However, prehospital intubation was associated with better functional outcome in patients with higher thorax and abdominal Abbreviated Injury Scale scores ($P = 0.009$ and $P = 0.02$, respectively), whereas in-hospital intubation was associated with better outcome in patients with lower Glasgow Coma Scale scores ($P = 0.01$): in-hospital intubation was associated with better functional outcome in patients with Glasgow Coma Scale scores of 10 or lower.

7.1.4 | Conclusion

The benefits and harms of tracheal intubation should be carefully evaluated in patients with TBI to optimise benefit. This study suggests that extracranial injury should influence the decision in the prehospital setting, and level of consciousness the decision in the in-hospital setting.

Editor's key points

1) It is difficult to know whether to intubate and institute mechanical ventilatory support for those with traumatic brain injuries.
2) This large observational study suggests that the indications for tracheal intubation in the setting of traumatic brain injury should be the extent of extracranial injury and the severity of brain injury.
3) Patients with extensive extracranial injury might benefit from intubation before arrival at the hospital.
4) Those with impaired level of consciousness as assessed by the Glasgow Coma Scale might benefit from tracheal intubation shortly after they arrive at the hospital.

7.2 | Introduction

The burden of traumatic brain injury (TBI) is high: it is a leading cause of injury-related death and disability [1]. TBI is estimated to be responsible for 287.2 hospital admissions and 11.7 deaths per 100,000 persons per year in Europe [2]. Mortality rates are higher for moderate and severe TBIs compared with mild TBIs. Although the primary injury arising at the time of impact cannot be mitigated, secondary brain injury arising from subsequent hypoxaemia and hypotension worsens outcome and should be prevented [3–5].

Hypoxaemia and hypotension are both influenced by intubation; tracheal intubation in patients who are not deeply comatose requires induction of anaesthesia and neuromuscular block [6, 7]. However, injudicious use of anaesthetics and positive pressure ventilation can cause hypotension, particularly in hypovolaemic trauma patients [8]. Meanwhile, inadequate depth of anaesthesia during laryngoscopy may precipitate hypertension and (further) increase of intracranial pressure (ICP) [9]. Drug-assisted intubation can be technically challenging in patients with TBI, particularly under prehospital conditions. Under these conditions, positioning and lighting may be suboptimal. If there is also associated facial injury present, the risks of a 'can't intubate can't ventilate' scenario, or oesophageal intubation, are not negligible. Failure to rapidly control the airway owing to delayed or unsuccessful intubation attempts may lead to, or worsen, hypoxia or hypercapnia. These secondary insults are associated with worse outcomes for TBI patients, and may be mitigated or contributed to by decisions to intubate [4, 10–13].

The international guidelines of the Brain Trauma Foundation on intubation in TBI [6] recommend intubation for patients with more severe in-

juries. However, the body of evidence underlying this recommendation consists of only class III evidence, mostly from small retrospective studies. The exception is a randomised trial by Bernard et al. [14] showing benefit of prehospital versus in-hospital intubation in injured prehospital patients with a Glasgow Coma Scale (GCS) score $\leq$ 9. These data have driven recommendations and practice: more severely injured patients, typically with a GCS score of 8 or lower, are intubated more often [15]. However, the primarily observational associations that underpin this practice recommendation are prone to 'confounding by indication' bias.

Possibly partly as a result of the low quality of evidence, guideline adherence varies [16]. For prehospital intubation (PHI), the estimate lies about 80% adherence, but a large range of 44% – 92% adherence is observed in the literature [17, 18]. There is a need for prospective evidence, sufficiently adjusting for confounding bias.

The aim of this prospective study was to improve evidence supporting the guideline recommendations regarding PHI and in-hospital intubation (IHI). Given the practice variation in intubation, we wanted to assess the effect of intubation both at the patient level and at the trauma system level. In addition, given the guideline recommendations to intubate more severely injured patients, we explored whether GCS score and extracranial injury influence the effect of intubation on functional outcome. Finally, we wanted to replicate the RCT by Bernard et al. [14] in the European setting, by comparing outcome of PHI versus intubation at the emergency department (ED) in patients whose tracheas were intubated.

7.3 | Methods

This study was reported according to STROBE (Strengthening The Reporting of OBservational Studies in Epidemiology) guidelines [19].

7.3.1 | Study population

We studied patients who were included in the European, prospective, longitudinal cohort study, Collaborative European NeuroTrauma Effectiveness Research for Traumatic Brain Injury (CENTER-TBI). In this study, data from 4509 all-severity TBI patients in 59 centres throughout Europe had been collected in the period 2014 - 2018 and were available for analysis. Further details of the CENTER-TBI study, including rationale for sample size, have been published elsewhere [20, 21]. A predetermined analysis plan was approved by the management committee before the actual analysis started.

7.3.2 | Patient selection

We excluded patients in whom intubation could not have been considered. For PHI, we therefore excluded patients who arrived to the study hospital without activating emergency medical services (self-presenters). For the IHI analysis, we excluded patients whose tracheas were already intubated on scene.

7.3.3 | Definitions

PHI was defined as intubation at the scene of injury. IHI was defined as intubation at the ED of the study hospital, or intubation at the referring

hospital if the patient was transferred. Intubation could include intubation with and without sedation. The best prehospital GCS score was used for the analysis of PHI and for the analysis of PHI versus IHI. The GCS score at ED arrival was used for the analysis of IHI. The baseline GCS score was defined as the last GCS score in the ED (after stabilisation). If this was missing, or when the patient was sedated or when the patient's trachea was intubated, a previous measurement moment was used: at ED arrival or prehospital, respectively. Outcome was measured using the Glasgow Outcome Scale – Extended (GOSE) at 6 months after injury, GOSE is an eight-point scale that measures functional outcome after TBI [22].

For risk adjustment, we used variables from the IMPACT (International Mission for Prognosis and Analysis of Clinical Trials in TBI) model [23] including age, GCS score, pupil reactivity, imaging characteristics (traumatic subarachnoid haemorrhage, epidural haematoma, Marshall CT class), physiological parameters at ED arrival (heart rate, systolic blood pressure, oxygen saturation), and also secondary insults during the ER treatment (hypoxia or hypotension at the ED). Hypoxia was defined as a documented PaO_2 below 8 kPa (60 mm Hg), a documented Sao2 below 90%, or both, or in case of clinical suspicion (e.g. cyanosis, apnoea, or respiratory distress) when not documented. Hypotension was defined as a documented systolic blood pressure below 90 mm Hg, or in case of clinical suspicion (e.g. shock or absent brachial pulse) when not documented. Moreover, because extracranial injury is also described as a confounder [24], we also included abbreviated injury severity (AIS) scores of head, spine/chest, abdominal (including pelvis), limbs, and face. Finally, as literature suggests differences in outcome between men and women [25], we assumed sex to be a potential confounder as well.

7.3.4 | Statistical analysis

For the patient-level descriptive analysis, baseline characteristics were compared between the PHI, IHI, and not-intubated (NI) group. Medians and inter-quartile ranges (IQRs) are reported for non-normally distributed variables; for normally distributed variables, means and standard deviations are reported.

Missing data were multiply imputed for the main analyses using the 'mice' package [26]. The missing pattern was assumed to be missing at random. Together with the potential confounders and intubation, GOSE was included in the imputation model. Five imputed datasets were obtained.

To assess the effect of intubation on outcome, proportional odds logistic regression was performed using intubation as independent variable and GOSE as dependent variable, with adjustment for confounders. We allowed for non-linear effects by using restricted cubic splines with three degrees of freedom for heart rate, systolic blood pressure, saturation, and age, and with second-degree polynomials for AIS scores. Finally, to assess whether GCS score, abdominal AIS, or thorax AIS influenced the effect of intubation, interaction terms between these characteristics and intubation were added in a consecutive model. We present the effect of intubation as odds ratios (ORs) for more unfavourable outcome and 95% confidence intervals (CIs). The exception is the presentation of the interaction effect: because the interaction effect is based on the combination of two coefficients (the main effect of intubation and the interaction with injury severity), the interpretation is more complex. Instead, we only present the P-value of the overall test (likelihood ratio test) for interaction.

To investigate the relationship between intubation practice and out-

come at the hospital level, we calculated the adjusted probabilities of intubation based on a multinomial mixed effects regression model. The covariates included in the model were based on previous work [27],and include age, GCS score, anatomical injury scales (head/neck/thorax/chest/face/abdomen), and pupil reactivity. A random intercept for centre, conditional on country, was used to adjust for random variation. Because we used multinomial regression, separate random intercepts for each centre were estimated for both outcomes (PHI and IHI). To define the outcome per centre, we calculated mean GOSE scores per centre. The association between intubation preference and outcome was estimated with linear regression with the random intercepts per centre for IHI and PHI, and IHI or PHI itself as an independent variable and mean GOSE per centre as a dependent variable. An interaction term between the intubation preference and PHI or IHI was included. The coefficient of the model was divided by 10 to calculate the coefficient per 10% increase in adjusted intubation rate. The coefficient for interaction between preference and intubation was added to the main effect. Only centres with more than 20 included patients were included in this analysis.

7.4 | Results

The CENTER-TBI database consists of 4509 patients, included across 59 centres in Europe. Information about intubation was present in a total of 3822 (85%) patients, who came from all participating centres (figure 7.1).

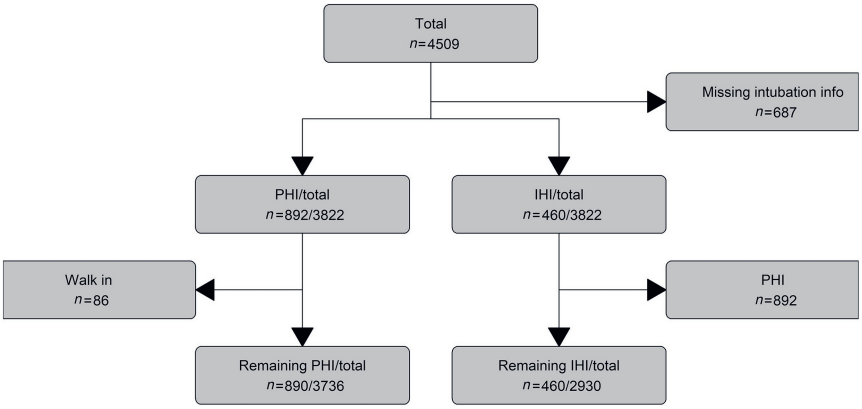


Figure 7.1: *Flowchart showing the number of patients excluded with each criterion. IHI: in-hospital intubation; PHI: prehospital intubation.*

7.4.1 | Prehospital intubation

In the PHI analysis, after excluding patients who self-presented at the ED (n=86), 3736 patients were included. Of these patients, 890 (24%) underwent tracheal intubation on scene. Of 3166 (85%) patients, a GOSE was obtained at 6 months follow-up.

In this PHI subset, 571 (72.4%) of the patients with a prehospital GCS score of 8 or lower had their tracheas intubated on scene and 212 (12%) of the patients with a prehospital GCS score higher than 8 had their tracheas intubated on scene (figure 7.2). On average, patients that had their tracheas intubated had lower baseline GCS score, were younger, and more often male. Furthermore, based on a threshold abbreviated injury scale (AIS) > 3, patients who were intubated had a higher proportion of head and cervical spine injury, major chest/spine injury, and abdominal injury. In addition, patients whose tracheas were intubated had more intracranial pathologies and suffered from more secondary hypoxic and hypotensive

insults in the ED (table 7.1). These differences were smaller when patients with GCS scores above 8 were excluded (Supplementary Table S1). The hospital stay of patients that required PHI was characterised by a longer total length of stay, and a longer ICU stay, and more days of mechanical ventilation and sedation. In addition, pneumonia was observed more often in these patients, and more extracranial and intracranial surgeries, including decompressive craniectomies. Although the absolute ICP values in patients in whom it was measured did not differ substantially on average, the therapy intensity that they received was higher in patients who required intubation. Finally, the blood glucose concentrations were higher in patients who required intubation, both at day 1 as during the entire stay.

Table 7.1: *Baseline table of characteristics of the studied cohort. * = Regional AIS score > 2. ASA: American Society of Anestehsiologists; NI: not intubated; PHI: prehospital intubation; IHI: in-hospital intubation; ISS: injury severity score; RTA: road traffic accident; GCS: Glasgow Coma Scale; mGCS: Glasgow Coma Scale: motor component; ED: emergency department; IQR: inter-quartile range; EDH: epidural haematoma; TSAH: traumatic subarachnoid haemorrhage; MLS: midline shift.*

	PHI (n=890)	NI (n=2846)	Missing (%)	IHI (n=460)	NI (n=2470)	Missing (%)
Age (median [IQR])	44 [25, 60]	52 [33, 68]	0	52 [31, 67]	53 [33, 68]	0
Male (%)	657 (73.8)	1895 (66.6)	0	334 (72.6)	1608 (65.1)	0
Pre-injury ASA physical status			2.6			1.7
1	545 (64.8)	1540 (55.1)		215 (48.8)	1368 (56.1)	
2	227 (27.0)	942 (33.7)		167 (37.9)	803 (32.9)	
3	68 (8.1)	291 (10.4)		49 (11.1)	251 (10.3)	
4	1 (0.1)	24 (0.9)		10 (2.3)	16 (0.7)	
Smoked any time before injury	273 (44.6)	979 (41.7)	20.7	157 (50.0)	851 (40.3)	17.2
Drank alcohol any time before injury	189 (31.3)	809 (34.8)	21.7	112 (36.5)	720 (34.4)	18.1
Major* head injury (%)	851 (95.6)	1960 (68.9)	0	441 (95.9)	1569 (63.5)	0
Major* chest/spine injury (%)	408 (45.8)	436 (15.3)	0	135 (29.3)	303 (12.3)	0
Major* face injury (%)	261 (29.3)	341 (12.0)	0	106 (23.0)	237 (9.6)	0
Major* abdominal injury (%)	139 (15.6)	148 (5.2)	0	40 (8.7)	108 (4.4)	0
Major* external injury (%)	40 (4.5)	45 (1.6)	0	12 (2.6)	33 (1.3)	0
Major* extremity injury (%)	235 (26.4)	356 (12.5)	0	80 (17.4)	277 (11.2)	0
Cause (%)			2			2
RTA	482 (55.5)	1059 (38.1)		173 (39.8)	903 (37.2)	
Fall	284 (32.7)	1306 (47.0)		184 (42.3)	1165 (48.0)	
Other	59 (6.8)	230 (8.3)		41 (9.4)	203 (8.4)	
Violence/suicide	44 (5.1)	186 (6.7)		37 (8.5)	155 (6.4)	
GCS score baseline (median [IQR])	4 [3, 8]	15 [13, 15]	2	8 [5, 13]	15 [14, 15]	2
GCS score prehospital (median [IQR])	6 [3, 9]	14 [13, 15]	36	10 [6, 14]	15 [14, 15]	40
GCS score at ED arrival (median [IQR])	3 [3, 3]	15 [14, 15]	17	8 [5, 12]	15 [14, 15]	12
mGCS score baseline (median [IQR])	1 [1, 4]	6 [6, 6]	1	5 [1, 6]	6 [6, 6]	1
mGCS score prehospital (median [IQR])	3 [1, 5]	6 [6, 6]	36	5 [3, 6]	6 [6, 6]	40
mGCS score at ED arrival (median [IQR])	1 [1, 1]	6 [6, 6]	16	5 [1, 6]	6 [6, 6]	12
Unreactive pupils, baseline (%)			4			5
0	592 (69.6)	2578 (94.7)		355 (81.1)	2293 (97.2)	
1	71 (8.4)	71 (2.6)		33 (7.5)	40 (1.7)	
2	187 (22.0)	74 (2.7)		50 (11.4)	26 (1.1)	

	PHI (n=890)	NI (n=2846)	Missing (%)	IHI (n=460)	NI (n=2470)	Missing (%)
Heart rate at ED arrival, mean (sd)	89 (24)	83 (18)	8	84 (21)	82 (17)	8
SBP at ED arrival, mean (sd)	129 (31)	141 (26)	7	140 (32)	141 (25)	7
Spo 2 at ED arrival, median [IQR]	100 [98, 100]	98 [96, 100]	12	98 [96, 100]	98 [97, 100]	12
Hypoxia at ED (%)	175 (20.6)	105 (3.9)	4	62 (14.9)	45 (1.9)	4
Hypotension at ED (%)	189 (22.2)	94 (3.4)	3	44 (10.4)	51 (2.1)	3
EDH (%)	133 (16.1)	253 (9.6)	7	78 (20.0)	182 (7.8)	6
TSAH (%)	606 (73.2)	1039 (39.3)	7	276 (70.8)	779 (33.5)	6
Marshall CT class (%)			10			9
No visible pathology on CT	77 (9.7)	1143 (44.6)		35 (9.4)	1151 (50.9)	
Cisterns present, MLS <5 mm	390 (48.9)	968 (37.8)		135 (36.1)	850 (37.6)	
Cisterns compressed or absent	110 (13.8)	74 (2.9)		31 (8.3)	43 (1.9)	
Mass lesion	220 (27.6)	376 (14.7)		173 (46.3)	217 (9.6)	
Arrival time (min)	20 [11, 30]	15 [10, 27]	44	14 [8, 24]	15 [10, 27]	44
On-scene time (min)	35 [25, 51]	20 [14, 30]	48	23 [15, 32]	20 [14, 30]	49
Travel time (min)	20 [12, 35]	16 [10, 25]	48	13 [9, 22]	16 [10, 25]	49

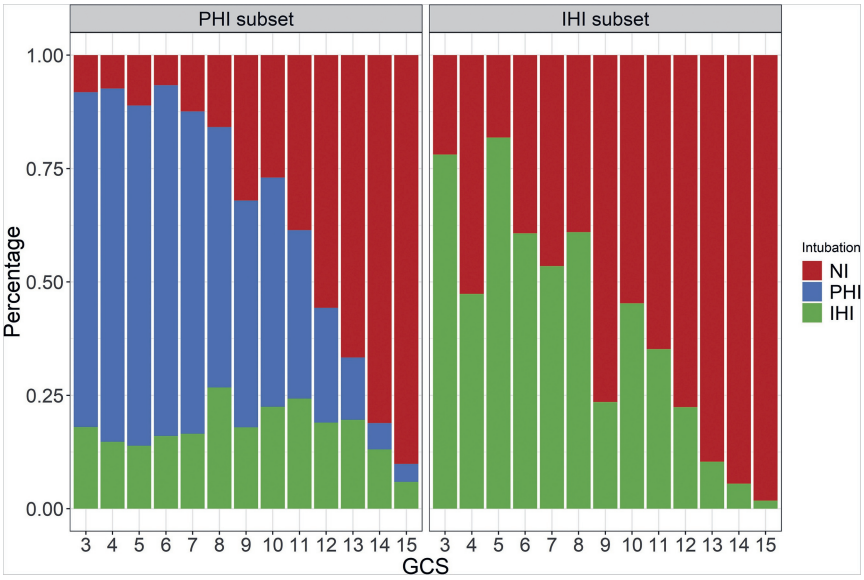


Figure 7.2: *Proportion of non-intubated (NI), prehospitally intubated (PHI) and in-hospital intubated (IHI) patients with a certain Glasgow Coma Scale (GCS) score.*

Before adjusting for possible confounders, PHI was associated with worse functional outcome (OR=6.70; 95% CI, 5.75 - 7.81; P<0.001). After adjustment, there was no evidence of an effect of PHI on functional outcome (OR=1.01; 95% CI, 0.79 - 1.28; P=0.96; table 7.2). The interaction with prehospital GCS score was not significant (P=0.32), but the effect with extracranial injury was significant: PHI was associated with better functional outcome in patients with higher thorax and abdominal AIS scores (P=0.009 for thorax AIS and P=0.02 for abdominal AIS; figure 7.3).

Table 7.2: *Effect of prehospital (PHI) and in-hospital intubation (IHI) on lower functional outcome (GOSE). An odds ratio greater than 1 indicates a higher probability of lower functional outcome (harmful). * For age, sex, baseline GCS, pupil reactivity, heart rate/systolic blood pressure/saturation at arrival, AIS scores of head/spine/abdominal/face regions, traumatic subarachnoid haemorrhage, epidural haematoma, CT class, hypoxia/hypotension at the emergency department. ** Only in patients with GCS $\leq$ 9, who received intubation. GCS, Glasgow Outcome Scale; GOSE, Glasgow Outcome Scale – Extended.*

Intubation	Unadjusted	Adjusted*
PHI	6.70 (5.75 – 7.81)	1.01 (0.79 – 1.28)
IHI	6.13 (5.05 – 7.44)	0.86 (0.65 – 1.13)
PHI versus IHI**	0.87 (0.66 - 1.15)	0.90 (0.65 - 1.23)

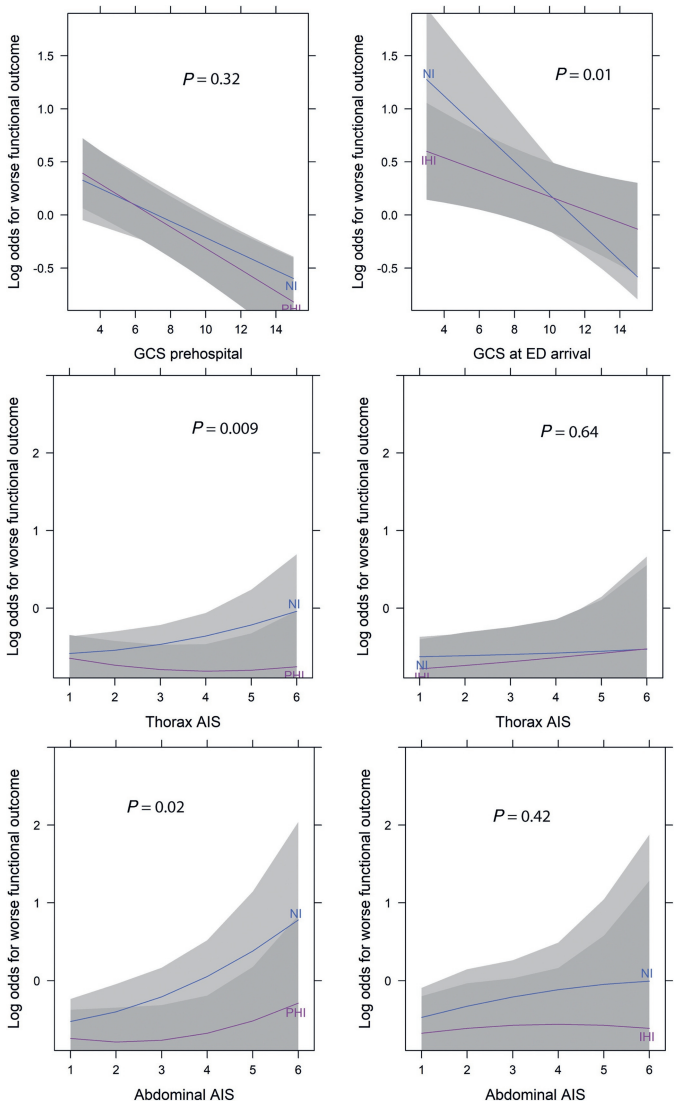


Figure 7.3: *Treatment effect estimates on functional outcome, allowing for interaction of intubation with GCS score, head AIS, and abdominal AIS. The left panel shows the results for prehospital intubation (PHI), and the right for in-hospital intubation (IHI). The effect is displayed for the statistically average patient, with the median (continuous) or mode (categorical) for all other characteristics.*

7.4.2 | In-hospital intubation

In the in-hospital analysis, after excluding patients whose tracheas were intubated on scene, 2930 patients were included (Fig. 1). Of these patients, 460 (16%) patients had their tracheas intubated at the ED. Of 2458 (84%) patients, a GOSE was obtained at 6 months follow-up.

In this IHI subset, 140 (65%) of the patients with a GCS score of 8 or lower at ED arrival had their tracheas intubated at the ED (41 [46%] of these had GOSE scores $\leq$ 4 at 6 months), and 127 (6%) of the patients with a GCS score higher than 8 at ED arrival. On average, they had lower baseline GCS score (figure 7.2). In addition, they were more often male, had a higher proportion of major head injury, and a higher proportion of major extracranial injury. Moreover, patients who had their tracheas intubated had more intracranial pathologies and suffered from more secondary insults (table 7.1). These differences were smaller when patients with GCS scores above 8 were excluded (Supplementary Table S1). The hospital stay of patients that required IHI was characterised by a longer total length of stay, and a longer ICU stay, and more days of mechanical ventilation and sedation. In addition, pneumonia was observed more often in these patients, and more extracranial and intracranial surgeries, including decompressive craniectomies. Although the absolute ICP value in patients in whom it was measured did not differ substantially on average, the therapy intensity that they received was higher in patients who required intubation. Finally, the blood glucose concentrations were higher in patients who required intubation, both at day 1 as during the entire stay.

Before adjusting for confounders, IHI was associated with worse functional outcome (OR=6.13; 95% CI, 5.05 - 7.44; P<0.001). After adjustment, there was no conclusive evidence of an effect of IHI functional outcome

(OR=0.86; 95% CI, 0.65 - 1.13; P=0.28; table 7.2). The interaction with extracranial injury was not significant, but the effect with GCS score was significant (P=0.01): IHI was associated with better functional outcome in patients with GCS scores of 10 or lower at ED arrival (figure 7.3).

7.4.3 | Prehospital *versus* in-hospital intubation

Compared with patients whose tracheas were intubated at the ED, patients with a GCS score ≤ 9 whose tracheas were intubated on scene were younger, had more extracranial injuries, had lower prehospital GCS scores, had more unreactive pupils, and suffered more from secondary insults. Moreover, the median arrival time was 18 min (IQR, 10 - 29), the median on-scene time was 30 min (IQR, 20 - 45), and the median travel time to the hospital was 18 min (IQR, 11 - 30; Table 1). The crude and adjusted effect of PHI versus IHI was beneficial, but not significant: the crude OR for lower GOSE was 0.87 (95% CI, 0.66 - 1.15), and the adjusted OR for a lower GOSE was 0.90 (95% CI, 0.65 - 1.23). The interaction with injury severity (both GCS score and extracranial injury) was not significant.

7.4.4 | Intubation practice

The intubation rates ranged from 0% to 60% for PHI, and from 2% to 56% for IHI (Supplementary Fig. S1). Higher adjusted intubation rates per hospital were associated with higher mean GOSE scores (figure 7.4). The relationship was not significantly different for PHI or IHI (P=0.34): for every 10% increase in PHI rate, the mean GOSE increased with 0.12 (95% CI, 0.01 - 0.22; P=0.04), whereas for every 10% increase in IHI rate, the mean GOSE increased with 0.19 (95% CI, 0.08 - 0.30; P=0.03).

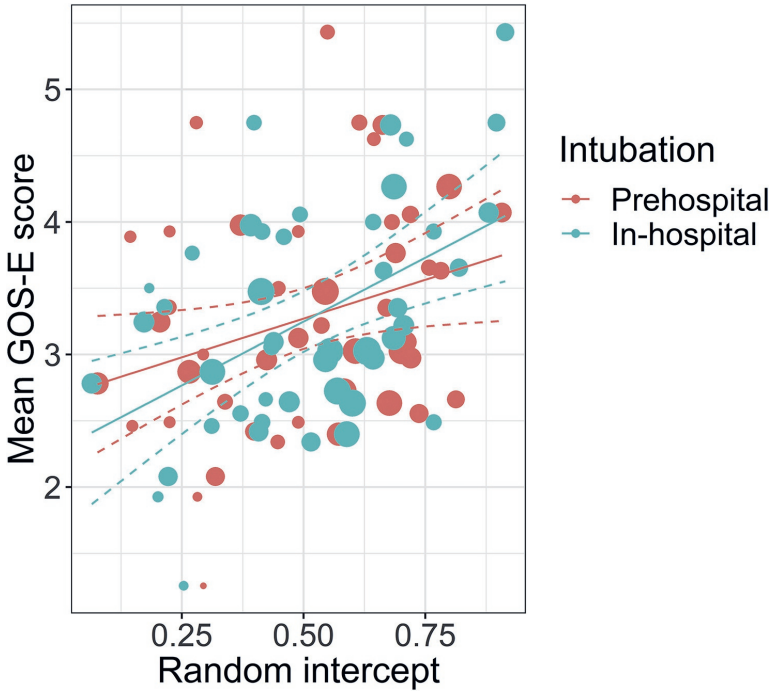


Figure 7.4: *Outcome with centres with different frequencies of intubation. On the x-axis, the values of the random intercept values of the mixed-effects multinomial model are displayed. These can be interpreted as the adjusted intubation rate (the higher the value, the higher the intubation rate). On the y-axis, the mean Glasgow Outcome Scale – Extended (GOSE) for the patients in that centre is displayed. Both prehospital and in-hospital intubation are shown. The sizes of the dots represent the sample size of the centres (corresponding to the inverse variance). The mean and 95% confidence interval (CI) are displayed.*

7.5 | Discussion

This study aimed to provide insight into the effect of intubation on outcome in TBI patients. We performed a patient-level analysis, which is com-

plicated because patients whose tracheas were intubated had sustained more severe trauma. After adjustment for possible confounders, there was no evidence for an overall effect of intubation on functional outcome in TBI patients. Although higher or lower GCS scores did not influence the effect of intubation in the prehospital setting, intubation at the ED seemed to have a more beneficial effect in patients with lower GCS scores. In contrast, higher extracranial injury AIS scores mainly influenced the effect of intubation in the prehospital setting, where intubation was associated with better functional outcome in patients with higher extracranial injury AIS scores. The findings of the RCT by Bernard et al. [14] were not reinforced by our results: PHI was not associated with better functional outcome than IHI. Finally, higher adjusted intubation rates per centre were associated with better functional outcomes.

At the patient level, previous observational studies that assessed the effect of intubation on outcome primarily counterintuitively showed a harm of intubation [19]. Observational studies are inherently prone to confounding bias. In an attempt to adjust for this bias, some recent studies used propensity score matching [28, 29]. These studies also showed an association of intubation with unwanted outcomes in severe TBI patients: these studies found worsened admission oxygenation and even higher mortality. A postintubation surge in ICP or occurrence of hypotension could increase mortality. However, interpreting this relationship as causal is not appropriate, because the purpose of intubation is to secure oxygenation. Rather, these studies are more likely to suffer from residual confounding bias. Our study extensively corrected for potential confounders, which resulted in a large apparent change in the effect of intubation before and after adjustment. Although the effect of intubation was not statistically significant overall, the effect of intubation, especially at the ED, appeared more

likely to be beneficial than harmful. This is in accordance with a study by Davis et al. [24]. This study found a small positive effect of intubation when adjusted for Trauma Score and Injury Severity Score (TRISS). This effect was particularly found in patients who would otherwise be expected to die: those with a very high TRISS score. The finding of a more beneficial effect for more severely injured patients is in accordance with our finding that the benefit of intubation is higher in patients with lower GCS scores and higher extracranial AIS scores. Although this was previously assumed from a physiological perspective [6], it has not been confirmed empirically extensively.

In TBI, particularly in patients with more severe TBI or with extracranial injury that impacts on respiratory physiology, the benefits of intubation appear to outweigh the harms. The potential harms of intubation are mostly associated with the administration of sedatives. These drugs are known to cause vasodilation and therefore hypotension. The latter is known to be associated with worse outcome [30]. In addition, patients whose tracheas are intubated are often hyperventilated [31], which again worsens outcomes [32, 33]. However, hypoxia and aspiration, also known to be harmful [34, 35], can be prevented through intubation. Our results, together with the data from Davis et al. [24], suggest that the prevention of hypoxia and aspiration apparently outweighs the harm of both hypotension and hyperventilation in more severe TBI. We found that the severity of both extracranial and intracranial injuries influence the benefit of intubation. Severity of extracranial injury primarily influences intubation in the prehospital setting, whereas in IHI intracranial injury seems more important: intubation was associated with better functional outcome in patients with a GCS score lower than 10. In our study, only a small proportion of patients with a GCS score higher than 8 received tracheal intubation.

This is in agreement with current Advanced Trauma Life Support (ATLS) guidelines and prior literature, which recommends intubation in patients with a GCS score of 8 or lower. [6] However, based on the current study, shifting the 'intubation threshold' to a GCS score of 10 or lower (especially at the ED) could be considered.

PHI was not found to be more beneficial than IHI, in contrast to the findings of Bernard et al. [14].On one hand, it is possible that our results are biased by confounding by indication and hence may not have been able to demonstrate the beneficial effect of PHI. On the other hand, the benefit of PHI demonstrated in an Australian setting by Bernard et al. [14] might not directly be generalisable to Europe. In Europe, the density of hospitals is higher, which probably results in shorter prehospital times: the travel time (time from departure from scene until arrival in a hospital) in particular was 10 min shorter in CENTER-TBI. The advantage of prehospital versus IHI is that the airway is secured at an earlier phase. In Europe, the difference in time between a secured airway because of PHI versus IHI might be too small to observe a benefit of PHI: the risks of intubating in a less-controlled environment might not be outweighed by the benefits of an earlier secured airway. This hypothesis, however, should be confirmed.

Higher rates of intubation were associated with more favourable outcome. However, this result is not directly applicable to patient-level decision making. Because of ecological bias [36], it should rather be explained by differences in resources. These differences in resources contribute to the large variation in intubation rates [27]. Therefore, this finding should stimulate support in improving current European trauma systems, especially in terms of coverage in appropriate intubation.

A limitation of our study is the observational aspect of our study. In the context of an observational study, it cannot be assumed that confound-

ing bias is entirely corrected for using covariate adjustment. There remains a possibility of unmeasured confounding, which is difficult to overcome. For PHI, in particular, we were not able to adjust for prehospital physiology. Therefore, we recommend future observational studies in this field to meticulously register prehospital physiology, including end-tidal CO2. Nevertheless, the estimates for in-hospital and PHI change similarly after adjustment, which supports our conclusion. The lack of details in the prehospital setting drives another limitation, because it complicates the adjustment for GCS score. For PHI, we adjust for the best prehospital GCS score. However, the most appropriate GCS score to account for the effect of intubation is the GCS score before intubation. There might be some subtle differences in adjustment that might have been missed because of that lack of details.

The size and international aspect of our study support generalisability. Our study also suggests a more liberal GCS score threshold should perhaps influence decisions regarding tracheal intubation, especially when considering IHI.

7.5.1 | Conclusion

At the systems level, higher intubation rates are associated with better functional outcome. This finding probably reflects that more resourced trauma systems have better outcomes. This finding warrants support for developing trauma systems throughout Europe.

At the patient level, intubation does not seem to be associated with better or worse outcome in the general TBI population. However, in more severely injured patients, intubation was associated with better functional outcome. Moreover, patients with TBI and significant extracranial injury

seemed to benefit most from prehospital intubation, whereas the impact of ED intubation was most influenced by GCS score. In addition, in this multicentre study, prehospital intubation was not associated with better functional outcome than ED intubation for patients with TBI.

7.6 | Supplementary Material

Supplementary materials can be found in the online version of this article: `https://tinyurl.com/n8p3hvtx`

References

1. Maas, A. I. R. *et al.* Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. *The Lancet Neurology* **4422.** ISSN: 14744422 (2017).
2. Majdan, M. *et al.* Epidemiology of traumatic brain injuries in Europe: a cross-sectional analysis. *The Lancet Public Health* **1,** e76–e83. ISSN: 2468-2667 (2016).
3. Advanced trauma life support (ATLS): The ninth edition (2013).
4. Green, R. S., Butler, M. B. & Erdogan, M. Increased mortality in trauma patients who develop postintubation hypotension. *Journal of Trauma and Acute Care Surgery* **83,** 569–574 (2017).
5. Manley, G. *et al.* Hypotension, Hypoxia, and Head Injury. *JAMA surgery* **136,** 1118–1123 (2001).
6. Badjatia, N. *et al.* Guidelines for Prehospital Management of Traumatic Brain Injury 2nd Edition. *Prehospital Emergency Care* **12,** S1–S52. ISSN: 1090-3127 (2008).
7. Carney, N. *et al.* Guidelines for the management of severe traumatic brain injury. *Neurosurgery* **80,** 6–15 (2017).

8. Shafi, S. & Gentilello, L. Pre-hospital endotracheal intubation and positive pressure ventilation is associated with hypotension and decreased survival in hypovolemic trauma patients: an analysis of the National Trauma Data Bank. *Journal of Trauma and Acute Care Surgery* **59,** 1140–1147 (2005).

9. Burney, R. G. & Winn, R. Increased cerbrospinal fluid pressure during laryngoscopy and intubation for induction of anesthesia. *Anesthesia and analgesia* **54,** 687–90. ISSN: 0003-2999 (1975).

10. Davis, D. P. *et al.* Both hypoxemia and extreme hyperoxemia may be detrimental in patients with severe traumatic brain injury. *Journal of neurotrauma* **26,** 2217–2223. ISSN: 0897-7151 (2009).

11. Davis, D. *et al.* The Impact of Hypoxia and Hyperventilation on Outcome after Paramedic Rapid Sequence Intubation of Severely Head-injured Patients. *The Journal of Trauma: Injury, Infection, and Critical Care* **57,** 1–10. ISSN: 0022-5282 (2004).

12. Marmarou, A. *et al.* IMPACT Database of Traumatic Brain Injury: Design And Description. *Journal of Neurotrauma* **24,** 239–250. ISSN: 0897-7151 (2007).

13. Stocchetti, N. & Maas, A. I. Traumatic Intracranial Hypertension. *New England Journal of Medicine* **370,** 2121–2130. ISSN: 0028-4793 (2014).

14. Bernard, S. A. *et al.* Prehospital Rapid Sequence Intubation Improves Functional Outcome for Patients With Severe Traumatic Brain Injury. *Annals of Surgery* **252,** 959–965. ISSN: 0003-4932 (2010).

15. Davis, D. P., Aguilar, S, Sonnleitner, C, Cohen, M & Jennings, M. Latency and loss of pulse oximetry signal with the use of digital probes during prehospital rapid-sequence intubation. English. *Prehospital Emerg Care* **15,** 18–22. ISSN: 1545-0066 (2011).

16. Cnossen, M. C. *et al.* Variation in structure and process of care in traumatic brain injury: Provider profiles of European Neurotrauma Centers participating in the CENTER-TBI study. *PLoS ONE* **11,** 1–21. ISSN: 19326203 (2016).

17. Cnossen, M. C. *et al.* Adherence to Guidelines in Adult Patients with Traumatic Brain Injury: A Living Systematic Review. *Journal of Neurotrauma* **14.** ISSN: 0897-7151 (2016).

18. Franschman, G *et al.* Prehospital endotracheal intubation in patients with severe traumatic brain injury: Guidelines versus reality. *Resuscitation* **80,** 1147–1151. ISSN: 0300-9572 (2009).

19. Von Elm, E. *et al.* Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *British Medical Journal* **335,** 806–8. ISSN: 1756-1833 (2007).

20. Maas, A. I. *et al.* Collaborative European neurotrauma effectiveness research in traumatic brain injury (CENTER-TBI): A prospective longitudinal observational study. *Neurosurgery* **76,** 67–80. ISSN: 15244040 (2015).

21. Steyerberg, E. W. *et al.* Case-mix, care pathways, and outcomes in patients with traumatic brain injury in CENTER-TBI: a European prospective, multicentre, longitudinal, cohort study. *The Lancet Neurology* **18,** 923–934 (2019).

22. Weir, J. *et al.* Does the Extended Glasgow Outcome Scale Add Value to the Conventional Glasgow Outcome Scale? *Journal of Neurotrauma.* ISSN: 0897-7151 (2012).

23. Steyerberg, E. W. *et al.* Predicting Outcome after Traumatic Brain Injury: Development and International Validation of Prognostic Scores Based on Admission Characteristics. *PLoS Medicine* **5** (ed Singer, M.) e165. ISSN: 1549-1676 (2008).

24. Davis, D. P. *et al.* Prehospital Airway and Ventilation Management: A Trauma Score and Injury Severity Score-Based Analysis. *The Journal of Trauma: Injury, Infection, and Critical Care.* ISSN: 0022-5282 (2010).

25. Pape, M. *et al.* Is there an association between female gender and outcome in severe trauma? A multi-center analysis in the Netherlands. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine* **27,** 16. ISSN: 1757-7241 (2019).

26. Van Buuren, S. *Flexible imputation of missing data* 415. ISBN: 9781138588318 (CRC Press, 2018).

27. Gravesteijn, B. Y. *et al.* Variation in the practice of tracheal intubation in Europe after traumatic brain injury: a prospective cohort study. *Anaesthesia,* anae.14838. ISSN: 0003-2409. `https://onlinelibrary.wiley.com/doi/abs/10.1111/anae.14838` (2019).

28. Haltmeier, T *et al.* Prehospital intubation for isolated severe blunt traumatic brain injury: worse outcomes and higher mortality. *Eur. j. trauma emerg. surg.* **43,** 731–739. ISSN: 1863-9941 (2017).

29. Karamanos, E. *et al.* Is Prehospital Endotracheal Intubation Associated with Improved Outcomes In Isolated Severe Head Injury? A Matched Cohort Analysis. *Prehospital and Disaster Medicine.* ISSN: 1049-023X (2014).

30. Spaite, D. W. *et al.* Association of out-of-hospital hypotension depth and duration with traumatic brain injury mortality. *Annals of emergency medicine* **70,** 522–530 (2017).

31. Davis, D. P. *et al.* Ventilation patterns in patients with severe traumatic brain injury following paramedic rapid sequence intubation. English. *Neurocrit Care* **2,** 165–171. ISSN: 1541-6933 (2005).

32. Gaither, J. B. *et al.* Balancing the potential risks and benefits of out-of-hospital intubation in traumatic brain injury: the intubation/hyperventilation effect. *Annals of emergency medicine* **60,** 732–736 (2012).

33. Warner, K. J., Cuschieri, J., Copass, M. K., Jurkovich, G. J. & Bulger, E. M. The impact of prehospital ventilation on outcome after severe traumatic brain injury. *Journal of Trauma and Acute Care Surgery* **62,** 1330–1338 (2007).

34. Chi, J. H. *et al.* Prehospital hypoxia affects outcome in patients with traumatic brain injury: A prospective multicenter study. *Journal of Trauma - Injury, Infection and Critical Care* **61,** 1134–1141. ISSN: 00225282 (2006).

35. Benjamin, E. *et al.* Witnessed aspiration in trauma: Frequent occurrence, rare morbidity-A prospective analysis. *Journal of Trauma and Acute Care Surgery* **79,** 1030–1036. ISSN: 21630763 (2015).

36. Greenland, S. & Morgenstern, H. Ecological bias, confounding, and effect modification. *International Journal of Epidemiology* **18,** 269–274. ISSN: 03005771 (1989).

8

Primary versus secondary referral to a specialized neurotrauma centre in patients with moderate/severe Traumatic Brain Injury: a CENTER-TBI study

Charlie A Sewalt, Benjamin Y Gravesteijn, David Menon, Hester F Lingsma, Andrew IR Maas, Nino Stocchetti, Esmee Venema, Fiona E Lecky and the CENTER-TBI Participants and Investigators

Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine

8.1 | Abstract

Prehospital care for patients with TBI varies with some emergency medical systems recommending direct transport of patients with moderate to severe Traumatic Brain Injury (TBI) to hospitals with specialist neurotrauma care (SNCs). The aim of this study is to assess variation in levels of secondary referral within European SNCs and to compare the outcomes of directly admitted and secondarily transferred patients.

Patients with moderate and severe TBI (Glasgow Coma Scale <13) in the prospective European CENTER-TBI study were included in this study. All participating hospitals were specialist neuroscience centres. First, adjusted between-country differences were analysed using random effects logistic regression where secondary referral was the dependent variable, and a random intercept for country was included. Second, the adjusted effect of secondary referral on survival to hospital discharge and functional outcome (6 months GOSE) was estimated using logistic and ordinal mixed effects models, respectively.

A total of 1347 moderate/severe TBI patients from 53 SNCs in 18 European countries were included. Of these 1347 patients, 195 (14.5%) were admitted after secondary referral. Secondarily referred moderate/severe TBI patients presented more often with a CT abnormality: mass lesion (52% versus 34%), midline shift (54% versus 36%) and acute subdural hematoma (77% versus 65%). After adjusting for case-mix, there was a large European variation in secondary referral, with a typical OR of 1.69 between countries. Secondary referral was not significantly associated with functional outcome (adjusted OR: 1.07, 95% CI: 0.78-1.69), nor with survival at discharge (1.05, 0.58 - 1.90).

Across Europe, substantial practice variation exists in the proportion

of secondarily referred TBI patients at SNCs that is not explained by case mix. Within SNCs secondary referral does not impact functional outcome and survival after stabilisation in a non-specialised hospital. Future research should identify which patients with TBI truly benefit from direct transportation.

8.2 | Introduction

Traumatic brain injury (TBI) remains an important cause of injury-related death and disability [1]. The incidence of TBI is increasing as the patient population becomes older [2, 3]. Care in specialized neurotrauma centres (SNC) with neurosurgical and neurocritical care expertise can reduce the incidence of death and disability from head injury, especially in more severe TBI [4–6]. However, not all TBI patients are directly transported to a SNC if this is not the nearest facility. In the prehospital setting Emergency Medical Services should decide whether these patients should be stabilized at the nearby non specialist acute hospital (NSAH) or directly transported to a more distant SNC, generally speaking. After stabilization and CT scan at a NSAH - the decision is made regarding the need for specialist neurotrauma care via secondary transfer. Stabilizing the patient at a nearby NSAH may cause an important time delay to critical neurosurgical and neurocritical care interventions which could adversely affect the outcome of TBI patients [7]. On the other hand prolonged primary transportation to a more distant specialist centre could delay direct access to critical interventions such as drug assisted intubation that can reduce secondary brain injury [8]. This is pertinent particularly to the majority of EMS staff who do not have this advanced airway skill [9]. Early neurosurgery might be a lower priority than early treatment of secondary insults such as hypoxia and hypotension [10] – the latter being addressed by hospital based damage control measures and balanced transfusion. The decision which patients should be conveyed directly to an SNC is made on-scene by EMS staff based on clinical parameters, injury characteristics and the local policy through trauma triage tools [9]. A systematic review on this issue failed to identify clear benefit from direct transportation to SNCs [11]. A recent

randomized trial also failed to identify benefit as the majority of patients who bypass the NSAH are subsequently shown not to have a brain injury on CT scan, diluting the impact of early access to neurotrauma care [12].

Notwithstanding this equivocal evidence base, several international guidelines recommend direct transportation of patients with moderate/severe TBI to hospitals with availability of neurosurgical care in order to reduce the time delay [13–15]. There might be substantial variation in referral practice between regions and countries. It remains unclear how long term outcomes of secondarily referred patients relate to outcomes of patients directly transported to a SNC.

Therefore, the aims of this CENTER-TBI study are:

1. to quantify European practice variation in secondary referrals;
2. to determine the association of arriving by secondary referral with survival at discharge and functional outcome at 6 months.

8.3 | Methods

8.3.1 | Study design

CENTER-TBI is a multicentre, longitudinal, prospective, observational study in 22 countries across Europe and Israel which enrolled patients between December 2014 and December 2017 [16]. All study sites are specialist neurotrauma sites. The core cohort includes patients presenting within 24 hours of injury, with a clinical diagnosis of TBI and indication for computed tomography (CT). Data for the CENTER-TBI study has been collected through the Quesgen e-CRF[1], hosted on the INCF platform and

[1]Quesgen Systems Inc, USA

extracted via the INCF Neurobot tool[2]. We repeated our analysis in the CENTER-TBI registry, comprising of all patients presenting at one of the study s between December 2014 and December 2017 with a clinical diagnosis of TBI and indication for CT scan. For the registry, informed consent was not necessary. Version 2.1 of the core and registry neurobot data sets were used for this study. Prehospital data was collected by physicians at the study sites. Policy and specific data was collected by provider profiling questionnaires, filled in by the leading researchers of each study [17]. Relevant questions from the provider profiling questionnaires to explain regional differences were the existence of a prehospital triage tool concerning direct transportation to more distant specialist neurotrauma centre, and level of education/skills training provided to prehospital staff.

Ethical approval was obtained for each recruiting site. Consent was obtained for all patients enrolled in the core study. The list of sites, Ethical Committees, approval numbers and approval dates can be found on the website: `https://www.center-tbi.eu/project/ethical-approval`.

8.3.2 | Patient selection

We included all patients with moderate/severe TBI (defined as a Glasgow Coma Scale (GCS) < 13 or intubated [18]) who were transported by ambulance or helicopter directly to a study site (SNC) or admitted after secondary referral. A sensitivity analysis was done by including all registry patients with moderate/severe TBI. This study was reported in accordance with the STROBE reporting guidelines [19].

[2]INCF, Sweden

8.3.3 | Definitions

The outcome measures to estimate the effect of secondary referral were survival at discharge and Glasgow Outcome Scale Extended (GOSE) at six months. For cases in which GOSE assessments had been performed outside the pre-specified window of 5 - 8 months (n=988), we used a multistate model to impute the 180-day GOSE [16, 20]. As confounders between the relationship of transfer status and outcome, the following baseline characteristics were used because they were associated with either arriving by secondary referral or part of the IMPACT model: age, GCS motor score, pupil inequality, hypoxia at ED arrival, hypotension at ED arrival, Injury Severity Score (ISS) and several CT abnormalities: traumatic subarachnoid haemorrhage (tSAH), epidural hematoma, mass lesion and acute subdural hematoma [21]. For the analysis in the registry, we used survival at discharge as primary outcome since longer term outcome data were not collected in the Registry.

8.3.4 | Statistical analysis

Continuous variables were described by the median and interquartile range (IQR). Categorical variables were described by the frequency and percentage. Missing data was imputed using multiple imputation, assuming missing at random. All variables, except for the outcome variables survival at discharge and the derived 6 months GOSE, were imputed. However, the outcome variables were included in the imputation model.

First, adjusted between-country differences were analysed by adding a random intercept for country to a logistic regression model with secondary referral as dependent variable. National variation or practice variation was

quantified using the Median Odds Ratio (MOR, typical odds ratio [OR] between two randomly picked countries/centres) [22].

Second, the effect of arriving by secondary referral on hypotension and hypoxia was estimated using random effects logistic regression models. We adjusted for age, GCS motor score, pupil inequality, ISS and a random intercept for study .

Third, the effect of arriving by secondary referral on survival at discharge and functional outcome (6 months GOSE) was estimated using random effects regression models. For in-hospital mortality, a random effects logistic regression model was used, which included the predefined confounders and a random intercept for study . For 6 months GOSE, a random effects ordinal regression model was used with similar structure. A subgroup analysis was done by including patients who presented with either a mass lesion or acute subdural hematoma on CT scan. As a secondary sensitivity analysis, the same analysis was repeated in the CENTER-TBI registry with survival at discharge as outcome measure. A random effects logistic regression model with the same case-mix variables was used with a random intercept for study. Finally, as sensitivity analysis, the main analyses were also repeated in the complete cases.

Statistical analyses were performed in R statistical software 3.5.1 (R Foundation for Statistical Computation, Vienna). The glmer function from the lme4 package was used for mixed effects logistic regression, the clmm function from the ordinal package was used for ordinal mixed effects logistic regression, and multiple imputation was performed using the MICE package.

8.4 | Results

8.4.1 | Patient characteristics

A total of 1347 patients with moderate/severe TBI were included in this study from 53 study centres in 18 European countries. Of these 1347 patients, 195 (14.5%) were transferred from another hospital. The proportion of TBI patients arriving through secondary referrals varied by study from 0% to 71%. The patients secondarily referred to the study were mostly male (146, 74.9%), with a median age of 52 years (IQR 29 - 67), a median GCS of 7 (IQR: 3 - 10), were not often intubated in the prehospital environment (37, 20.7%) and their median ISS was 26 (25 - 41). The patients who were primarily transported to study centres were also mostly male (837, 72.7%), young to middle aged (median age was 47, IQR 28 - 65), with a GCS of 7 (4 - 10), however they were often intubated on-scene (701, 62.1%);their median ISS was 34 (25 - 45) (Table 1, Figure S1 available online). Mode of injury differed between both patient groups where road traffic incidents with extracranial injury were more common in TBI patients arriving by primary referral. When looking at the prehospital characteristics, patients secondarily referred had fewer on scene interventions (e.g intubation, IV fluids) compared to primarily transported patients (table 8.1).

Table 8.1: *Patient characteristics, continuous: median (IQR), categorical: number (%) including percentage missingness for patient characteristics from core dataset ($N = 1347$).*

	Primary referral (N=1152)	% missing	Secondary referral (N=195)	% missing
Patient characteristics				
Male (%)	837 (72.7)	0.0%	146 (74.9)	0.0%
Age (median [IQR])	47 [28, 65]	0.0%	52 [29, 67]	0.0%
Alcohol usage (%)	295 (29.0)	11.6%	71 (42.3)	13.8%
Drugs usage (%)	50 (5.4)	20.3%	10 (7.4)	30.3%
Injury characteristics				
Cause of Injury(%)		10.9%		24.1%
Fall	420 (40.9)		84 (51.2)	
Road Traffic incident	551 (53.7)		60 (36.6)	
Suicide	26 (2.5)		3 (1.8)	
Violence	30 (2.9)		17 (10.4)	
Area of injury = Urban (%)	837 (75.0)	3.1%	143 (76.5)	4.1%
Place of injury (%)		2.3%		5.1%
Home	275 (24.4)		50 (27.0)	
Public location	66 (5.9)		35 (18.9)	
Sport	56 (5.0)		7 (3.8)	
Street	663 (58.9)		86 (46.5)	
Work	66 (5.9)		7 (3.8)	
GCS at arrival ED (median [IQR])	6 [3, 9]	0.0%	7 [3, 10]	0.0%
Hypoxia at arrival ED	186 (16.9)	0.0%	18 (10.3)	0.0%
Hypotension at arrival ED	196 (17.9)	0.0%	20 (11.7)	0.0%
Total ISS (median [IQR])	34 [25, 45]	0.7%	26 [25, 41]	1.0%
Major extracranial injury (AIS >3) (%)	652 (56.6)	0.7%	83 (42.6)	1.0%
Pupil differences at ED (%)		3.4%		6.7%
No pupil difference	809 (72.7)		146 (80.2)	
One pupil not reactive	94 (8.4)		12 (6.6)	
Two pupils not reactive	210 (18.9)		24 (13.2)	

	Primary referral (N=1152)	% missing	Secondary referral (N=195)	% missing
Prehospital care				
Intubation (%)	701 (62.1)	2.1%	37 (20.7)	8.2%
Ventilation (%)	646 (58.0)	3.4%	35 (19.7)	8.7%
CPR (%)	35 (3.0)	0.0%	2 (1.0)	0.0%
Oxygen supply (%)	853 (78.8)	6.1%	97 (65.5)	24.1%
IV Fluids (%)	727 (63.1)	0.0%	65 (33.3)	0.0%
Physician on scene (%)	858 (74.7)	0.3%	86 (44.6)	1.0%
Mode of transport (%)		0.0%		0.0%
Ambulance	769 (66.8)		160 (82.1)	
Helicopter	271 (23.5)		16 (8.2)	
Medical mobile team	112 (9.7)		19 (9.7)	
Time from leaving scene to first hospital (median [IQR])	20 [12, 39]	41.8%	11 [9, 24]	88%
Time to study (median [IQR])	20 [12, 39]	41.8%	205[160, 286]	53%
Local policy characteristics				
Prehospital triage protocol favours direct admission (%)	334 (46.7)	37.9%	28 (27.2)	47.2%
Training of prehospital staff (%)		19.4%		24.1%
BLS only	168 (18.1)		22 (14.9)	
Emergency Medical Technician	590 (63.6)		99 (66.9)	
Nurse	170 (18.3)		27 (18.2)	

	Primary referral (N=1152)	% missing	Secondary referral (N=195)	% missing
Imaging characteristics				
Acute subdural hematoma (%)	700 (64.6)	5.9%	131 (77.1)	12.8%
Traumatic subarachnoid hemorrhage (%)	747 (73.6)	11.9%	114 (76.5)	23.6%
Epidural hematoma (%)	165 (16.2)	11.7%	28 (18.7)	23.1%
Skull fracture (%)	617 (59.9)	10.6%	104 (65.8)	19.0%
Midline shift (%)	377 (35.6)	8.0%	91 (53.8)	13.3%
Cisternal compression (%)	435 (41.6)	9.3%	81 (50.6)	17.9%
Mass lesion (%)	347 (34.1)	11.6%	79 (52.0)	22.1%
Intraventricular hemorrhage (%)	290 (28.6)	12.0%	41 (27.0)	22.1%
Contusion (%)	734 (69.3)	8.1%	134 (79.3)	13.3%
Emergency intracranial surgical intervention (%)	286 (24.9)	0.3%	62 (32.1)	1.0%
Time from injury to emergency surgery (median [IQR])	210 [150, 348]	37%	345 [259, 479]	20.6%
Outcome				
6 Months GOSE (median [IQR])	4.00 [1.00, 6.00]	12.7%	4.00 [1.00, 7.00]	14%
Unfavourable outcome (GOSE <5, %)	512 (44.4)	12.7%	64 (32.8)	14%
In-hospital mortality (%)	190 (21.2)	17.7%	30 (19.4)	16.9%

A total of 1347 patients with moderate/severe TBI were included in this study from 53 study centres in 18 European countries. Of these 1347 patients, 195 (14.5%) were transferred from another hospital. The proportion of TBI patients arriving through secondary referrals varied by study centre from 0% to 71%. The patients secondarily referred to the study centre were mostly male (146, 74.9%), with a median age of 52 years (IQR 29 - 67), a median GCS of 7 (IQR: 3 - 10), were not often intubated in the prehospital environment (37, 20.7%) and their median ISS was 26 (25 - 41). The patients who were primarily transported to study centres were also mostly male (837, 72.7%), young to middle aged (median age was 47, IQR 28 - 65), with a GCS of 7 (4 - 10), however they were often intubated on-scene (701, 62.1%); their median ISS was 34 (25-45) (Table 1, Figure S1 available online).

The median 6 months GOSE was 4 (IQR: 1-6) among primary referred patients and 4 (IQR: 1 - 7) among secondary referred patients. In-hospital mortality was 21.2% among primary referred patients and 19.4% among secondarily referred patients.

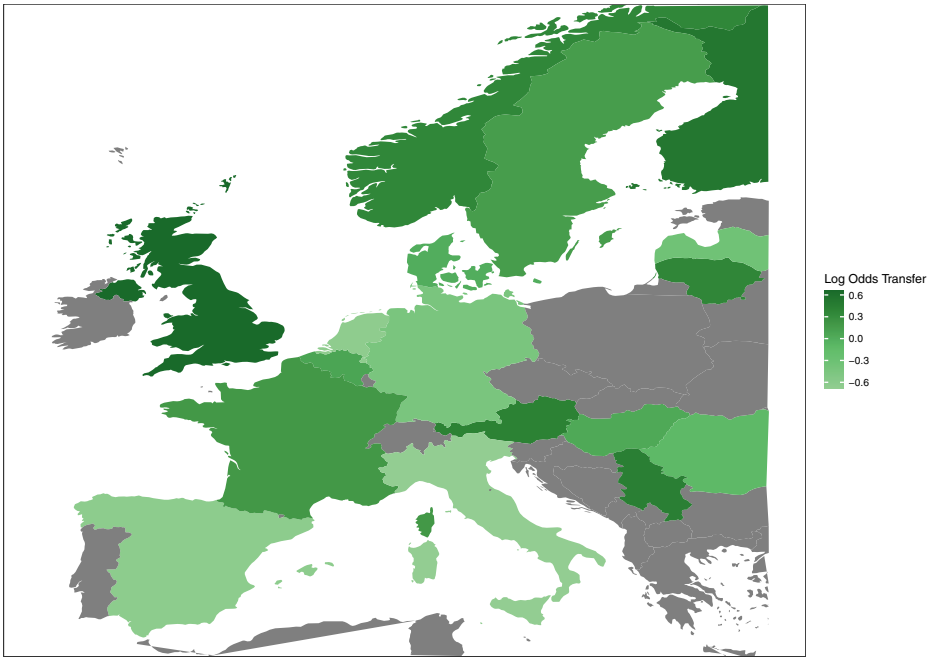


Figure 8.1: *European practice variation in secondary referrals, adjusted for extended IMPACT model (age, GCS motor score, pupil inequality, hypoxia, hypotension, ISS, CT lesions: tSAH, epidural hematoma, mass lesion, acute subdural hematoma). Log Odds represents the chance of arriving by secondary referral for the mean moderate/severe TBI patient compared to the mean European chance of being referred. A log-odds above 0 means more chance than average of arriving by secondary referral, a log odds below 0 means less chance than average of arriving by secondary referral.*

8.4.2 | European practice variation of secondary referrals

When analysing European practice variation, patients admitted to specialist neurotrauma centres in Scandinavian countries, Austria and England were more often secondarily referred (figure 8.1). Patients in the Netherlands and Italy had relatively lower adjusted chance of arriving by secondary referral. The MOR is 1.69 which means that the OR between two randomly picked countries is 1.69 for the average TBI patient included in our study.

8.4.3 | Effect of secondary referral on outcome

There was a non-significant association between type of referral and hypotension and hypoxia at arrival at the SNC (OR 0.57 with direct admission as reference, 95% CI 0.28 - 1.15 for hypoxia and OR 0.72 with direct admission as reference, 95% CI 0.38 – 1.38 for hypotension, table 8.2). Arriving by secondary referral as moderate/severe TBI patient was not significantly associated with 6 month GOSE compared to being directly admitted (multivariable adjustment, OR 1.07 with direct admission as reference, 95% CI 0.78 – 1.46) and there was no significant association between secondary referral and survival at discharge (OR 1.05 with direct admission as reference, 95% CI 0.58 – 1.90). Subgroup analysis of patients with a mass lesion or acute subdural hematoma and patients needing emergency intracranial surgical intervention showed similar magnitude and direction of the effects (table 8.3).

Table 8.2: *Effect of secondary referral on hypotension and hypoxia at arrival at the Emergency Department of the Specialized Neurotrauma centre.*

	Hypoxia OR (95% CI)	Hypotension OR (95% CI)
Unadjusted	0.53 (0.27 – 1.02)	0.65 (0.36 – 1.19)
Multivariable adjustment*	0.57 (0.28 – 1.15)	0.72 (0.38 – 1.38)

*adjusted for: age, GCS motor score, pupil inequality, ISS and a random intercept for centre.

Table 8.3: *Effect of secondary referral on GOSE and survival at discharge. Higher OR for 6 months GOSE means better outcome, while higher OR for survival at discharge means higher chance of survival.*

	6 months GOSE OR (95% CI)	Survival at discharge OR (95% CI)
Unadjusted	1.13 (0.82 – 1.55)	1.04 (0.65 – 1.62)
Multivariable adjustment*	1.07 (0.78 - 1.46)	1.05 (0.58 – 1.90)
Subgroup: patients with mass lesion/ASDH		
Unadjusted	1.64 (1.10 – 2.44)	1.24 (0.67 – 2.32)
Multivariable adjustment**	1.28 (0.86 – 1.93)	1.02 (0.64 - 1.64)
Subgroup: patients with emergency intracranial surgical intervention		
Unadjusted	1.51 (0.85 – 2.69)	1.59 (0.68 – 3.72)
Multivariable adjustment**	1.56 (0.89 – 2.74)	1.61 (0.56 – 4.76)

* Adjusted for: age, GCS motor score, pupil inequality, hypoxia, hypotension, ISS, CT lesions: tSAH, epidural hematoma, mass lesion, acute subdural hematoma, and a random intercept for centre.
** Adjusted for: age, GCS motor score, pupil inequality, hypoxia, hypotension, ISS and a random intercept for centre.

8.4.4 | Sensitivity analysis in the registry

A total of 2150 moderate/severe TBI patients were included in the registry of which 25% arrived by secondary transfer, the characteristics of both groups were similar to patients in the core study (Table S1 available online). Secondarily referred patients had craniotomy for hematoma more often as emergency intervention (171 (10.5%) of directly admitted and 164 (31.4%) of secondarily referred patients). Also, the CT scans of secondarily referred patients more frequently showed midline shift (54.5% for secondarily referred versus 37.5% for directly admitted). There was no associa-

tion between arriving by secondary referral and survival at discharge after adjustment for confounders (OR 1.21 95% CI 0.84 – 1.73, Table S2 available online).

8.5 | Discussion

This study showed that variation in the proportion of moderate and severe TBI patients who have been secondarily referred to European specialist centres varied significantly by country after adjusting for case-mix factors. The secondarily referred TBI patients received less prehospital interventions. However, they had more serious abnormalities at CT scanning. Secondarily referred TBI patients were non-significantly associated with fewer secondary insults (hypoxia and hypotension at ED arrival). We found no association between secondary referral and clinical long term outcomes. These findings were confirmed in the registry database, including a larger and more heterogenous population.

The European variation in the proportion of secondary referrals to specialist centres is large, and only partly confirmed in previous literature. The likelihood of arriving by secondary referral was lowest in the Netherlands and Italy. A previous Italian study showed that 58% of the TBI patients presenting at SNCs in the whole country were referred from another peripheral hospital [23]. However, Italian centres that contributed to CENTER-TBI were mainly situated in Northern Italy. Fifteen years ago, an English study found that one third of the severe head injury patients were treated in non-neurotrauma centres which was associated with higher mortality [24]. A study from Greece found that around half of the TBI patients in specialist centres were secondarily referred, higher than our findings. Sec-

ondary referral increased the travel time to a neurosurgical centre by 3.5 hours [25]. The percentage of secondary referrals seems to be decreasing when comparing our sample of moderate/severe TBI patients to older European studies. The percentage of secondarily referred patients was highest in Scandinavian countries, Austria and the UK. This is in line with their geography, less densely populated areas with long distances and the consequent need to stabilise their patients at closer non-specialised acute hospitals in order to avoid secondary insults. Earlier research suggested that arriving by secondary referral is associated with worse outcomes in severe TBI patients [7, 23, 26, 27]. It was suggested that one of the most important explanations for worse outcomes was time delay [28]. Also, care in centres that practice high-volume protocol-driven therapy, like ICP monitoring, is associated with better outcomes especially when neurocritical interventions are necessary [29, 30]. However, we could not find an effect of secondary referral on long term outcomes. A meta-analysis including eleven studies found comparable results [11]. This is in line with previous research, suggesting that time interval to surgery was not associated with outcomes in patients with acute subdural hematomas requiring surgery [31]. Since subdural hematomas were the most prevalent CT abnormality in secondary referred patients, these data suggest that these patients can safely be stabilised in non-specialised centres.

Our study shows that the impact of time to emergency surgery on outcomes becomes less critical when secondary insults (hypoxia and hypotension) are avoided. Hypoxia and hypotension are although less frequently observed over time in TBI patients still strongly associated with worse long term outcomes [32, 33]. We found that secondarily referred TBI patients are less likely to arrive with hypoxia or hypotension compared to directly admitted TBI patients at the Specialised Neurotrauma Centre. This is in

line with previous research which shows that interventions to treat life-threatening events may significantly decrease mortality [34].

This study has several strengths. CENTER-TBI is a multicentre study in 22 European countries, which increases external validity. External validity is further increased because we were able to validate our findings for the effect of secondary referral on outcome in the CENTER registry. Moreover, the precision of our results are high because of the large sample size. We could rigorously adjust for potential case-mix differences due to the broad data collection of CENTER-TBI, and assess both survival and long term functional outcome.

However, our study also has several limitations. First, we could only include patients that were referred to a neurosurgical study centre within 24 hours after injury. Some moderate/severe TBI patients who may have benefited from specialised care might not have been transferred, or might have been transferred after 24 hours. Late secondary transfers are associated with worse outcomes [35]. Second, inevitably our large multicentre prospective observational study meant data was missing for some variables. For example, time to first hospital was missing in 50% of the cases. This was addressed by using multiple imputation, a method proven to give valid estimates under the missing at random assumption [36]. Third, the between-country and between-centre differences could not be explained by the captured policy and care characteristics [17]. Last, geographical differences like the distance from scene to the specialised centre, or the number of specialised neurosurgical centres per km^2 were not measured at a patient level.

The debate about whether or not to transport TBI patients directly to specialist neurotrauma centres – past closer non specialist hospitals - has not yet been concluded. We were not able to find an association between

secondary referral and outcome. Intuitively, arriving by secondary referral with extended time from injury to definitive treatment remains undesirable. One could look for alternatives. An English study shows for example that observation in a non-specialised hospital with neurosurgical consult by e-health and repeated CT scanning was not associated with worse outcomes for TBI patients [37]. However, this could lead to extra transfers between hospitals and increasing health care costs.

Once moderate/severe TBI patients are stabilised (on-scene or at the first hospital), it is possible that there is no effect of the time delay on outcome anymore. Patients arriving by secondary referral receive less interventions on-scene, but do have more serious CT brain scan abnormalities, highlighting the limitations of current prehospital triage tools. Future research in this area also needs to include patients with TBI admitted to non-specialist hospitals. This will enable assessment of subgroups of TBI patients with benefit from direct transport to SNCs. Consequently, this would also allow further evaluation of the cost-effectiveness of direct transport to SNCs which was recently shown to be equivocal [38].

8.5.1 | Conclusion

Across Europe, substantial practice variation exists in the proportion of secondarily referred moderate/severe TBI patients within specialised neurotrauma centres. Patients who are secondarily referred present less often with secondary insults, although they have more serious CT abnormalities. In moderate/severe patients with TBI treated at specialised neurotrauma centres, we did not find a harmful effect of secondary referral. Future research should focus upon which on scene characteristics identify TBI patients that benefit from direct transportation to distant specialist neuro-

trauma centres in order to improve guidelines and outcomes for patients with TBI.

8.6 | Supplementary Material

Supplementary materials can be found in the online version of this article: `https://tinyurl.com/258ps23n`

References

1. Maas, A. I. R. *et al.* Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. *The Lancet Neurology* **4422.** ISSN: 14744422 (2017).
2. Roozenbeek, B., Maas, A. I. & Menon, D. K. Changing patterns in the epidemiology of traumatic brain injury. *Nature Reviews Neurology* **9,** 231–236. ISSN: 17594758 (2013).
3. Feigin, V. L. *et al.* Global, regional, and national burden of neurological disorders, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Neurology* **18,** 459–480. ISSN: 14744465 (2019).
4. McConnell, K. J., Newgard, C. D., Mullins, R. J., Arthur, M. & Hedges, J. R. Mortality benefit of transfer to level I versus level II trauma centers for head-injured patients. *Health services research* **40,** 435–458 (2005).
5. Mendeloff, J. M. & Cayten, C. G. Trauma Systems and Public Policy. *Annual Review of Public Health* **12,** 401–424. ISSN: 0163-7525 (1991).
6. DuBose, J. J. *et al.* Effect of trauma center designation on outcome in patients with severe traumatic brain injury. *Archives of surgery* **143,** 1213–1217 (2008).

7. Härtl, R. *et al.* Direct transport within an organized state trauma system reduces mortality in patients with severe traumatic brain injury. *Journal of Trauma and Acute Care Surgery* **60,** 1250–1256 (2006).

8. Bernard, S. A. *et al.* Prehospital Rapid Sequence Intubation Improves Functional Outcome for Patients With Severe Traumatic Brain Injury. *Annals of Surgery* **252,** 959–965. ISSN: 0003-4932 (2010).

9. Cnossen, M. C. *et al.* Prehospital trauma care among 68 European neurotrauma centers: Results of the CENTER-TBI Provider Profiling Questionnaires. *Journal of neurotrauma* **36,** 176–181 (2019).

10. Helling, T. S., Davit, F. & Edwards, K. First echelon hospital care before trauma center transfer in a rural trauma system: Does it affect outcome? *Journal of Trauma - Injury, Infection and Critical Care* **69,** 1362–1366. ISSN: 00225282 (2010).

11. Pickering, A. *et al.* Impact of prehospital transfer strategies in major trauma and head injury: systematic review, meta-analysis, and recommendations for study design. *Journal of Trauma and Acute Care Surgery* **78,** 164–177 (2015).

12. Lecky, F. E. *et al.* Bypassing nearest hospital for more distant neuroscience care in head-injured adults with suspected traumatic brain injury: Findings of the head injury transportation straight to neurosurgery (HITS-NS) pilot cluster randomised trial. *BMJ Open* **7.** ISSN: 20446055 (2017).

13. Hoogmartens, O. *et al.* Evidence-based prehospital management of severe traumatic brain injury: A comparative analysis of current clinical practice guidelines. *Prehospital Emergency Care* **18,** 265–273. ISSN: 15450066 (2014).

14. Badjatia, N. *et al.* Guidelines for Prehospital Management of Traumatic Brain Injury 2nd Edition. *Prehospital Emergency Care* **12,** S1–S52. ISSN: 1090-3127 (2008).

15. Gabriel, E. J. *et al.* Guidelines for prehospital management of traumatic brain injury. *Journal of neurotrauma* **19,** 111–174 (2002).

16. Maas, A. I. *et al.* Collaborative European neurotrauma effectiveness research in traumatic brain injury (CENTER-TBI): A prospective longitudinal observational study. *Neurosurgery* **76,** 67–80. ISSN: 15244040 (2015).

17. Cnossen, M. C. *et al.* Variation in structure and process of care in traumatic brain injury: Provider profiles of European Neurotrauma Centers participating in the CENTER-TBI study. *PLoS ONE* **11,** 1–21. ISSN: 19326203 (2016).

18. Teasdale, G & Jennett, B. Assessment of coma and impaired consciousness. A practical scale. *The Lancet* **2,** 81–84. ISSN: 0140-6736 (Print) (1974).

19. Von Elm, E. *et al.* Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *British Medical Journal* **335,** 806–8. ISSN: 1756-1833 (2007).

20. Steyerberg, E. W. *et al.* Case-mix, care pathways, and outcomes in patients with traumatic brain injury in CENTER-TBI: a European prospective, multicentre, longitudinal, cohort study. *The Lancet Neurology* **18,** 923–934 (2019).

21. Dijkland, S. A. *et al.* Prognosis in Moderate and Severe Traumatic Brain Injury: A Systematic Review of Contemporary Models and Validation Studies. *Journal of Neurotrauma,* neu.2019.6401. ISSN: 0897-7151 (2019).

22. Merlo, J. *et al.* A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *Journal of Epidemiol Community Health* **60,** 290–297 (2006).

23. Citerio, G. *et al.* Application of guidelines for severe head trauma: data from an Italian database. *European journal of emergency medicine : official journal of the European Society for Emergency Medicine* **10,** 68–72. ISSN: 09699546 (2003).

24. Patel, H. C. *et al.* Trends in head injury outcome from 1989 to 2003 and the effect of neurosurgical care: An observational study. *The Lancet* **366,** 1538–1544. ISSN: 01406736 (2005).

25. Stranjalis, G. *et al.* Outcome in 1,000 head injury hospital admissions: The athens head trauma registry. *Journal of Trauma - Injury, Infection and Critical Care* **65,** 789–793. ISSN: 00225282 (2008).

26. Joosse, P. *et al.* Impact of secondary transfer on patients with severe traumatic brain injury. *Journal of Trauma and Acute Care Surgery* **72,** 487–490. ISSN: 21630755 (2012).

27. Tepas, J. J., Pracht, E. E., Orban, B. L. & Flint, L. M. High-volume trauma centers have better outcomes treating traumatic brain injury. *Journal of Trauma and Acute Care Surgery* **74,** 143–148. ISSN: 21630755 (2013).

28. Haselsberger, K., Pucher, R. & Auer, L. M. Prognosis after acute subdural or epidural haemorrhage. *Acta Neurochirurgica* **90,** 111–116. ISSN: 00016268 (1988).

29. Elf, K., Nilsson, P. & Enblad, P. Outcome after traumatic brain injury improved by an organized secondary insult program and standardized neurointensive care. *Critical Care Medicine* **30,** 2129–2134. ISSN: 00903493 (2002).

30. Fuller, G. *et al.* The effect of specialist neurosciences care on outcome in adult severe head injury: a cohort study. *Journal of neurosurgical anesthesiology* **23,** 198–205. ISSN: 0898-4921 (2011).

31. Walcott, B. P. *et al.* Time interval to surgery and outcomes following the surgical treatment of acute traumatic subdural hematoma. *Journal of Clinical Neuroscience* **21,** 2107–2111 (2014).

32. Gravesteijn, B. *et al.* Prehospital management of traumatic brain injury across Europe: a CENTER-TBI study. *Prehospital Emergency Care,* 1–22. ISSN: 1090-3127 (2020).

33. Spaite, D. W. *et al.* Association of out-of-hospital hypotension depth and duration with traumatic brain injury mortality. *Annals of emergency medicine* **70,** 522–530 (2017).

34. Gomes, E. *et al.* The importance of pre-trauma centre treatment of life-threatening events on the mortality of patients transferred with severe trauma. *Resuscitation* **81,** 440–445. ISSN: 0300-9572 (2010).

35. Harrison, D. A. *et al.* Risk Adjustment In Neurocritical care (RAIN)–prospective validation of risk prediction models for adult patients with acute traumatic brain injury to use to evaluate the optimum location and comparative costs of neurocritical care: a cohort study. *Health technology assessment (Winchester, England)* **17,** vii–350. ISSN: 1366-5278 (2013).

36. Griffith, D. A., Bennett, R. J. & Haining, R. P. Statistical analysis of spatial data in the presence of missing observations: a methodological guide and an application to urban census data. *Environment and Planning A* **21,** 1511–1523. ISSN: 0308-518X (1989).

37. Fabbri, A., Servadei, F., Marchesini, G., Stein, S. C. & Vandelli, A. Observational approach to subjects with mild-to-moderate head injury and initial non-neurosurgical lesions. *Journal of Neurology, Neurosurgery & Psychiatry* **79,** 1180–1185. ISSN: 0022-3050 (2008).

38. Lecky, F. *et al.* The Head Injury Transportation Straight to Neurosurgery (HITS-NS) randomised trial: a feasibility study. *Health technology assessment* **20.** ISSN: 1366-5278 (2016).

9

Cost-effectiveness of extracorporeal cardiopulmonary resuscitation after in-hospital cardiac arrest: A Markov decision model

Benjamin Y Gravesteijn, Marc Schluep, Daphne C Voormolen, Anna C van der Burgh, Dinís Dos Reis Miranda, Sanne E Hoeks, Henrik Endeman

Resuscitation

9.1 | Abstract

9.1.1 | Background

This study aimed to estimate the cost-effectiveness of extracorporeal cardiopulmonary resuscitation (ECPR) for in-hospital cardiac arrest treatment.

9.1.2 | Method

A decision tree and Markov model were constructed based on current literature. The model was conditional on age, Charlson Comorbidity Index (CCI) and sex. Three treatment strategies were considered: ECPR for patients with an Age-Combined Charlson Comorbidity Index (ACCI) below different thresholds (2–4), ECPR for everyone (EALL), and ECPR for no one (NE). Cost-effectiveness was assessed with costs per quality-of-life adjusted life years (QALY).

9.1.3 | Results

Treating eligible patients with an ACCI below 2 points costs €8394 (95% CI: €4922–€14,911) per extra QALY per IHCA patient; treating eligible patients with an ACCI below 3 costs €8825 (95% CI: €5192–€15,777) per extra QALY per IHCA patient; treating eligible patients with an ACCI below 4 costs €9311 (95% CI: €5478–€16,690) per extra QALY per IHCA patient; treating every eligible patient with ECPR costs €10,818 (95% CI: €6357–€19,400) per extra QALY per IHCA patient. For WTP thresholds of €0–€9500, NE has the highest probability of being the most cost-effective strategy. For WTP thresholds between €9500 and €12,500, treating eligible patients with an ACCI below 4 has the highest probability of being

the most cost-effective strategy. For WTP thresholds of €12,500 or higher, EALL was found to have the highest probability of being the most cost-effective strategy.

9.1.4 | Discussion

Given that conventional WTP thresholds in Europe and North-America lie between €50,000–€100,000 or U.S. dollars, ECPR can be considered a cost-effective treatment after in-hospital cardiac arrest from a healthcare perspective. More research is necessary to validate the effectiveness of ECPR, with a focus on the long-term effects of complications of ECPR.

9.2 | Introduction

Cardiac arrest, cardiopulmonary arrest, or circulatory arrest is the loss of effective blood circulation, which inevitably leads to death if cardiopulmonary resuscitation (CPR) is not started. Cardiac arrest is usually divided based on location into out-of-hospital cardiac arrest (OHCA) and in-hospital cardiac arrest (IHCA). OHCA is described to occur around 19–104 times per 100,000 population per year and results in 10% survival at hospital discharge [1]. The incidence of IHCA is 1–6 events per 1000 hospital admissions [2–4] and recent meta-analyses showed a pooled survival to discharge of 15% (ranging from 3% to 40%) and a one-year survival of 13% (ranging from 4% to 69%) [5, 6]. Patient-specific factors associated with survival are age [7, 8], comorbidities [7, 9–11] and presence of shockable rhythm [12].

A possible advantage for patients suffering IHCA versus OHCA is that hospitals are equipped with advanced life support teams, who could employ extracorporeal cardiopulmonary resuscitation (ECPR) using veno-arterial extracorporeal membrane oxygenation (VA-ECMO). This technique has seen an increase in use over the last decades [13, 14]. By taking over cardiac and respiratory function, VA-ECMO ensures oxygenation and circulation [15]. Although evidence from randomized controlled trials is lacking [16], observational studies have repeatedly shown an increase in survival after ECPR compared to conventional CPR [16–18]. Furthermore, the American Heart association recommends the in-hospital use of ECPR in patients with a reversible cause of CA (e.g.: acute coronary syndrome).

When assessing whether or not to implement ECPR, cost-effectiveness should be taken into account. Ethical and economic considerations are of increasing importance in decision making pertaining to intensive care al-

location [19]. Financial resources are limited and health care should be focused more on therapies that do not only extend life, but rather offer a reasonable health-related quality of life (HRQoL). This study was designed to provide cost-effectiveness evidence for international comparison and to provide an overview of current knowledge of the economic aspects of ECPR.

Two small observational studies (US and Australia) have shown indications of cost-effectiveness of ECPR for both OHCA and IHCA [20, 21]. There are however several caveats. Because of low sample size and estimates pertaining to local situations these studies are not likely to be generalisable to all settings. Furthermore, for the in-hospital and out-of-hospital setting, effectiveness should be assessed separately.

The primary aim of this study was therefore to assess the cost-effectiveness of ECPR treatment after IHCA based on current literature. By using all available evidence, this modelling approach would ensure a high generalisability of our results. For this purpose, a decision tree and Markov model were developed. Both models are frequently used in health-economic evaluations, because they are able to calculate quality of life adjusted life years (QALY) [22, 23]. The secondary aim was to assess in which patient group ECPR is most likely to be cost-effective.

9.3 | Methods

This cost-effectiveness evaluation is reported according to the CHEERS reporting guidelines [24]. We searched PubMed for relevant studies to inform on all parameters used for the models. We used the search terms “in-hospital cardiac arrest” and “extracorporeal cardiopulmonary resus-

citation" in combination with the specific parameter of interest. Furthermore, we found literature using the reference list of already found studies.

A three-strategy decision tree was created, which encompasses the in-hospital phase. This type of model uses known absolute and relative risks to calculate the probability of an outcome. The decision tree calculates the probability of dying before discharge. The strategies considered were ECPR for no one (NE), ECPR for every eligible patient (EALL) and ECPR for eligible patients with an Age-Combined Charlson Comorbidity Index (ACCI) score below a certain threshold (EACCI_lo). The thresholds for the ACCI analysed ranged from two to four: patients with an ACCI above the threshold did not receive ECPR. The ACCI thresholds have been based on best available ECPR guidelines to exclude patients with a terminal illness, comorbidities that form a contraindication for ICU admission or for intravascular cannulation [25]. Furthermore patients > 75 years of age are generally not considered eligible. The ACCI score is described in Table 1 in Supplement 1 available online.

The ACCI threshold can be illustrated by the following example: a patient of 50 years old with moderate renal disease (GFR < 40 mL/min/1.73 m2) will have an ACCI of 3. If the patient would suffer a myocardial infarction the score will rise to 4.

The decision tree consists of multiple nodes with probability estimates found in literature (figure 9.1 and table 9.1). The first node represents patients with a Do-Not-Resuscitate (DNR) status. This is an agreement between a patient and a health care professional not to attempt cardiopulmonary resuscitation in case of cardiac arrest. Since a DNR status is more often agreed upon by patients with higher age [26], we assumed higher probabilities for higher aged patients. We assumed that for patients who suffered cardiac arrest with a DNR status, no CPR would be attempted and

death is certain. When patients did not have a DNR status, CPR would be attempted. The next node represents the probability of having a contra-indication for ECPR. Having a contra-indication, e.g. refractory cardiac disease or metastatic cancer, was assumed to increase the risk of dying after CPR. If CPR was started and no contra-indication was present, the next node represents the probability of having return of spontaneous circulation (ROSC) within 20 min after cardiac arrest [27]. If ROSC would not be achieved within 20 min, ECPR could be started and could increase the remaining survival probability [28]. The probability of having a complication of ECPR and the probability of subsequent death are also taken into account [29–31]. These probabilities were calculated from the ELSO database [14]. The extra probability of mortality, given that the patient had a complication was: the mortality rate of patients with a complication minus the overall mortality rate. Finally, the mortality rate after CPR increases with increasing Age-Combined Charlson Comorbidity Index (ACCI) [7, 9].

Figure 9.1: *Decision tree of the in-hospital phase of the model. For the assumed probabilities (P), odds ratio's (OR), relative risks (RR), and betas, see Table 1. DNR = do-not-resuscitate; CPR = cardiopulmonary resuscitation; ROSC = return of spontaneous circulation.*

Table 9.1: *Assumed estimates and their distributions for the decision tree in the probabilistic sensitivity analysis.*

Abbr.	Parameter	Median (IQR)	Distribution	Source
Decision tree				
P1	Probability of having DNR status if <75 years	0.05 (0.02–0.10)	Beta(5,95)	Clinical insight
P2	Probability of ECPR contraindication	0.19 (0.11–0.32)	Beta(10,40)	Clinical insight
P3	The probability of dying after CPR	0.85 (0.83–0.87)	Beta(850,150)	Zhu and Zhang [5]
P4	The probability of having ROSC within 20 min	0.38 (0.35–0.41)	Beta(338,556)	Khan et al. [27]
P5	Probability of complication	0.38 (0.29–0.47)	Beta(38,62)	Sheu et al., Muller et al. and Sakamoto et al. [29–31]
P6	Probability of dying because of complication	0.2 (0.1–0.32)	Beta(10,40)	Clinical insight
RR1	The relative risk of dying, ECPR vs non ECPR	0.43 (0.3–0.62)	Lognormal(-0.85, 0.19)	Chen et al. [28]
OR1	OR of dying when contraindication for ECPR	2.00 (1.40–2.93)	Lognormal(0.69, 0.2)	Clinical insight
OR2	OR of having DNR status if 75 - 84 years, compared to <75 years	1.71 (1.23–2.32)	Lognormal(0.53, 0.16)	Cook et al. [26]
OR3	OR of having DNR status if >85 years, compared to <75 years	2.98 (2.38–3.75)	Lognormal(1.09, 0.12)	Cook et al. [26]
Beta1	The log odds increase in dying per ACCI increase	0.09 (0.03–0.14)	Log-Lognormal(0.09, 0.03)	Hirlekar et al. [7]
Costs and utilities				
	In-hospital incremental cost of ECPR after cardiac arrest	51756.66 (31377.83–73978.21)	Normal(51997, 10767)	Oude Lansink-Hartgring et al. [32]
	Utility score for men	0.79 (0.69–0.87)	Triangle(a = 0.66, b = 0.89, c = 0.82)	Israelsson et al. [33]
	Utility score for women	0.74 (0.62–0.81)	Triangle(a = 0.58, b = 0.82, c = 0.81)	Israelsson et al. [33]

ECPR = extracorporeal cardiopulmonary resuscitation; CPR = cardiopulmonary resuscitation; ROSC = return of spontaneous circulation; ACCI = Age-Combined Charlson Comorbidity Index; DNR = do not resuscitate.

The prevalence of DNR status below 75 years was assumed to be around 5% (range 2–10%), based on experience in our hospital: the Erasmus Medical centre, Rotterdam. The probability of having a contra-indication for ECPR was also based on experience in our hospital, where we implemented ECPR in 2016. We assumed that 20% (range 10–30%) of the patients have the contra-indications described by Makdisi et al. [34]. Since the described contra-indications (e.g. refractory cardiac disease or metastatic cancer) are severe conditions, the risk of dying was assumed to double (OR: 2.0, with a minimum of 1.4, and a maximum 2.9).

9.3.1 | Markov model

For the calculation of long-term outcomes, a Markov model was used. A Markov model uses states and transition probabilities to calculate long-term outcomes [22]. We propose a model consisting of two states: an alive state (with decreased HRQoL) and a dead state (the absorbing state). Markov models can be used to calculate the time spent in each state. Therefore, QALYs can be calculated, making this type of model useful for cost-effectiveness analysis [23]. Each individual probability of dying at the end of the decision tree described above is used as input in the subsequent Markov model. The model simulated 20 years of follow-up and the model cycles were one year long. The data on age and sex specific mortality rates were provided by Statistics Netherlands (CBS) [35]. We did not assume a lasting effect of IHCA on long-term survival [36]. The amount of life-years were then multiplied by the sex-specific utility score after IHCA to obtain QALYs for men and women [33] (table 9.1).

As an example, consider a patient with a 100% chance of surviving the in-hospital phase: the Markov model will calculate the amount of life years

this patient will spend after discharge. For a patient with 0% chance of surviving the in-hospital phase, the Markov model will estimate 0 life years after discharge. For chances between 0% and 100%, the model calculates the average life years that patients with the same characteristics will spend after discharge.

9.3.2 | Cost-effectiveness analysis

The total costs of ECPR were calculated based on how many patients received ECPR following the decision tree outcomes: a patient received ECPR according to the treatment strategy if they did not have a DNR status, no contra-indication, and no ROSC within 20 min (figure 9.1 and table 9.1).

Only direct additional costs of ECPR treatment were taken into account, taking a health care's perspective. The average additional costs of ECPR described in the literature were used in the model. A detailed description of the items included in the total costs has been described by Lansink-Hartgring et al. [32]. A discount rate of 4% was applied, the appropriate rate for cost-effectiveness analyses in the Netherlands [37]. To assess cost-effectiveness of the strategies, incremental cost-effectiveness ratios (ICER) were calculated, where NE serves as the reference category. The ICER informs about how many extra € per QALY a strategy costs, compared to NE. The incremental costs and QALYs were plotted and the cost-effectiveness acceptability curves were calculated and drawn to obtain the most cost-effective strategy.

Important to take into account is that the calculated costs for ECPR are notably lower than the costs of ECMO. This is due to the model structure, in which costs are calculated for an average patient who suffers IHCA, thereby including also patients who do not receive ECPR.

9.3.3 | Probabilistic sensitivity analysis

To take the uncertainty of our model parameters into account, a probabilistic sensitivity analysis (PSA) was performed. A PSA repeats the model a large number of times with different (but probable) parameters. The type of distributions that were used were beta distributions for probabilities, log-normal distributions for the odds ratios and relative risks, and log–log-normal distribution for the log-odds increase in mortality for an ACCI point increase. The characteristics of the distributions were adjusted so that the median and interquartile range were identical to the estimate and 95% confidence interval. The type and characteristics of the distributions of the parameters are described in Table 9.1. From these distributions, 1000 random samples were drawn, resulting in 1000 replicates of the model. Additionally, a representative cohort of 1000 patients was randomly sampled (table 9.2) [9, 38]. After running the 1000 replicates of the model in this cohort, outcomes were calculated 1000 times. We calculated the QALYs and costs per strategy. The median was taken as the most probable estimate of the model. The 2.5th and 97.5th percentile were calculated, which indicated the borders of the 95% credibility interval.

To estimate whether the conclusions were affected by the parameters that were not found in literature, linear regression was performed. As the dependent variable, the ICER of the EALL strategy per iteration was used. As predictors, the standardized parameter values were used. The coefficients of the model could therefore be interpreted as "with one standard deviation (SD) increase in the parameter, the ICER for the EALL strategy increases with x".

All analyses were performed using R[1]. For the Markov model, the "dampack"

[1]R Core Team (2013). R: A language and environment for statistical computing. R

Table 9.2: *Patient characteristics of the simulated cohort, based on literature [9, 38]*

Characteristic	N = 1000
Age (mean (sd))	65.49 (15.71)
Male (%)	578 (57.80)
CCI (%)	
0	373 (37.30)
1	230 (23.00)
2	183 (18.30)
3	107 (10.70)
4	43 (4.30)
5	40 (4.00)
6	15 (1.50)
7	4 (0.40)
8	5 (0.50)

CCI = Charlson Comorbidity Index

package was used [39]. The code of the model is online available in Supplement 2 available online, for transparency and reproducibility [40].

9.4 | Results

In the decision tree, survival rates between 9% and 13% were observed for the NE strategy, and between 30% and 35% for the EALL strategy (Fig. 1, Supplement 1 available online). After applying a Markov model, expected life years after CPR per patient for the NE strategy ranged from 0.79 to 2.48 and for the EALL strategy from 2.57 to 6.55 years (Fig. 2 in Supplement 1 available online).

The expected costs per IHCA patient for treating eligible patients below an ACCI of 2 points with ECPR are €3,975 (95% CI: €2,418–€5,780), and increased to €23,272 (95% CI: €14,159–€33,838) for treating all eli-

Foundation for Statistical Computing, Vienna, Austria

gible patients (table 9.3). The associated QALYs for treating no patients with ECPR are 1.2 (95% CI: 1.0–1.5); for treating eligible patients below an ACCI of 2 points 1.7 (95% CI: 1.4–2.0); for treating eligible patients below an ACCI of 3 points 2.1 (95% CI: 1.7–2.6); for treating eligible patients below an ACCI of 4 points 2.6 (95% CI: 2.0–3.2); and for treating all eligible patients 3.4 (95% CI: 2.4–4.2).

Table 9.3: *The health economic evaluation for each strategy.*

Strategy	Costs*	QALY	ICER**
NE	–	1.2 (1.0–1.5)	–
ACCI <2	3,975 (2,418–5,780)	1.7 (1.4–2.0)	8,394 (4,922–14,911)
ACCI <3	8,066 (4,909–11,731)	2.1 (1.7–2.6)	8,825 (5,192–15,777)
ACCI <4	12,942 (7,881–18,829)	2.6 (2.0–3.2)	9,311 (5,478–16,690)
EALL	23,272 (14,159–33,838)	3.4 (2.4–4.2)	10,818 (6,357–19,400)

The strategies are nobody ECPR (NE), treating everyone with an Age-Combined Charlson Comorbidity Index (ACCI) of 2, 3, or 4 or less, and treating everyone with ECPR (EALL).
The ranges indicate 95% credibility intervals (CI)
* In €, only direct additional ECPR costs
** The incremental cost-effectiveness ratio (ICER) is calculated with the most conservative methos (NE: nobody ECPR) as the reference strategy. It represents the costs per extra QALY.

Compared to treating NE, the expected incremental costs per extra QALY (ICER) for treating eligible patients with an ACCI below 2 points is €8,394 (95% CI: €4,922–€14,911) per extra QALY; for treating eligible patients with an ACCI below 3, the ICER is €8,825 (95% CI: €5,192–€15,777) per extra QALY compared to NE; for treating eligible patients with an ACCI below 4, the ICER is €9,311 (95% CI: €5,478–€16,690) per extra QALY; for treating all eligible patients, the ICER was €10,818 (95% CI: €6,357–€19,400) per extra QALY. Table 9.3 displays an overview of the

economic evaluation. The considered strategies are comparable in terms of mean ICER, but the incremental costs and incremental QALYs vary significantly between the considered strategies (Fig. 3 in Supplement 1 available online).

The cost-effectiveness acceptability curves depicted in figure 9.2 show that for WTP thresholds of €0–€9,500, NE has the highest probability of being the most cost-effective strategy. For WTP thresholds between €9,500 and €12,500, treating eligible patients with an ACCI below 4 has the highest probability of being the most cost-effective strategy. For WTP thresholds of €12,500 or higher, EALL was found to have the highest probability of being the most cost-effective strategy.

The only parameter that was found to influence the cost-effectiveness significantly was the relative risk of dying of ECPR (effect of one unit increase of the parameter on the ICER was €−255 (€−481 to €−28) per incremental QALY), see Table 2 in Supplement 2 available online.

9.5 | Discussion

In this study we found that the expected costs per IHCA patient of treating each eligible IHCA patient with ECPR are approximately €23,000. A patient was eligible when no contra-indications was present, and in whom ROSC cannot be achieved within 20 min after cardiac arrest. Per QALY increase, the associated costs were around €15,000. The Willingess-To-Pay tresholds in Europe and North-America are between €50,000–€100,000 per incremental QALY. Within this range, performing ECPR in every eligible IHCA patient, is likely to be costs-effective.

The use of ECMO has steadily increased from 2007 onwards [13]. Posi-

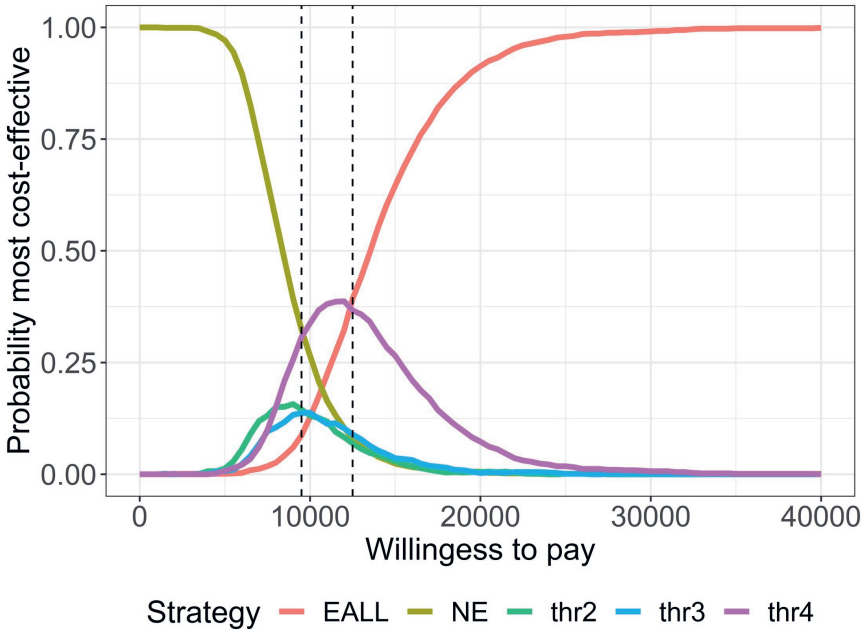


Figure 9.2: *Cost-effectiveness acceptability curves. For given willingness to pay (WTP) thresholds, the probability of being the most cost-effective strategy is plotted. The strategies are nobody ECPR (NE), treating everyone with an Age-Combined Charlson Comorbidity Index (ACCI) of 2, 3 or 4 or less (thr2, thr3, thr4 respectively), and treating everyone with ECPR (EALL). The dotted lines indicates the WTP thresholds of €9,500 and €12,500.*

tive results from observational studies and increasing clinical applicability led to the inclusion of ECPR in the Advanced Life Support Guidelines by the European Resuscitation Counsil [41]. However, ECPR is costly and labour-intensive and careful economic evaluation was still lacking.

Because ECPR was found to be cost-effective, this study substantiates its increased implementation and inclusion as possible treatment in the guidelines. The allocation of intensive care treatments should be critically

evaluated, especially when financial resources are limited [19]. The difference in survival probability after ECPR seems to be sufficient to render the therapy cost-effective. Because we performed an analysis taking all uncertainties of parameters into account, we believe that we reliably estimated the average cost per IHCA patient when every eligible patient is treated with ECPR: around €11,000 per extra QALY.

Our cost-effectiveness analysis based on literature supports findings of empirical studies. Firstly, our study confirms the results of a recent small retrospective study in the United States that suggested that ECPR after IHCA is cost-effective, considering only in-hospital costs [20]. This study suggested that the costs per extra QALY saved is around 56,000 U.S. dollars. This estimate is larger than our estimate of €11,000, but health care expenditures in the United States tend to be higher than in Europe [42]. Nevertheless, it is reassuring that both studies conclude that ECPR after IHCA is cost-effective, since they both assess primarily in-hospital costs. Secondly, our study confirms the results of Dennis et al. This study showed that for IHCA, €15,000 (25,000 AUD) per extra QALY was expected, which is similar to our estimate [21].

The results of our study are also similar to results of the cost-effectiveness of a mobile ECPR team [43]. This team is able to treat patients with ECPR in multiple centres, and its application was found to be potentially cost-effective. The application could benefit centres that do not have the resources for ECPR or lack experience with its application. Centres that often use ECPR rely on perfusionists for aid in initiation and maintenance of treatment, which enhances the costs. Therefore, it could well be that ECPR is mostly cost-effective when there is no need for these extra costs. This hypothesis, however, warrants further investigation.

The range of costs of ECMO found in the literature is large [44]. Mostly

because studies inconsistently report their results, there are no factors described that explain this variation. We used a structured Dutch study as input for our cost-effectiveness analysis, since it describes clearly the incremental costs for ECPR [32]. This study found that the majority of the costs are composed of nursing days. Being able to shorten the length of ICU stay would therefore enhance cost-effectiveness of ECPR after IHCA.

We did not find that treating a subgroup of IHCA patients with ECPR based on Age-Combined Charlson Comorbidity Index affected cost-effectiveness. Since others described that cost-effectiveness depends on patient characteristics [43], we consider this to be attributed to two factors. First, the effect of comorbidity on survival of CPR is uncertain [7, 9]. More research into this relationship is necessary. Second, if there is an effect of comorbidity, this effect is more likely to be significant in a cohort with a high prevalence of comorbidities. The prevalence in our representative cohort, however, was low [9, 38].

This study has several limitations. Unfortunately, not all information needed for the model could be found in the literature. The lack of evidence had two consequences. First, it was necessary to base some of the parameters on clinical knowledge; e.g., for the probability of having a contraindication for ECPR. However, a sensitivity analysis showed that these parameters were not likely to influence the overall cost-effectiveness of ECPR. Second, cost-effectiveness might be somewhat overestimated. Evidence from randomized controlled trials was unfortunately absent at this moment [16]. Observational studies could have overestimated the effect of ECPR on survival because of confounding bias [17, 28]. An overestimated effect of ECPR would result in an overestimated cost-effectiveness. Additionally, we were not able to model long-term effects of complications of ECPR: the extra health care costs and lower quality of life after major

complications of ECPR (stroke, acute kidney injury) could decrease overall cost-effectiveness.

Although we did not take non-direct costs of ECPR into account, we still believe this study provides a valid economic evaluation. Other identifiable costs are costs of rehabilitation, future health care costs and non-medical costs such as loss of participation in working life. However, these costs are more interesting from a societal perspective than a health care perspective. Other costs that are not taken into account are the costs of implementation. These expenses are large and could explain the stagnating increase in the use of ECPR [45, 46]. Therefore, we believe that our findings are most applicable to large hospitals in western countries, which often do have access to these resources to overcome the first barrier to an apparent cost-effective therapy.

We believe future studies should have three goals. First, to identify patients who could benefit most from ECPR. Second, randomized controlled trials are necessary, as indicated in the advanced life support guidelines [41]. Fortunately, five ongoing randomized controlled trials will hopefully fill this knowledge gap in the upcoming years [18]. Third, the long-term effects of complications of ECPR should be investigated, since they could decrease the cost-effectiveness of the intervention. The knowledge gained from further research could improve implementation and cost-effectiveness of this costly and labour-intensive intervention.

9.5.1 | Conclusion

For in-hospital cardiac arrest patients, extracorporeal cardiopulmonary was demonstrated to be cost-effective from a healthcare perspective given that conventional WTP thresholds lie between €50,000–€100,000 or U.S. dol-

lars. More research is necessary to validate the effectiveness of ECPR, with a focus on the long-term effects of complications of ECPR.

9.6 | Supplementary Material

Supplementary materials can be found in the online version of this article: `https://tinyurl.com/39xs8eac`

References

1. Gräsner, J.-T. *et al.* EuReCa ONE—27 Nations, ONE Europe, ONE Registry: A prospective one month analysis of out-of-hospital cardiac arrest outcomes in 27 countries in Europe. *Resuscitation* **105,** 188–195 (2016).
2. Skogvoll, E, Isern, E, Sangolt, G. K. & Gisvold, S. E. In-hospital cardiopulmonary resuscitation. 5 years' incidence and survival according to the Utstein template. *Acta anaesthesiologica Scandinavica* **43,** 177–84. ISSN: 0001-5172 (1999).
3. Hodgetts, T. J. *et al.* Incidence, location and reasons for avoidable in-hospital cardiac arrest in a district general hospital. *Resuscitation* **44,** 115–123 (2002).
4. Sandroni, C, Nolan, J, Cavallaro, F & Antonelli, M. In-hospital cardiac arrest: Incidence, prognosis and possible measures to improve survival. *Intensive Care Medicine* **33,** 237–245 (2007).
5. Zhu, A. & Zhang, J. Meta-analysis of outcomes of the 2005 and 2010 cardiopulmonary resuscitation guidelines for adults with in-hospital cardiac arrest. *The American journal of emergency medicine* **34,** 1133–1139 (2016).
6. Schluep, M., Gravesteijn, B. Y., Stolker, R. J., Endeman, H. & Hoeks, S. E. One-year survival after in-hospital cardiac arrest: A systematic review and meta-analysis. *Resuscitation* **132,** 90–100. ISSN: 1873-1570 (2018).

7. Hirlekar, G. *et al.* Survival and neurological outcome in the elderly after in-hospital cardiac arrest. *Resuscitation* **118,** 101–106. ISSN: 0300-9572 (2017).

8. Skrifvars, M. B. *et al.* Do patient characteristics or factors at resuscitation influence long-term outcome in patients surviving to be discharged following in-hospital cardiac arrest? *Journal of Internal Medicine* **262,** 488–495. ISSN: 0954-6820 (2007).

9. Andrew, E, Nehme, Z, Bernard, S & Smith, K. The influence of comorbidity on survival and long-term outcomes after out-of-hospital cardiac arrest. *Resuscitation* **110,** 42–47. ISSN: 1873-1570 (2017).

10. Piscator, E., Hedberg, P., Göransson, K. & Djärv, T. Survival after in-hospital cardiac arrest highly associated to age-adjusted Charlson co-morbidity index a cohort study from a two-sited Swedish University hospital. *Resuscitation* **1,** 109–110 (2015).

11. Schluep, M., Rijkenberg, S., Stolker, R. J., Hoeks, S. & Endeman, H. One-year mortality of patients admitted to the intensive care unit after in-hospital cardiac arrest: a retrospective study. *Journal of Critical Care* **48,** 345–351. ISSN: 15578615 (2018).

12. Taglieri, N. *et al.* Prognostic significance of shockable and non-shockable cardiac arrest in ST-segment elevation myocardial infarction patients undergoing primary angioplasty. *Resuscitation* **123,** 8–14. ISSN: 0300-9572 (2018).

13. Karagiannidis, C. *et al.* Extracorporeal membrane oxygenation: evolving epidemiology and mortality. *Intensive Care Medicine* **42,** 889–896. ISSN: 0342-4642 (2016).

14. Extracorporeal life support organisation. *ECLS registry report* tech. rep. (2019).

15. Massetti, M *et al.* Back from irreversibility: Extracorporeal life support for prolonged cardiac arrest. *Annals of Thoracic Surgery* **79,** 178–183. ISSN: 0003-4975 (2005).

16. Tramm, R. *et al.* Extracorporeal membrane oxygenation for critically ill adults. *Cochrane Database of Systematic Reviews* (2015).

17. Ouweneel, D. M. *et al.* Extracorporeal life support during cardiac arrest and cardiogenic shock: a systematic review and meta-analysis. *Intensive Care Medicine* **42,** 1922–1934. ISSN: 0342-4642 (2016).

18. Holmberg, M. J. *et al.* Extracorporeal cardiopulmonary resuscitation for cardiac arrest: A systematic review. *Resuscitation* **131,** 91–100. ISSN: 1873-1570 (2018).

19. Luce, J. M. & Rubenfeld, G. D. Can Health Care Costs Be Reduced by Limiting Intensive Care at the End of Life? *American Journal of Respiratory and Critical Care Medicine* **165,** 750–754. ISSN: 1073-449X (2002).

20. Bharmal, M. I. *et al.* Cost-utility of extracorporeal cardiopulmonary resuscitation in patients with cardiac arrest. *Resuscitation* **136,** 126–130 (2019).

21. Dennis, M. *et al.* Cost effectiveness and quality of life analysis of extracorporeal cardiopulmonary resuscitation (ECPR) for refractory cardiac arrest. *Resuscitation* **139,** 49–56. ISSN: 18731570 (2019).

22. Sonnenberg, F. A. & Beck, J. R. Markov Models in Medical Decision Making. *Medical Decision Making* **13,** 322–338. ISSN: 0272-989X (1993).

23. Gold, M., Siegel, J., Russell, L. & Weinstein, M. *Cost-effectiveness in Health and Medicine.* (Oxford University Press, New York, 1996).

24. Husereau, D. *et al.* Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *European Journal of Health Economics.* ISSN: 16187598 (2013).

25. Callaway, C. W. *et al.* Part 4: Advanced Life Support. *Circulation* **132,** S84–S145. ISSN: 0009-7322 (2015).

26. Cook, I. *et al.* End of life care and do not resuscitate orders: how much does age influence decision making? A systematic review and meta-analysis. *Gerontology and geriatric medicine* **3,** 2333721417713422 (2017).

27. Khan, A. M. *et al.* Age, sex, and hospital factors are associated with the duration of cardiopulmonary resuscitation in hospitalized patients who do not experience sustained return of spontaneous circulation. *Journal of the American Heart Association* **3,** e001044 (2014).

28. Chen, Y. S. *et al.* Comparison of outcome of extracorporeal cardiopulmonary resuscitation for out-of-hospital and in-hospital cardiac arrest. *Circulation* **128** (2013).

29. Sheu, J.-J. *et al.* Early extracorporeal membrane oxygenator-assisted primary percutaneous coronary intervention improved 30-day clinical outcomes in patients with ST-segment elevation myocardial infarction complicated with profound cardiogenic shock. *Critical care medicine* **38,** 1810–7. ISSN: 1530-0293 (2010).

30. Muller, G. *et al.* The ENCOURAGE mortality risk score and analysis of long-term outcomes after VA-ECMO for acute myocardial infarction with cardiogenic shock. *Intensive Care Medicine* **42,** 370–378 (2016).

31. Sakamoto, S., Taniguchi, N., Nakajima, S. & Takahashi, A. Extracorporeal life support for cardiogenic shock or cardiac arrest due to acute coronary syndrome. *The Annals of thoracic surgery* **94,** 1–7. ISSN: 1552-6259 (2012).

32. Lansink-Hartgring, A. O. *et al.* Hospital Costs of Extracorporeal Life Support Therapy. *Critical Care Medicine.* ISSN: 15300293 (2016).

33. Israelsson, J. *et al.* Health status and psychological distress among in-hospital cardiac arrest survivors in relation to gender. *Resuscitation* **114,** 27–33. ISSN: 0300-9572 (2017).

34. Makdisi, G. & Wang, I.-W. Extra Corporeal Membrane Oxygenation (ECMO) review of a lifesaving technology. *Journal of thoracic disease* **7,** E166–76. ISSN: 2072-1439 (2015).

35. Centraal Bureau voor de Statistiek. *Levensverwachting; geslacht, leeftijd (per jaar en periode van vijf jaren)* 2018.

36. Feingold, P. L. *et al.* Long-term survival following in-hospital cardiac arrest: A matched cohort study. *Resuscitation* **99,** 72–78 (2016).

37. Zorginstituut Nederland. Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg (2016).

38. Girotra, S. *et al.* Trends in survival after in-hospital cardiac arrest. *New England Journal of Medicine* **367,** 1912–1920 (2012).

39. Alarid-Escudero, F. dampack: an R package with useful functions to develop and analyze decision-analytic models (2018).

40. Jalal, H. *et al.* An Overview of R in Health Decision Sciences. *Medical Decision Making* **37,** 735–746. ISSN: 0272-989X (2017).

41. Soar, J. *et al.* European Resuscitation Council Guidelines for Resuscitation 2015. *Resuscitation* **95,** 100–147. ISSN: 03009572 (2015).

42. Garber, A. M. & Skinner, J. Is American Health Care Uniquely Inefficient? *Journal of Economic Perspectives.* ISSN: 0895-3309 (2008).

43. Aubin, H. *et al. A Suprainstitutional Network for Remote Extracorporeal Life Support A Retrospective Cohort Study* tech. rep. (2016).

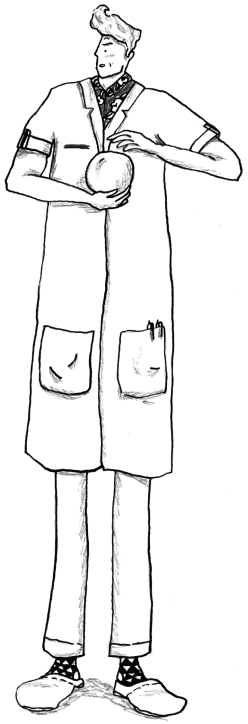
44. Harvey, M. J., Gaies, M. G. & Prosser, L. A. US and international in-hospital costs of extracorporeal membrane oxygenation: a systematic review. *Applied health economics and health policy* **13,** 341–357 (2015).

45. Karagiannis, C., Georgiou, M., Kouskouni, E., Iacovidou, N. & Xanthos, T. Association of lactate levels with outcome after in-hospital cardiac arrest. *Resuscitation* **83,** e175–e176 (2012).

46. Hastings, S. L., Pellegrino, V. A., Preovolos, A. & Salamonsen, R. F. Survey of adult extracorporeal membrane oxygenation (ECMO) practice and attitudes among Australian and New Zealand intensivists. *Critical care and resuscitation : journal of the Australasian Academy of Critical Care Medicine* **10,** 46. ISSN: 1441-2772 (2008).

Part III

Identifying patients at risk



10

Towards a New Multi-Dimensional Classification of Traumatic Brain Injury: A Collaborative European NeuroTrauma Effectiveness Research for Traumatic Brain Injury Study

Benjamin Y Gravesteijn, Charlie A Sewalt, Ari Ercole, Cecilia Akerlund, David Nelson, Andrew IR Maas, David K Menon, Hester F Lingsma, Ewout W Steyerberg, and the CENTER-TBI Collaborators

Journal of Neurotrauma

10.1 | Abstract

Traumatic brain injury (TBI) is currently classified as mild, moderate, or severe TBI by trichotomizing the Glasgow Coma Scale (GCS). We aimed to explore directions for a more refined multidimensional classification system. For that purpose, we performed a hypothesis-free cluster analysis in the Collaborative European NeuroTrauma Effectiveness Research for TBI (CENTER-TBI) database: a European all-severity TBI cohort (n = 4509). The first building block consisted of key imaging characteristics, summarized using principal component analysis from 12 imaging characteristics. The other building blocks were demographics, clinical severity, secondary insults, and cause of injury. With these building blocks, the patients were clustered into four groups. We applied bootstrap resampling with replacement to study the stability of cluster allocation. The characteristics that predominantly defined the clusters were injury cause, major extracranial injury, and GCS. The clusters consisted of 1451, 1534, 1006, and 518 patients, respectively. The clustering method was quite stable: the proportion of patients staying in one cluster after resampling and reclustering was 97.4% (95% confidence interval [CI]: 85.6–99.9%). These clusters characterized groups of patients with different functional outcomes: from mild to severe, 12%, 19%, 36%, and 58% of patients had unfavourable 6 month outcome. Compared with the mild and the upper intermediate cluster, the lower intermediate and the severe cluster received more key interventions. To conclude, four types of TBI patients may be defined by injury mechanism, presence of major extracranial injury and GCS. Describing patients according to these three characteristics could potentially capture differences in etiology and care pathways better than with GCS only.

10.2 | Background

The global burden of traumatic brain injury (TBI) is high: it is a leading cause of injury-related death and disability [1]. Although the rates vary among countries, TBI is estimated to be responsible for around 300 hospital admissions and 12 deaths per 100,000 persons per year in Europe [2]. TBI is currently classified using the baseline Glasgow Coma Scale (GCS) [3]. Although there is variation [4], TBI is usually divided according to GCS scores 3–8 (severe), 9–12 (moderate), and 13–15 (mild).

The current classification, based on only GCS, does not fully capture the multidimensionality of TBI [5, 6]. TBI is defined as an alteration in brain function, or other brain pathology, following an external force [7]. However, the manifestation of TBI is heterogeneous: a variety of pathoanatomical lesions can be present as the result of a multitude of trauma mechanisms [5]. A novel multidimensional classification of TBI could potentially be used for improving the efficiency of care pathways. Additionally, the classification could increase understanding of the divergent clinical courses of TBI patients.

This study aimed to explore directions for a more refined multidimensional classification system, capturing the heterogeneity throughout the entire spectrum of TBI severity. For that purpose, a hypothesis-free cluster analysis was performed.

10.3 | Methods

10.3.1 | Study population

Data from the Collaborative European NeuroTrauma Effectiveness Research for TBI (CENTER-TBI) was used for this analysis. This prospective cohort study comprised 4509 patients with all-severity TBI. The patients were included in 59 centres from 18 countries across Europe. Inclusion criteria were a clinical diagnosis of TBI, presentation within 24 h, and clinical indication for computed tomographic (CT) scanning. The exclusion criterion of CENTER TBI was pre-existing neurological disease. For this study, the total CENTER TBI cohort was used. The study design was previously published [8]. Version 1.0 of the database was used.

10.3.2 | Variable selection

The cluster analysis was hypothesis free, as we did not assume any relationship, weights, or importance among the variables, or a role such as exposure, confounder, or outcome. However, to arrive at a set of variables to be used by the algorithm, a starting point was that the classification should be implementable, including characteristics that are generally available at any emergency department. Additionally, we wanted to use prognostically relevant characteristics: the characteristics of which the International Mission for Prognosis and Analysis of Clinical Trials in TBI (IMPACT) and Corticosteroid Randomization After Significant Head Injury (CRASH) prediction models are composed [9, 10]. Finally, we included variables describing the mechanism of injury.

The prognostic and mechanistic relevant variables were aggregated in

"building blocks": groups of variables describing similar information of a patient. The building blocks that were used for the exploratory clustering were: (1) demographics: age; (2) clinical severity: baseline GCS score, baseline pupil score, and major extracranial injury (defined as an abbreviated injury scale [AIS] >3 in a body region other than neck and head); (3) second insults: hypoxia and hypotension in the emergency department; (4) cause of injury: road traffic incident (RTI), all falls, violence, or suicide, or other; and (5) imaging characteristics: all imaging characteristics available in the database, which are the presence of epidural hematoma, subdural mixed density collection, skull fracture, subacute subdural hematoma, midline shift (> 5 mm), traumatic subarachnoid haemorrhage, any mass lesion, intraventricular hemorrhage, subdural hematoma, or cisternal compression. Imaging characteristics were obtained through a central reviewing process [11].

10.3.3 | Clustering

First, the key imaging characteristics were extracted. The imaging characteristics comprised 12 binary variables, which are not easily handled by a clustering algorithm. Therefore, to increase efficiency of the clustering algorithm, we described all those binary variables using principal components: the primary principal component is a continuous variable capturing the most information across the included variables. The second principal component captures somewhat less, and subsequent principal components capture progressively less. The PCAmixdata package was used, because this version of a PCA can handle non-continuous data [12]. Consecutively, the first four principal components (dimensions) were included in the clustering algorithm. We included four principal components, because these

described the majority (> 70%) of the variability in the imaging characteristics. Although principal components themselves are not clinically applicable, they can be easily calculated from all binary imaging variables.

The selected clinical and injury severity variables (n = 8), together with the four imaging dimensions were included in a clustering algorithm. The cluster package was used. First, the metric on which the data are grouped is calculated. Because we are using both categorical and numerical data, the Gower's distance was calculated with the daisy function [13]. Using this distance metric, four clusters in the data were identified using the partition around medoids (pam) function.

Clustering studies with mixed data may optimize the silhouette value to arrive at an optimal number of clusters [14]. It is a measure of the similarity to its own cluster (cohesion), compared with other clusters (separation).

Stability of the clustering was assessed using the same variables and a bootstrapping procedure to repeatedly resample with replacement and recluster the patients. The proportion of patients who stayed in a cluster after resampling was calculated per repetition. The median and 95% credibility interval, defined by the 2.5th to the 97.5th percentile, was calculated with 999 repetitions.

To assess the importance of the clustering variables, we used multinomial regression. The independent variables of this regression were the four clusters, and the dependent variables were the clustering variables. We assumed linear effects, and we did not allow for any statistical interaction. The partial Nagelkerke R^2 was calculated for each variable by comparing the Nagelkerke R^2 of the model without the variable to the Nagelkerke R^2 of the model with the variable.

10.3.4 | Cluster description

The clusters were described based on the clustering variables. Additionally, gender, motor GCS score, as well as clinical course characteristics (receiving intracranical pressure [ICP] monitoring, intracranial or extracranial surgery, length of intensive care unit [ICU] stay) were described across the clusters. We then examined the outcome of the patients within the clusters.

First, the 6 months Glasgow Outcome Scale – Extended (GOS-E) was used to describe the functional outcome. The GOS-E score was imputed exactly at 180 days, using a multi-state model. Subsequently, outcomes among the clusters were compared, and used to rank the clusters based on the proportion of favourable outcomes in the following order: "mild," "upper intermediate," "lower intermediate," or "severe." This order resembles the GOS-E, in which "lower" refers to the more severe category (e.g.: "lower severe disability" versus "upper severe disability"). The clusters were named accordingly to enable easier interpretation of the characteristics of clusters. Second, using all baseline characteristics in a logistical regression model, the predicted probability of 6 months unfavourable outcome (GOS-E < 5) was calculated. The observed and predicted probabilities were compared to assess the calibration of the model within the four clusters.

Further, the most important classification strategies, as defined by the partial R^2, were used to describe the patients. The GOS of all combinations of possible characteristics was visually assessed.

Finally, we assessed whether the baseline characteristics included in the clustering algorithm were prognostically relevant. Ordinal logistical regression with GOS-E as outcome variable was used. The area under the receiver operating characteristic (ROC) curve was used to describe the dis-

crimination of the models. The following models were compared:

1. GCS
2. GCS + most important clustering variables (defined by the partial R^2)
3. GCS + pupils + age (core version of the IMPACT model [9])

All analyses were performed using R.[1] The published code can be found online.[2]

10.4 | Results

The 4509 patients in the CENTER-TBI study were on average 50 (interquartile range [IQR]: 30–66) years old, and predominantly male (67%). The most important causes of injury were road traffic incident (RTI) (39%) and incidental falls (47%). The majority of patients were classified as having mild TBI: the median GCS in the cohort was 15 (IQR: 10–15) (table 10.1).

10.4.1 | Imaging characteristics

The first four dimensions of the principal component analysis (PCA) explained 68% of the variation in all imaging characteristics. In the first dimension, the dimension explaining most of the variability in all imaging characteristics (34%), the most important imaging characteristics were the absence or presence of traumatic axonal injury, midline shift, and subdural mixed density (Figure 1 in the supplementary material available online).

[1] R Core Team (2013). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria

[2] `https://github.com/ErasmusCMB/CENTER-TBI/blob/master/code_classification_TBI.R`.

10.4.2 | Clustering analysis

We restricted the number of clusters to four for easy interpretation, and thereby used a silhouette value similar to the maximum silhouette value (0.21 with 3, 0.24 with 4, and 0.25 with 5). The most important building blocks of the clusters were injury cause, major extracranial injury, and GCS, respectively: the partial R^2, indicating relative importance in the clusters, were 13%, 5%, and 2%, respectively. The key imaging characteristics and age also were relatively important clustering characteristics (figure 10.1).

The clustering method was quite stable: the proportion of patients staying in one cluster after resampling and reclustering was 97% (95% CI: 86–100%). Four examples of resampling and recluster iterations are shown in Figure 2 in the supplementary material available online.

From mild to severe, 12%, 19%, 36%, and 58% of patients had unfavourable outcome in the four clusters (figure 10.2). The same pattern was seen for mortality, where 1%, 4%, 8%, and 17% mortality rates were observed. Based on the model with the IMPACT variables fitted on the data, the severe cluster had 1.5 times worse functional outcome than expected (calibration intercept: 0.4, 95% CI: 0.2–0.6; observed to expected ratio 1.5; Figure 3 in the supplementary material available online). From mild to severe, the four clusters consisted of 1451, 1534, 1006, and 518 patients respectively (table 10.2).

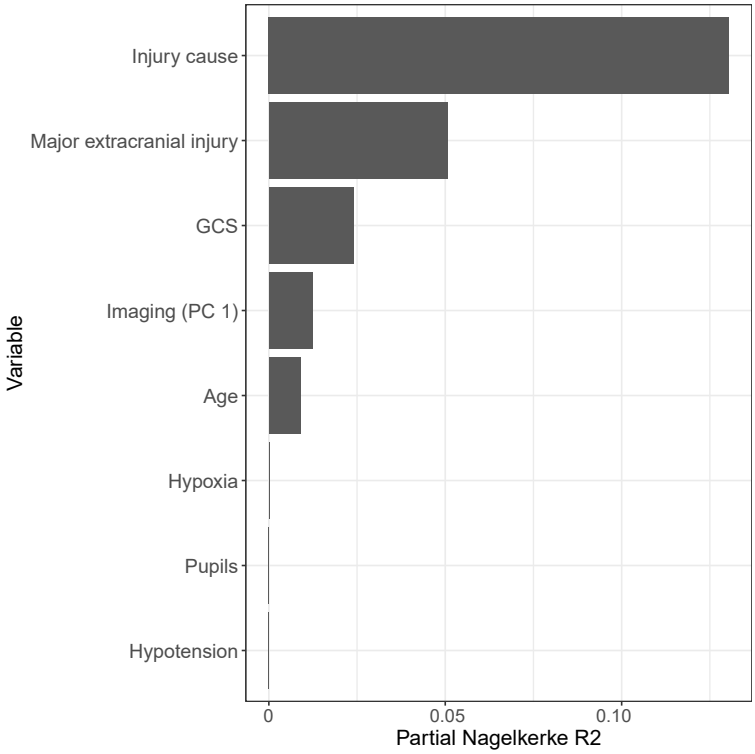


Figure 10.1: *The importance of the variables to identify the four clusters, quantified by the partial Nagelkerke R^2 value of the multinomial model predicting class. The R^2 is a measure for the proportion of the variation in outcome (class) explained by the predictors (the clustering variables). Imaging is displayed, which is the first principal component (PC 1) of the imaging characteristics.*

Table 10.1: *Baseline Characteristics Used for the Clustering, as Well as the Six Month Outcome*

In k-mode clustering	n = 4509	Missing
Age (median [IQR])	50 [30, 66]	0.0
Injury cause (%)		3.7
RTI	1682 (38.7)	
Fall	2024 (46.6)	
Other	343 (7.9)	
Violence/suicide	293 (6.7)	
GCS Motor (median [IQR])	6.0 [5.0, 6.0]	2.5
GCS Score (median [IQR])	15.0 [10.0, 15.0]	4.0
Pupils (%)		5.8
Both reactive	3802 (89.5)	
One reactive	164 (3.9)	
None reactive	281 (6.6)	
ED hypoxia (%)	299 (7.0)	5.6
ED hypotension (%)	297 (6.9)	4.7
Major extracranial injury* (%)	668 (14.8)	0.0
PCA before clustering		
Axonal injury (%)	324 (9.4)	23.2
Contusion (%)	1087 (31.4)	23.2
Subdural hematoma subacute chronic (%)	17 (0.5)	23.2
Traumatic subarachnoid hemorrhage (%)	1531 (44.2)	23.2
Epidural hematoma (%)	373 (10.8)	23.2
Subdural hematoma acute (%)	943 (27.2)	23.2
Skull fracture (%)	1266 (36.6)	23.2
Subdural collection mixed density (%)	82 (2.4)	23.2
Cisternal compression (%)	494 (14.3)	23.2
Midline shift (%)	380 (11.0)	23.2
Mass lesion (%)	579 (16.7)	23.2
Intraventricular hemorrhage (%)	453 (13.1)	23.2
Stratum (%)		0.0
ER	848 (18.8)	
Admission	1523 (33.8)	
ICU	2138 (47.4)	
6 month outcome		
GOSE (%)		15.7
1	475 (12.5)	
2**	370 (9.7)	
4	110 (2.9)	
5	198 (5.2)	
6	401 (10.6)	
7	725 (19.1)	
8	1520 (40.0)	

** Defined as non-head Abbreviated Injury Scale (AIS) *leq*3
** GOSE 2 & 3 are combined into 2
IQR, interquartile range; RTI, road traffic incident; GCS, Glasgow Coma Scale; ED, emergency department; PCA, principal component analysis; ER, emergency room; ICU, intensive care unit; GOS-E, Glasgow Outcome Scale–Extended

Table 10.2: *The Characteristics of the Four Clusters*

Cluster n =	Mild 1451	Upper intermediate 1534	Lower intermediate 1006	Severe 518
Age (median [IQR])	38 [24, 53]	61 [42, 73]	57 [42, 70]	40 [25, 57]
Male (%)	1000 (68.9)	909 (59.3)	708 (70.4)	405 (78.2)
Injury cause (%)				
RTI	1095 (78.2)	0 (0.0)	194 (20.2)	393 (79.1)
Fall	0 (0.0)	1354 (91.3)	617 (64.2)	53 (10.7)
Other	138 (9.9)	90 (6.1)	91 (9.5)	24 (4.8)
Violence/suicide	168 (12.0)	39 (2.6)	59 (6.1)	27 (5.4)
GCS Motor (median [IQR])	6 [6, 6]	6 [6.0, 6.0]	6 [4.0, 6.0]	1 [1, 4]
GCS Score(median [IQR])	15 [14, 15]	15 [14, 15]	13 [8, 15]	3 [3,7]
Pupils (%)				
Both reactive	1298 (94.7)	1339 (93.1)	829 (87.8)	336 (67.9)
One reactive	30 (2.2)	45 (3.1)	38 (4.0)	51 (10.3)
None reactive	42 (3.1)	54 (3.8)	77 (8.2)	108 (21.8)
ED hypoxia (%)	60 (4.4)	59 (4.0)	57 (6.1)	123 (25.4)
ED hypotension (%)	76 (5.4)	51 (3.5)	45 (4.8)	125 (25.8)
Major extracranial injury (%)	145 (10.0)	87 (5.7)	52 (5.2)	384 (74.1)

IQR, interquartile range; RTI, road traffic incident; GCS, Glasgow Coma Scale; ED, emergency department; ICP, intracranial pressure; LOS, length of stay; LOICUS, length of ICU stay; ICU, intensive care unit.

Cluster n =	Mild 1451	Upper intermediate 1534	Lower intermediate 1006	Severe 518
Subdural hematoma Subacute chronic (%)	2 (0.2)	9 (0.8)	6 (0.8)	0 (0.0)
Traumatic subarachnoid hemorrhage (%)	269 (24.0)	302 (25.7)	640 (83.6)	320 (80.4)
Epidural hematoma (%)	77 (6.9)	55 (4.7)	169 (22.1)	72 (18.1)
Subdural hematoma acute (%)	130 (11.6)	206 (17.5)	428 (55.9)	179 (45.0)
Skull fracture (%)	232 (20.7)	216 (18.4)	584 (76.2)	234 (58.8)
Subdural collection mixed density (%)	8 (0.7)	30 (2.5)	34 (4.4)	10 (2.5)
Cisternal compression (%)	39 (3.5)	92 (7.8)	220 (28.7)	143 (35.9)
Midline shift (%)	28 (2.5)	94 (8.0)	180 (23.5)	78 (19.6)
Mass lesion (%)	34 (3.0)	113 (9.6)	328 (42.8)	104 (26.1)
Intraventricular hemorrhage (%)	66 (5.9)	64 (5.4)	150 (19.6)	173 (43.5)
Axonal injury (%)	77 (6.9)	51 (4.3)	78 (10.2)	118 (29.6)
Contusion (%)	49 (4.4)	0 (0.0)	765 (99.9)	273 (68.6)
ICP monitoring (%)	99 (6.9)	92 (6.0)	245 (24.5)	308 (59.9)
Intracranial surgery (%)	75 (5.2)	116 (7.6)	214 (21.4)	116 (22.4)
Extracranial surgery (%)	129 (9.0)	49 (3.2)	37 (3.7)	134 (25.9)
LOS (median [IQR])	5 [2, 12]	4 [2, 10]	11 [6, 23]	31 [18, 47]
LOICUS (median [IQR])				
Stratum (%)				
ED	361 (24.9)	440 (28.7)	46 (4.6)	1 (0.2)
Admission	603 (41.6)	660 (43.0)	258 (25.6)	2 (0.4)
ICU	487 (33.6)	434 (28.3)	702 (69.8)	515 (99.4)

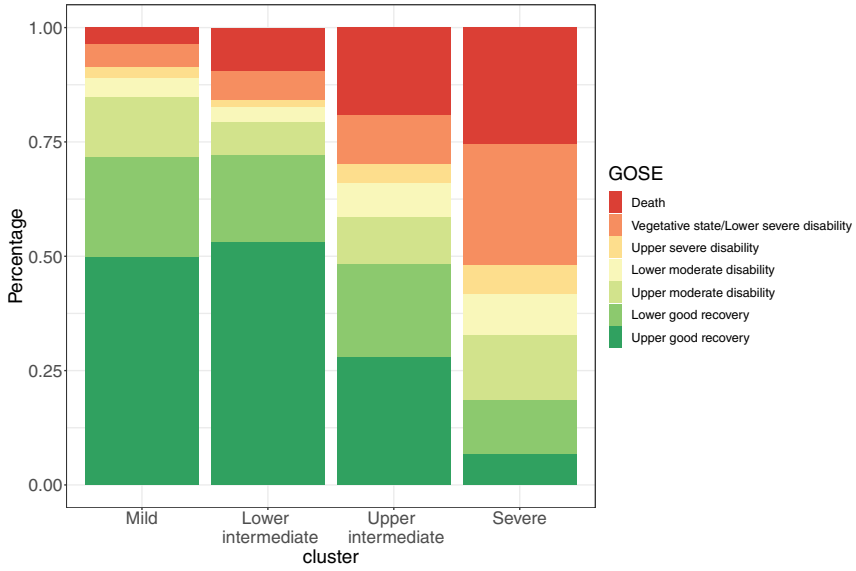


Figure 10.2: *Outcome of the four clusters. The stacked bar chart shows the distributions of Glasgow Outcome Scale – Extended (GOS-E) in the four identified clusters.*

The mild and the severe cluster consisted of younger patients (median of 38 [24–53] and 40 [25–57] years old, compared with 61 [42–73] in the upper intermediate cluster and 57 [42–70] in the lower intermediate cluster). In these younger patients, the trauma was predominantly caused by road traffic incidents, instead of incidental falls. The lower intermediate and the severe cluster consisted of patients with a median GCS < 15, and more unreactive pupils.

The different clusters were also characteriezd by different care pathways and disease evolutions. In the severe cluster, 515 (99%) patients were admitted to the ICU, whereas only 702 (70%) of the patients in the lower intermediate cluster were admitted to the ICU. Compared with the mild and

the upper intermediate cluster, the lower intermediate and the severe cluster received more key interventions, such as ICP monitoring, intracranial surgery, and extracranial surgery. However, the severe cluster consisted of more patients requiring extracranial surgery: 134 (26%) versus 37 (3.7%) in the lower intermediate cluster. Although the length of (ICU) stay was longer in the lower intermediate and severe cluster, the length of (ICU) stay was longest in the severe cluster: the median length of ICU stay was 15 (IQR: 7– 24) days in the severe cluster, whereas in the lower intermediate cluster, the median was 3 (IQR: 0–10); the length of hospital stay was on average 30.5 (IQR: 18–47) days in the severe cluster, compared with a median length of hospital stay of 11 (IQR: 6–23) days in the lower intermediate cluster. Although some of the patients in the upper intermediate and the mild cluster were admitted to the ICU, the median ICU length was 0 (IQR 0–2 for the mild cluster, and 0–1 for the upper intermediate cluster).

All these characteristics are also presented for the current classification based on GCS in Table 1 in the supplementary material available online. In comparison to the four clusters, the groups based on GCS scaled less well with demographic differences and cause of injury: the median age was 46 (IQR: 25–64) in the severe group, 53 (IQR: 34–69) in the moderate group, and 51 (IQR 31–67) in the mild group. The proportions of road traffic accidents were 47%, 36%, and 35% in the three groups, respectively. The treatment intensity and presence of imaging abnormalities differed across the three groups.

Based on the most important clustering variables, the patients were described again on outcome (figure 10.3). The distribution of GOS-E scores was mainly different for patients with lower GCS scores. The largest group consisted of low energy (no road traffic incident), mild (GCS 13–15) TBI with major extracranial injury (1125 [25%]). The smallest groups were high

energy moderate (GCS 9–12) and severe (GCS <9) TBI without major extracranial injury: 44 (1%) and 80 (2%) patients, respectively.

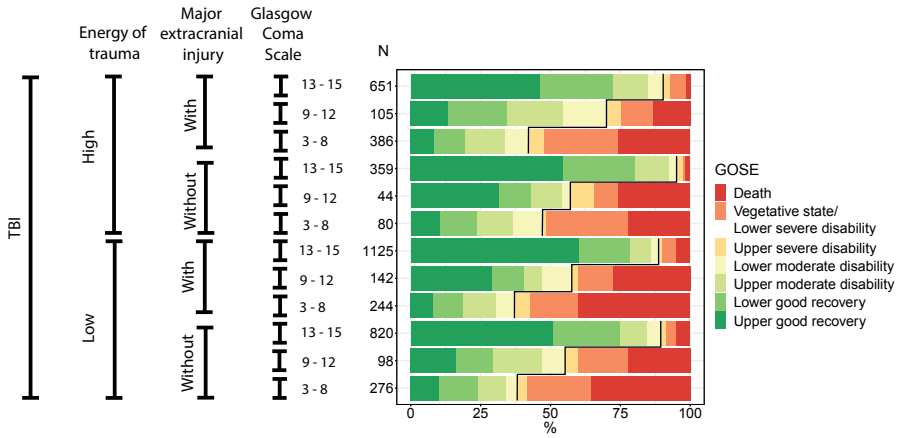


Figure 10.3: *The proposed classification system for traumatic brain injury (TBI) patients and their observed Glasgow Outcome Scale – Extended (GOS-E) scores. The classification is based on the characteristics that mostly defined the clustering algorithm. The black line in the stacked bar chart indicates the border of unfavourable and favourable.*

For the prediction of functional outcome, the model with only GCS had an area under the ROC curve of 0.72 (95% CI 0.71–0.72). Adding major extracranial injury and cause of injury (as most relevant clustering variables), did not improve the discrimination of the model. In contrast, adding age and pupils did increase the area under the ROC curve to 0.75 (95% CI 0.74–0.75).

10.5 | Discussion

This study was a hypothesis-free exploration of cluster analysis in TBI to inform development of a new, multidimensional classification for TBI. We

clustered TBI patients into four groups. The most defining building blocks of the clustered groups were injury cause, major extracranial injury, and GCS. With these three most defining characteristics, patients could be classified into 12 groups, ranging from high energy mild TBI with major extracranial injury, to low energy severe TBI without major extracranial injury.

Our proposed classification might capture differences in required treatment approaches, irrespective of differences in prognosis. Patients with similar risk of outcome could still require different treatment approaches [15]. As an illustration, an elderly patient with multiple comorbidities who fell at home might, according to the IMPACT model, be at equal risk of dying and unfavourable outcome within 6 months compared with a younger patient with TBI caused by a road traffic incident (figure 10.4) [9]. Even though their risk would be equal, they would need different approaches of care: our study suggests that the first patient would more likely require intracranial surgery and would have a relatively short ICU stay, and the latter would require extracranial surgery and ICP monitoring with a long ICU stay.

Additionally, the characteristics identified by our study relate to care pathways. This is because they are already used to hand over trauma patients. This is the experience in our hospitals. A possible reason is that the widely used format for handovers, the "Situation, Background, Assessment, Recap or treatment" (acronym: S-BAR) [16], dictates including background information: this is typically described by the mechanism of injury, and whether the patient has major extracranial injury. Clinical experience has led to the description of these characteristics, because they apparently impact care pathways.

Describing TBI patients based on energy of trauma and major extracra-

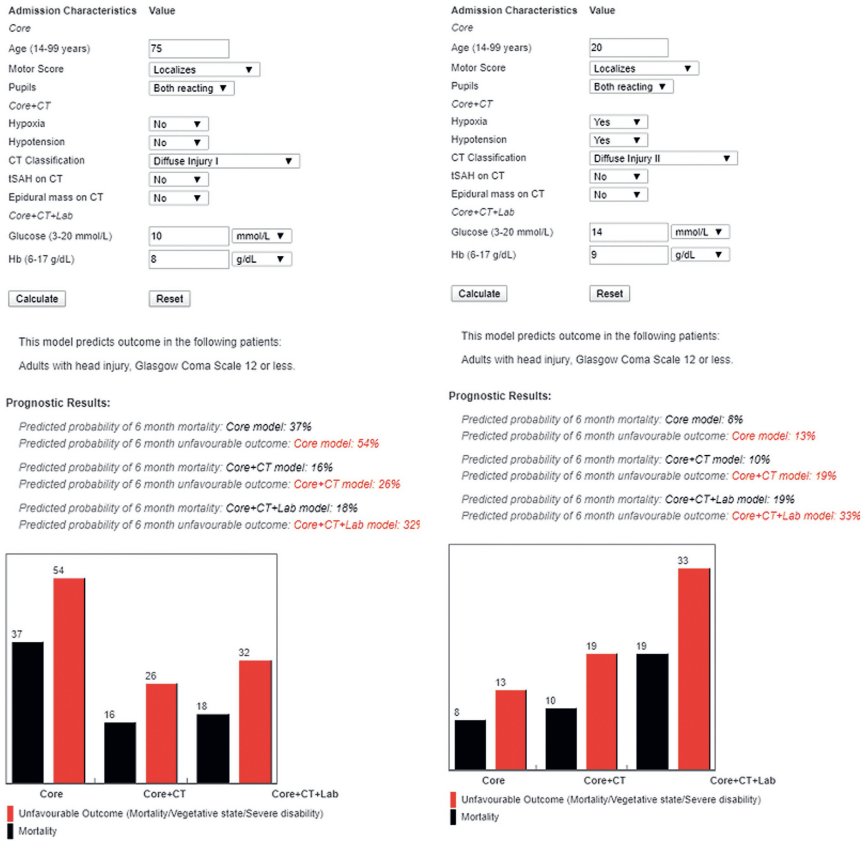


Figure 10.4: *Two exemplary patients: an elderly patient who fell and a younger patient who was in a road traffic accident. Their predicted risk of 6 month mortality and 6 month unfavourable outcome is similar in both cases. These figures are made from:* `http://www.tbi-impact.org/?p=impact/calc`.

nial injury potentially may capture etiological differences and could possibly improve the development of new treatments and subsequent clinical trials in the TBI field. It has been suggested that the traditional classification of TBI is one of the causes of a history of negative trials in TBI [5,

17]. A classification that better integrates the pathological differences in the heterogeneous TBI patient population could enable more focused, and therefore potentially more positive, trials.

It could be argued that imaging characteristics, which we included in our analysis, are not always available at the emergency department: only selected TBI patients should be scanned, to avoid unnecessary oncogenic risk of radiation, costs, and productivity loss [18]. However, in contrast to novel biomarkers, or characteristics visible on magnetic resonance imaging (MRI) scan, CT characteristics are usually available. Moreover, imaging characteristics are key to discerning different TBI pathologies, such as epidural versus subdural hematoma. Our aim was to explore a classification that better describes the variation in TBI pathologies. Therefore, it was considered essential to include this type of information.

The fact that this study has applied hypothesis-free analyses in a large TBI database is both a limitation and a strength. On the one hand, a data-driven approach to clustering could lead to poor generalizability. Moreover, critique on clustering algorithms often involves low interpretability of the clusters, because they are not based on pre-existing subject knowledge [19]. In our case, the clustering approach revitalized the importance of describing patients using major extracranial injury and mechanism of injury. This is in contrast with previous research, which mainly has focused on a prognostic, instead of a mechanistic, description of TBI patients [1].

Another limitation is that we did not take biomarker profiles into account. Currently, there is not enough knowledge about longitudinal biomarker profiles. Implementing these profiles could improve the classification, and more research is necessary to know what precisely should be included in such a classification.

Finally, another limitation of our study is that the current analysis is biased toward classifying more severe injuries. The majority of the used variables are known to be prognostically relevant for moderate to severe TBI [9]. Further, ICU patients were preferentially included in the core CENTER-TBI database. This resulted in a somewhat selected TBI sample. However, 2310 (51%) of the patients in our sample were non-ICU patients. Moreover, most heterogeneity is to be expected among those patients with severe TBI [5]. Therefore, it can be argued that analyzing a cohort with an over-representation of the most heterogeneous subgroup can assist in better characterizing the disease. However, we recognize that other variables might be more appropriate for clustering milder TBI patients.

10.5.1 | Conclusion

After unsupervised, hypothesis-free clustering, four clusters were identified, which were mainly defined by injury mechanism, presence of major extracranial injury, and GCS. Describing patients with these three characteristics could potentially capture more differences in etiological and care pathway aspects than based on GCS alone. Our proposed classification should be validated and extended upon; in particular, we feel that biomarkers could play an important role.

10.6 | Supplementary Material

Supplementary materials can be found in the online version of this article: `https://tinyurl.com/dzb273r6`

References

1. Maas, A. I. R. *et al.* Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. *The Lancet Neurology* **4422.** ISSN: 14744422 (2017).
2. Majdan, M. *et al.* Epidemiology of traumatic brain injuries in Europe: a cross-sectional analysis. *The Lancet Public Health* **1,** e76–e83. ISSN: 2468-2667 (2016).
3. Teasdale, G & Jennett, B. Assessment of coma and impaired consciousness. A practical scale. *The Lancet* **2,** 81–84. ISSN: 0140-6736 (Print) (1974).
4. Foks, K. A. *et al.* Management of Mild Traumatic Brain Injury at the Emergency Department and Hospital Admission in Europe: A Survey of 71 Neurotrauma Centers Participating in the CENTER-TBI Study. *Journal of Neurotrauma* **34,** 2529–2535. ISSN: 0897-7151 (2017).
5. Saatman, K. E. *et al.* Classification of traumatic brain injury for targeted therapies. *Journal of neurotrauma* **25,** 719–738 (2008).
6. Maas, A. I. *et al.* Re-orientation of clinical research in traumatic brain injury: report of an international workshop on comparative effectiveness research. *Journal of neurotrauma* **29,** 32–46 (2012).
7. Menon, D. K., Schwab, K., Wright, D. W. & Maas, A. I. Position Statement: Definition of Traumatic Brain Injury. *Archives of Physical Medicine and Rehabilitation* **91,** 1637–1640 (2010).
8. Maas, A. I. *et al.* Collaborative European neurotrauma effectiveness research in traumatic brain injury (CENTER-TBI): A prospective longitudinal observational study. *Neurosurgery* **76,** 67–80. ISSN: 15244040 (2015).
9. Steyerberg, E. W. *et al.* Predicting Outcome after Traumatic Brain Injury: Development and International Validation of Prognostic Scores Based on Admission Characteristics. *PLoS Medicine* **5** (ed Singer, M.) e165. ISSN: 1549-1676 (2008).

10. Collaborators, M. *et al.* Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients. *British Medical Journal* **336,** 425–429 (2008).

11. Vande Vyvere, T. *et al.* Central versus Local Radiological Reading of Acute Computed Tomography Characteristics in Multi-Center Traumatic Brain Injury Research. *Journal of Neurotrauma*, neu.2018.6061. ISSN: 0897-7151 (2018).

12. Chavent, M., Kuentz-Simonet, V., Labenne, A. & Saracco, J. Multivariate analysis of mixed data: The R Package PCAmixdata. *arXiv preprint arXiv:1411.4911* (2014).

13. Gower, J. C. A General Coefficient of Similarity and Some of Its Properties. *Biometrics* **27,** 857. ISSN: 0006341X (1971).

14. Rousseeuw, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* **20,** 53–65. ISSN: 03770427 (1987).

15. Roozenbeek, B. *et al.* Baseline characteristics and statistical power in randomized controlled trials: Selection, prognostic targeting, or covariate adjustment? *Critical Care Medicine* **37,** 2683–2690. ISSN: 15300293 (2009).

16. Thomas, C. M., Bertram, E. & Johnson, D. The SBAR communication technique: teaching nursing students professional communication skills. *Nurse educator* **34,** 176–80. ISSN: 1538-9855 (2009).

17. Pineda, J. A. *et al.* Re-Orientation of Clinical Research in Traumatic Brain Injury: Report of an International Workshop on Comparative Effectiveness Research. *Journal of Neurotrauma*. ISSN: 0897-7151 (2011).

18. Foks, K. A. *et al.* External validation of computed tomography decision rules for minor head injury: prospective, multicentre cohort study in the Netherlands. *British Medical Journal* **362** (2018).

19. Feher, M. D., Munro, N., Russell-Jones, D., de Lusignan, S. & Khunti, K. *Novel diabetes subgroups* 2018.

11

Machine learning algorithms performed no better than regression models for prognostication in traumatic brain injury

Benjamin Y Gravesteijn, Daan Nieboer, Ari Ercole, Hester F Lingsma, David Nelson, Ben van Calster, Ewout W Steyerberg, and the CENTER-TBI collaborators

Journal of Clinical Epidemiology

11.1 | Abstract

11.1.1 | Objective

We aimed to explore the added value of common machine learning (ML) algorithms for prediction of outcome for moderate and severe traumatic brain injury.

11.1.2 | Study design and setting

We performed logistic regression (LR), lasso regression, and ridge regression with key baseline predictors in the IMPACT-II database (15 studies, n = 11,022). ML algorithms included support vector machines, random forests, gradient boosting machines, and artificial neural networks and were trained using the same predictors. To assess generalizability of predictions, we performed internal, internal-external, and external validation on the recent CENTER-TBI study (patients with Glasgow Coma Scale <13, n = 1,554). Both calibration (calibration slope/intercept) and discrimination (area under the curve) was quantified.

11.1.3 | Results

In the IMPACT-II database, 3,332/11,022 (30%) died and 5,233 (48%) had unfavourable outcome (Glasgow Outcome Scale less than 4). In the CENTER-TBI study, 348/1,554 (29%) died and 651 (54%) had unfavourable outcome. Discrimination and calibration varied widely between the studies and less so between the studied algorithms. The mean area under the curve was 0.82 for mortality and 0.77 for unfavourable outcomes in the CENTER-TBI study.

11.1.4 | Conclusion

ML algorithms may not outperform traditional regression approaches in a low-dimensional setting for outcome prediction after moderate or severe traumatic brain injury. Similar to regression based prediction models, ML algorithms should be rigorously validated to ensure applicability to new populations.

What is new?

Key findings
Considering discrimination and calibration and overall performance, no clear difference was seen in performance between machine learning (ML) algorithms or regression-based models.
More variability in performance (both discrimination and calibration) was seen between study populations than between algorithms.
What this adds to what was known?
A recent systematic review showed that studies that suggested superior performance of ML methods are more prone to bias. However, these studies mainly focused on comparing discriminative performance of these models. Our study also focused on performance in terms of calibration and generalizability.
What is the implication and what should change now?
Using novel ML algorithms will likely not improve outcome prediction. Instead, prediction research should focus including predictors with substantial incremental prognostic value.
Prediction models, based on both ML algorithms and regression-based methods, need continuous validation and updating to ensure applicability to new populations.

11.2 | Introduction

Traumatic brain injury (TBI) is a common disease, with a significant societal burden [1]: TBI is estimated to be responsible for around 300 hospital admissions and 12 deaths per 100,000 persons per year in Europe [2]. TBI is a heterogeneous disease in terms of phenotype and prognosis [3]. Therefore, prognostic models, which predict outcome for a patient, given a particular combination of baseline characteristics, are important: they may give us insight in mechanisms of disease that lead to poor outcome and allow for risk-based stratification of patients for logistic, research, and clinical reasons.

A large number of prediction models have been developed to predict outcome for patients with TBI, mostly using traditional regression techniques [4]. However, these models have not yet been widely implemented in clinical practice. In recent years, more flexible machine learning (ML) algorithms have enjoyed enthusiasm as potentially promising techniques to improve outcome prognostication [5]. Frequently used methods are support vector machines (SVMs) [6], deep neural networks (NNs) [7], random forests (RFs) [8], and gradient boosting machine (GBM) [9]. Some of these algorithms have been used to develop prediction models on small data sets ($<$ 200 events) [10–12]. Because ML algorithms are more prone to overfitting [13], it remains unclear what the impact on prognostication is of these novel techniques.

Although the incremental value of flexible ML methods has been previously assessed, these comparisons were potentially subject to bias [14]. The incremental value of ML algorithms is potentially overrated because studies up to this point mainly focused on the ability of the methods to discriminate between patients with good and poor outcomes [15–19]. Per-

formance of prediction models is however commonly measured across at least two dimensions: calibration and discrimination [20, 21]. Calibration refers to the agreement of predicted probabilities of a model and observed outcomes (e.g., "if the risk of death is x%, do x% of the patients with this prediction actually die?"). Poor calibration of prediction models may lead to harmful decision-making when applying these models [22–24].

One of the more thoroughly validated prediction models with good performance exists in the field of TBI: the IMPACT model [25]. This model comprises baseline clinical characteristics, presence of secondary insults, imaging findings, and laboratory characteristics. Using the variables of this model, the present study aims to fairly assess the potential incremental value of flexible ML methods beyond classical regression approaches.

11.3 | Methods

This study was reported to conform with the TRIPOD guidelines [23].

11.3.1 | Study population

We included 15 studies from the IMPACT-II database. These include four observational studies and eleven randomised controlled trials on moderate to severe TBI (Glasgow Coma Scale [GCS] $\leq$ 12), which were conducted between 1984 and 2004 [26]. Furthermore, we validated models in the patients with moderate to severe TBI (GCS $\leq$ 12) from the CENTER-TBI core study. This is a recent prospective study, which included patients from 2014 to 2018 [1]. Data for the CENTER-TBI study have been collected

through the Quesgen e-CRF[1], hosted on the INCF platform, and extracted via the INCF Neurobot tool (INCF, Sweden). Version 1.0 of the CENTER-TBI data was used for this analysis.

11.3.2 | Model specification

The outcomes which were predicted were 6 months mortality and unfavourable outcome (Glasgow Outcome Scale < 3, or Glasgow Outcome Scale–Extended < 5). The predictors included in the models were 11 predictors of the IMPACT laboratory model [25]. Continuous variables were included as continuous variables in the model (no categorisation). An overview of the included variables, and their specifications, is shown in table 11.1. The baseline GCS score was defined as the last GCS in the emergency department ("poststabilisation"). If this score was missing, the nearest GCS at an earlier moment was used. In total, eleven predictors were included, representing 19 parameters (or degrees of freedom [df]). In the case of mortality, 3,491 events (or 184 events per parameter) were on average present in our database for each training. The variables were normalised or one-hot encoded because this is standard practice for training algorithms which use gradient descent optimisation.

11.3.3 | Regression techniques

The regression techniques which were compared with the ML algorithms included standard logistic regression, but also penalised regression: lasso and ridge regressions [21]. These algorithms were developed to improve the performance of logistic regression models by shrinking the coefficients

[1]Quesgen Systems Inc, Burlingame, CA, USA

Table 11.1: *Model specification: 11 predictors, with 19 degrees of freedom*

Variable in the model	Characteristics
Age	Continuous
Motor GCS score	Categorical, 1–6
Pupils	Categorical, 3 levels: Both reactive, One reactive, Two reactive
CT class	Categorical, 5 levels: No visible pathology Diffuse injury Diffuse injury with swelling Diffuse injury with shift Mass
Traumatic subarachnoid hemorrhage	Binary
Epidural hematoma	Binary
Hypoxia	Binary
Hypotension	Binary
Glucose, first measured	Continuous
Sodium, first measured	Continuous
Hemoglobin, first measured	Continuous

Abbreviations: GCS: Glasgow Coma Scale;
CT: computed tomography.

during estimation [27, 28]. The objective is to obtain models that are less prone to making too extreme predictions (overfitting). The glmnet function from the glmnet package was used (alpha = 0 for ridge, and alpha = 1 for lasso). No nonlinear or interaction terms were included in the regression models.

11.3.4 | Machine learning algorithms

All analyses were performed using R[2]. The script can be found online.[3]

The flexible ML algorithms that were compared with logistic regression were SVM, NN, RF, and GBM. All these algorithms have so-called "hyperparameters," which need to be optimised for the algorithms to work optimally. To select the optimal hyperparameters, the framework of the caret package was used. The best combination of hyperparameters of the algorithms was chosen based on the highest log likelihood. The average log likelihood over 10 repetitions of tenfold cross-validation was used to select the optimal parameters (Fig. 11.1). For a detailed description of what algorithms were used and what hyperparameters were considered, see Appendix B available online.

The included flexible ML methods, just like regression, do not allow for missing values. Unlike regression, however, they are not readily compatible with multiple imputation: not every algorithm uses weights as core operators. Moreover, for the algorithms that use weights, there is no implementation of pooling these weights over multiple data sets using Rubin's rules [29]. Therefore, multiple imputation using the mice package was performed [30], but only one imputed data set was used to train the models. The outcome and all predictors were included in the imputation model. To check for stability of results, a sensitivity analysis was performed with a different imputed data set.

[2]R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

[3]`https://github.com/bgravesteijn/ML_baseline_pred_code.`

11.3.5 | Cross-validation

The models were validated using three different strategies. First, they were cross-validated per study: the algorithms were trained on all but one study, and calibration and discrimination were assessed by applying the models to the study not used at model development. This procedure has been referred to as "Internal-external cross-validation" [31, 32]. For an overview of the analytical steps of internal-external cross-validation, see figure 11.1. Second, internal validation was performed in the IMPACT-II database using 10 times 10-fold random cross-validation. For this method, the data were randomly divided by deciles. The model was developed on 9/10 and validated on 1/10 of the data. This process was repeated until all patients were used once as validation sample. Finally, a fully external validation was performed, with training of the models in the IMPACT-II database and validating in CENTER-TBI.

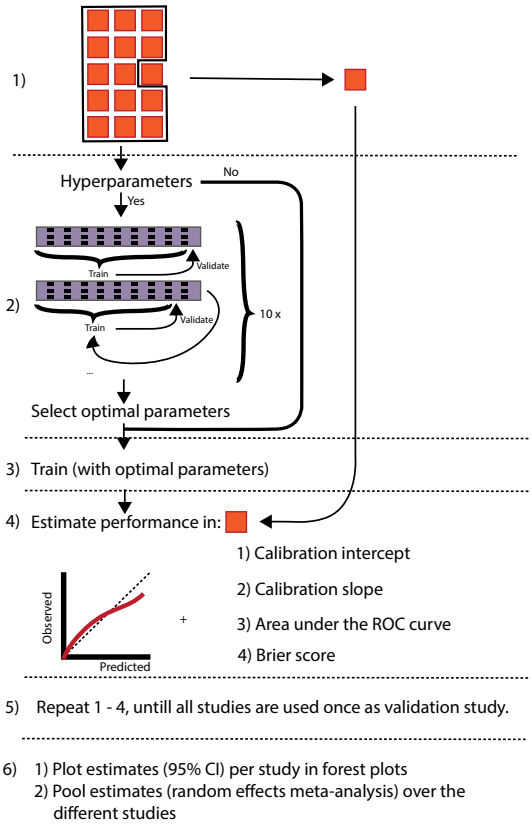


Figure 11.1: *Overview of the experimental setup. Step 1 is selecting a study as a validation study. Step 2 is selecting the optimal hyperparameters through 10 times 10-fold cross-validation. If the algorithm did not require hyperparameters, this step was skipped. Step 3 is the training of the final model with optimal hyperparameters on the full training data. The model of step 3 was validated in step 4 with the study that was left out of the training set. Step 5 is repeating step 1–4 until all studies are used once as validation study. Finally step 6 is the presentation of the results, and pooling the results over the different studies.*

The performance was assessed in three domains. First, calibration was examined graphically and quantified using a calibration slope and the calibration intercept: the calibration test proposed by Cox [33]. Second, discrimination was quantified using the c-statistic, also known as the area under the receiver operating curve. The confidence intervals of the c-statistic were obtained using the DeLong et al. method [34], using the ci.auc function from the pROC package. Third, as a measure of overall performance, the Brier score was calculated [35]. More extensive descriptions of these metrics can be found in Appendix B available online.

The estimates and 95% confidence intervals were plotted in forest plots, to visually inspect the variation. To obtain estimates per model and outcome, the estimates (and standard errors) in every validation were pooled using a random effects meta-analysis, using the DerSimonian and Laird estimator for τ^2 [36]. Because the CENTER-TBI database is a recent study, unlike the IMPACT-II studies, the estimates obtained from validating in this study were presented separately.

To compare whether observed variation of the performance measures can be attributed to differences in performance across study population or type of model used, we used mixed effects linear regression. This was performed in the internal-external validation framework. The performance measure was used as dependent variable, and two random intercepts were included in the model: one for what algorithm was used and one for what study the models were validated in. These random intercepts were assumed to follow a normal distribution with mean 0 and variance τ^2. The percentage variation in performance attributable to in which study the model was validated was calculated by dividing the τ^2 of study by the total variance (the sum of the variance of the random intercepts of study

and algorithm, and the residuals):$\tau^2_{study}/(\tau^2_{study}+\tau^2_{algorithm}+\tau^2_{residuals})$. Similarly, the percentage variation in performance attributable to what algorithm was trained was calculated.

11.4 | Results

11.4.1 | Patient characteristics

The baseline characteristics differed substantially between the IMPACT-II and the CENTER-TBI data. In the IMPACT-II database, patients were younger (35 versus 47.4 years), had less traumatic subarachnoid hemorrhages (4,016 [45%] versus 759 [74%]), and presented less often with a motor GCS of one (1,565 [16%] versus 615 [45%]). However, the patients showed similar Glasgow Outcome Scale in the two studies: In the IMPACT-II database, 3,332 (30%) died and 5,233 (48%) had an unfavourable outcome, and in the CENTER-TBI study, 348 (29%) died and 651 (54%) had unfavourable outcome (table 11.2). For an overview of the patient characteristics per study in IMPACT-II and CENTER-TBI, see Table 1 appendix A available online.

11.4.2 | Discrimination

At internal-external validation, the difference between maximum and minimum c-statistic of the algorithms was only 0.02 for mortality and unfavourable outcome. The discriminatory performance of the implementation of RF was suboptimal: the median and IQR of c-statistic of the RF were 0.79 (0.77–0.82) for mortality (the overall average was 0.81) and 0.79 (0.76–0.81) for unfavourable outcome (the overall average was 0.80). The

Table 11.2: *Baseline characteristics of the CENTER-TBI and IMPACT-II databases*

Characteristic	IMPACT-II	CENTER-TBI	Missing data, total %
N	11,022	1,375	
Age (median [IQR])	31 [22, 46]	48 [28, 65]	0.0
Hypoxia (%)	1,707 (22)	217 (16.8)	26.3
Hypotension (%)	1,518 (17.2)	205 (15.9)	18.3
Marshall CT class (%)			40.6
1	379 (5.9)	81 (8.3)	
2	2,281 (36)	428 (43.9)	
3	1,259 (20)	86 (8.8)	
4	248 (3.9)	19 (2.0)	
5	2,223 (35)	360 (37.0)	
Traumatic subarachnoid hemorrhage (%)	4,016 (44.6)	759 (73.6)	19.1
Epidural hematoma (%)	1,275 (13.4)	172 (16.7)	14.8
Glucose (median mmol/L (SD))	8.84 (3.46)	8.18 (2.95)	44.5
Hemoglobin (mean g/dL (SD))	12.46 (2.42)	7.96 (2.36)	52.2
GCS motor (%)			7.4
1	1,565 (15.5)	615 (44.7)	
2	1,285 (12.7)	77 (5.6)	
3	1,362 (13.5)	80 (5.8)	
4	2,438 (24.1)	136 (9.9)	
5	2,791 (27.6)	357 (26.0)	
6	658 (6.5)	110 (8.0)	
Pupil (%)			12.8
Both reactive	6,292 (66.3)	973 (73.7)	
One reactive	1,192 (12.6)	110 (8.3)	
None reactive	2,010 (21.2)	238 (18.0)	
Glasgow outcome scale (%)			1.4
2	3,322 (30.1)	348 (29.0)	
3	1,911 (17.3)	303 (25.2)	
4	2,262 (20.5)	246 (20.5)	
5	3,527 (32.0)	303 (25.2)	

Abbreviations: CT: computed tomography; GCS: Glasgow Coma Scale; SD: standard deviation; IQR: interquartile range.

discriminative performances varied substantially per study (figure 11.2A and table 11.3). At internal validation in IMPACT-II, a similar pattern was seen, but the c-statistics were somewhat higher. For example, the GBM showed a c-statistic of 0.81 (0.79–0.83) at internal-external validation and 0.83 (0.82–0.84) at internal validation. When performing external valida-

tion in CENTER-TBI, this pattern was also seen: the RF showed a median and 95% CI for the c-statistic of 0.81 (0.78–0.84) for mortality (overall average was 0.82) and 0.76 (0.74–0.79) for unfavourable outcome (overall average was 0.77). Similar results were observed over a different imputed set, see Table 5 appendix A available online.

Table 11.3: *Results for discriminative performance of all algorithms, in all three validation strategies: internal-external (per-study CV), internal (10-fold CV), and external (CENTER-TBI) validation*

Algorithm	Outcome	Internal-external	Internal	External
Logistic regression	Mortality	0.81 (0.79–0.84)	0.82 (0.81–0.83)	0.82 (0.79–0.84)
Support vector machine		0.81 (0.78–0.83)	0.82 (0.82–0.83)	0.81 (0.79–0.84)
Random forest		0.79 (0.77–0.82)	0.79 (0.78–0.81)	0.81 (0.78–0.84)
Neural network		0.81 (0.79–0.84)	0.82 (0.81–0.83)	0.82 (0.79–0.84)
Gradient boosting machine		0.81 (0.79–0.84)	0.83 (0.82–0.84)	0.83 (0.81–0.86)
Lasso regression		0.81 (0.79–0.84)	0.82 (0.82–0.83)	0.82 (0.79–0.84)
Ridge regression		0.81 (0.79–0.84)	0.82 (0.82–0.83)	0.82 (0.79–0.84)
Logistic regression	unfavourable outcome	0.81 (0.79–0.83)	0.82 (0.81–0.82)	0.77 (0.75–0.80)
Support vector machine		0.80 (0.79–0.82)	0.81 (0.81–0.82)	0.78 (0.75–0.80)
Random forest		0.79 (0.76–0.81)	0.79 (0.78–0.80)	0.76 (0.74–0.79)
neural network		0.80 (0.79–0.82)	0.81 (0.81–0.82)	0.78 (0.76–0.80)
Gradient boosting machine		0.80 (0.78–0.82)	0.81 (0.80–0.82)	0.78 (0.76–0.80)
Lasso regression		0.81 (0.79–0.83)	0.81 (0.80–0.82)	0.77 (0.75–0.80)
Ridge regression		0.81 (0.79–0.83)	0.81 (0.80–0.82)	0.77 (0.75–0.80)

Abbreviation: CV: cross-validation.

Figure 11.2: *Results of the internal-external cross-validation for mortality. Panel A shows the results of the c-statistic/area under the ROC curve, panel B shows the calibration intercept, panel C shows the calibration slope, and panel D shows the Brier score. The validation results are displayed per study (left: observational, right: randomised controlled trials), and per algorithm. Abbreviations: LR: logistic regression; SVM: support vector machine; RF: random forest; NN: neural network; GBM: gradient boosting machine; ROC: receiver operating curve.*

11.4.3 | Calibration

At internal-external validation, the average calibration intercepts across the algorithms did not vary substantially: the range of calibration intercepts was -0.08 to -0.02 for mortality, and for unfavourable outcome, the calibration intercepts were 0.02 (figure 11.2B and Table 2 appendix A available online). The range of calibration slopes was larger: 0.85–1.05 for mortality and 0.89-1.06 for unfavourable outcome (figure 11.2C and Table 3 appendix A available online). The RF made too extreme predictions, with a median (95% CI) calibration slope of 0.85 (0.77–0.93) for mortality, whereas the overall mean was 0.97, and 0.89 (0.82–0.96) for unfavourable outcome, whereas the overall mean was 0.99. At internal validation in IMPACT-II, calibration slopes and intercepts were similar. In external validation in CENTER-TBI, the RF had again a too low calibration slope (0.88, 95% CI: 0.77-0.99 for mortality).

The calibration intercept for mortality was generally low in CENTER-TBI: the overall mean was -0.58, indicating that the 6-month mortality was lower than expected in CENTER-TBI.

11.4.4 | Overall predictive ability

The Brier score was very similar at internal-external validation, internal, and external validation for both outcomes (Table 4 appendix A available online). The Brier score was somewhat higher at external validation but consistent for all methods (e.g., 0.19 versus 0.18 for logistic regression to predict unfavourable outcome).

11.4.5 | Explained heterogeneity

At internal-external validation, variation in c-statistic, calibration intercept, and the Brier score was mainly attributable to the study in which the algorithm was validated (table 11.4): for mortality, the variation in c-statistic was 97% attributable to the study in which the algorithm was validated (versus 2.0% to what algorithm was used), whereas the variation in calibration intercept was 98% attributable to the study in which the algorithm was validated (versus 0.3% to what algorithm was used); and variation in Brier score was 96% attributable to the study in which the algorithm was validated (versus 2.0% to what algorithm was used). Variation in calibration slope was slightly more attributable to what algorithm was used, compared with the other metrics (Fig. A1). For mortality, the variation in calibration slope was 11% attributable to the algorithm used and 86% attributable to the study in which the algorithm was validated. This was mostly caused by the low calibration slope of the RF algorithm. This algorithm displayed the worst calibration slope, as indicated in figure 11.2C. For unfavourable outcome, the results were similar.

Table 11.4: *Percentage of variation in performance attributable to what study the algorithms were validated in*

Outcome	C-statistic	Calibration intercept	Calibration slope	Brier score
Mortality				
Algorithm	2.0	0.3	11	2.0
Study	97	98	86	96
unfavourable				
Algorithm	2.9	0.0	12	2.5
Study	96	99	85	97

11.4.6 | Nonadditivity and nonlinearity

To explore whether nonadditive and nonlinear effects were frequently appropriate to assume in our data, we performed a post hoc analysis. Per study, logistic regression models allowing for nonadditivity and nonlinearity were tested with likelihood ratio tests (omnibus tests) to the model which did not allow for relaxation of those assumptions [20]. It was observed that the model predicting mortality had a better fit when nonlinearity was allowed for in 7 (44%) studies. Less often, the assumption of nonadditivity improved the model fit (Table 6 appendix A available online).

11.5 | Discussion

This study aimed to compare flexible ML algorithms to more traditional logistic regression in contemporary patient data. We trained the algorithms to obtain a model with both high discrimination and good calibration. This was achieved by optimising the log-likelihood for both regression and ML algorithms. All models and algorithms were developed and validated in large data sets, including the recent prospective cohort study CENTER-TBI [1]. Performance was assessed in terms of both discrimination and calibration, which are both important characteristics to be assessed in algorithm validation [22, 24, 37]. Similar performance of most methods was found across a large number of studies from different time periods.

The algorithm that relatively underperformed was the RF: the discrimination was somewhat lower, but it clearly underperformed in terms of calibration. In particular, the RF showed a calibration slope that was far below one. This indicates overfitting, a problem often arising in small data

sets [35]. According to theoretical arguments, the RF algorithm should not overfit [38]. The discrepancy between the theory and the empirical evidence of our study should be explored further. There could be a role for the selection of hyperparameters, in particular the number of random variables at the split, and the fraction of observations in the training sample [39]. Because the RF shows signs of overfitting, even in large data sets, the discriminative performance should be interpreted with caution: due to optimism, the discrimination in new data sets can be lower [21]. As a contrast, this method was one of the better performing methods in other studies [15, 40], which however did not assess calibration. Because calibration is a crucial step before implementation of a prediction model in clinical practice [20, 37, 41], our study encourages the use of other modelling techniques compared with RFs for outcome prediction.

The variation in observed performance was more explained by the cohorts where the algorithms were validated than by which algorithms were used. This implies that prediction models need continuous updating and validation because their performance is often worse in new cohorts [42]. This is a limitation which needs to be addressed, to effectively use these models in clinical practice [43]. This finding does raise concerns about the validity of individual patient data meta-analysis in the context of prediction modelling.

A recent systematic review compared flexible ML methods to traditional statistical techniques in relatively small data sets (median sample size was 1,250) and did not find incremental value [14]. This was perhaps to be expected because modern ML methods are known to be data hungry compared with classical statistical techniques [13, 44]. However, due to the increased sharing of data, international collaborations, and the availability of data from electronical health records and other data sets with routinely

collected data, data sets are becoming increasingly large [45–47]. Our study shows that in this situation, flexible ML methods are not improving outcome prognostication as well.

A limitation of our study is that we only used a linear kernel function of SVM. Other kernels could have increased the performance of the algorithm because the performance of the algorithm is substantially dependent on its hyperparameters [39]. Unfortunately, the computation time increased drastically when this kernel was implemented (the expected running time for one series of cross-validation was 21 days). Because the first six iterations did not show substantial increase in discriminative performance, we decided to use the linear basis function instead.

Second, we only considered a relatively small number of predictors (11 predictors, with 19 df). The reason for not including more predictors is that there were no other common data elements between all databases. This potentially limits the performance of ML techniques because it has been suggested that flexible ML techniques perform better than traditional regression techniques when a large number of predictors are being considered, that is, high-dimensional data [48, 49]. A reason for such presumed superiority is the flexibility of these algorithms, enabling them to capture complex nonlinear and interaction effects. It should be noted that regression-based techniques can also be extended by nonlinear and interaction effects [20]. Given that ML algorithms did not outperform regression, these effects are not likely to be essential in the field of outcome prediction in patients with TBI. Our study was not able to fully use the potential benefit of multidimensional data because of a phenomenon that is expected in big data research: larger volumes of data for better models may come at the price of less detailed or lower quality data.

We do believe that although we could perhaps not use the full poten-

tial performance of ML algorithms, our comparison is just as relevant. Published ML-based prediction algorithms often include a large number of predictors, sometimes with the suggestion to result in high discriminative performance [50, 51]. We note that external validation of these high-dimensional prediction algorithms is challenging because availability of predictors may differ from one setting to the other. For prediction with genomics data, this may be feasible if sufficient standardisation and harmonisation was performed [52]. However, clinical variables often have different definitions, notations, or units, which complicate the validation procedure with a large number (say n > 50) of predictors. External validation remains an essential step before implementing prediction algorithms in clinical practice. To train and validate high-dimensional data, a sophisticated IT environment is necessary [53]. Therefore, we believe that the low-dimensional setting, such as our study, might be more relevant for clinical practice, also for the near future. Powerful predictions for outcome after TBI can apparently be made with linear effects which are captured with simple algorithms.

Finally, this study should be replicated in other fields than TBI to ensure the generalisability of our findings, again from a largely neutral perspective [52]. Preferably, a wide range of studies should be used, representing different settings in terms of study design (randomised controlled trials versus observational), geography (different countries), types of centres (level I trauma centres versus other), and so forth. Most studies that compared algorithms used only one or a limited number of study populations [15–19]. Because the performance heavily relies on the study population, comparing the methods in multiple populations is recommended.

11.5.1 | Conclusion

In a low-dimensional setting, flexible ML algorithms do not perform better than more traditional regression models in outcome prediction after moderate or severe TBI. This is potentially explained by the most important prognostic effects acting as linear additive effects. Predictive performance is more dependent on the population in which the model is applied than the type of algorithm used. This finding has strong implications: continuous validation and updating of prediction models is necessary to ensure applicability to new populations of both ML algorithms and regression-based models. To improve prognostication for TBI, future studies should extend current prognostic models with new predictors (biomarkers, imaging, genomics) with strong incremental value, for the reliable identification of patients with poor versus good prognosis.

11.6 | Supplementary Material

Supplementary materials can be found in the online version of this article: `https://tinyurl.com/k77tcbyt`

References

1. Maas, A. I. R. *et al.* Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. *The Lancet Neurology* **4422.** ISSN: 14744422 (2017).
2. Majdan, M. *et al.* Epidemiology of traumatic brain injuries in Europe: a cross-sectional analysis. *The Lancet Public Health* **1,** e76–e83. ISSN: 2468-2667 (2016).

3. Saatman, K. E. *et al.* Classification of traumatic brain injury for targeted therapies. *Journal of neurotrauma* **25,** 719–738 (2008).

4. Lingsma, H. *Measuring quality of care : methods and applications to acute neurological diseases* ISBN: 9789077283110 (Erasmus University Rotterdam, 2010).

5. Liu, N. T. & Salinas, J. Machine Learning for Predicting Outcomes in Trauma. *SHOCK* **48,** 504–510. ISSN: 1073-2322 (2017).

6. Burges, C. J. C. *A Tutorial on Support Vector Machines for Pattern Recognition* tech. rep. (1998), 121–167.

7. Jain, A., Jianchang Mao & Mohiuddin, K. Artificial neural networks: a tutorial. *Computer* **29,** 31–44. ISSN: 00189162 (1996).

8. Afanador, N. L., Smolinska, A., Tran, T. N. & Blanchet, L. Unsupervised random forest: a tutorial with case studies. *Journal of Chemometrics* **30,** 232–241. ISSN: 08869383 (2016).

9. Natekin, A. & Knoll, A. Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics* **7,** 21. ISSN: 1662-5218 (2013).

10. Rau, C.-S. *et al.* Mortality prediction in patients with isolated moderate and severe traumatic brain injury using machine learning models. *PloS One* **13,** e0207192 (2018).

11. Matsuo, K. *et al.* Machine Learning to Predict In-Hospital Morbidity and Mortality after Traumatic Brain Injury. *Journal of Neurotrauma,* neu.2018.6276. ISSN: 0897-7151 (2019).

12. Feng, J.-z. *et al.* Comparison between logistic regression and machine learning algorithms on survival prediction of traumatic brain injuries. *Journal of Critical Care* **54,** 110–116. ISSN: 08839441 (2019).

13. Van der Ploeg, T., Austin, P. C. & Steyerberg, E. W. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. *BioMed Central: medical research methodology* **14,** 1–13 (2014).

14. Christodoulou Evangelia *et al.* A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of Clinical Epidemiology* **0.** ISSN: 08954356 (2019).

15. Van Os, H. J. A. *et al.* Predicting Outcome of Endovascular Treatment for Acute Ischemic Stroke: Potential Value of Machine Learning Algorithms. *Frontiers in Neurology* **9,** 784. ISSN: 1664-2295 (2018).

16. Churpek, M. M. *et al.* Multicenter Comparison of Machine Learning Methods and Conventional Regression for Predicting Clinical Deterioration on the Wards. *Critical care medicine* **44,** 368–74. ISSN: 1530-0293 (2016).

17. Lee, H.-C. *et al.* Prediction of Acute Kidney Injury after Liver Transplantation: Machine Learning Approaches vs. Logistic Regression Model. *Journal of Clinical Medicine* **7,** 428. ISSN: 2077-0383 (2018).

18. Bisaso, K. R., Karungi, S. A., Kiragga, A., Mukonzo, J. K. & Castelnuovo, B. A comparative study of logistic regression based machine learning techniques for prediction of early virological suppression in antiretroviral initiating HIV patients. *BioMed Central: Medical Informatics and Decision Making* **18,** 77. ISSN: 1472-6947 (2018).

19. Decruyenaere, A. *et al.* Prediction of delayed graft function after kidney transplantation: comparison between logistic regression and machine learning methods. *BioMed Central: Medical Informatics and Decision Making* **15,** 83. ISSN: 1472-6947 (2015).

20. Harrell, F. E. *Regression Modeling Strategies* ISBN: 978-1-4419-2918-1 (Springer New York, New York, NY, 2001).

21. Steyerberg, E. W. *Clinical Prediction Models* ISBN: 978-3-030-16398-3. `http://link.springer.com/10.1007/978-3-030-16399-0` (Springer International Publishing, Cham, 2019).

22. Van Calster, B. *et al.* A calibration hierarchy for risk models was defined: from utopia to empirical data. *Journal of Clinical Epidemiology* **74,** 167–176. ISSN: 0895-4356 (2016).

23. Collins, G. S., Reitsma, J. B., Altman, D. G. & Moons, K. G. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. *Annals of Internal Medicine* **162,** 55. ISSN: 0003-4819 (2015).

24. Moons, K. G. M. *et al.* Risk prediction models: II. External validation, model updating, and impact assessment. *Heart (British Cardiac Society)* **98,** 691–8. ISSN: 1468-201X (2012).

25. Steyerberg, E. W. *et al.* Predicting Outcome after Traumatic Brain Injury: Development and International Validation of Prognostic Scores Based on Admission Characteristics. *PLoS Medicine* **5** (ed Singer, M.) e165. ISSN: 1549-1676 (2008).

26. Marmarou, A. *et al.* IMPACT Database of Traumatic Brain Injury: Design And Description. *Journal of Neurotrauma* **24,** 239–250. ISSN: 0897-7151 (2007).

27. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* **58,** 267–288. ISSN: 00359246 (1996).

28. FIRTH, D. Bias reduction of maximum likelihood estimates. *Biometrika* **80,** 27–38. ISSN: 0006-3444 (1993).

29. Rubin, D. B. *Multiple imputation for nonresponse in surveys* (John Wiley & Sons, 2004).

30. Van Buuren, S. *Flexible imputation of missing data* 415. ISBN: 9781138588318 (CRC Press, 2018).

31. Royston, P., Parmar, M. K. B. & Sylvester, R. Construction and validation of a prognostic model across several studies, with an application in superficial bladder cancer. *Statistics in Medicine* **23,** 907–926. ISSN: 02776715 (2004).

32. Steyerberg, E. W. & Harrell, F. E. Prediction models need appropriate internal, internal-external, and external validation HHS Public Access. *Journal of Clinical Epidemiology* **69,** 245–247 (2016).

33. Cox, D. Two further applications of a model for binary regression. *Biometrika* **45,** 562–565 (1958).

34. DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* **44,** 837–45. ISSN: 0006-341X (1988).

35. Steyerberg, E. W. *et al.* Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology (Cambridge, Mass.)* **21,** 128–38. ISSN: 1531-5487 (2010).

36. DerSimonian, R. & Laird, N. Meta-analysis in clinical trials. *Controlled Clinical Trials* **7,** 177–188. ISSN: 0197-2456 (1986).

37. Steyerberg, E. W. & Vergouwe, Y. Towards better clinical prediction models: Seven steps for development and an ABCD for validation. *European Heart Journal* **35,** 1925–1931. ISSN: 15229645. arXiv: `arXiv:1011.1669v3` (2014).

38. Breiman, L. *Random Forests* tech. rep. (2001), 5–32.

39. Probst, P., Boulesteix, A.-L. & Bischl, B. *Tunability: Importance of Hyperparameters of Machine Learning Algorithms* tech. rep. (2018). arXiv: `1802.09596v3`.

40. Sakr, S. *et al.* Comparison of machine learning techniques to predict all-cause mortality using fitness data: the Henry ford exercIse testing (FIT) project. *BioMed Central: medical informatics and decision making* **17,** 174. ISSN: 1472-6947 (2017).

41. König, I. R. *et al.* Practical experiences on the necessity of external validation. *Statistics in medicine* **26,** 5499–511. ISSN: 0277-6715 (2007).

42. Thelin, E. P. *et al.* Evaluation of novel computerized tomography scoring systems in human traumatic brain injury: an observational, multicenter study. *PLoS medicine* **14,** e1002368 (2017).

43. Ngiam, K. Y. & Khor, W. Big data and machine learning algorithms for healthcare delivery. *The Lancet Oncology* **20,** e262–e273 (2019).

44. Van der Ploeg, T., Nieboer, D. & Steyerberg, E. W. Modern modeling techniques had limited external validity in predicting mortality from traumatic brain injury. *Journal of Clinical Epidemiology* **78,** 83–89. ISSN: 0895-4356 (2016).

45. Poldrack, R. A. & Gorgolewski, K. J. Making big data open: data sharing in neuroimaging. *Nature Neuroscience* **17,** 1510–1517. ISSN: 1097-6256 (2014).

46. Neurology, T. L. The changing landscape of traumatic brain injury research. *The Lancet. Neurology* **11,** 651. ISSN: 1474-4465 (2012).

47. Charles, D., Gabriel, M. & Searcy, T. *Adoption of Electronic Health Record Systems among U.S. Non-Federal Acute Care Hospitals: 2008-2014* tech. rep. 23 (2015).

48. Rajkomar, A. *et al.* Scalable and accurate deep learning with electronic health records. *NPJ Digital Medicine* **1,** 1–10 (2018).

49. Beam, A. L. & Kohane, I. S. Big Data and Machine Learning in Health Care. *Journal of the American Medical Association* **319,** 1317. ISSN: 0098-7484 (2018).

50. Delahanty, R. J., Kaufman, D. & Jones, S. S. Development and Evaluation of an Automated Machine Learning Algorithm for In-Hospital Mortality Risk Adjustment Among Critical Care Patients. *Critical Care Medicine.* ISSN: 0090-3493 (2018).

51. Desautels, T. *et al.* Prediction of early unplanned intensive care unit readmission in a UK tertiary care hospital: A cross-sectional machine learning approach. *British Medical Journal Open.* ISSN: 20446055 (2017).

52. Klau, S., Jurinovic, V., Hornung, R., Herold, T. & Boulesteix, A.-L. Priority-Lasso: a simple hierarchical approach to the prediction of clinical outcome using multi-omics data. *BioMed Central: Bioinformatics* **19,** 322. ISSN: 1471-2105 (2018).

53. Reps, J. M., Schuemie, M. J., Suchard, M. A., Ryan, P. B. & Rijnbeek, P. R. Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. *Journal of the American Medical Informatics Association* **25,** 969–975 (2018).

12

Blood biomarkers on admission in acute traumatic brain injury: Relations to severity, CT findings and care path in the CENTER-TBI study

Endre Czeiter*, Krisztina Amrein*, Benjamin Y Gravesteijn*, Fiona E Lecky,David K Menon, Stefania Mondello, Virginia FJ Newcombe, Sophie Richter, Ewout W Steyerberg, Thijs Vande Vyvere, Jan Verheyden, Haiyan Xu, Zhihui Yang, Andrew IR Maas, Kevin KW Wang, András Buki, and the CENTER-TBI collaborators **both authors contributed equally*

EBioMedicine

12.1 | Abstract

12.1.1 | Background

Serum biomarkers may inform and improve care in traumatic brain injury (TBI). We aimed to correlate serum biomarkers with clinical severity, care path and imaging abnormalities in TBI, and explore their incremental value over clinical characteristics in predicting computed tomographic (CT) abnormalities.

12.1.2 | Methods

We analyzed six serum biomarkers (S100B, NSE, GFAP, UCH-L1, NFL and t-tau) obtained <24 h post-injury from 2867 patients with any severity of TBI in the Collaborative European NeuroTrauma Effectiveness Research (CENTER-TBI) Core Study, a prospective, multicentre, cohort study. Univariable and multivariable logistic regression analyses were performed. Discrimination was assessed by the area under the receiver operating characteristic curve (AUC) with 95% confidence intervals.

12.1.3 | Findings

All biomarkers scaled with clinical severity and care path (ER only, ward admission, or ICU), and with presence of CT abnormalities. GFAP achieved the highest discrimination for predicting CT abnormalities (AUC 0.89 [95% CI: 0.87 - 0.90]), with a 99% likelihood of better discriminating CT-positive patients than clinical characteristics used in contemporary decision rules. In patients with mild TBI, GFAP also showed incremental diagnostic value: discrimination increased from 0.84 [95% CI: 0.83 - 0.86] to 0.89 [95% CI: 0.87

- 0.90] when GFAP was included. Results were consistent across strata, and injury severity. Combinations of biomarkers did not improve discrimination compared to GFAP alone.

12.1.4 | Interpretation

Currently available biomarkers reflect injury severity, and serum GFAP, measured within 24 h after injury, outperforms clinical characteristics in predicting CT abnormalities. Our results support the further development of serum GFAP assays towards implementation in clinical practice, for which robust clinical assay platforms are required.

12.2 | Research in context

12.2.1 | Evidence before this study

Blood-based biomarkers hold potential for informing and substantially improving the clinical management of patients with traumatic brain injury (TBI). To date, however, only one biomarker (S100B) has been integrated into some national guidelines for triaging the need to perform computerized tomography scanning (CT scan) of the brain in patients with mild TBI. The superiority or incremental benefit of biomarkers beyond canonical clinical variables used in CT prediction rules has not convincingly been demonstrated. Moreover, uncertainty exists as to how single or multi-biomarker tests perform. We conducted a living systematic review and meta-analysis to quantify the ability of blood biomarkers with advanced analytical and clinical validity (GFAP, UCH-L1, NSE, S100B, t-tau and NFL) to predict the presence of traumatic abnormalities on head CT scanning in the acute clinical setting. We screened MEDLINE, Embase, EBM Reviews and Cochrane Library for relevant articles on October 25, 2016, and, subsequently updated the results by monitoring the literature every 3 months. Synthesis of these data indicated that only S100B had high sensitivity and negative predictive value (NPV) and could be used to rule out the need for acute CT scanning. The evidence to support use of other emerging markers was limited and insufficient to warrant clinical application. Recently, the ALERT-TBI trial (1959 participants), published in 2018, showed that GFAP and UCH-L1, in combination, discriminated between patients with and without CT abnormalities. Results from the TRACK-TBI study in the US, highlighted the potential of biomarkers as a screening tool for MRI abnormalities in patients with normal CT findings. However, the clinical utility

of these biomarkers still remains uncertain. Further work is needed to determine the most effective and efficient biomarker or multimarker strategy for integration into clinical care.

12.2.2 | Added value of this study

We use data from a well-characterized multicentre cohort of 2867 patients with TBI to provide the first evidence of medical utility of blood biomarkers beyond standard of care-based clinical characteristics. Our results support potential for their adoption into clinical use to inform and improve decision making in current practice. In particular, we corroborate and extend recent results regarding the diagnostic performance of GFAP, showing that GFAP captures the greatest discriminatory information, performing as well as a combination of all markers and adding value to clinical characteristics. We believe this is the largest study to simultaneously assess and compare the diagnostic performance of a panel of 6 biomarkers reflecting distinct types of injury and pathophysiological mechanisms, across a population of patients with a full range of TBI severities and wide range of injury patterns.

The assessment of the utility of biomarker measurements in individual care pathways, as defined in CENTER-TBI, allows us to explore their use in a range of contexts of care. This ensures the robustness and generalizability of our estimates derived from real-world patient population and clinical scenarios, while confirming the generation of actionable information for clinicians. Comparison of our results with those from other studies shows similar trends, but also highlights between-platform inconsistencies in assay calibration and reported biomarker values.

12.2.3 | Implications of all the available evidence

This study provides the most exhaustive and comparable assessment to date of the six best-validated TBI biomarkers, demonstrating their potential utility in refining diagnosis, triage, injury characterization, and clinical care in TBI, beyond currently established clinical variables. We highlight the potential role of GFAP as part of a comprehensive triage strategy and consider it to be best positioned for implementation into medical practice and incorporation in clinical decision algorithms. Robust clinical use assay platforms are a prerequisite for such clinical implementation.

12.3 | Introduction

The delivery of precision medicine for traumatic brain injury (TBI) requires objective tools to identify disease phenotypes and to guide clinical decisions [1]. Clinical assessment and computerized tomography (CT) of the head form the diagnostic cornerstone in clinical practice, but a need remains for more detailed disease classification utilizing a multidimensional approach. Moreover, indiscriminate use of CT, resulting in high costs, and increased recognition of risks of radiation exposure have called for more selective use of CT scanning in patients with milder forms of TBI [1, 2]. Various clinical decision rules (CDR) have been developed for this purpose [3–6], but their adoption in clinical practice is variable.

A major focus of recent research has been on the potential of biomarkers to improve diagnosis and patient characterization, and enable tailored management [7]. Several publications have provided extensive evidence of analytical and early clinical validity of various biomarkers, and documented efforts to achieve regulatory clearance. However, the development of clinical algorithms and guidelines which integrate biomarker measurements to inform decision-making has been inconsistent, partial and inconclusive [8]. S100 calcium-binding protein B (S100B), a biomarker of astroglial breakdown, has been implemented in the Scandinavian TBI Guidelines [9], but is seldom used outside the Nordic countries and with suboptimal performance in real-world conditions [10]. The pivotal ALERT-TBI study showed high sensitivity for the combination of astroglial (glial fibrillary acidic protein [GFAP]) and neuronal (ubiquitin C-terminal hydrolase L1 [UCH-L1]) biomarker blood levels measured within 12 h after injury in triaging the need for CT scanning [11], but did not address the added value compared to clinical characteristics used in CDRs, or explore

its value relative to S100B [12]. More in general, few studies have examined the incremental value of biomarkers beyond clinical characteristics.

As a consequence, uncertainty exists how biomarkers, either singly or in combination, can best improve existing decision-making and processes of care, creating a barrier to widespread implementation and adoption of these tests in medical practice. Nevertheless, blood-based biomarkers provide objective information, offer additional risk stratification and hold potential to inform personalised interventions.

The CENTER-TBI Core study (Collaborative European NeuroTrauma Effectiveness Research: www.center-tbi.eu) was designed to advance multimodal characterization and classification in TBI [13, 14]. Within this unique framework, in which patients were stratified by care path, we aimed to determine the relation – and their relative performance – of a panel of biomarkers, assessed within 24 h of injury, with clinical severity, care pathways and presence of CT abnormalities across the entire injury spectrum of TBI. We further aimed to explore the incremental value of biomarkers compared to established clinical characteristics in predicting the presence of CT abnormalities.

12.4 | Materials and methods

12.4.1 | Study design and participants

The CENTER-TBI Core study is a prospective observational clinical and biomarker study of patients with TBI, conducted in 65 clinical sites from 17 European countries and Israel between December 19, 2014, and December 17, 2017. The study was registered with ClinicalTrials.gov (NCT02210221). Details of protocol and clinical data have been previously published [15,16].

In brief: patients with all severities of TBI presenting to a study centre within 24 h of injury and scheduled for CT scanning were enrolled, stratified by care path (emergency department [ER], admission [Adm] and intensive care unit [ICU]). The only exclusion criterion was severe pre-existing neurological disorder.

The study was conducted in accordance with all relevant laws of the EU if directly applicable or of direct effect and all relevant laws of the country where the Recruiting sites were located, including but not limited to, the relevant privacy and data protection laws and regulations (the "Privacy Law"), the relevant laws and regulations on the use of human materials, and all relevant guidance relating to clinical studies from time to time in force including, but not limited to the ICH Harmonized Tripartite Guideline for Good Clinical Practice ("ICH GCP") and the World Medical Association Declaration of Helsinki entitled "Ethical Principles for Medical Research Involving Human Subjects". Informed Consent by the patients and/or the legal representative/next of kin was obtained for all patients and documented in the e-CRF. The use of the biological samples was in accordance with the terms of the informed consent. The list of sites, Ethical Committees, approval numbers and approval dates can be found online.[1]

The study is reported in accordance with the STROBE recommendations (see Supplementary material available online).

In this analysis, which was pre-specified in the Description of Work for CENTER-TBI, we focused on a cohort of patients in whom

1. blood sampling within 24 h of injury, and
2. an early CT scan were available (figure 12.1).

[1] `https://www.center-tbi.eu/project/ethical-approval`

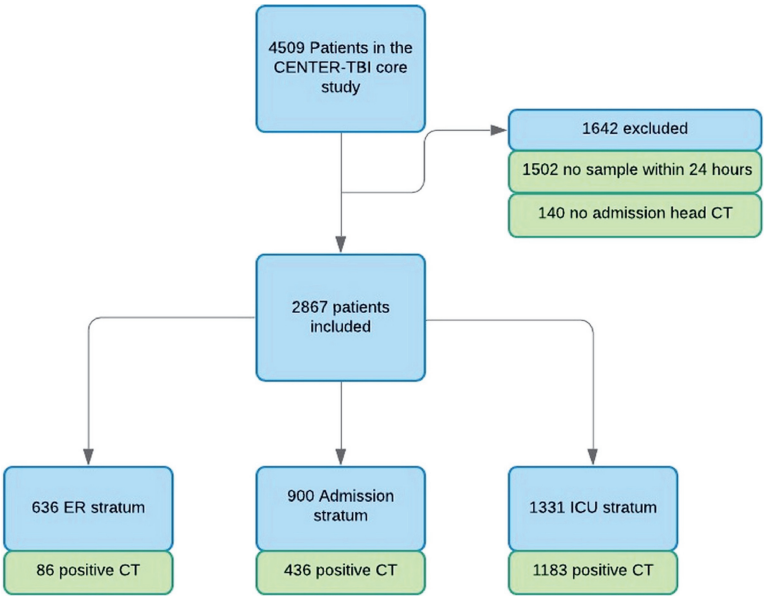


Figure 12.1: *Flow chart for biomarker cohort in the CENTER TBI core study.*

12.4.2 | Procedures

Blood samples for determination of biomarkers were collected using gel-separator tubes for serum, and centrifuged within 60 minutes. Serum was processed, aliquoted (8 × 0.5 ml), and stored at -80°C locally until shipment on dry ice to the central CENTER-TBI biobank (Pécs, Hungary).

We assayed six biomarkers: S100B, neuron-specific enolase (NSE), GFAP, UCH-L1, neurofilament protein-light (NFL), and total tau (t-tau). S100B and NSE were measured with a clinical-use automated system, using an electrochemiluminescence immunoassay kit (ECLIA) (Elecsys S100 and Elecsys NSE assays) run on the e 602 module of Cobas 8000 modular analyzer (Roche Diagnostics, Mannheim, Germany) at the University of Pécs

(Pécs, Hungary). Serum GFAP, UCH-L1, NFL and t-tau were analyzed with an ultrasensitive immunoassay using digital array technology (Single Molecule Arrays, SiMoA)-based Human Neurology 4-Plex B assay (N4PB) run on the SR-X benchtop assay platform (Quanterix Corp., Lexington, MA) at the University of Florida (Gainesville, Florida). Unique aliquots were used for analyses on the two platforms to avoid repeated freeze-thaw cycles, and analyzed in one round of experiments using the same batch of reagents by qualified laboratory technicians blinded to clinical information. Replicate assays were performed on a subset of samples with a balanced distribution across strata. The percent of replicates performed were 4.3% for the SiMoA platform and 5.7% for the Roche platform – these numbers were selected to fit within the assaying logistics and assay work flow for the full assay runs on the respective platforms.

Clinical data, including variables used in clinical decision rules (CDRs) for triaging CT scanning (Supplementary Table 1), were collected using a web-based electronic case report form (eCRF), with variables coded in accordance with the Common Data Elements (CDE) scheme[2].

All patients underwent head CT examinations according to local protocols. Imaging studies were transmitted to a central repository (Icometrix, Leuven) for structured reporting according to the NINDS TBI-CDEs. Central reviewers were blinded to clinical information, except gender, age, and care path. The presence of any traumatic intracranial abnormality on CT was considered a positive scan. Skull fractures in isolation were not considered as intracranial abnormality.

MR scans performed according to study protocol were obtained in a subgroup of 502 patients. We report on 152 of these patients who had a

[2] `https://commondataelements.ninds.nih.gov/`

negative CT scan on presentation.

12.4.3 | Statistical analysis

Baseline characteristics were summarized using standard descriptive statistics. Continuous variables are presented as median (interquartile range) and categorical variables as absolute frequencies and percentages. Bland–Altman plots were made for replicate assays. The coefficient of variation was calculated on the log transformed values and expressed as percentage.

Relations of biomarkers to clinical severity (GCS) and care path were displayed in tabular and graphical formats. Correlations between biomarkers were visualized by scatterplots and quantified using Spearman's rank correlation coefficients. Distributions of biomarker levels for patients with intracranial abnormalities were compared to those without abnormalities with Mann–Whitney U tests. We explored adjusting for multiple testing (Bonferroni, false discovery rate), but only reported these results when they changed the interpretation of the results.

Multiple imputation of missing characteristics was performed using the mice package [15], assuming a missing at random mechanism. Clinical characteristics with $> 50\%$ completion, CT positivity and biomarkers were included in the imputation model. Predictive mean matching was used for continuous data, logistic regression for binary data, and polytomous regression for categorical data. Fifty imputed datasets were created, with results summarized according to Rubin's rules [16]. The diagnostic performance of biomarkers, separately and in combination to identify patients with positive CT findings was assessed with logistic regression. We allowed for non-linear effects of log transformed biomarkers using restricted cubic splines with 3 degrees of freedom. From these models, we

derived estimates of the area under the ROC curves (AUC, or c-statistic). A bootstrap resampling procedure with 200 repetitions was applied to calculate 95% confidence intervals (CIs). Univariable and multivariable analysis, adjusting for clinical characteristics, was performed.

The clinical characteristics derived from CDRs were included in the multivariable analysis as they are presented in Supplementary Table 2 available online. Continuous variables were included without categorization to fully capture the diagnostic information they contain. In case variables were non-informative (GCS in the GCS 15 sensitivity analysis, or depressed skull fracture in the ER stratum), they were excluded from the model.

The performance of biomarkers compared to clinical characteristics in the univariable analysis was explored by a bootstrapping procedure that included 1000 repetitions. The percentage of repetitions where a univariable model with the biomarker outperformed (higher c-statistic) a multivariable model with all clinical characteristics used in current CDRs was calculated.

Predictions based on clinical characteristics were compared to predictions with biomarkers added to the clinical variables and results visualized using reclassification plots [17].

Sensitivity analyses were carried out on patients with GCS 13 - 14, in those with GCS 15, with major extracranial injuries (defined as AISleq3), and according to stratum (ER, admission, and ICU). In addition, AUCs were generated for samples collected at different times post-injury (in 6-h intervals) to explore possible influence of sampling time. Statistical analysis was performed using R[3] in RStudio[4].

[3] `http://www.r-project.org`, version 3.5.1
[4] `http://www.rstudio.com`, version 1.1.456

12.5 | Results

12.5.1 | Patient cohort and sampling

Data on 2867/4509 (64%) patients analyzed in the CENTER-TBI Core study were available for analysis of biomarkers in serum samples obtained within 24 h of injury (figure 12.1). The time between injury and sampling was shortest in the ER stratum (5.1 h; IQR [3.4 - 9.73]), in contrast to the admission (15.7 h; IQR [9.8 - 20.2]) and ICU (14.3 h; IQR [7.5 - 19.7]) strata (Supplementary Fig. 1 available online). The median needle to freezer times was 1.08 h (IQR [0.92 - 1.33]), with no substantial differences across strata. Agreement between replicates of biomarker assessments was good for the clinical platform assays of S100B and NSE (CV: 7% for both on a log transformed scale), but poorer for the research-use only (RUO) assays of GFAP, UCH-L1, NFL and t-tau (CV: 22-30%) (Supplementary Fig. 2 available online).

Clinical characteristics of the study cohort, differentiated by stratum, are summarized in table 12.1 and the frequency of specific characteristics, contained in CDRs for predicting CT abnormalities in patients with mild TBI, presented in Supplementary Table 2 available online. Characteristics of patients excluded (n=1642; see Fig. 12.1) were largely similar to those analyzed (n=2867), although the median GCS was lower (14 versus 15) and the percentage of non-reacting pupils higher (9.0% versus 5.3%) (Supplementary Table 3 available online).

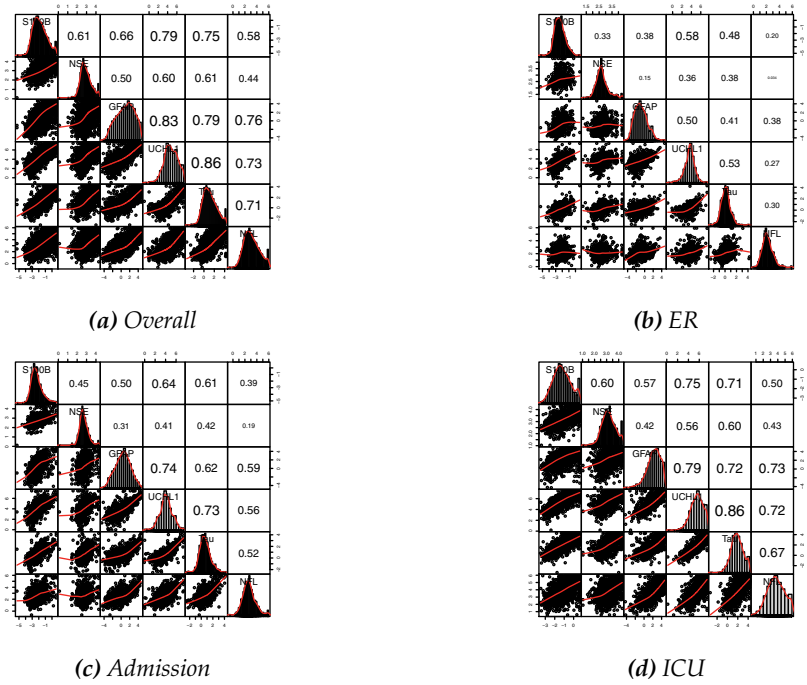
12.5.2 | Biomarker values by stratum and clinical severity

The median values of the six biomarkers displayed a clear association with injury severity (classified according to the GCS) and care path (Table 12.1, and Supplementary Table 4 available online). Within the group of mild TBI (GCS 13 - 15), median values were higher in patients with a GCS of 13 - 14 compared to 15.

All biomarkers showed correlations across all strata, except for NSE in the ER stratum with NFL and GFAP (figure 12.2). Correlations were strongest in the ICU stratum, likely reflecting greater differences in case-mix. The strongest correlation was found between UCH-L1 (neuronal marker) and t-tau (axonal marker) varying from 0.53 (ER) to 0.86 (ICU). The correlation between GFAP and S100B, both glial markers, was relatively weak, varying between strata from 0.38 (ER) to 0.57 (ICU).

12.5.3 | Biomarkers and traumatic intracranial CT abnormalities

Biomarker levels were higher in patients with traumatic abnormalities on CT scanning compared to those without (figure 12.3 and Supplementary Table 5 available online). Differences in biomarker levels between CT+ and CT- patients were greater when analyzed by clinical severity (GCS) than by care path. Differences were most pronounced for GFAP.



(a) *Overall* ***(b)*** *ER*

(c) *Admission* ***(d)*** *ICU*

Figure 12.2: *Correlation plots displaying associations between biomarkers in each stratum. The diagonal part with the name of the biomarker contains the distribution plot specific for the log-transformed biomarker. Scatter plots of correlations between biomarkers are presented below the diagonal, and Spearman correlation coefficients above the diagonal. The font size is indicative of the strength of correlation.*

Table 12.1: *Characteristics of biomarker cohort (n=2867) in the CENTER-TBI core study.*

N complete	Overall (n=2867)	ER (n=636)	Admission (n=900)	ICU (n=1331)	
Demographic characteristics					
Age (median [IQR])	2867	49 [30, 66]	48 [30, 65]	53 [33, 68]	48 [30, 64]
>65 years (%)		735 (25.6)	153 (24.1)	266 (29.6)	316 (23.7)
Male sex (%)	2867	1948 (67.9)	354 (55.7)	619 (68.8)	975 (73.3)
Cause of injury	2711				
Road traffic incident (%)		1098 (38.3)	204 (32.1)	295 (32.8)	599 (45.0)
Incidental fall (%)		1264 (44.1)	317 (49.8)	436 (48.4)	511 (38.4)
Clinical presentation					
GCS baseline (median [IQR])	2775	15 [10, 15]	15 [15, 15]	15 [14, 15]	10 [4, 14]
Severe (3 - 8) (%)		601 (21.0)	1 (0.2)	7 (0.8)	593 (44.6)
Moderate (9 - 12) (%)		222 (7.7)	2 (0.3)	28 (3.1)	192 (14.4)
Mild (13 - 14) (%)		457 (15.9)	38 (6.0)	202 (22.4)	217 (16.3)
Mild (15) (%)		1494 (52.1)	589 (92.6)	643 (71.4)	262 (19.7)
Pupillary reactivity	2732				
One pupil unreactive (%)		97 (3.4)	2 (0.3)	14 (1.6)	81 (6.1)
Two pupils unreactive (%)		144 (5.0)	7 (1.1)	4 (0.4)	133 (10.0)
Hypoxia (prehospital/ER) (%)	2709	184 (6.4)	1 (0.2)	15 (1.7)	168 (12.6)
Hypotension (prehospital/ER) (%)	2735	177 (6.2)	3 (0.5)	12 (1.3)	162 (12.2)
Any major extracranial injury (AIS >=3) (%)	2867	1032 (36.0)	22 (3.5)	262 (29.1)	748 (56.2)
CT characteristics					
Any intracranial abnormality at central reading (%)	2867	1705 (59.5)	86 (13.5)	436 (48.4)	1183 (88.9)
Biomarker (median [IQR])					
S100B, µg/L	2861	0.15 [0.08, 0.33]	0.09 [0.05, 0.15]	0.09 [0.06, 0.16]	0.28 [0.16, 0.58]
NSE, ng/ml	2858	17.08 [12.49, 25.85]	14.02 [11.14, 18.09]	14.22 [11.18, 19.39]	23.14 [16.24, 34.10]
GFAP, ng/ml	2850	3.14 [0.53, 15.07]	0.30 [0.11, 0.94]	1.51 [0.39, 5.28]	12.92 [4.22, 34.92]
UCH-L1, pg/ml	2846	94.74 [35.54, 307.12]	35.48 [17.63, 63.34]	51.44 [24.92, 109.66]	274.35 [119.23, 622.66]
Tau, pg/ml	2851	2.79 [1.23, 7.67]	1.16 [0.71, 1.84]	1.81 [1.05, 3.45]	7.08 [3.21, 16.61]
NFL, pg/ml	2849	18.55 [8.40, 49.74]	7.90 [5.09, 13.29]	12.88 [7.15, 24.55]	42.88 [19.49, 104.59]

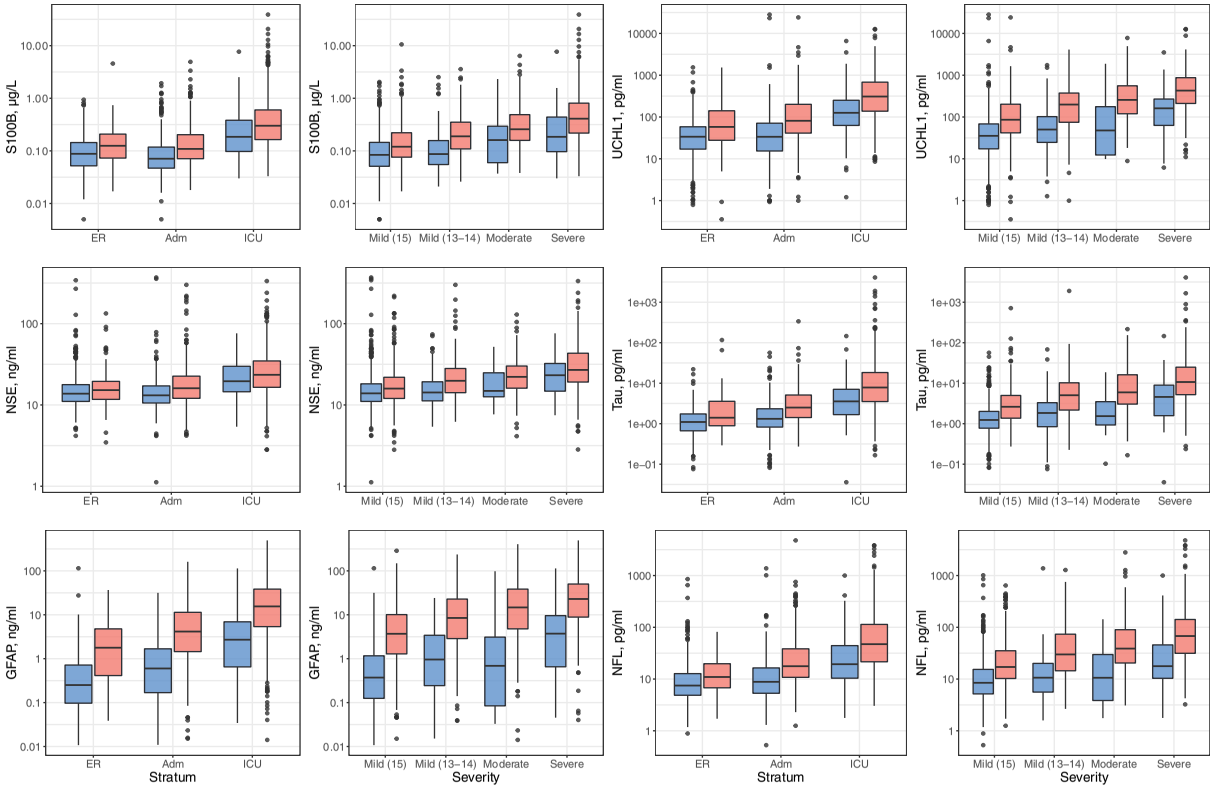


Figure 12.3: *Biomarker values by stratum and by clinical severity, differentiated for the absence (blue) or presence (pink) of traumatic intracranial CT abnormalities. P-values of the Mann–Whitney U tests are presented in the Supplemental material available online, Table 11.*

Univariable analysis confirmed that GFAP had the highest discrimination for predicting the presence of CT abnormalities (AUC 0.89 [95% CI: 0.87 - 0.90]), and performed as well, and even better in the admission stratum, as clinical characteristics (Supplementary Fig. 3 and Supplementary Table 6 available online). Most other biomarkers showed substantial discrimination, but performed poorer than clinical characteristics. This finding was confirmed in the bootstrap analysis: the chance that GFAP outperformed clinical characteristics was >99% in all subgroups, except for the ER stratum (figure 12.4). In the admission stratum, UCH-L1, NFL and t-tau outperformed clinical characteristics in less than 95% of the bootstrap samples. Combining all biomarkers showed slightly higher discrimination compared to GFAP alone. We found no clear benefit of any other combination of biomarkers, including the combination of GFAP and UCH-L1 (Supplementary Fig. 4 and Supplementary Table 7 available online).

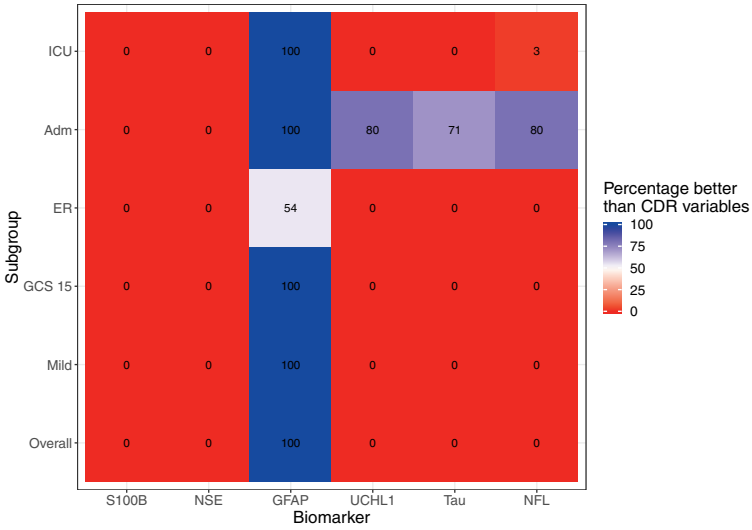


Figure 12.4: *Heat map demonstrating the discriminative ability of single biomarkers in comparison to a regression model that includes clinical characteristics contained in CT decision rules. The heat map summarizes the percentage of bootstrap replicates in which the model with the biomarker outperforms (higher c-statistic) the model with CT decision rule variables. The lower number of positive replicates for GFAP in the ER stratum may be due to lower number of events in this stratum (86/636 CT positive).*

Multivariable analysis adjusted for clinical characteristics incorporated in CT rules confirmed that GFAP provided incremental discriminative ability over clinical characteristics (increase from 0.89 [95% CI: 0.88 - 0.90] to 0.92 [95% CI: 0.91 - 0.93] when GFAP was included; figure 12.5 and Supplementary Table 8 available online). The incremental value was most pronounced in the admission stratum (increase in AUC from 0.72 [95% CI: 0.69 - 0.75] to 0.84 [95% CI: 0.81 - 0.86]), and was consistent in patients with a GCS of 15 (increase in AUC from 0.83 [95% CI: 0.80 - 0.85] to 0.88 [95% CI: 0.86 - 0.89]), and in those with a GCS of 13 - 14 (increase in AUC from 0.84 [95% CI: 0.80 - 0.88] to 0.90 [95% CI: 0.86 - 0.92]). Combinations of biomarkers showed no clear increase in discrimination compared to GFAP alone on multivariable analysis (Supplementary Fig. 5 and Supplementary Table 9 available online).

Figure 12.5: *Incremental discriminative ability of biomarkers to predict CT positivity. Plots show the difference in Area under the ROC curve (AUC) with 95% confidence intervals (bars) of logistic regression models combining age, time interval (injury to needle time) as well as clinical parameters included in current CT rules with and without biomarkers. Panels are presented for the overall sample (n=2867), according to stratum (ER, Admission, ICU), and for the mild (GCS 13 - 15; GCS 13 - 14; GCS 15) groups. Six biomarkers are considered separately and in combination ("all"). The dotted line indicates the predictive value of the clinical parameters and serves as a reference. The absolute values are presented in the Supplementary material available online, Table S7.*

While profound differences in the performance of biomarkers were not observed across different time intervals within 24 h of injury, GFAP levels displayed the greatest incremental predictive power when measured 12 - 18 h post-injury. (Supplementary Fig. 6 available online).

The incremental value of biomarkers was similar across patients with major extracranial injury and those without major extracranial injury (Supplementary Fig. 7 available online).

Reclassification plots confirmed the incremental value of GFAP across all strata compared to clinical characteristics (figure 12.6). Adding GFAP correctly reclassified 71% of patients with a negative CT scan, and 70% of the patients with a positive CT scan. The plots show that when added to the decision rule, GFAP levels often provide higher risk estimates to patients who had lower predicted risks by clinical characteristics. The same pattern, and similar extent of correct reclassification was seen when all biomarkers were combined (Supplementary Fig. 8 available online).

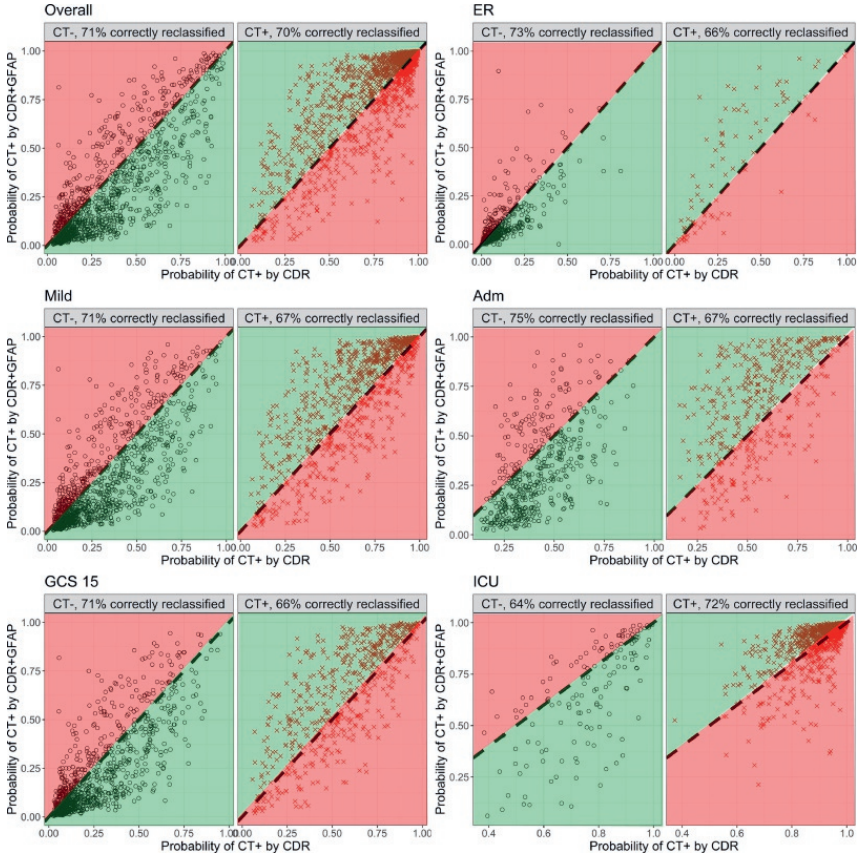


Figure 12.6: *Reclassification plots illustrating superior performance of GFAP compared to clinical characteristics in predicting traumatic intracranial CT abnormalities. GFAP is added to a logistic regression model that includes clinical parameters from current CT rules. The model with GFAP (y-axis), is compared to the model without GFAP (x-axis). Panels are presented for the overall sample (n=2867), for the strata (ER, Admission, ICU), and for the mild (GCS 13 - 15), and GCS 15 subgroups. Black dots represent patients with a negative CT scan, red crosses represent patients with a positive CT scan. The percentage of correct reclassification, indicated with green (higher probability of CT positivity if CT is positive, and lower probability of CT positivity if CT is negative), is displayed in the top of each plot.*

We further explored the discriminative value of biomarkers for predicting abnormalities on MR imaging performed within 3 weeks of injury in a subgroup of 152 patients who had a negative CT scan on presentation and underwent subsequent MR imaging. Traumatic MR abnormalities were found in 44/152 cases. The estimates were uncertain, but GFAP showed the highest discriminative ability (c-statistic: 0.76, 95% CI: 0.67 - 0.85, Supplementary Fig. 9 and Supplementary Table 10 available online).

12.6 | Discussion

We studied the relation of serum biomarkers to clinical severity, care path and the presence of traumatic intracranial abnormalities on CT scanning. We found that all biomarkers studied scaled with injury severity, classified according to the GCS and with care path intensity. Within each of these stratifications, biomarkers were higher in patients with abnormalities on head CT imaging compared to those without. Serum GFAP levels in the first 24 h post-injury were highly predictive for CT positivity, outperforming other markers and adding value to clinical variables considered in contemporary CT decision rules. Combining results of all biomarkers did not clearly improve discrimination compared to GFAP alone.

To the best of our knowledge, this is the first large scale study across all injury severities, evaluating systematically a panel of 6 biomarkers, and quantifying their performance relative to clinical characteristics in predicting CT abnormalities in patients with mild TBI. Recently, Thelin et al. reported on the same panel of biomarkers in a cohort of 172 patients with TBI [18]. The focus of this study was, however, on prognosis. Other previous studies have shown relations of biomarkers to clinical severity and

to the presence of CT abnormalities [19–22]. Most, however, mainly focused on one or two biomarkers and have seldom addressed performance of biomarkers relative to clinical characteristics. Moreover, a living systematic review identified serious problems in the design, analysis and reporting of many of the studies [7]. Until recently, the strongest evidence for a role in triaging CT scanning existed for S100B. This marker is incorporated in the Scandinavian guidelines for the management of moderate to minimal TBI [9]. Previously, a study including 397 patients with general trauma, of whom 209 had mild TBI, reported that GFAP performed better than S100B, in particular in patients with extracranial injuries [19]. Our results expand these earlier data, convincingly showing that GFAP outperforms S100B both in the overall group of mild TBI and in the subset of patients with a GCS of 15. However, we believe that a decision to replace S100B by GFAP in such guidelines may be premature. Assays for S100B are commercially available with high reproducibility. Currently, no GFAP assays are available as a commercialized clinical assay platform. Indeed, the platform we used for analysis of GFAP is a research-use-only (RUO) platform, and our replication assays showed substantial variation (28% on a log transformed scale). Factors that may have contributed to poor reproducibility include: not fully automated analyses, and that replicate samples had undergone a second freeze-thaw cycle. There is no evidence that refrigerated storage of samples for up to 72 h has a significant effect on GFAP values [23], but a decrease of GFAP levels has been reported in CSF after two freeze-thaw cycles [24].

The clear effects of GFAP and other biomarkers in the presence of assay heterogeneity speaks to the robustness of our findings. The role of biomarkers goes beyond triage for CT scanning in patients with mild TBI. Our results in this specific subgroup confirm the potential of GFAP to pre-

dict presence of MRI abnormalities in patients with normal CT findings after TBI [2].

12.6.1 | Assay reproducibility and thresholds

Reliability and reproducibility of biomarker assays are fundamental for clinical implementation. Absolute GFAP levels in our study were much higher than those reported in the ALERT-TBI study, which used a different platform [11]. Whilst different reference values between platforms may be acceptable, insight into comparability of values obtained with different platforms is desirable. Variation between platforms also precludes the concept of determining a universal cut-off value. Moreover, cut-off values are generally derived from reference values, obtained from healthy controls, whilst "action thresholds" are needed in diseased patients, which may be very different from reference values [25]. Further, it should be recognized that any biomarker represents a continuous variable and that use of a threshold value leads to loss of information. We suggest that biomarkers may be combined with clinical characteristics for risk estimation, and then continuous values may be retained. We have hence refrained from suggesting threshold values, a decision which was reinforced by the variability in replicate assays.

12.6.2 | GFAP versus a multi-marker approach

Contrary to our expectations, a multi-marker approach applying combinations of biomarkers did not increase the diagnostic value for CT positivity, compared to GFAP alone. GFAP showed similar discrimination as all biomarkers combined when analyzed versus clinical severity and care

path. Nevertheless, these observations do not preclude potential usefulness of combinational approaches in terms of outcome prediction and/or tracking the disease process over time. Conceptually, different biomarkers should differentially reflect specific aspects of the disease process of TBI [19, 26]. However, in this study we did not find any benefit of combining acute GFAP levels with UCH-L1 levels, a combination of (presumed) glial and neuronal markers used in the recent ALERT-TBI trial, which provided the basis for a recent FDA marketing authorization [11]. Our results provide no support for implementation of this combined assay into clinical practice. We do note, however, that a key difference between our study and ALERT-TBI is that the time window for blood sampling was 24 h in our study and 12 h in ALERT-TBI. This may be relevant as the half-life for UCH-L1 is short [27]. However, sensitivity analysis differentiated for time of sampling (Supplementary Fig. 6 available online) did not show superior performance of UCH-L1 in the first six hours after trauma. The finding that a single marker approach (GFAP) may be sufficient in the acute phase to inform diagnosis and care path is of particular relevance for low and middle income countries (LMICs) and other austere environments, where even basic imaging is inaccessible, or too expensive and unevenly distributed, with limited opportunities for patient transfer.

12.6.3 | Strengths and limitations

Strengths of our study include the large number of patients, analyses across all severities of TBI, the use of a comprehensive panel of biomarkers that addresses current clinical interest and the focus on the incremental value of biomarkers in predicting CT positivity compared to clinical characteristics used in CDRs. These strengths support the generalizability of our

findings, obtained in the "real-world" situation of an observational study.

Several limitations of our study should be acknowledged:

1. Our study should be considered as an exploratory diagnostic accuracy study [25], and was not designed to seek regulatory approval.

2. We were able to analyze samples from 2867/4509 (64%) patients available in the CENTER-TBI database. Although baseline characteristics of the study cohort were very similar to those reported for the Core study (n=4509) [13], we here included slightly less severe patients.

3. We utilized a research-use only (RUO) platform for assays of four biomarkers (GFAP, UCH-L1, NFL and t-tau), and coefficients of variation in replicate samples were relatively high.

4. The inclusion criteria for CENTER-TBI included the intent to perform a CT scan. As a consequence, the patient population may have been biased towards inclusion of more patients with CT abnormalities. However, overall 40% of patients were CT negative (86% in ER, 52% in Admission and 11% in the ICU stratum).

5. The reported interpretation of results is only valid for the biomarkers studied and cannot be extrapolated to other biomarkers. CENTER-TBI has prepared for facilitating legacy research on other markers or on clinical-use platform(s) by reserving a number of pristine aliquots for future studies.

6. The permitted time window of 24 h may have affected the diagnostic accuracy of biomarkers with short circulating half-lives. Understanding the kinetics of such biomarkers may inform optimization of time windows for improving diagnostic performance [28, 29].

7. We did not explore possible gender effects.

8. We did not take the clustered structure of the data into account, because it was statistically not feasible to adjust both for clinical characteristics as well for between centre variations.

9. As explained above, we deliberately refrained from attempting to identify action thresholds (cutoff values) in a post-hoc analysis.

In conclusion, each of the six investigated biomarkers scaled with the severity of TBI and with care path. GFAP serum levels obtained within 24 h post-injury predict CT positivity across the full range of injury severities. In patients with mild TBI and in patients with a GCS of 15, GFAP adds value to clinical characteristics and outperforms other markers, including S100B. No clear additional value for predicting CT positivity was found when combining GFAP with other biomarkers. Our evidence supports development of novel CT decision rules, combining serum GFAP with clinical characteristics, for triaging patients with mild TBI for CT scanning. To this purpose, validated clinical-use assays are required.

12.7 | Supplementary Material

Supplementary materials can be found in the online version of this article: `https://tinyurl.com/wmbtmyz9`

References

1. Maas, A. I. R. *et al.* Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. *The Lancet Neurology* **4422.** ISSN: 14744422 (2017).
2. Yue, J. K. *et al.* Association between plasma GFAP concentrations and MRI abnormalities in patients with CT-negative traumatic brain injury in the TRACK-TBI cohort: a prospective multicentre study. *The Lancet Neurology* **18,** 953–961 (2019).
3. Stiell, I. G. *et al.* The Canadian CT Head Rule for patients with minor head injury. *The Lancet* **357,** 1391–1396 (2001).
4. Foks, K. A. *et al.* External validation of computed tomography decision rules for minor head injury: prospective, multicentre cohort study in the Netherlands. *British Medical Journal* **362** (2018).
5. Smits, M. *et al.* Predicting intracranial traumatic findings on computed tomography in patients with minor head injury: the CHIP prediction rule. *Annals of Internal Medicine* **146,** 397–405 (2007).
6. Haydel, M. J. *et al.* Indications for computed tomography in patients with minor head injury. *New England Journal of Medicine* **343,** 100–105 (2000).
7. Mondello, S. *et al.* Blood-based protein biomarkers for the management of traumatic brain injuries in adults presenting to emergency departments with mild brain injury: a living systematic review and meta-analysis. *Journal of neurotrauma* (2018).
8. Mondello, S. & Hayes, R. L. in *Handbook of clinical neurology* 245–265 (Elsevier, 2015).
9. Undén, J., Ingebrigtsen, T., Romner, B., Committee, S. N., *et al.* Scandinavian guidelines for initial management of minimal, mild and moderate head injuries in adults: an evidence and consensus-based update. *BioMed Central: medicine* **11,** 50 (2013).

10. Minkkinen, M. *et al.* Prospective validation of the scandinavian guidelines for initial management of minimal, mild, and moderate head injuries in adults. *Journal of neurotrauma* **36,** 2904–2912 (2019).

11. Bazarian, J. J. *et al.* Serum GFAP and UCH-L1 for prediction of absence of intracranial injuries on head CT (ALERT-TBI): a multicentre observational study. *The Lancet Neurology* **17,** 782–789 (2018).

12. Maas, A. I. & Lingsma, H. F. ALERT-TBI study on biomarkers for TBI: has science suffered? *The Lancet Neurology* **17,** 737–738 (2018).

13. Maas, A. I. *et al.* Collaborative European neurotrauma effectiveness research in traumatic brain injury (CENTER-TBI): A prospective longitudinal observational study. *Neurosurgery* **76,** 67–80. ISSN: 15244040 (2015).

14. Steyerberg, E. W. *et al.* Case-mix, care pathways, and outcomes in patients with traumatic brain injury in CENTER-TBI: a European prospective, multicentre, longitudinal, cohort study. *The Lancet Neurology* **18,** 923–934 (2019).

15. Van Buuren, S. *Flexible imputation of missing data* 415. ISBN: 9781138588318 (CRC Press, 2018).

16. Rubin, D. B. *Multiple imputation for nonresponse in surveys* (John Wiley & Sons, 2004).

17. Steyerberg, E. W. *et al.* Graphical assessment of incremental value of novel markers in prediction models: from statistical to decision analytical perspectives. *Biometrical Journal* **57,** 556–570 (2015).

18. Thelin, E. *et al.* A serum protein biomarker panel improves outcome prediction in human traumatic brain injury. *Journal of neurotrauma* **36,** 2850–2862 (2019).

19. Papa, L. *et al.* GFAP out-performs S100β in detecting traumatic intracranial lesions on computed tomography in trauma patients with mild traumatic brain injury and those with extracranial lesions. *Journal of neurotrauma* **31,** 1815–1822 (2014).

20. Czeiter, E. *et al.* Brain injury biomarkers may improve the predictive power of the IMPACT outcome calculator. *Journal of neurotrauma* **29,** 1770–1778 (2012).

21. Mondello, S. *et al.* Clinical utility of serum levels of ubiquitin C-terminal hydrolase as a biomarker for severe traumatic brain injury. *Neurosurgery* **70,** 666 (2012).

22. McMahon, P. J. *et al.* Measurement of the glial fibrillary acidic protein and its breakdown products GFAP-BDP biomarker for the detection of traumatic brain injury compared to computed tomography and magnetic resonance imaging. *Journal of neurotrauma* **32,** 527–533 (2015).

23. Rezaii, P. G. *et al.* Stability of Blood biomarkers of traumatic brain injury. *Journal of neurotrauma* **36,** 2407–2416 (2019).

24. Abdelhak, A. *et al.* Glial activation markers in CSF and serum from patients with primary progressive multiple sclerosis: potential of serum GFAP as disease severity marker? *Frontiers in Neurology* **10,** 280 (2019).

25. Bossuyt, P. M., Olsen, M., Hyde, C. & Cohen, J. F. An analysis reveals differences between pragmatic and explanatory diagnostic accuracy studies. *Journal of Clinical Epidemiology* **117,** 29–35 (2020).

26. Wang, K. K. *et al.* An update on diagnostic and prognostic biomarkers for traumatic brain injury. *Expert review of molecular diagnostics* **18,** 165–180 (2018).

27. Papa, L. *et al.* Time course and diagnostic accuracy of glial and neuronal blood biomarkers GFAP and UCH-L1 in a large cohort of trauma patients with and without mild traumatic brain injury. *Journal of the American Medical Association: Neurology* **73,** 551–560 (2016).

28. Brophy, G. M. *et al.* Biokinetic analysis of ubiquitin C-terminal hydrolase-L1 (UCH-L1) in severe traumatic brain injury patient biofluids. *Journal of neurotrauma* **28,** 861–870 (2011).

29. Welch, R. D. *et al.* Modeling the kinetics of serum glial fibrillary acidic protein, ubiquitin carboxyl-terminal hydrolase-L1, and S100B concentrations in patients with traumatic brain injury. *Journal of neurotrauma* **34,** 1957–1971 (2017).

13

Missing data in prediction research: A five step approach for multiple imputation, illustrated in the CENTER-TBI study

Benjamin Y Gravesteijn, Charlie A Sewalt, Esmee Venema, Daan Nieboer, Ewout W Steyerberg

Journal of Neurotrauma

13.1 | Abstract

In medical research, missing data is common. In acute diseases such as traumatic brain injury (TBI), even well conducted prospective studies may suffer from missing data in baseline characteristics and outcomes. Statistical models may simply drop patients with any missing values, potentially leaving a selected subset of the original cohort. Imputation is widely accepted by methodologists as an appropriate way to deal with missing data. We aim to provide practical guidance on handling missing data for prediction modelling. We hereto propose a five-step approach, centred around single and multiple imputation:

1. explore the missing data patterns;
2. choose a method of imputation;
3. perform imputation;
4. assess diagnostics of the imputation;
5. analyse the imputed datasets.

We illustrate these 5 steps with the estimation and validation of the IMPACT prognostic model in 1375 patients from the CENTER-TBI database, included in 53 centers across 17 countries, with moderate or severe TBI in the prospective European CENTER-TBI study. Future prediction modelling studies in acute diseases may benefit from following the suggested 5 steps for optimal statistical analysis and interpretation, after maximal effort have been made to minimise missing data.

13.2 | Background

Missing data is a common problem in medical research [1]. Missing data occur in studies with routinely collected data, and even if data are prospectively collected with attempts to minimise the occurrence of missing data (Box 1).

Box 1 – Exemplary causes of missing data

Missing data occurs for example when there is no response to surveys or follow-up appointments; or if laboratory tests or imaging were not ordered for all patients; or when a score or test was simplified instead of properly executed to save time at a busy emergency department. For example, the Glasgow Coma Scale may be noted as 13 instead of the detailed score, e.g. E5M6V2; or "Pupils Equal And Reactive to Light" instead of the exact diameters of the pupils with and without exposure to light.

Prediction modelling is central to many domains of medicine, such as screening, diagnostics, and therapy. It is a growing research area [2, 3]. Predictions are based on the combination of characteristics for a diagnostic outcome (e.g. presence of abnormalities at CT scan) or prognostic outcome (e.g. Glasgow Outcome Scale at 6 months) [3]. Potential predictors commonly relate to the characteristics of the patient, disease, or treatment [4]. Statistical models typically drop patients with any missing values from analyses. Prediction research is particularly sensitive to the occurrence of missing data, because it relies on the statistical combination of multiple variables, which may each have missing values.

Analysing only the available data (often referred to as a "complete case analysis") is the most basic approach to deal with missing data. It de-

creases the available information for statistical analyses. Moreover, it could potentially introduce selection bias since patients with observed characteristics may be systematically different from patients with missing values [2, 5, 6]. For example, patients with missing baseline characteristics may be a selected subgroup, because laboratory tests or imaging might not be ordered for less severe cases.

Although the problem of missing data is complex, some methodological standards to deal with missing data are available. General recommendations to handle missing data have been suggested [7]. One generic recommendation is that multiple imputation is the method of choice in many areas of medical research [8]. Imputation exploits the availability of information from non-missing predictors for partly complete patients rather than discarding these patients. A recent systematic review showed that missing data is reported inconsistently and often handled suboptimally, underscoring the importance of a practical framework [9].

We aim to provide guidance to deal with missing data in prediction research. Some nuances and methodological considerations are provided in text boxes. We first address various general issues in prediction modelling and handling of missing data, specifically with imputation procedures. We then propose a five-step approach to perform imputation as a way to deal with missing data, illustrated with a case study in a multinational prospective cohort study for traumatic brain injury (TBI): CENTER-TBI [10, 11].

13.2.1 | Principles of prediction modelling

Prediction modelling entails the prediction of a clinically relevant outcome based on the combination of multiple predictors effects [12]. Prediction models typically provide estimates of the relative effect of predictors in the

model, and absolute risk predictions for individual patients. The CRASH and IMPACT models are examples of such prediction models in traumatic brain injury (TBI) [13, 14]. These models predict mortality and 6 months unfavourable outcome according to the Glasgow Outcome Scale (GOS) [14–16].

The performance of such models is ideally assessed in external validation studies, where predictions of an existing model are compared to observed outcomes in a new setting. For binary outcomes, the performance is commonly assessed using discrimination and calibration [8]. Discrimination refers to the ability of a prediction model to discriminate between patients with and without the outcome of interest. It is commonly assessed by the area under the Receiver Operating Characteristic curve (AUC, or c-statistic). Calibration refers to the agreement of predicted probabilities of a model and observed outcomes (e.g. "if the risk of death is x, do x% of the patients with this prediction actually die?"). Calibration can be assessed graphically, and can be summarised according to calibration-in-the-large (measuring under- or over-estimation of the average predicted risk) and calibration slope (measuring the average strength of overall predictor effects) [17].

The current tutorial gives guidance on how to deal with missing data to use all available information in a data set to develop a prediction model and to validate an existing model. Effectively, this boils down to correct estimation of the parameters in the model at model development. At validation, we apply the model to new data and compare the observed outcome to the predicted risk.

13.2.2 | Mechanisms of missing data

To describe missing data, a paradigm of three distinctive mechanisms for missing data has been established (table 13.1) [5]. First, Missing Completely At Random (MCAR) arises when missingness is not associated with observed or unobserved variables. For example, this missingness arises when administrative errors or accidents occur. Second, Missing At Random (MAR) is defined as missingness that is associated with observed variables. For example, patients with lower injury severity scores may have more missing computed tomography (CT) scans. Finally, Missing Not At Random (MNAR) arises when the missingness is associated with unobserved variables, or the value of the variable itself. For example, patients may be less likely to fill in their income in a survey if their income is substantially higher than average. Both missingness in the predictors and the outcome can be categorised according to these missing data mechanisms. An example for MNAR in the outcome is that patients may not return for follow-up visits if they are doing very poorly or very well.

Table 13.1: *missing data mechanisms for predictors with examples [2].*

Label	Missing Mechanism	Description	Clinical example
MCAR	Missing completely at random	Administrative errors, accidents	A batch of blood samples got lost
MAR	Missing at random	Missingness related to known patient characteristics	Older patients have more difficulty to come to follow-up visits
MNAR	Missing not at random	Missingness related to either the value of the predictor, or to variables not available in the analysis.	Self-report of income in a survey: lower incomes are less likely to be reported. Patients with impaired cognition (not measured) cannot understand and fill in a neuropsychological test.

13.2.3 | Methods for imputation

The most basic method for dealing with missing data is a complete case analysis. This leads to loss of statistical power, and potentially to bias in risk predictions [18]. The most problematic consequence of less statistical power is a higher risk of overfitting of a prediction model. The model perform much better in the original dataset (too "optimistic"), than the model will perform when it is used to calculate risk for new patients. Bias in risk prediction as a result of a complete case analysis may also result in a systematically over- or underestimated risk (poor calibration-in-the-large). Therefore, many methodologists and journal editors nowadays recommend to statistically impute missing data [8, 19]. For such imputation, we have to assume that the mechanism of missingness is MCAR or MAR, not MNAR. And we assume that the imputation model, which only includes observed variables, is valid. We will discuss variants of single imputation, and multiple imputation (table 13.2). Multiple imputation, because of its theoretical attractiveness, has become a methodological standard. But as every method it has its limitations [20], an educated and balanced choice based on the research aim should be made.

Table 13.2: *Methods of dealing with missing data.*

Method	Description	Valid* under
Complete case	Drop cases with missing values	MCAR
Average imputation	Replace the missing values by the average	-
Single imputation by conditional estimation	Replace the missing values by the most likely value based on the observed data.	MCAR, MAR
Multiple imputation	Replace the missing values by the most likely values based on the observed data, multiple times	MCAR, MAR
Multiple imputation, including the outcome	Multiple imputation, also imputing the outcome	MCAR, MAR

* Valid here is defined as providing unbiased estimates of the final parameters in the model

13.2.3.1 | Single imputation

Replacing the missing value by a different value is referred to as single imputation. We describe two distinct methods for replacing this value: average imputation, and single imputation using conditional estimates. The theoretical disadvantage of single imputation in general is that the uncertainty in imputation is not incorporated in the final analysis (Box 2).

Box 2 – Single imputation and uncertainty

The main theoretical disadvantage of single imputation is that the uncertainty of the imputed values is not fully taken into account in the estimation of the final model at development: the imputed values are used as observed values in the final analysis. The result is that the standard errors (a measure of uncertainty) are too low. However, it can be argued that this disadvantage is not relevant. If our main aim is to predict the most likely risk for a new patient, the reliability of that prediction is more important than the uncertainty in the estimated parameters. Notably, machine learning algorithms do not incorporate uncertainty into their parameter estimates, and are only compatible with single datasets. Single imputation may also be considered when working with very large sample sizes to reduce the computational time.

Imputing the value by the average value as the best estimate for the missing value may translate to taking the mean for normally distributed variables, the median for non-normally distributed variables, and the mode for categorical variables. This method typically results in biased estimates of parameters under any missing data mechanism at model development (Box 3).

Single imputation by conditional estimates translates into using predictions of the most likely value of the missing observation from a regression

model. Both categorical and continuous predictors may be estimated by a variety of regression models (see step 3: performing imputation). Such conditional imputation provides better point estimates of the prediction model than average imputation. We may also draw imputed values from a distribution of likely values.1 Different imputed datasets then yield different estimates in the statistical analysis. Although the random drawing of values now incorporates some uncertainty, the final analysis still treats these imputed values as observed. Because this technique incorporates existing correlations in observed data, it is not only valid under MCAR but also under MAR.

Box 3 – Average imputation and bias towards the null

Average imputation seems like a non-educated guess for the missing observation: the distribution can create a non-natural spike of the mean. This simple imputation approach leads to bias in the estimated predictor effects. This bias is usually towards the null, which makes the approach generally conservative. This behavior is similar to shrinkage of regression coefficients towards the null, as can be achieved with penalised regression methods [21]. Average imputation biases the parameters towards the null even when the sample size is large. Therefore, penalised regression models are preferred to address overfitting rather than average imputation.

13.2.3.2 | Multiple imputation

Multiple imputation is an extension of single imputation using conditional estimates, and is valid under the same missing data mechanisms (MCAR, MAR) [1]. Instead of imputing one value per patient per missing value, multiple values are drawn with models fitted on the observed variables and stored in multiple completed datasets. The statistical analysis is then

performed on each imputed dataset separately and the results can be pooled using specific rules ("Rubin's rules", Box 4) [22].

Box 4 – Rubin's rules for prediction research
Estimates of regression coefficients and performance measures are simply averaged. For the standard error, the between-imputation variance of the coefficients and performance measures is added to the average variance within imputation sets. Therefore, multiple imputation leads to a better estimation of variability in the parameters than single imputation [5, 6, 22]. Modern software makes it relatively easy to implement multiple imputation (MI) for prediction models based on regression techniques.

13.3 | Methods: case study

As a case study we consider updating (or re-estimation) and external validation of the IMPACT prognostic model in the CENTER-TBI study. The IMPACT model aims to predict 6-month mortality and unfavourable outcome in moderate and severe TBI patients [14]. With good discriminatory performance (area under the ROC-curve around 0.8017) and multiple external validation studies [23–25], the IMPACT model might be considered a rather robust prediction model. The full model contains 10 predictors, with 18 logistic regression coefficients (Table 1 in appendix A available online) [14]. It was developed in the IMPACT data base, a collection of 8 trials and 3 surveys, with a total of 8509 patients [26, 27].

We validated this model in the CENTER-TBI data base, a European multinational prospective cohort study including all severities of TBI [10, 11]. We selected 1375 patients with moderate (GCS 9-12) and severe (GCS 3-8) TBI (Table 1 appendix A available online), included in 53 centers across

17 countries.

The analyses described in this paper are all performed in the freely available R software (A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria). The code with explanations is available in the online appendix C available online.

13.4 | Results: 5 steps for missing values

We consider 5 steps to deal with missing data in prediction research (Table 13.3).

Table 13.3: *A 5 step approach to deal with missing data in prediction research.*

Step	Action	Main objective
1	Explore missingness	Assess the quantity and pattern of missingness.
2	Choose imputation method	Balance the benefits and harms of alternative imputation methods in relation to the prediction question.
3	Perform imputation	To produce one or multiple imputed datasets.
4	Diagnostics	Assess the convergence of the iteration process, and compare the imputed data with the original dataset.
5	Analyse	Estimate the regression coefficients in a prediction model, or assess performance for an existing prediction model.

13.4.1 | Step 1 – Explore missingness

The first step in handling missing data is inspecting the dataset and explore missingness of variables (table 13.3). Recommended steps are to assess:

1. The quantity of missingness: the proportion of missing observations in each variable;

2. Patterns of missingness: the frequency of specific missing data patterns per patient;

3. Correlation between variables;

4. Associations between missingness in one variable and other variables.

One of the main considerations for missing data is what proportion of missingness in a predictor is still acceptable when imputing missing data. The stability of imputation declines with higher proportions of missingness [28]. A general advice is that the researcher should decide what proportion can be accepted based on their research question and their context (Box 5).

The quantity and pattern of missingness can be visualised (figure 13.1). If a particular pattern per patient arises more frequently than other patterns, a reason should be hypothesises. In our example, glucose and haemoglobin are often missing together in our case study. This might be MCAR: if no blood sample was taken (perhaps by mistake), both glucose and haemoglobin are missing. Similarly, imaging characteristics are also missing together frequently: if the patient did not undergo CT scanning, all these variables are missing together. This could be considered as MAR: less severe patients (higher GCS scores) are less likely to undergo a CT scan. Missing data mechanisms should always be assessed critically. Evaluating specific combinations of missingness may help to do so.

Imputation is more efficient if variables are correlated. A correlation plot can be useful (figure 13.2). When highly correlated variables are identified (r close to 1, e.g. haemoglobin and haematocrit), only one of the

highly collinear variables is sufficient to use, preferably the variable with the lowest proportion of missingness.

Box 5 – Maximum proportion of missingness to impute

No consensus has been reached about a maximum limit of missingness per variable or per patient. Limits may depend on the specific research question and context [2]. For example, when we are interested in the diagnostic value of a specific biomarker, we would probably not impute missing values for the biomarker, and be liberal in imputing missing values for other covariates that may potentially act as confounders. In contrast, when we consider prediction based on the combination of predictors, we may focus on including strong predictors with few missings (which are imputed). Depending on what assumption the researcher is willing to make, multiple imputation can be used for different situations: if MCAR or MAR is plausible, multiply imputing about 10% to 20% missing values per variable is generally acceptable, while under MCAR perhaps 50% missing values can be imputed without much instability in predicted values [2]. The larger the proportion of missing values, the more important that you specified the imputation model correctly, or else the final parameters might be biased [29]. Nevertheless, when the model is correctly specified, multiple imputation still reduces bias compared to a complete case analysis [30]. Multiple imputation accounts for the uncertainty of the imputed values in the standard errors of the final model but relies on the validity of the imputation model. Higher proportions of missingness motivate a higher number of imputation rounds [1].

Associations between missingness in one variable and other variables can be investigated to test the MCAR assumption. This can be done by logistic regression analysis with missingness of a variable as the dependent variable. All other variables are the predictors. For example, we observe that the variable 'pupils' is more often missing in patients with higher mo-

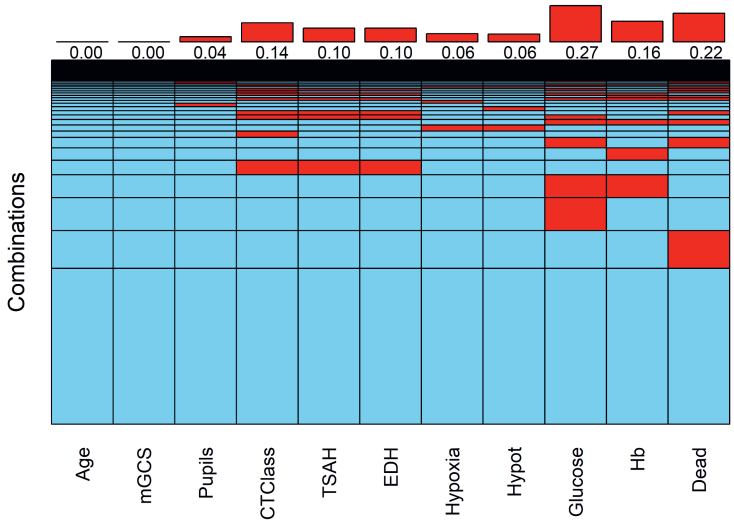


Figure 13.1: *Quantity and pattern of missingness for each variable in the CENTER-TBI data base (n=1375). On the top of the graph, the proportion of missingness is shown. In the table, the patterns of missingness are shown, with cell size proportional to number of patients. Each row represents a patient group, and the red blocks indicate combinations of variables that are missing. A total of 589/1375 (43%) had fully complete data.*

tor GCS (Figure 1 in appendix A available online). Although MCAR versus non-MCAR can be tested, the practical value remains unclear: there is no consequence for the consecutive strategy. What would be useful instead is testing MAR versus MNAR. However, this is impossible, because the information to test MAR versus MNAR is missing. Therefore, we still need to make the assumption that the data is MAR, not MNAR, when performing imputation.

To increase the likelihood for the MAR assumption to hold, auxiliary variables may be added to the imputation model. Auxiliary variables should

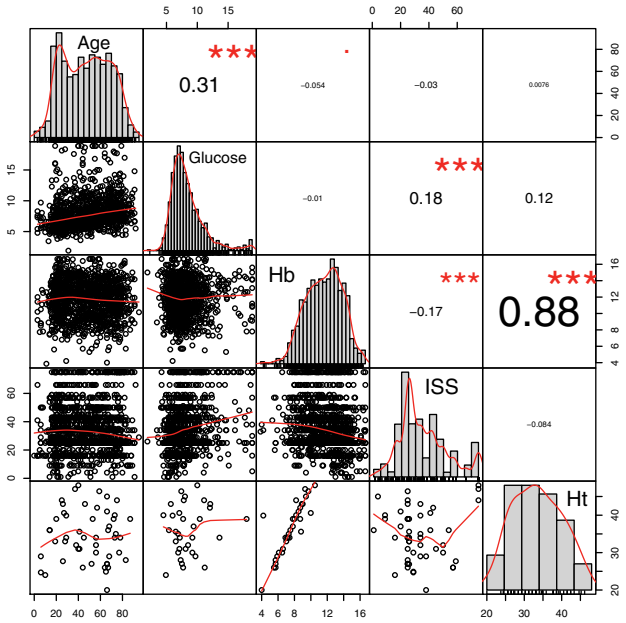


Figure 13.2: *Correlation plot between continuous variables, with Spearman rank correlation coefficients (numbers with p-values: * <0.05, ** <0.01, *** <0.001). Hb= hemoglobin; Ht= hematocrit; ISS= injury severity score. A strong correlation is observed between Hb and Ht (r=0.88), while glucose increases somewhat with age (r=0.31).*

meet two criteria. First, they should be frequently observed: if the auxiliary variable is often missing together with the variables of interest, there is no information gained by adding this variable. Secondly, the auxiliary variable should contain statistical information about the variables of interest: variables that are not associated with the predictor variables do not add information to the imputation model. An example of a potential auxiliary in our dataset is injury severity score (ISS): it was present for 1365/1375 patients (99.3%) in our dataset and was correlated with some of the IMPACT

predictors (figure 13.2). Moreover, the ANOVA test of ISS across motor GCS, or pupil categories was statistically significant (both: $p < 0.001$). Additionally, adding variables which correlate with other variables might increase the adequacy of the imputation model [29, 31].

In conclusion, exploring missing data should at least assess the quantity and patterns of missingness. This exploration should assess all variables that will be added to the imputation model: all predictors and the outcome. Additionally, we should identify highly correlated variables, and select the one that is most relevant and most frequently observed to add to the imputation and prediction model. Moreover, we can test MCAR versus non-MCAR. For imputation of predictors, the result of this test may not impact the strategy to deal with missing values, because we still need to assume that the data is MAR. Imputing the outcome might be especially beneficial when the outcome is non-MCAR (Box 6). Finally, we should focus on adding auxiliary variables in the imputation model, such as ISS in this case study. Adding these variables increases the likelihood of the MAR assumption to hold.

13.4.2 | Step 2 – Choose method of imputation

The approach to deal with missing data may depend on the size of the data set and the proportion of missing values (table 13.4).

With smaller data sets (say total n<1000, or <100 events), multiple imputation is the default method of choice to estimate the relative effects of predictors, with the appropriate associated standard error. However, we might also consider single imputation for prediction of absolute risk, especially if the sample size is large (Box 1).

Variability between imputations may be substantial, especially with a

Table 13.4: *Attractiveness of imputation method in prediction research for scenarios with small or large data sets with small or large proportions of missing data*

Method	Small dataset, large proportion of missingness	Small dataset, small proportion of missingness	large dataset, large proportion of missingness	large dataset, small proportion of missingness
Complete case	-	-	-	+/-
Average imputation	-	+/-	-	+/-
Single imputation	-	+/-	+/-	+
Multiple imputation	+	+	+	+/-

large proportion of missing values. Some researchers might find average imputation attractive, which may be defendable in situations with a very small proportion of missingness.

13.4.3 | Step 3 – Performing imputation

13.4.3.1 | Single and multiple imputation

For single and multiple imputation, flexible imputation using chained equations is the current standard, for example with the MICE package [1] or the mi impute chained command in Stata. The procedure uses an iterative imputation method, to arrive at stable estimates of imputations for unobserved variables. There are various models within the MICE framework to impute with, the most standard options are Bayesian linear regression (“norm”) for normal distributed variables, predictive mean matching (“pmm” or “midastouch”) for non-normally distributed continuous variables, logistic regression (“logreg”) for binary categorical, polytomous regression (“polyreg”) for categorical variables with more than 2 categories, and proportional odds logistic regression (“polr”) for ordered categorical variable is ordered (such as GOS-E).

We draw one imputed value for a single completed data set (single imputation), or create multiple copies of the dataset (multiple imputation). Although Rubin famously claimed that three copies was enough for stability [22], five datasets is now seen as the minimum. Creating even more imputed datasets often might be more beneficial, for example when instability is expected due to substantial proportions of missing values [1], for example 20 imputed sets if 20% of the values of a predictor are missing.

Finally, it is important to include the outcome in the imputation model. Not including the outcome in the imputation model can result in biased coefficients of the prediction model (Box 6) [32].

13.4.4 | Step 4 – Diagnostics

After imputation, the adequacy of the imputation procedure needs to be checked. For any imputation method, it is advised to compare the distribution of imputations with the distribution of the observed values dataset (figure 13.3). If distributions do not match, the question is what causes this difference. The distributions of average imputed values are different when compared to the original distribution in our case study. The distribution of multiple imputed data corresponds well with the distribution of the observed variables. Moreover, convergence of the algorithm may be checked (Box 7).

13.4.5 | Step 5 - Analysis

The final step is to use the (imputed) dataset for the analysis. As stated before, we consider model development and model validation.

At model development, we fit the model of interest (Box 8) for complete case analysis, average imputation, and single imputation by conditional estimates. For multiple imputed datasets the analyses are performed on each of the separate imputed datasets and the results are pooled using Rubin's rules [22] (See appendix D available online).

When we re-estimated the IMPACT model, we found that most imputation methods led to similar estimates of the regression coefficients (figure 13.4). The only imputation method with markedly different estimates was the complete case analysis. Indeed, relatively few patients were in the complete case dataset (n=589) compared to the total cohort (n=1375). Estimation of predictor effects suffers from small numbers in the complete case analysis. Interestingly, the refitted coefficients were generally more extreme (further away from 0) compared to the estimates from the IMPACT data sets.

The estimated prediction model should be validated to ensure reliable risk prediction for new patients [33]. At the developmental stage, one should at least perform some form of internal validation. A recommended method is a bootstrap procedure (Box 9) [34].

Box 6 – Imputation for predictors and/or outcome

A distinction should be made for imputation of missing values of predictors versus outcomes. It is generally accepted that it is important to include the outcome when imputing predictor values.

More controversial is the possibility to impute the outcome. A recent paper discusses handling missing outcome data in TBI studies in general [9]. From a predictive modelling perspective, imputation seems counterintuitive: we aim to predict the outcome, how could imputation help in this case?

First and foremost, the average risk estimate might be biased if we exclude patients with missing outcomes: if those patients had on expectation worse outcomes, the prediction model will underestimate the average risk. An exemplary study is the external validation of various CT-decision rules by Foks et al. [35], where the outcome (intracranial abnormality) was imputed to arrive at unbiased estimates of sensitivity and specificity. Excluding patients without CT scans, which are likely less severe patients, led to a higher estimated sensitivity and lower specificity of the decision rules.
A second reason for imputation of the outcome is that there is more information available to accurately predict the outcome. This may be the case when there are variables measured after baseline that are related to the outcome, or when the outcome is measured repeatedly. For example we may have assessed quality of life at 3, 6, and 12 months. If quality of life score is 50% at 3 and 12 months, it is likely 50% at 6 months as well. We note that repeated assessments of outcomes can also be assessed with mixed effects models or state transition models [36].
Patients with imputed outcomes can be excluded from the final analysis, while they were included in the imputation process. This has been labelled multiple-impute-and-delete (MID) [37]. The final analysis is performed only on the set of patients with observed outcomes. The advantages of MID over MI may be debated [2, 37, 38].
In conclusion, it is important to include the outcome as predictor in the imputation model. It is also statistically possible to impute the outcome. Whether or not to include patients with imputed outcomes in the final analysis depends on the research question: at model development, regression coefficients may generally be similar. But imputation of outcome may be beneficial for better estimation and assessment of average risk, when the outcome is non-MCAR.

A more rigorous test for model performance is external validation: the model is applied to a new dataset of patients not used at model development. The previously fitted model (with one set of coefficients) can be applied to predict the outcome in each imputed dataset. In each dataset, the performance of the model can be tested, and consecutively pooled. For

pooling of the calibration intercept and calibration slope, Rubin's Rules (appendix D available online) can be used to estimate the mean performance and the variance. For the c-statistic, the approach described above (pooling the bootstrap replicate results) can be used. These approaches were performed for validation of the IMPACT model in CENTER-TBI: it was confirmed that the complete case strategy results in different estimates for discrimination and calibration (figure 13.6). The model calibration was worse in this small subset. In contrast, discrimination was better and more uncertain. Interestingly, the multiple imputed datasets with imputed outcome (MI+y) has slightly worse calibration than the datasets where the patients with missing outcome were excluded. Interestingly, most imputations of the outcome resulted in a positive outcome (Figure 2 in appendix A available online).

Box 7 – Convergence of imputation with chained equations

Single and multiple imputation using the MICE algorithm is an iterative process where missing values in multiple predictors are imputed sequentially and multiple times. This iterative process starts with a 'best guess' of a missing value and refines this in subsequent iterations. This iterative process is finished when the algorithm that imputes missing values is converged. Whether or not convergence is reached can be studied using diagnostic plots (Appendix B available online). If these diagnostic plots show non-convergence of the algorithm the number of iterations can be increased or the imputation model can be refined. Convergence problems can occur especially when imputing missing values for variables that are functionally dependent on each other. For instance, when imputing missing values for weight using the body mass index, and subsequently imputing missing values for body mass index using weight [1].

Finally, when calibration is poor, the model can be recalibrated by re-estimating the model parameters in the new dataset, and again validated. This iterative process is inherent to prediction research to ensure applicability and reliability of a prediction model [3].

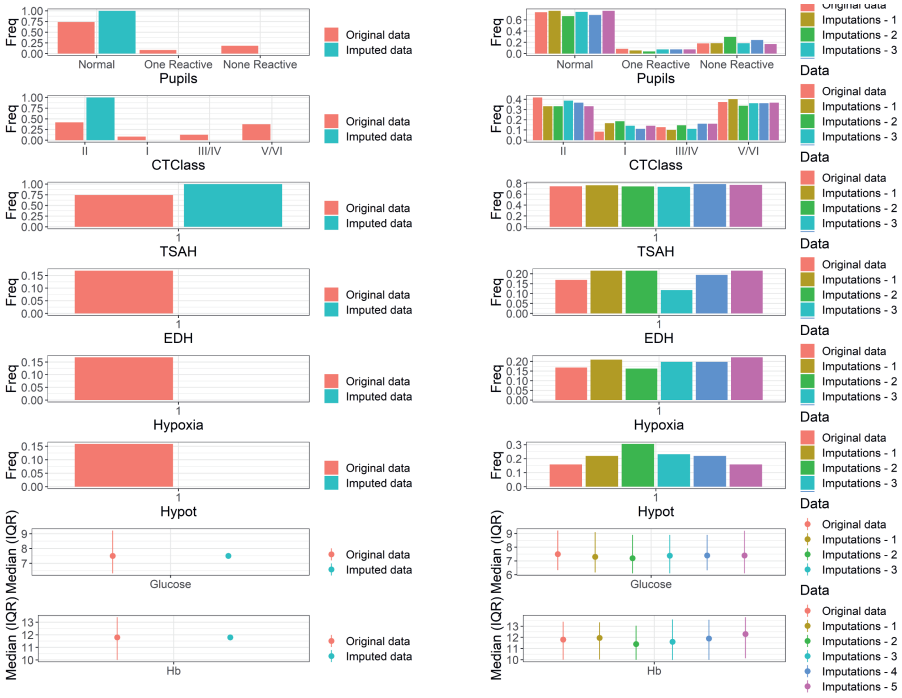
Box 8 – Model selection and multiple imputation

Model selection often precedes model estimation. Two strategies are common. First, stepwise selection may be performed with a forward selection strategy. The computer selects predictors based on a significant contribution over a smaller model. If stepwise selection is used, backward elimination is preferable. Predictors are dropped from a full model if this does not result in significantly worse model fit.

The second strategy bases selection on subject knowledge, from earlier studies and or from medical experts. This may be the most advantageous in terms of external validity [39, 40]: by selecting a model with well-known predictors, the predictions become is less dependent on the specifics of the current dataset at hand.

For complete case analysis and single imputation, these selection strategies can be implemented directly. For multiple imputed datasets, the last strategy (fitting a model based on prior knowledge) can be directly implemented. For model selection using multiple imputed data there are several potential approaches [41]:

1. Majority: Select variables that are included in the majority of the methods
2. Stack: Stack the imputed datasets into a single dataset and use weighting to adjust for multiple occurrences of the same patient and apply the usual variable selection methods.
3. Pool and test: Perform stepwise selection based on the pooled regression coefficients and associated standard errors using the Wald test ("D1" function in MICE) or a likelihood ratio test ("D3" function in MICE).



(a) *Average imputation* ***(b)*** *Multiple imputation*

Figure 13.3: *Distribution of variables before and after imputation (panel A: average imputation; panel B: multiple imputation). For categorical variables, the proportion of the total number of observed values is shown, and for continuous variables, the median and interquartile range is shown. Pupils: pupillary reactivity; CT class: Marshall CT class; Hypot: Hypotension*

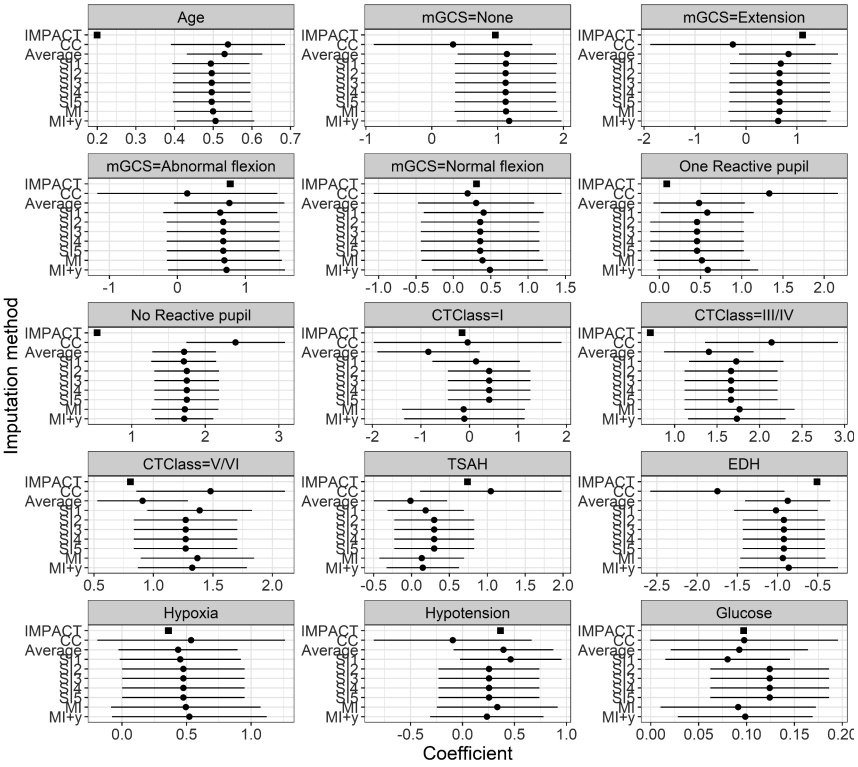


Figure 13.4: *Re-estimation of the IMPACT model for mortality in the completed CENTER-TBI datasets. The IMPACT square represents the point estimate of the IMPACT model in the original dataset (18). The complete case data set includes 589 patients, the MI+y dataset contains 1375 patients, while the other imputed datasets contain 1077 patients. CC: complete case; SI: single imputation by conditional estimates; MI: multiple imputation; MI+y multiple imputation, also the outcome.*

Box 9 – Bootstrap procedure for internal validation

The possible bootstrap validation procedure is visually displayed in figure 13.5: after sampling patients with replacement, the model is fitted on that sample (the bootstrap sample), and its performance assessed in the bootstrap sample and the original data. The difference indicates the optimism in performance. This procedure is repeated many times (e.g. 200 times), and the average is taken [42]. This procedure is less straightforward after multiple imputation, since we have multiple datasets on which we can perform bootstrapping and different approaches to pool results. Simply taking the average of the optimism to correct the average of the apparent performance over multiple imputed data sets may be reasonable. Simulation studies show that this approach produces quite valid estimates [43].

13.5 | Discussion

This tutorial considered the role of imputation approaches to alleviate the problem of missing data in prediction research in acute medicine. There has been much debate about the reliability and applicability of prediction model studies, even when published in high-impact journals, because of their non-adherence to methodological principles [44]. In this context, we recommend multiple imputation, because it is a readily implementable procedure and generally superior to simple approaches, such as a complete case analysis. However, simpler imputation methods (single imputation by conditional estimates, average imputation) can be considered when the proportion of missing data is small and the database large (Table 13.4). Technically, it is important to include the outcome in the imputation model. For prediction of absolute risk, it may even be considered to also impute missing outcomes. Although this approach is counterintuitive, it can increase statistical efficiency and avoid biased average risk estimates. It should be noted that the guidance provided in our tutorial needs fur-

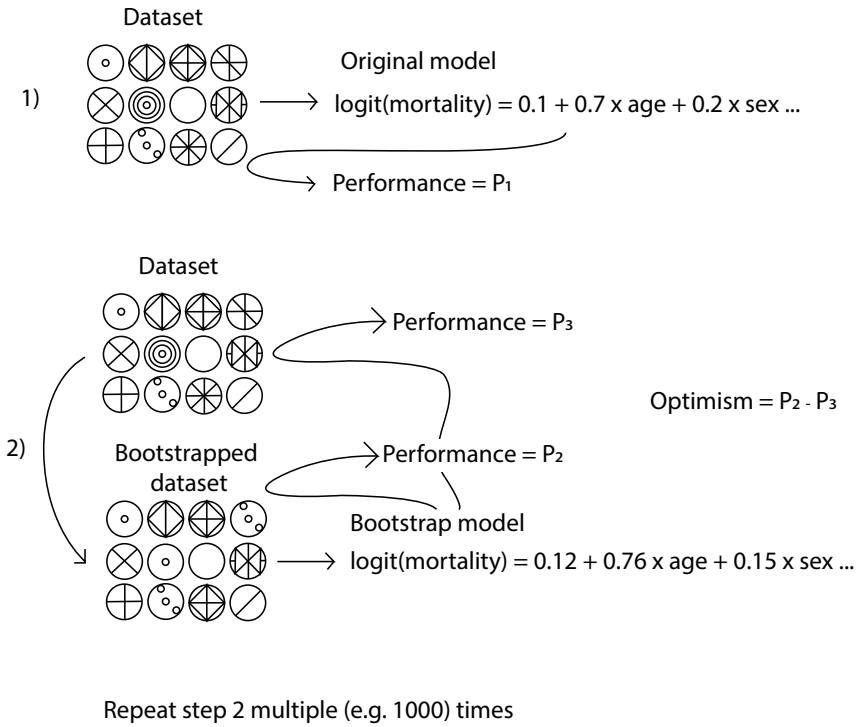


Figure 13.5: *Diagram showing the recommended technique for internal validation, bootstrapping. First, the model is fitted on the original data and the performance is determined in this dataset. Step two is the bootstrap procedure: patients are drawn with replacement from the original dataset to arrive at the bootstrapped dataset. The model is again fitted on the bootstrapped dataset. Finally, the performance in the bootstrapped dataset and the original dataset is obtained. The difference between performance is the optimism. Step two is repeated a number of times (e.g. 1000) to obtain multiple estimates of the optimism. Step three is to correct the originally obtained performance by the mean of the optimism obtained in step two.*

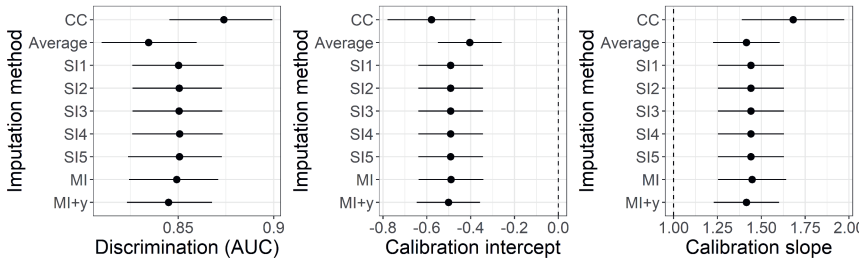


Figure 13.6: *External validation of the IMPACT model in variants of the CENTER-TBI dataset. Estimates and 95% confidence intervals are shown. CC: complete case; SI: single imputation; MI: multiple imputation; MI+y multiple imputation, also the outcome.*

ther underpinning. Specifically, more evidence from simulation studies is needed.

13.5.1 | Other approaches

Next to imputation, there are other relevant approaches to deal with missing data. We will discuss two of these approaches.

A promising strategy to deal with missing data at external validation relies on pattern submodels [45]. Models are fitted in a development setting for each pattern of missingness. For example, one submodel might fit all predictors except for age in patients with non-observed age values, one submodel might fit all predictors except for sex and age in patients with non-observed age and sex values, and so forth. These models might be more robust, since in contrast to imputation, they do not assume MAR [45]. Furthermore, these models are practical in the real-world scenario: when a prediction needs to be made for a new patient, but not all predictors are measured, these submodels can still be applied. When imputation is used to develop a model, as we have advocated, either all predictors need to

be collected, or imputed again at validation. The sample sizes needed for validation with pattern submodels are larger than with imputation, since a reasonable number of patients are needed for each missing data pattern.

The second other strategy to deal with missing data is inverse probability weighting [46]. In this approach, only complete cases are used, but weighted with the inverse of the probability of being a complete case. The more likely all variables are observed in a patient, the less they contribute: they are already represented well in the dataset. This approach aims to mitigate selection bias associated with complete case analysis. However, because it uses only a subset of all patients, it may be less efficient than multiple imputation, which exploits the availability of information on non-missing predictors for partly complete patients [47].

13.5.2 | Conclusion

When imputation is the strategy of choice, the suggested five-step approach emphasises key considerations and pitfalls which might be encountered. Additionally, the online vignette provides directly implementable code in the freely available R software (Appendix C available online), such that the discussed approaches can easily be applied for prediction modelling in acute medicine.

13.6 | Supplementary Material

Supplementary materials can be found in the online version of this article: `https://tinyurl.com/5zccjhef`

References

1. Van Buuren, S. *Flexible imputation of missing data* 415. ISBN: 9781138588318 (CRC Press, 2018).
2. Steyerberg, E. W. *Clinical Prediction Models* ISBN: 978-3-030-16398-3. `http://link.springer.com/10.1007/978-3-030-16399-0` (Springer International Publishing, Cham, 2019).
3. Steyerberg, E. W. *et al.* Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Medicine* **10,** e1001381 (2013).
4. Riley, R. D. *et al.* Prognosis Research Strategy (PROGRESS) 2: Prognostic Factor Research. *PLoS Medicine* **10,** e1001380. ISSN: 1549-1676 (2013).
5. Molenberghs, G., Kenward, M. & Ebrahim, G. J. Missing Data in Clinical Studies. *Journal of Tropical Pediatrics* **53,** 294–294. ISSN: 0142-6338 (2007).
6. Little, R. J. A. & Rubin, D. B. Statistical Analysis with Missing Data. *Journal of Educational Statistics* **16,** 150. ISSN: 03629791 (1991).
7. Nielson, J. L. *et al.* Statistical Guidelines for Handling Missing Data in Traumatic Brain Injury Clinical Research. *Journal of Neurotrauma.* ISSN: 0897-7151 (2020).
8. Leisman, D. E. *et al.* Development and Reporting of Prediction Models: Guidance for Authors From Editors of Respiratory, Sleep, and Critical Care Journals. *Critical care medicine,* 1. ISSN: 1530-0293 (2020).
9. Richter, S. *et al.* Handling of missing outcome data in traumatic brain injury research – a systematic review. *Journal of Neurotrauma,* neu.2018.6216. ISSN: 0897-7151 (2019).
10. Maas, A. I. *et al.* Collaborative European neurotrauma effectiveness research in traumatic brain injury (CENTER-TBI): A prospective longitudinal observational study. *Neurosurgery* **76,** 67–80. ISSN: 15244040 (2015).

11. Steyerberg, E. W. *et al.* Case-mix, care pathways, and outcomes in patients with traumatic brain injury in CENTER-TBI: a European prospective, multicentre, longitudinal, cohort study. *The Lancet Neurology* **18,** 923–934 (2019).

12. Shmueli, G. To explain or to predict? *Statistical Science* **25,** 289–310. ISSN: 08834237. arXiv: 1101.0891 (2010).

13. Collaborators, M. *et al.* Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients. *British Medical Journal* **336,** 425–429 (2008).

14. Steyerberg, E. W. *et al.* Predicting Outcome after Traumatic Brain Injury: Development and International Validation of Prognostic Scores Based on Admission Characteristics. *PLoS Medicine* **5** (ed Singer, M.) e165. ISSN: 1549-1676 (2008).

15. Dijkland, S. A. *et al.* Prognosis in Moderate and Severe Traumatic Brain Injury: A Systematic Review of Contemporary Models and Validation Studies. *Journal of Neurotrauma,* neu.2019.6401. ISSN: 0897-7151 (2019).

16. MRC CRASH Trial Collaborators. Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients. *British Medical Journal* **336,** 425–429. ISSN: 1756-1833 (2008).

17. Cox, D. Two further applications of a model for binary regression. *Biometrika* **45,** 562–565 (1958).

18. White, I. R. & Carlin, J. B. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. *Statistics in Medicine* **29,** 2920–2931. ISSN: 02776715 (2010).

19. Collins, G. S., Reitsma, J. B., Altman, D. G. & Moons, K. G. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. *Annals of Internal Medicine* **162,** 55. ISSN: 0003-4819 (2015).

20. Hughes, R. A., Heron, J., Sterne, J. A. & Tilling, K. Accounting for missing data in statistical analyses: Multiple imputation is not always the answer. *International Journal of Epidemiology* **48,** 1294–1304. ISSN: 14643685 (2019).

21. Van Houwelingen, H. C. & Sauerbrei, W. Cross-Validation, Shrinkage and Variable Selection in Linear Regression Revisited. *Open Journal of Statistics* **3,** 79–102 (2013).

22. Rubin, D. B. *Multiple imputation for nonresponse in surveys* (John Wiley & Sons, 2004).

23. Lingsma, H. *Measuring quality of care : methods and applications to acute neurological diseases* ISBN: 9789077283110 (Erasmus University Rotterdam, 2010).

24. Roozenbeek, B. *et al.* Prediction of outcome after moderate and severe traumatic brain injury. *Critical Care Medicine* **40,** 1609–1617. ISSN: 0090-3493 (2012).

25. Lingsma, H. *et al.* Prognosis in moderate and severe traumatic brain injury. *Journal of Trauma and Acute Care Surgery* **74,** 639–646. ISSN: 2163-0755 (2013).

26. Marmarou, A. *et al.* IMPACT Database of Traumatic Brain Injury: Design And Description. *Journal of Neurotrauma* **24,** 239–250. ISSN: 0897-7151 (2007).

27. Steyerberg, E. W., Nieboer, D., Debray, T. P. & Houwelingen, H. C. Assessment of heterogeneity in an individual participant data meta-analysis of prediction models: An overview and illustration. *Statistics in Medicine* **38,** 4290–4309. ISSN: 0277-6715 (2019).

28. Lee, K. J. & Carlin, J. B. Recovery of information from multiple imputation: a simulation study. *Emerging themes in epidemiology* **9,** 1–10 (2012).

29. Von Hippel, P. & Lynch, J. Efficiency gains from using auxiliary variables in imputation. *arXiv preprint arXiv:1311.5249* (2013).

30. Madley-Dowd, P., Hughes, R., Tilling, K. & Heron, J. The proportion of missing data should not be used to guide decisions on multiple imputation. *Journal of Clinical Epidemiology* **110,** 63–73. ISSN: 18785921 (2019).

31. Mustillo, S. The Effects of Auxiliary Variables on Coefficient Bias and Efficiency in Multiple Imputation. *Sociological Methods & Research* **41,** 335–361. ISSN: 0049-1241 (2012).

32. Moons, K. G., Donders, R. A., Stijnen, T. & Harrell, F. E. Using the outcome for imputation of missing predictor values was preferred. *Journal of Clinical Epidemiology* **59,** 1092–1101. ISSN: 08954356 (2006).

33. Steyerberg, E. W. & Harrell, F. E. Prediction models need appropriate internal, internal-external, and external validation HHS Public Access. *J Clin Epidemiol* **69,** 245–247 (2016).

34. Steyerberg, E. W. *et al.* Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. *Journal of Clinical Epidemiology* **54,** 774–781. ISSN: 08954356 (2001).

35. Foks, K. A. *et al.* External validation of computed tomography decision rules for minor head injury: prospective, multicentre cohort study in the Netherlands. *British Medical Journal* **362** (2018).

36. Kunzmann, K. *et al.* Imputation of ordinal outcomes: a comparison of approaches in traumatic brain injury. *Journal of Neurotrauma,* neu.2019.6858. ISSN: 0897-7151 (2020).

37. Kontopantelis, E., White, I. R., Sperrin, M. & Buchan, I. Outcome-sensitive multiple imputation: A simulation study. *BioMed Central: Medical Research Methodology* **17,** 2. ISSN: 14712288 (2017).

38. Sullivan, T. R., Salter, A. B., Ryan, P. & Lee, K. J. Bias and Precision of the "multiple Imputation, Then Deletion" Method for Dealing with Missing Outcome Data. *American Journal of Epidemiology* **182,** 528–534. ISSN: 14766256 (2015).

39. Wählby, U., Jonsson, E. N. & Karlsson, M. O. Comparison of stepwise covariate model building strategies in population pharmacokinetic-pharmacodynamic analysis. *Aaps Pharmsci* **4,** 68–79 (2002).

40. Derksen, S. & Keselman, H. J. Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology* **45,** 265–282. ISSN: 20448317 (1992).

41. Wood, A. M., White, I. R. & Royston, P. How should variable selection be performed with multiply imputed data? *Statistics in Medicine* **27,** 3227–3246. ISSN: 02776715 (2008).

42. Efron, B. & Tibshirani, R. *An Introduction to the Bootstrap* (CRC Press, 1994).

43. Schomaker, M. & Heumann, C. Bootstrap inference when using multiple imputation. *Statistics in Medicine.* ISSN: 02776715 (2018).

44. Bouwmeester, W. *et al.* Reporting and Methods in Clinical Prediction Research: A Systematic Review. *PLoS Medicine* **9** (ed Macleod, M. R.) e1001221. ISSN: 1549-1676 (2012).

45. Fletcher Mercaldo, S. & Blume, J. D. Missing data and prediction: the pattern submodel. *Biostatistics (Oxford, England)* **21,** 236–252. ISSN: 14684357 (2020).

46. Seaman, S. R. & White, I. R. Review of inverse probability weighting for dealing with missing data. *Statistical methods in medical research* **22,** 278–95. ISSN: 1477-0334 (2013).

47. Perkins, N. *et al.* Principled Approaches to Missing Data in Epidemiologic Studies. English. *American Journal of Epidemiology* **187,** 568–575. ISSN: 1476-6256 (2018).

Part IV

Closing

14

General discussion

In this thesis, I aimed to contribute to more efficient and effective clinical decision making for traumatic brain injury (TBI) and in-hospital cardiac arrest (IHCA), based on observational studies. The following three specific questions were addressed:

1. Given that current practice is described by given interventions and their respective outcome,

   - what is the variation in (prehospital) interventions for TBI, and
   - what is the expected long-term outcome after IHCA, and how much variation exists?

   This descriptive question was explored in the first part, *current practice*. This thesis identified large variation in interventions for TBI. This variation in intervention was primarily found in the prehospital setting, where the use of intubation, helicopter, IV fluids, and transport times varied substantially, which was not attributable to

differences in patient characteristics (**chapter 2 and 3**). We also found substantial variation in outcome in IHCA patients worldwide, which was generally poor (13.4% survived to 1 year) (**chapter 4**). In contrast, the outcome after IHCA patients who undergo ECPR (**chapter 5**), and IHCA patients in the Netherlands (**chapter 6**), is higher and more homogeneous. These two populations have one characteristic in common, compared to the general IHCA population: they are both more selected. Finally, I explored how care for IHCA patients can be improved in the Netherlands. Mainly the neurological outcome after IHCA can be improved by improving care, for example by more frequent CPR training of personnel (**chapter 6**).

2. What is the best practice, or what is the (cost-)effectiveness of currently performed interventions for TBI and IHCA?
   This causal question was assessed in part two, *best practice*. It was shown that there was no overall beneficial or harmful effect for either intubation (**chapter 7**) or direct transfer to a specialised neurotrauma centre (SNC) in patients with TBI (**chapter 8**). However, the intubation of the tracheas of more severely injured patients was associated with better outcome. More specifically, in patients with severe extracranial injury, intubation in the prehospital setting was associated with better outcome; while in patients with severe intractranial injury (a lower level of consciousness), intubation in the in-hospital setting was associated with better outcome. On the contrary, no subpopulation of TBI patients was found in which direct transfer to a SNC was associated with better outcome. Finally, using a decision model for ECPR after IHCA, the most likely cost-effective strategy was to treat every eligible IHCA patient with ECPR (**chapter 9**). This

approach was expected to cost €10,818 per extra year spent in perfect health (QALY).

3. How to better characterise and predict outcome in TBI?
   This predictive question was discussed in the last part, *identifying patients at risk*. I first presented an exploratory clustering analysis in TBI (**chapter 10**). The aim was to use this hypothesis-free analysis to cluster all TBI patients into clinical phenotypes. The most important characteristics to describe the TBI clusters were injury mechanism, presence of major extracranial injury, and GCS. Nevertheless, we showed that such a clustering strategy will not likely improve identifying patients at risk of poor outcome. These characteristics do not capture differences in prognosis, but differences in aetiology and care pathways. A statistically more efficient way to identify patients at risk is by predictive modelling. This thesis shows that for building a clinical prediction model, any algorithm can be used (**chapter 11**). More importantly, a model should be rigorously validated in the population in which it will be used. Moreover, predictions can be improved upon by including strong new predictors: we showed that in TBI, GFAP improves prediction of CT abnormalities (**chapter 12**). Finally, we proposed a 5-step approach to deal with missing data in prediction research (**chapter 13**):

   a) explore the missing data patterns;
   b) choose a method of imputation;
   c) perform imputation;
   d) assess diagnostics of the imputation;
   e) analyse the imputed datasets.

In this final chapter, these results are discussed, together with the implications for both the clinical practice and future research in this field.

14.1 | Current practice

The premise of studying current clinical practice, is to be able to identify clinical areas which need improvement. Clinical practice can be defined by the interventions given to patients, and the outcome that follows (figure 14.1). In order to quantitatively assess practice, variation can be used as measure of interest. Two types of practice variation can be determined: variation in interventions given to patients, and variation in outcome. If there is large variation in either used interventions or outcome, we call this heterogeneous. If not, the practice can be called homogeneous. When evaluating this variation, homogeneity seems most desirable: patients with the same disease should receive the same interventions, which should result in the same outcome, no matter in which hospital/country the patient is treated. In the next sections, I will discuss that the interpretation of quality of care based on variation is somewhat more complex.

In order to be able to attribute the observed variation to factors which can be improved upon (quality of care or differences in resources), the variation should be adjusted for irrelevant reasons of variation. These reasons are often differences in case-mix and statistical uncertainty [1]. These are irrelevant, because removing differences in case-mix or statistical uncertainty between centres does not directly result in improved care for patients. The effect of case-mix and statistical uncertainty on outcome after stroke can be explored using a shiny app I developed[1].

[1] `https://bgravesteijn.shinyapps.io/app_scbh/`

Case-mix describes the average characteristics of a patient population. Populations from different studies, centres, or countries can be more or less severely injured or frail. These characteristics impact both the decision to intervene (the propensity for an intervention), as well as the outcome (the risk of outcome). When there are differences in case-mix between centres or studies, the variation in interventions or outcome is often larger. Therefore, it is recommended to include characteristics that define case-mix in statistical models to adjust for them. The resulting "adjusted" variation is not attributable to differences in case-mix anymore.

Statistical uncertainty arises when sample sizes per studied group are small: the same mortality rate can be estimated in a sample of patients where 3/5 die, as in a sample of patients where 60/100 die. The latter estimate, however, is much more certain. The problem of statistical uncertainty in different groups can be eloquently solved by the random-effects paradigm [1].

14.1.1 | Variation in interventions

To assess the variation in interventions, we used regression models incorporating random effects on observational data. These regression models are able to adjust for differences in case-mix between populations by adding relevant prognostic factors in the model. The statistical uncertainty is adjusted for by using a random intercept for centre, country, or both (centre conditional on country). Because the dependent variable is the intervention, these models are similar to propensity score models: the predicted probability of the model is the propensity of receiving the intervention. The variation of the random intercept is the main measure of interest, and can be easier interpreted by calculating the median odds ratio (MOR)

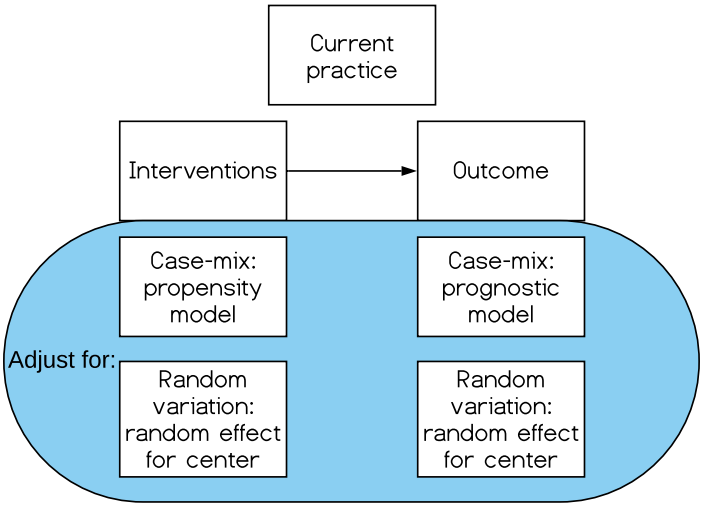


Figure 14.1: *Theoretical framework for assessing current practice, described in section 14.1. Current practice can be described by the interventions given to patients, and the outcome associated with this practice. First, the appropriate analysis in one study is a regression model incorporating a random intercept for the different cohorts within that study (a mixed model). If interventions are assessed, the researcher needs to adjust for the risk of receiving the intervention based on case-mix (a propensity model). If outcome is assessed, the researcher needs to adjust for the risk on outcome (a prognostic model).*

[2, 3]: the expected odds ratio between two randomly picked clusters.

There was substantial heterogeneity in the use of interventions in the field of TBI, where few established effective treatments exist [4, 5]. We found substantial variation -irrespective of patient characteristics- for multiple prehospital interventions for TBI (intubation [6], IV fluids, helicopter, (**chapter 2 and 3**)). It has been described that a major driver of heterogeneity in interventions in this field is low guideline adherence [2]. This low adherence is probably a result of the low quality evidence base for these

guidelines [7]. This thesis again highlights that more evidence is needed to establish effective treatments. This evidence can be used to write sensible and applicable evidence-based guidelines [8, 9]. The premise of improved guidelines is that they can create "order" where there is "disorder" [10]. With these guidelines implemented, clinical practice could be more homogeneous and of better quality, effectively producing better results.

However, heterogeneity in the use of interventions can also be explained by variation in resources. For example in Europe, the median total health care expenditure per capita was $3,428 in 2018, ranging from $1,749 to $5,986 (figure 14.2). Because resources are limited, health care systems need to be pragmatic. Systems with less resources necessarily need to be more pragmatic than others. These differences need to be acknowledged when interpreting heterogeneity in interventions at a large scale, such as the European scale. Not only should observing heterogeneity stimulate higher quality research and implementation of better guidelines, it also should stimulate policy makers to financially support the development of underachieving areas.

14.1.2 | Variation in outcome

Compared to assessing heterogeneity in interventions, assessing heterogeneity in outcome generally provides insight in the overall state of a field. The more evolved a field is, the better the outcome of patients. Better outcomes can be expected if treatment strategies improve, or if better (secondary) prevention decreases overall severity of disease (for example, the introduction of seat belts can reduce the severity of TBI [11]).

In this thesis, variation in outcome was mainly compared between studies using meta-analyses (**chapter 4 and 5**). The included studies are differ-

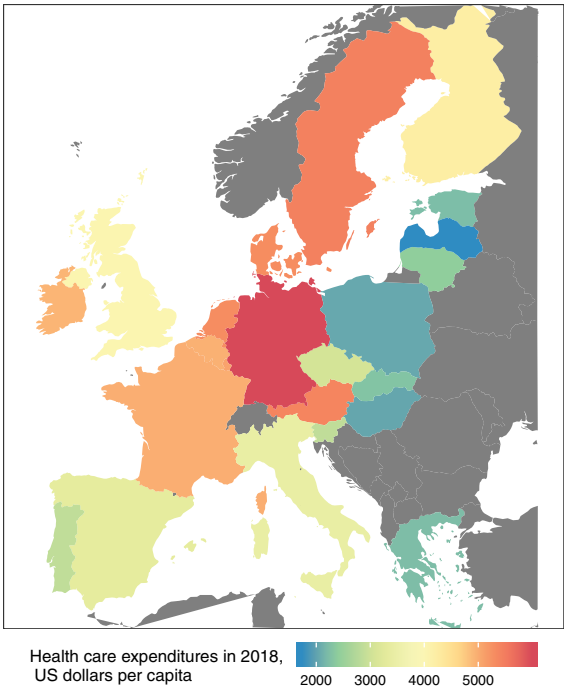


Figure 14.2: *Total health care expenditures in EU countries in 2018, in US dollars per capita (data from* `https://data.oecd.org/healthres/health-spending.html`*)*

ent cohorts from different time periods and countries. We estimated the overall outcome and assessed the variation in outcome. This information provides insight in the overall state of the IHCA field, but it is not directly applicable to all settings. Rather, comparing centres on outcome can result in a more relevant local insight in the quality of care, such as for IHCA in the Netherlands (**chapter 6**). When comparing either centres or studies, there is a multilevel structure in the data. Therefore, the random-effects paradigm can be used to eloquently arrive at an overall estimate.

Our studies in IHCA show that restrictive access to an intervention can also result in homogeneity in outcome (**chapter 5**). In IHCA, we found an overall 1 year survival rate of 13.4%, with significant heterogeneity between studies. In contrast, patients who underwent ECPR have a survival estimate of 31%, with limited heterogeneity between studies. Although their survival could be higher because of ECPR increases survival, they are also a much more selected group: I identified a large list of contra-indications for ECPR in these studies, which all aimed to exclude patients at risk for poor outcome (e.g.: age, excessive bleeding, pre-existing severe neurological damage). Because of this selection, patients are much more similar. The same selection argument may hold for the situation in the Netherlands (**chapter 6**). Our health system excels in defending patient rights, and providing patients with relevant information [12]. In this health care culture, advanced care planning is considered good medical practice [13]. Because of our proactive approach to discuss end-of-life care [14], but probably also because of high-quality care, the overall survival in the Netherlands was estimated at 32%. There was no substantial variation in outcome between centres. To conclude, heterogeneity in survival does not only depend on differences in quality of care, but also depends on selection.

We were not able to address the issue of case-mix in our meta-analysis for outcome after IHCA. To better adjust for differences in case-mix, we would advocate the use of individual patient data meta-analysis to assess current outcome of a disease [15]. Such a study has been performed in the field of TBI [16], albeit the authors could only adjust for age.

To conclude, to assess current practice in terms of outcome, both the main estimate as well as the variation between cohorts should be evaluated. More relevant estimates of overall outcome can be obtained from

comparing centres, rather than by comparing studies. However, the latter might be more feasible. We advocate the use of individual patient data (random effects) meta-analyses to assess the current practice in terms of outcome. When substantial heterogeneity is found, it is important to evaluate whether adequate adjustment for case-mix is performed. Note that adjusting for case-mix is important for assessing both variation in outcome, as well as for variation in interventions. When the variation in outcome is homogeneous, researchers should evaluate whether this is not due to restrictive access to care.

14.1.3 | Linking variation in interventions to outcome

The relationship between variation in interventions and variation in outcome is complex. Two studies in quality of care of stroke in the Netherlands also explored this relationship [17, 18]. Both studies found that variation in outcome was primarily explained by differences in case-mix, rather than processes or structures of care. This was also found in our study in IHCA (**chapter 6**). The differences in outcome after IHCA between centres could also be mostly explained by differences in case-mix.

It is interesting, possibly also discouraging, to find that the quality of care for patients does not seem to explain variation in outcome. Lingsma et al. [17] argue that this can partly be explained by the fact that "treatment effects are generally modest". Moreover, they argue that "not all items of care or treatments apply to all patients and so cannot be expected to have a large impact on aggregated outcomes made up of all patients".

Importantly, these studies were performed in one country (the Netherlands), characterised by a densely connected network of centres. Within our country, protocols are exchanged between centres, facilities are shared,

centres collaborate in networks, and physicians are frequently relocated during their training. Therefore, practices within a country are likely more homogeneous than practices in different countries. Differences in case-mix might therefore be a more important explanation for differences in outcome than differences in quality of care. It would be interesting to repeat such a study in CENTER-TBI [19]. In the field of TBI, practices are known to differ significantly throughout Europe, also resulting in substantial variation in outcome [4, 5, 19–21].

Nevertheless, it is possible to show that improving on relevant process indicators improves patient outcome in stroke [22]. This was possible by analysing a large enough sample size from a large number of centres. This finding underscores the importance of identifying relevant and robust, evidence based process indicators and sensitive outcomes, and act on these to improve quality of care.

One of the improvable structure of care indicators we did identify was frequent CPR training (**chapter 6**). Patients treated in centres where they trained personnel twice per year for CPR had on average better neurological outcome. Similar to the time-to-groin time in stroke [18], this is a relevant quality of care indicator to monitor and improve upon.

Another initiative to develop process indicators in this field is a newly developed quality indicator set for ICU care for TBI patients [23]. At validation in the CENTER-TBI study, nine structure and five processes of care were considered statistically robust indicators, but their relevance (in terms of explaining variation in outcome) remains to be assessed [24].

14.2 | Best practice

The next part of this thesis evaluated what "best practice" would be. This is also known as Comparative Effectiveness Research (CER) [25]: to investigate what already implemented interventions actually improve outcome. CER aims to better understand what existing interventions are beneficial, primarily on a group level. This was one of the primary aims of the CENTER-TBI study [19]. To identify practices that result in better outcome, they have to be compared to practices that result in worse outcomes. Therefore, the field of TBI is seen as ideal for CER, because substantial variation in practices exist (**chapter 2 and 13**) [4, 21, 26].

The initial aim of CER in CENTER-TBI was to use hospital-level treatment preference as instrument to estimate causal effects of treatments on outcome. However, as I will discuss later, complex methodological issues arise when using treatment preference as instrument. Nevertheless, variation in practice does ensure sufficient sample size when different practices are compared: some areas do not intubate patients with GCS scores above 8, and some do. If the research question entails what the effect of intubating patients with a GCS below 8 is compared to intubating patients with a GCS above 8, you need to include patients in areas where they actually intubate patients with a GCS above 8: It is important to have data on the groups which are compared. If practice is homogeneous, it might be harder to compare the practices of interest.

We evaluated for TBI whether being primarily transferred is more effective than being secondarily transferred (**chapter 7**), and whether prehospital and in-hospital intubation are beneficial (**chapter 8**). After correcting for observed confounders, we did not find an overall effect in any of the two interventions. With hindsight, this could have been expected.

The decision whether to intubate or whether to transport patients directly towards a neurosurgical site heavily depends on clinical insight. Intubation should be performed in patients whose airways are compromised, or whose airways are going to be compromised (e.g.: loss of consciousness is expected in patients with internal bleeding) [7]. Direct transfer to a neurosurgical centre is required for patients who actually need neurosurgery [27]. Both these decisions, therefore, heavily rely on clinical expertise and experience. Of course, patients whose tracheas are intubated, or who are directly transported to a neurosurgical centres, have poorer outcome because they are more severely injured. However, after carefully adjusting for these confounding factors, it concurs with this reasoning that there is no overall effect on outcome: it does not make sense to apply these interventions to every patient.

More importantly, specific subgroups which benefit more from these interventions should be identified. For intubation, extracranial injury might indicate benefit of intubation in the prehospital setting, whereas level of consciousness seemed to be more important in the in-hospital setting (**chapter 7**). Moreover, intubating patients with a GCS lower than 10 was associated with better outcome, which conflicts with previous beliefs: intubation used to be recommended in patients with a GCS lower than 8 [7, 28]. For direct transfer to a neurosurgical centre, we explored whether patients who eventually received neurosurgery had better outcome when they were directly transported to a neurosurgical centre (**chapter 8**). This hypothesis was not confirmed by our data, which is also an important finding: it is likely not harmful to first stabilise patients at a non-neurosurgical centre. These analyses are more informative than the overall effect of the intervention.

Nevertheless, subgroup analyses are often unreliable [29]. There are multiple ways of improving reliability of these subgroup effects. We mainly

adhered to recommendations by Sun et al. [30] and Peter Rothwell [31]:

1. Evaluate only subgroup effects with prior evidence of relevance, and use expert clinical input to define relevance;
2. Obtain optimal power, either through sample size or by limiting categorisation of continuous variables (e.g.: continuous GCS instead of categorizing patients into mild/moderate/severe).

We recommend to adhere to these principles in future research with these type of data.

In the context of CER, absolute sample size is often large enough. However, when performing subgroup analyses in the context of CER, another consideration becomes important. Some centres might select patients differently for treatment. For example in case of ECPR, some centres might exclude patients above 70 years, and some above 80 years. It is important to assess whether the subgroups that are compared (e.g. above versus under 80 years) actually exist within the data per centre.

A potentially relevant approach which we did not explore, is to evaluate multiple subgroups simultaneously [32]. In this approach, predictive analytical strategies are used to model heterogeneous treatment effects. Through this novel approach we gain the ability to predict an individual's benefit of treatment: a more statistically efficient way towards individualised medicine.

14.2.1 | Bias

Identifying best practice is a causal question. If practice B is better than practice A, we need evidence to be able to claim that the following statement is true: "if patients were treated by practice B, instead of practice A,

their outcome would have been better" [33]. To be able to claim this, the estimated effect of practice B (versus practice A) on outcome needs to be without random or systematic error (figure 14.3). With large enough data (which was often the case in this thesis), random error is negligible. Systematic error however remains relevant in the context of large datasets. There are various subtypes of systematic errors, often called biases, in observational studies. However, they can be aggregated into three main types of biases [34]:

1. *Information bias* occurs when exposures or outcomes are measured differently in the compared groups. This would have been the case in CENTER-TBI if for example the GOS-E would have been obtained by postal questionnaires in the Netherlands and by in-person visits in Italy.

2. *Selection bias* occurs when different patients are included in the compared groups. This would have been the case in CENTER-TBI if for example some centres included all TBI patients, and other centres only patients with findings on the CT scan.

3. *Confounding bias* occurs when another factor than the exposure or outcome distorts the effect of the intervention or practice on outcome. For example, more severe TBI patients underwent endotracheal intubation. Therefore, the outcome of patients who underwent intubation is worse than patients who did not need this treatment.

Although all types of biases may occur in large observational data, confounding bias is arguably the hardest and most important bias to address. Statistical methods to adjust for confounding bias exist, but they all need

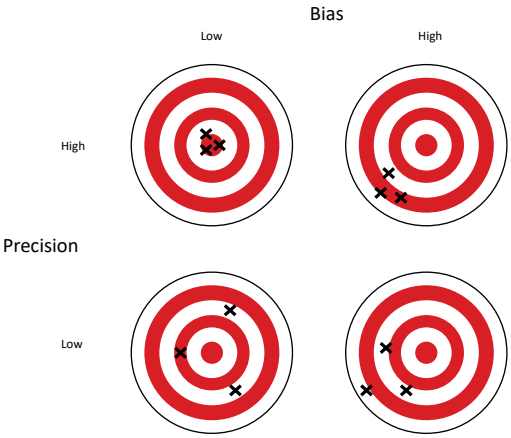


Figure 14.3: *Illustration of the concepts random error (precision) and systematic error (bias), using three arrows and a target.*

extensive subject-specific knowledge to collect the data needed for adjustment. If not all data required is collected, it is impossible to adjust completely for confounding bias. Therefore, some bias will remain, often referred to as residual confounding [35]. Two approaches which in theory avoid any bias, are instrumental variable approach, and directed acyclic graph (DAG) informed regression approaches.

Instrumental variable approach can hypothetically correct for both observed, as well as unobserved confounders [36] (21). In the field of TBI, it has been proposed to use treatment preference as instrumental variable [37]. However, this method requires the researcher to make a set of assumptions which are often hard to defend [38]. We have tried using treat-

ment preference of centres as an instrumental variable to assess the effect of intubation on outcome (**chapter 7**), as previously proposed [37]. However, the assumption of monotonicity [39] is likely violated: it is conceivable that a TBI patient can be intubated when brought to a centre where patients are not often intubated *and* not be intubated if he/she would have been brought to a centre where patients are frequently intubated. If the assumption does not hold, the estimated effect of intubation on outcome would only be applicable to a subset of patients. This subset of patients would be those who would have been intubated when brought to centres where patients are often intubated *and* not have been intubated when brought to centres where patients are not often intubated [40]. However, it is not possible to identify this subgroup of patients, because we only can observe one of the two possibilities: they are either brought to a centre where patients are frequently intubated, or not. This theoretical problem likely holds for many preference-based instrumental variables, and limits their applicability.

Therefore, in this thesis, we mainly used DAG-informed regression approaches. As explained in the introduction (**chapter 1**), directed acyclic graphs (DAGs) are visual representations of the expected (causal) relationships between variables in the study [41]. These DAGs can inform on what factors need to be adjusted for. As an illustration, see figure 14.4, which shows the DAG we propose for intubation in TBI. Pathways between intubation and outcome which are not of interest need to be "blocked" in the main analysis (here: the pathway through injury severity/demographics). A pathway can be blocked by conditioning on that variable, either by stratification or by adjusting in a regression model. Since there is no pathway via in-hospital care after intubation, it does not need to be adjusted for (even though it likely influences outcome). The statistical analysis that

such a DAG informs is often a regression based method. In this thesis, the main approach was to add these variables into a regression model together with the exposure of interest. Others favour to use propensity score based methods [42]. I think that it seems more important to ensure that the appropriate variables are adjusted for than to choose the best regression based method for the analysis.

In reality, it is often not possible to measure the appropriate variable entirely without measurement error. Instead of measuring the actual factor, we often measure proxies of that factor, e.g. number of prescribed drugs as proxy for frailty [43–45]. Even if the assumed causal model is correct, using variables with measurement error might not resolve confounding bias entirely [46, 47].

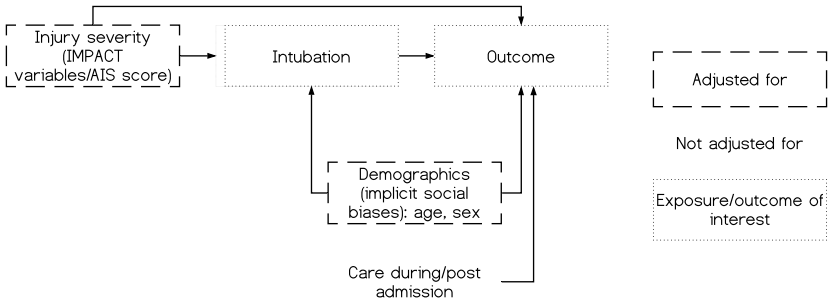


Figure 14.4: *Directed acyclic graph (DAG) framework for intubation in TBI. Factors which need to be adjusted for (injury severity or demographics) are presented, as well as factors which do not need to be adjusted for (care during or post admission) are presented.*

Instrumental variable approach and DAG based regression methods have in common that they require subject-specific knowledge to be used validly. This subject-specific knowledge should inform on what variables should be collected and adjusted for. To evaluate new interventions, ran-

domised controlled trials (RCTs) remain the gold standard. However, when used transparently and with enough expertise, these approaches have the potential to evaluate existing interventions on effectiveness in a real-world setting. This is the ultimate goal of CER.

14.2.2 | Cost-effectiveness

After having established the benefit of an intervention, it remains important to evaluate whether the benefit is worth it. As resources vary substantially (figure 14.2), it is important to weigh the benefit of interventions against the costs. It was already discussed that varying amounts of financial resources is a driver for undesired variation in the use of interventions. If, however, an intervention costs less per extra year spent in perfect health (QALY) than society is willing to pay, policy makers have more reason to globally implement these beneficial interventions.

I evaluated the cost-effectiveness of a costly intervention for IHCA, extracorporeal cardiopulmonary resuscitation (**chapter 9**). This intervention has been increasingly implemented over the past decade [48, 49]. Because of the high costs and the high labour-intensity, this increase in use has been the subject of many debates. Therefore, we investigated whether the benefits outweigh the costs of ECPR, which they apparently do. We found that an extra QALY costs €10,818 (95% CI: €6,347 - €19,400). This is acceptable for a life-saving intervention in the west: in the US, the willingness to pay (WTP) threshold lies between $50,000 - $100,000([50–52]); in the UK, **£**20,000 - **£**30,000 are considered reasonable [53]. In the Netherlands, we are more willing to pay for diseases with disease larger burden [54]. Therefore, our willingness to pay is high for IHCA patients, who have tremendous amount of disease burden.

However, there is one big caveat to address: The beneficial treatment effect has mostly been evaluated in observational studies [55, 56]. We showed that there is a large set of contra-indications (e.g.: metastasised malignant disease) for ECPR (**chapter 5**). Because these contra-indications mainly prevent patients with poor outcome from undergoing ECPR, the benefit is likely exaggerated due to selection bias. Nevertheless, assuming this evidence is correct, the intervention seems very cost-effective, concurring with other cost-effectiveness studies [57, 58]. A more careful interpretation of all these findings is that this is a best-case scenario estimate. We would however still recommend implementing ECPR, because even a lower cost-effectiveness is likely reasonable: if the ICER is 100% underestimated (€21,636 per QALY instead of €10,818 per QALY), it would be comparable to other procedures which are considered cost effective (vascular surgery: €16,500; interferon for skin cancer €15,000; liver transplantation: €19,000) [59]. Moreover, IHCA has a high disease burden. Since the willingness to pay per QALY scales with disease burden, the willingness to pay is likely much higher than our estimated costs per QALY.

This cost-effectiveness study confirms earlier monocentre observational studies [57, 58]. Because of the stacked evidence in favour of cost-effectiveness, the increase in use of this labour-intensive treatment over the last decades can now be seen as advantageous. More and more patients have enjoyed the benefits of ECPR. Nevertheless, we show in **chapter 5** that the treatment is still only used in a subset of patients. Patients with a high risk of poor outcome are excluded for treatment, contributing to a high survival rate (30% at 1-year [60], versus 13.4 % [61] for conventional CPR). The experience of physicians with ECPR increases, and therefore it is likely that the exclusion criteria will become less stringent over time. Within that context, we should monitor and ensure acceptable cost-effectiveness and functional

outcome of this costly and labour-intensive treatment.

14.3 | Identifying patients at risk

The last part of the thesis explored ways to identify patients who are at risk. This is especially useful for acute care, because it would enable triaging of patients, thereby optimising workflow and optimally distributing attention and resources. The higher the risk on poor outcome, the more invasive and aggressive treatments are generally needed and warranted. This research question was mainly explored in TBI.

The first approach we explored was to identify clusters of patients with similar baseline characteristics (**chapter 10**). Based on known predictors of outcome (not the outcome itself), we clustered TBI patients in 4 clusters. The characteristics that mostly define these clusters were GCS, major extracranial injury (MEI), and mechanism of injury. Unfortunately, these clusters remain hard to use in clinical practice. Although the average outcome differed substantially between these clusters, the range of plausible outcomes in each cluster was broad. A new individual might be roughly classified to one of these clusters, but prediction of outcome based on these clusters remains unreliable. However, these characteristics did separate patients with different care pathways: intracranial surgery, longer length of ICU stay, and ICP monitoring, was to be expected when some of these characteristics were present. Therefore, these characteristics can be building blocks of a classification that capture aetiology. This was exactly what was recommended by Saatman et al [62], as a way forward to better characterise TBI.

Rather, we explored ways to improve development of prediction mod-

els. One way forward that has been proposed is using machine learning algorithms instead of classical statistics (regression). A systematic review, however, found no difference in discriminative performance [63]. Moreover, machine learning algorithms are known to require more data to perform optimally [64]. The calibration of regression and machine learning algorithms might be similar, and deteriorates similarly over time [65]. Large-scale comparisons of performance however focus on discrimination, instead of calibration [66, 67]. Our study compared various algorithms. Similar performance between algorithms was found across a large number of studies from different time periods. However, the variation in discrimination and calibration between studies was substantial. (**chapter 11**). This result underscores the importance of validation [68]. For reliable prediction, ensuring validity in the population in which the model is used is more important than the choice of algorithm.

In the introduction, I hypothesised that the more flexible an algorithm is (measured partly by the amount of parameters), the higher its performance in highly complex, high volume data. Even though the clinical data used in this thesis is highly voluminous, the complexity is apparently limited. It can be argued that most natural relationships in clinical data are relatively simple, meaning additive and linear associations, which can easily be captured by regression methods. The added benefit of using regression techniques, is that they fit into an epidemiological framework that has learned to cope with common problems such as incomplete data, measurement error, and confounding bias [69]. These problems occur in many epidemiological studies. Therefore it is crucial to be able to address these problems. In case any of these problems apply to data used in clinical research, I would therefore advice to use more traditional statistical approaches to address these problems.

It is worth noting that the applications of machine learning algorithms are very similar (perhaps equal) to the regression framework, although they are primary used by researchers from different backgrounds. Machine learning algorithms are more often used by researchers who identify as engineers or data scientists, whereas regression is more often used by researchers who identify as statisticians or epidemiologists. As a result, different terminology is used by researchers who use similar approaches for similar research questions, hampering scientific cross- contamination.

Another important way to improve prediction is by including strong, new predictors. We investigated the added value of biomarkers to commonly used decision rules for prediction of a positive CT scan in TBI patients (**chapter 12**). The most promising biomarker was GFAP. This biomarker consistently showed to add new predictive information to clinical characteristics which are already known to predict a positive CT scan. Moreover, this result was consistent in a large number of sensitivity analysis, including an analysis in mild TBI patients. This finding has major impact on the future treatment of mild TBI patients: a point-of-care GFAP test will most likely be used together with clinical decision rules to send patients home without the need for a CT scan [70–72]. However, the replacement of S100B in the guidelines is still premature, since the assay reliability of GFAP needs to be improved first. Nevertheless, inclusion of this biomarker in clinical decision rules has the potential to lower costs and decreases exposure to radiation [73].

Finally, we propose a 5-step approach to deal with missing data, an ubiquitous problem in medical research (**chapter 13**):

1. explore the missing data patterns;
2. choose a method of imputation;

3. perform imputation;

4. assess diagnostics of the imputation;

5. analyse the imputed datasets.

By following these steps, and by reflecting on the considerations discussed in **chapter 13**, future development of prediction models might be more robust. Later, the STRATOS initiative[2] also suggested a framework for handling missing data in observational studies in general [74]. The most important difference is that during the first step of their framework is answering the question whether complete case analysis is likely valid. As discussed in our paper, complete case analysis is rarely valid in the case of prediction modelling.

Some methodological hurdles still exist which need to be addressed. Most importantly:

- evidence is needed describing how imputing the outcome influences external validity;

- more guidance is needed when average or single imputation can be relevant alternatives to multiple imputation;

- a framework to deal with missing data for machine learning needs to be developed.

[2]STRengthening Analytical Thinking for Observational Studies, see www.https://stratos-initiative.org/

14.4 | Exploiting large observational data

In the discussion above, multiple opportunities and pitfalls in using large observational datasets were described, compared to interventional studies. I will summarise them here, and add some more general reflections.

- Opportunities:
  - Generalisability; Observational studies often have limited exclusion criteria, thereby ensuring the findings in these studies to be applicable for the complete population of interest. Moreover, because they are collected in multiple locations, estimates from these data are not sensitive to centre-specific contexts.
  - Sample size is often large, which ensures high precision. However, perhaps even more interesting is the statistical power to detect heterogeneity in effects. I do want to underscore the importance of having a strong hypothesis prior to analysing the subgroups of interest.
  - Feasibility is higher than for randomised controlled trials, both from an ethical as financial perspective. Often the largest ethical hurdle to overcome when setting up these studies are the privacy issues associated with collecting patient data. This hurdle, however, is inherent to all medical research.
- Pitfalls:
  - Bias; The precision in estimates, however, does not prevent researchers from drawing wrong conclusion about effects. Causality, as discussed above, is possible to ascertain in observational

data. However, extensive subject-specific knowledge about the problem at hand is required to ensure a valid DAG-informed regression approach. Being able to shoot consistently in a particular area in a shooting target (precision), does not help the researcher if this area is at the edge of the target (biased, figure 14.3).

- Needs to be tailored to the research question: whereas randomised controlled trials often answer one main research question, large observational data often answer multiple. Experience from CENTER-TBI teaches us that if research questions are not formulated specifically enough, local investigators are required to report a large number of data points. This lead to lower data quality, and investigators were required to retrieve their data in a framework which was cumbersome and oaf. As a result, a substantial part of the analysis took place before actually starting the statistical analysis. This part of the analysis is often referred to as data preparation or cleaning. However, decisions are being made in this process that impact the actual analysis (E.g.: should I combine these two categories? Should I use the central or local reading of the CT scans?). Therefore, the term initial data analysis has been coined to describe this process [75]. It requires transparent reporting, to ensure reproducible analyses. The main approach in this thesis for the reporting of the initial data analysis was by sharing the code.

14.5 | Recommendations

Since this thesis explored three questions in two diseases, a wide range of recommendations can be formulated for future research, as well as clinical care. Here, I will formulate the three most important recommendations for IHCA and TBI, and for the overall field of acute care research (table 14.1, 14.2, and 14.3).

Table 14.1: *The main recommendations for the treatment and research of IHCA.*

1. Improving treatment for IHCA should focus on improving neurological outcome rather than mortality: only the former might likely be improved upon.
2. Hospitals should focus on preventing IHCA, for example via higher implementation of early warning scores.
3. Widespread implementation of ECPR is defendable, but the high proportion of favourable outcome should be maintained when the indication for ECPR extends to a more fragile population.

Table 14.2: *The main recommendations for the treatment and research of TBI.*

1. To improve the decision to justifiably discharge mild TBI patients from the ED without CT scan, improve GFAP assay reliability and include this biomarker in CT decision rules.
2. The use of prehospital interventions in TBI patients, as well as the outcome of these patients, are heterogeneous in Europe. Next to establishing effective interventions, variation in financial resources might also contribute to variation in outcome: effective investments in prehospital care throughout Europe might be warranted.
3. Previously, it was thought that the benefit of intubation in TBI patients only outweighs the harm in patients with a GCS lower or equal to 8, but intubation might already be considered when GCS drops below 10.

Table 14.3: *The main recommendations for improving clinical decision making in acute care using observational data.*

1. Assessing quality of care should take into account differences in case-mix and statistical uncertainty, but should be interpreted in the context of patient selection: a more selected group of patients might result in a better and homogeneous outcome overall, but this is not directly attributable to higher quality of care.
2. It might be possible to assess causal questions in observational data, but this requires extensive subject-specific knowledge about the problem at hand.
3. Predictive research in this field should focus on external validation of predictive algorithms and adding new, strong predictors, instead of exploring to fit the model with a different type of algorithm.

14.6 | Future research

As can be read in this methodological and clinically oriented thesis, more questions than answers arise. Going forward, I would like to keep combining descriptive, causal, and predictive questions. By trying to answering these questions, hopefully some answers will arise to improve selection and treatment of patients. Moreover, I will keep focusing on the economical or feasibility aspect of medicine. The ultimate goal is to contribute to a more effective, efficient, and evidence-based health care that creates the most health for the overall population.

14.7 | Overall summary

This thesis aimed to explore ways to translate data from large observational studies in acute medicine to knowledge. Recommendations pertaining to clinical practice as well as research were discussed. The largest opportunities described in this thesis for large observational studies in acute care lie in describing current practices to identify areas of improvement, and in prediction research by developing reliable prediction models with strong predictors. Large observational data can be used to address the question what interventions are beneficial, but this requires extensive subject-specific knowledge feeding into valid DAG-informed regression models. Acute medicine is a field where decisions need to be made under great uncertainty. I feel confident that this thesis has contributed to address how to best describe where the most uncertainty is, what interventions to use, and who to prioritise.

References

1. Lingsma, H. F. *et al.* Comparing and ranking hospitals based on outcome: Results from The Netherlands Stroke Survey. *Quarterly Journal of Medicine* **103,** 99–108. ISSN: 14602725 (2009).
2. Cnossen, M. C. *et al.* Adherence to Guidelines in Adult Patients with Traumatic Brain Injury: A Living Systematic Review. *Journal of Neurotrauma* **14.** ISSN: 0897-7151 (2016).
3. Merlo, J. *et al.* A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *Journal of Epidemiol Community Health* **60,** 290–297 (2006).
4. Cnossen, M. C. *et al.* Prehospital trauma care among 68 European neurotrauma centers: Results of the CENTER-TBI Provider Profiling Questionnaires. *Journal of neurotrauma* **36,** 176–181 (2019).
5. Lingsma, H. F. *et al.* Large Between-Center Differences in Outcome After Moderate and Severe Traumatic Brain Injury in the International Mission on Prognosis and Clinical Trial Design in Traumatic Brain Injury (IMPACT) Study. *Neurosurgery* **68,** 601–608. ISSN: 0148-396X (2011).
6. Franschman, G *et al.* Prehospital endotracheal intubation in patients with severe traumatic brain injury: Guidelines versus reality. *Resuscitation* **80,** 1147–1151. ISSN: 0300-9572 (2009).
7. Badjatia, N. *et al.* Guidelines for Prehospital Management of Traumatic Brain Injury 2nd Edition. *Prehospital Emergency Care* **12,** S1–S52. ISSN: 1090-3127 (2008).
8. Bal, J. *Mediating Uncertainties in Times of a Pandemic* MA thesis (Maastricht, 2020).
9. Berg, M. & Mol, A. *Differences in medicine: Unraveling practices, techniques, and bodies* (Duke University Press, 1998).

10. Berg, M. Order (s) and disorder (s): of protocols and medical practices. *Differences in medicine: Unraveling practices, techniques, and bodies,* 226–46 (1998).

11. Kim, J.-M. *et al.* Preventive effects of seat belts on traumatic brain injury in motor vehicle collisions classified by crash severities and collision directions. *European Journal of Trauma and Emergency Surgery.* ISSN: 1863-9941. `https://doi.org/10.1007/s00068-019-01095-4` (2019).

12. Björnberg, A. & Phang, A. Y. *Euro Health Consumer Index 2018 Report* ISBN: 978-91-980687-5-7 (2018).

13. Rietjens, J. A. C. *et al.* Definition and recommendations for advance care planning: an international consensus supported by the European Association for Palliative Care. *The Lancet. Oncology* **18,** e543–e551. ISSN: 1474-5488 (2017).

14. Schluep, M., Van Limpt, G. J. C., Stolker, R. J., Hoeks, S. E. & Endeman, H. Cardiopulmonary resuscitation practices in the Netherlands: Results from a nationwide survey. *BioMed Central: Health Services Research* **19.** ISSN: 14726963 (2019).

15. Riley, R. D., Lambert, P. C. & Abo-Zaid, G. Meta-analysis of individual participant data: Rationale, conduct, and reporting. *British Medical Journal (Online)* **340,** 521–525. ISSN: 17561833 (2010).

16. Majdan, M. *et al.* Epidemiology of traumatic brain injuries in Europe: a cross-sectional analysis. *The Lancet Public Health* **1,** e76–e83. ISSN: 2468-2667 (2016).

17. Lingsma, H. *et al.* Netherlands stroke survey investigators. Variation between hospitals in patient outcome after stroke is only partly explained by differences in quality of care: results from the Netherlands stroke survey. *J Neurol Neurosurg Psychiatry* **79,** 888–94 (2008).

18. Amini, M. *et al.* Improving quality of stroke care through benchmarking center performance: why focusing on outcomes is not enough. *BioMed Central: health services research* **20,** 1–10 (2020).

19. Maas, A. I. *et al.* Collaborative European neurotrauma effectiveness research in traumatic brain injury (CENTER-TBI): A prospective longitudinal observational study. *Neurosurgery* **76,** 67–80. ISSN: 15244040 (2015).

20. Maas, A. I. R. *et al.* Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. *The Lancet Neurology* **4422.** ISSN: 14744422 (2017).

21. Cnossen, M. C. *et al.* Variation in structure and process of care in traumatic brain injury: Provider profiles of European Neurotrauma Centers participating in the CENTER-TBI study. *PLoS ONE* **11,** 1–21. ISSN: 19326203 (2016).

22. Cadilhac, D. A. *et al.* Quality of Acute Care and Long-Term Quality of Life and Survival: The Australian Stroke Clinical Registry. *Stroke.* ISSN: 15244628 (2017).

23. Huijben, J. A. *et al.* Development of a quality indicator set to measure and improve quality of ICU care for patients with traumatic brain injury. *Critical Care* **23.** ISSN: 1466609X (2019).

24. Huijben, J. A. *et al.* Quality indicators for patients with traumatic brain injury in European intensive care units: A CENTER-TBI study. *Critical Care* **24.** ISSN: 1466609X (2020).

25. Of Medicine, I. *Initial National Priorities for Comparative Effectiveness Research* ISBN: 978-0-309-13836-9 (National Academies Press, Washington, D.C., 2009).

26. Lingsma, H. *et al.* Prognosis in moderate and severe traumatic brain injury. *Journal of Trauma and Acute Care Surgery* **74,** 639–646. ISSN: 2163-0755 (2013).

27. DuBose, J. J. *et al.* Effect of trauma center designation on outcome in patients with severe traumatic brain injury. *Archives of surgery* **143,** 1213–1217 (2008).

28. Bernard, S. A. *et al.* Prehospital Rapid Sequence Intubation Improves Functional Outcome for Patients With Severe Traumatic Brain Injury. *Annals of Surgery* **252,** 959–965. ISSN: 0003-4932 (2010).

29. Brookes, S. T. *et al. Subgroup analysis in randomised controlled trials: quantifying the risks of false-positives and false-negatives* 2001.

30. Sun, X., Ioannidis, J. P., Agoritsas, T., Alba, A. C. & Guyatt, G. How to use a subgroup analysis: users' guide to the medical literature. *Journal of the American Medical Association* **311,** 405–411 (2014).

31. Rothwell, P. M. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *The Lancet* **365,** 176–186 (2005).

32. Kent, D. M. *et al.* The Predictive Approaches to Treatment effect Heterogeneity (PATH) statement. *Annals of Internal Medicine* **172,** 35–45. ISSN: 15393704 (2020).

33. Pearl, J. & Mackenzie, D. *The book of why: the new science of cause and effect* ISBN: 0465097618 (Basic Books, 2018).

34. Grimes, D. A. & Schulz, K. F. Bias and causal associations in observational research. *The Lancet* **359,** 248–252 (2002).

35. Becher, H. The concept of residual confounding in regression models and some applications. *Statistics in medicine* **11,** 1747–1758 (1992).

36. Swanson, S. A. & Hernán, M. A. Think globally, act globally: an epidemiologist's perspective on instrumental variable estimation. *Statistical science: a review journal of the Institute of Mathematical Statistics* **29,** 371 (2014).

37. Cnossen, M. C. *et al.* Adjusting for confounding by indication in observational studies: a case study in traumatic brain injury. *Clinical epidemiology* **10,** 841–852. ISSN: 1179-1349 (2018).

38. Swanson, S. A. & Hernán, M. A. Commentary: How to Report Instrumental Variable Analyses (Suggestions Welcome). *Epidemiology* **24,** 370–374 (2013).

39. Lousdal, M. L. An introduction to instrumental variable assumptions, validation and estimation. *Emerging themes in epidemiology* **15,** 1. ISSN: 1742-7622 (2018).

40. Swanson, S. A., Miller, M., Robins, J. M. & Hernán, M. A. Definition and evaluation of the monotonicity condition for preference-based instruments. *Epidemiology (Cambridge, Mass.)* **26,** 414 (2015).

41. VanderWeele, T. J., Hernán, M. A. & Robins, J. M. Causal directed acyclic graphs and the direction of unmeasured confounding bias. *Epidemiology* **19,** 720–728. ISSN: 10443983 (2008).

42. ROSENBAUM, P. R. & RUBIN, D. B. The central role of the propensity score in observational studies for causal effects. *Biometrika* **70,** 41–55. ISSN: 0006-3444 (1983).

43. Jamsen, K. M. *et al.* Effects of changes in number of medications and drug burden index exposure on transitions between frailty states and death: the concord health and ageing in men project cohort study. *Journal of the American Geriatrics Society* **64,** 89–95 (2016).

44. Bonaga, B. *et al.* Frailty, polypharmacy, and health outcomes in older adults: the frailty and dependence in albacete study. *Journal of the American Medical Directors Association* **19,** 46–52 (2018).

45. Veronese, N. *et al.* Polypharmacy Is Associated With Higher Frailty Risk in Older People: An 8-Year Longitudinal Cohort Study. *Journal of the American Medical Directors Association* **18,** 624–628. ISSN: 1525-8610. `https://www.sciencedirect.com/science/article/pii/S1525861017301044` (2017).

46. Hernán, M. A. & Cole, S. R. Invited commentary: causal diagrams and measurement bias. *American journal of epidemiology* **170,** 959–962 (2009).

47. Pearl, J. On measurement bias in causal inference. *arXiv preprint arXiv:1203.3504* (2012).

48. Karagiannidis, C. *et al.* Extracorporeal membrane oxygenation: evolving epidemiology and mortality. *Intensive Care Medicine* **42,** 889–896. ISSN: 0342-4642 (2016).

49. Extracorporeal life support organisation. *ECLS registry report* tech. rep. (2019).

50. Kaplan, R. M. & Bush, J. W. Health-related quality of life measurement for evaluation research and policy analysis. *Health psychology* **1,** 61 (1982).

51. Ubel, P. A., Hirth, R. A., Chernew, M. E. & Fendrick, A. M. What is the price of life and why doesn't it increase at the rate of inflation? *Archives of internal medicine* **163,** 1637–1641 (2003).

52. Shiroiwa, T. *et al.* International survey on willingness-to-pay (WTP) for one additional QALY gained: what is the threshold of cost effectiveness? *Health economics* **19,** 422–437 (2010).

53. National Institute for Clinical Excellence and others. Guide to the methods of technology appraisal (2008).

54. Zorginstituut Nederland. Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg (2016).

55. Chen, Y. S. *et al.* Comparison of outcome of extracorporeal cardiopulmonary resuscitation for out-of-hospital and in-hospital cardiac arrest. *Circulation* **128** (2013).

56. Ouweneel, D. M. *et al.* Extracorporeal life support during cardiac arrest and cardiogenic shock: a systematic review and meta-analysis. *Intensive Care Medicine* **42,** 1922–1934. ISSN: 0342-4642 (2016).

57. Bharmal, M. I. *et al.* Cost-utility of extracorporeal cardiopulmonary resuscitation in patients with cardiac arrest. *Resuscitation* **136,** 126–130 (2019).

58. Dennis, M. *et al.* Cost effectiveness and quality of life analysis of extracorporeal cardiopulmonary resuscitation (ECPR) for refractory cardiac arrest. *Resuscitation* **139,** 49–56. ISSN: 18731570 (2019).

59. RVS. *Zinnige en duurzame zorg* `https://www.raadrvs.nl/documenten/publicaties/2006/06/07/zinnige-en-duurzame-zorg` (2021).

60. Gravesteijn, B. Y. *et al.* Neurological outcome after extracorporeal cardiopulmonary resuscitation for in-hospital cardiac arrest: a systematic review and meta-analysis. *Critical Care* **24,** 1–12 (2020).

61. Schluep, M., Gravesteijn, B. Y., Stolker, R. J., Endeman, H. & Hoeks, S. E. One-year survival after in-hospital cardiac arrest: A systematic review and meta-analysis. *Resuscitation* **132,** 90–100. ISSN: 1873-1570 (2018).

62. Saatman, K. E. *et al.* Classification of traumatic brain injury for targeted therapies. *Journal of neurotrauma* **25,** 719–738 (2008).

63. Christodoulou Evangelia *et al.* A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of Clinical Epidemiology* **0.** ISSN: 08954356 (2019).

64. Van der Ploeg, T., Austin, P. C. & Steyerberg, E. W. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. *BioMed Central: medical research methodology* **14,** 1–13 (2014).

65. Davis, S. E., Lasko, T. A., Chen, G., Siew, E. D. & Matheny, M. E. Calibration drift in regression and machine learning models for acute kidney injury. *Journal of the American Medical Informatics Association* **24,** 1052–1061 (2017).

66. Christodoulou, E. *et al.* A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of clinical epidemiology* **110,** 12–22 (2019).

67. Fleuren, L. M. *et al.* Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive care medicine* **46,** 383–400 (2020).

68. Steyerberg, E. W. & Harrell, F. E. Prediction models need appropriate internal, internal-external, and external validation HHS Public Access. *Journal of Clinical Epidemiology* **69,** 245–247 (2016).

69. Van Smeden, M., Penning de Vries, B. B., Nab, L. & Groenwold, R. H. Approaches to Addressing Missing Values, Measurement Error and Confounding in Epidemiologic Studies. *Journal of Clinical Epidemiology* **0.** ISSN: 08954356 (2020).

70. Papa, L. *et al.* GFAP out-performs S100β in detecting traumatic intracranial lesions on computed tomography in trauma patients with mild traumatic brain injury and those with extracranial lesions. *Journal of neurotrauma* **31,** 1815–1822 (2014).

71. Mondello, S. *et al.* Clinical utility of serum levels of ubiquitin C-terminal hydrolase as a biomarker for severe traumatic brain injury. *Neurosurgery* **70,** 666 (2012).

72. Czeiter, E. *et al.* Blood biomarkers on admission in acute traumatic brain injury: Relations to severity, CT findings and care path in the CENTER-TBI study. *EBioMedicine* **56,** 102785 (2020).

73. Foks, K. A. *et al.* External validation of computed tomography decision rules for minor head injury: prospective, multicentre cohort study in the Netherlands. *British Medical Journal* **362** (2018).

74. Lee, K. J. *et al.* Framework for the treatment and reporting of missing data in observational studies: The Treatment And Reporting of Missing data in Observational Studies framework. *Journal of clinical epidemiology* **134,** 79–88 (2021).

75. Huebner, M., Vach, W. & le Cessie, S. A Systematic Approach to Initial Data Analysis Is Good Research Practice. *The Journal of thoracic and cardiovascular surgery* **151,** 25–27 (2016).

15

Summary

Acute medicine is a distinct field of medicine, which is characterized by swift medical decision making under great uncertainty with large impact on the survival and quality of life of patients. This thesis studies two acute diseases and applies and compares modern approaches to learn from observational data.

The first studied disease is traumatic brain injury (TBI), which results in a large burden of disease worldwide. TBI is often referred to as "the most complex disease, in the most complex organ", and is very heterogeneous. This heterogeneity is currently insufficiently captured in the most used classification system, based on Glasgow Coma Scale (GCS). We need better characterization, classification, and prediction models to understand which patients require what intervention. Partly due to this lack of sufficient characterization, the evidence base underlying many of the interventions for TBI is thin. As a consequence, self-reported non-adherence to guidelines is low, and variation in outcome is high. The variation in interventions, and especially in the prehospital setting, is less well studied.

Similarly, the effectiveness of prehospital interventions such as intubation and direct transfer to a specialist neurotrauma care (SNC) centre is not established.

The second studied disease is in-hospital cardiac arrest (IHCA), which is a dramatic adverse event during hospitalization with major impact on short-term survival. However, the long-term outcome is poorly described compared to out-of-hospital cardiac arrest. To improve survival and potentially quality of life as well, extracorporeal cardiopulmonary resuscitation (ECPR) using extracorporeal membrane oxygenation (ECMO) has been increasingly used to temporarily take over cardiac and pulmonary functions. This costly and labor intensive intervention has shown effective in observational studies, but the cost-effectiveness still needs to be established.

1. Given that current practice is described by given interventions and their respective outcome,

    - what is the variation in (prehospital) interventions for TBI, and
    - what is the expected long-term outcome after IHCA, and how much variation exists? – a descriptive question.

2. What is the best practice, or what is the (cost-)effectiveness of currently performed interventions for TBI and IHCA? — a causal question.

3. How to better characterise and predict outcome in TBI? — a predictive question.

15.1 | Part I – Current practice

In **chapter 2 & 3**, I studied the prehospital management of Traumatic Brain Injury across Europe in the CENTER-TBI study. There is substantial variation in the prehospital care across Europe, regardless of injury severity. Although prehospital management of TBI aims to prevent hypotension and hypoxia, the prevalence of these secondary insults is still 13% and 7% in severe TBI, respectively. The main determinant for secondary insults was extracranial injury. In **chapter III**, I also performed an in-depth analysis of intubation practice across Europe, where GCS was found to be the main driver of the decision to intubate. Nevertheless, there was significant variability in intubation practice between countries: the decision to intubate was statistically equally dependent on the geographical location in Europe as on the effect of unreactive pupils.

In **chapter 4 & 5**, I studied the long-term survival after cardiac arrest. In the two chapters, I assessed the overall survival after IHCA in general and the neurological outcome after ECPR, respectively. The overall 1-year survival after IHCA is low (13.4%), but similar to the survival at discharge. Moreover, a modest trend of increasing survival over the period 1985 - 2018 was found. Compared to the general IHCA population, the patients receiving ECPR are more selected. The survival after ECPR was more homogeneous, and higher (30%). The neurological outcome in survivors was also excellent on average (84% recovered to favourable outcome).

In **chapter 6**, I compared structure of care, processes of care, and outcome between hospitals for in-hospital cardiac arrest in the Netherlands. We found moderate variation in outcome, which was for the most part explained by differences in case-mix. Only for neurological outcome some unexplained variation existed, but the rankability of this outcome indica-

tor was low. There was moderate variation in structure and processes of care, and for most of them no evidence was found that they affect neurological outcome. Only for CPR training twice per year, a positive effect on neurological outcome was found.

15.2 | Part II – Best practice

In **chapter 7**, I studied the effectiveness of intubation on outcome in TBI. An association was found between higher intubation rates per system and better functional outcome. Moreover, the results of the Australian trial by Bernard et al. (showing benefit of prehospital intubation over in-hospital intubation in patients with a GCS below 10) could not be replicated to the European setting. Finally, there was no evidence that prehospital or in-hospital intubation affects functional outcome overall. However, in patients with more severe injury, intubation was associated with better functional outcome: in patients with lower GCS scores, in-hospital intubation was associated with better functional outcome, and in patients with more severe abdominal or thoracal injury, prehospital intubation was associated with better outcome.

In **chapter 8**, direct transfer of moderate to severe TBI patients to a specialist neurotrauma care (SNC) centre was evaluated. There was substantial variation in the transfer of TBI patients to SNC centres, unexplained by patient characteristics. No association between secondary referral and clinical long term outcomes was found, even though patients who were referred had more serious CT abnormalities. With current triage strategies, it can be considered safe to first stabilize patients at centres without SNC.

In **chapter 9**, a cost-effectiveness analysis for ECPR was presented. Ag-

gregating best available evidence with a decision model and Markov state transition model, we compared the scenarios where ECPR was used in no patient, in all patients, and in patients with an age-combined Charlson comorbidity (ACCI) score below a threshold. We found that the most likely cost-effective intervention given a willingness to pay between €50.000 and €100.000, was to use ECPR in all patients. The incremental cost-effectiveness ratio for this scenario was of 10.818 €/QALY.

15.3 | Part III – Identyifying patients at risk

In **chapter 10**, an exploratory analysis towards a new multidimensional classification of traumatic brain injury was performed. Using unsupervised clustering methods with clinical characteristics, we clustered the entire span of TBI into 4 subgroups. The main determinants of these clusters were injury mechanism, presence of major extracranial injury, and GCS. Except for GCS, these factors are not prognostically relevant, but they are potentially relevant to capture differences in etiology and care pathways.

In **chapter 11**, I compared various machine learning algorithms and regression methods for prognostication in traumatic brain injury in the IMPACT database. After optimising all algorithms based on a metric which is important for discrimination and calibration (log-likelihood), we performed extensive internal-external cross validation in all studies included in the IMPACT database. Similar performance was found across a large number of studies from different time periods. However, the variation in discrimination and calibration between studies was substantial. Therefore, it is more beneficial to ensure optimal validity in the population were predictive algorithms are applied, than to use a specific algorithm.

In **chapter 12**, another strategy for increasing predictive performance was assessed in TBI: the prediction of intracranial findings on the CT scan by including blood biomarkers to existing decision rules. I analysed six serum biomarkers (S100B, NSE, GFAP, UCH-L1, NFL, and t-tau), and found that GFAP consistently improved discrimination of the decision rules. Unfortunately, agreement between replicates of biomarker assessments was poor for the research-use only assay for GFAP, and therefore needs to be improved before GFAP can be used in clinical practice.

In **chapter 13**, I propose a 5-step approach for handling missing data in predictive research. The five steps are

1. explore the missing data pattern;
2. chose a method of imputation;
3. perform imputation;
4. assess diagnostics of the imputation;
5. analyse the imputed datasets.

Although the tutorial is mainly focused on multiple imputation, we also discuss when other imputation methods can be considered or even required.

15.4 | General discussion

Studying current clinical practice might enable the identification of areas which need improvement. Current practice can be studied by assessing variation in interventions and their consecutive outcome. This is ideally

done within the random effects framework. Substantial variation, or heterogeneity, in the administration of interventions was found for prehospital interventions for TBI. This situation is not ideal, since similar patients should receive similar (effective) interventions. This variation is either attributable to the lack of evidence resulting in varying national guidelines for prehospital care, or differences in context and resources for prehospital care. We mainly assessed outcome in in-hospital cardiac arrest, where both the main estimate as the variation around this estimate was important to study current practice. We found that not only quality of care improves outcome, but also a selection of low-risk patients.

The assessment of best practice should focus on identifying effective treatments and to establish their cost-effectiveness. The effectiveness of interventions can be investigated in observational data, but extensive prior subject-specific knowledge is necessary. This knowledge is necessary to inform data collection and design analysis which adjust for specific biases. The technique mostly used in this thesis is adjustment by fulfilling the back-door criterion. However, other types of adjustments can be considered given the assumed causal model underlying the data. After establishing effectiveness, the clinical relevance and usefulness of interventions should be assessed in cost-effectiveness analyses. To identify patients at risk, clustering methods are not the ideal methods to use. They can provide insight into specific clinical phenotypes, but they do not efficiently contribute to outcome prognostication. Rather, prediction models should be developed to model prognosis of patients. When clinical data is used, the often associated limitations and the moderate degree of complexity currently warrant the use of traditional statistical techniques over machine learning algorithms. It can be useful to include new strong predictors, for example adding GFAP to predict intracranial abnormalities in TBI. Finally,

it is recommended to adhere to the 5-step approach for missing data in predictive research.

Finally, the opportunities of observational data include the generalizability, ethical and financial feasibility and power. However, to establish causality requires extensive prior subject-specific knowledge and the data need to be tailored to your question.

16

Samenvatting

Acute geneeskunde is een specifieke vorm van geneeskunde, die wordt gekarakteriseerd door beslisvorming onder tijdsdruk met grote onzekerheid en grote impact op de overleving en kwaliteit van leven van patiënten. Dit proefschrift bestudeert twee acute ziektes en gebruikt en vergelijkt nieuwe methoden om te leren van observationele data. Data die informatie bevat over de klinische praktijk zoals zij is, zonder erin in te grijpen of iets te veranderen.

De eerste onderzochte ziekte is traumatisch hersenletsel, wat een belangrijke oorzaak is van ziektelast wereldwijd. De ziekte wordt beschreven als "de meest complexe ziekte in het meest complexe orgaan" en is daarom ook erg heterogeen: traumatisch hersenletsel is een parapluterm voor een grote verscheidenheid aan verschillende soorten hersenletsel. Deze heterogeniteit wordt echter slecht samengevat in de meest gebruikte classificatie, gebaseerd op een schaal die iets zegt over hoe wakker iemand is (de Glasgow Coma Scale, GCS). We moeten de ziekte beter karakteriseren, classificeren of de uitkomsten van deze patiënten beter kunnen voorspellen, om

te kunnen begrijpen welke patiënt welke behandeling nodig heeft. Deels vanwege het maar matig kunnen karakteriseren van deze ziekte, is de bewijslast van veel behandelingen voor deze patiëntengroep erg dun. Vanwege het weinig doorslaggevende bewijs, houden artsen zich vaak niet aan internationale richtlijnen en zijn de uitkomsten van behandelen erg afhankelijk van waar je behandeld wordt. Het is minder vaak onderzocht welke behandelingen vaak worden gegeven en waar ze het meest worden gegeven. Dit is met name het geval voor behandelingen die op straat worden gegeven. Ook de effectiviteit van behandelingen zoals intubatie en direct transport naar een gespecialiseerd centrum, is niet goed genoeg onderzocht.

Reanimaties in het ziekenhuis zijn een dramatische complicatie van een ziekenhuisopname, die niet veel mensen overleven. De lange termijn uitkomsten zijn echter slecht beschreven, zeker vergeleken met reanimaties buiten het ziekenhuis. Om overleving en kwaliteit van leven na reanimatie in het ziekenhuis te verbeteren, wordt er de laatste decennia steeds vaker gebruikt gemaakt van een ECMO (een soort hart-long machine) tijdens reanimatie. Door de functie van het hart en de longen over te nemen, krijgen deze rust en hebben ze meer kans om te herstellen. Het is een kostbare en arbeidsintensieve behandeling. Echter, in studies wordt wel de meerwaarde ervan gezien, ook al zijn deze studies niet ideaal om effectiviteit aan te tonen. Of de kosten echter opwegen tegen de baten is nog niet duidelijk.

Het overkoepelende doel van dit proefschrift is om bij te dragen aan een efficiëntere en effectievere beslisvorming voor traumatisch hersenletsel en reanimaties in het ziekenhuis. Hiervoor zal ik een verscheidenheid aan methodes gebruiken om te leren van grote observationele datasets. Er zijn drie specifieke doelen die in dit proefschrift aan bod komen:

1. Gegeven dat de huidige praktijk kan worden beschreven aan de hand van behandelingen en de uitkomsten,

   - wat is de variatie in behandelingen die op straat worden gedaan bij traumatisch hersenletsel, en
   - wat is de te verwachten lange-termijn uitkomsten na reanimaties in het ziekenhuis en hoeveel variatie bestaat hierbij?

   – Descriptieve vragen

2. Welke behandelingen verbeteren de uitkomsten patiënten met traumatisch hersenletsel en van patiënten die in het ziekenhuis gereanimeerd worden, of welke behandelingen zijn het echt waard? — Een causale vraag.

3. Hoe kunnen we traumatisch hersenletsel beter karakteriseren en hoe kunnen we hun uitkomst beter voorspellen? — Een predictie vraag.

16.1 | Deel I – Huidige praktijk

In **hoofdstuk 2 & 3** bestudeer ik hoe traumatisch hersenletsel wordt behandeld voordat de patiënt in het ziekenhuis is. In heel Europe wordt deze zorg heel gevarieerd uitgevoerd, onafhankelijk van patiëntfactoren: de kans dat je een behandeling krijgt, of de tijd die erover gedaan wordt om je bij het ziekenhuis te brengen, is heel anders in Scandinavië dan in Zuid-Europa. Ook al is deze zorg erop gericht om secundaire schade aan de hersenen te voorkomen door een te lage bloeddruk of te weinig zuurstof in het bloed, toch komt dit nog respectievelijk voor in 13% en 7% van de ernstige gevallen van traumatisch hersenletsel. Met name patiënten met

veel andere verwondingen naast het hoofdletsel hebben kans op een te lage bloeddruk of zuurstofspanning. In **hoofdstuk 3** bestudeer ik ook nog een bepaalde behandeling die vaak wordt gedaan op straat: het plaatsen van een beademingsbuis in de luchtpijp van patiënten om ze te helpen met ademen. Deze handeling heet intubatie. Intubatie wordt met name vaak verricht als het bewustzijn van patiënten daalt. Er was veel variatie in hoe vaak er wordt geïntubeerd in Europa, onafhankelijk van hoe ernstig het letsel was: de beslissing om te intuberen was statistisch even afhankelijk van waar de patiënt het ongeluk krijgt, als van de reactiviteit van de pupillen van de patiënt.

In **hoofdstuk 4 & 5** bestudeer ik de lange termijn overleving na reanimaties in het ziekenhuis. In deze twee hoofdstukken bestudeer ik respectievelijk de algemene overleving na reanimaties in het ziekenhuis en de neurologische uitkomst als ECMO wordt gebruikt bij de reanimatie. De algemene 1-jaarsoverleving na reanimatie in het ziekenhuis is laag (slechts 13,4%), maar vergelijkbaar met de overleving tot ontslag. Bovendien is er een kleine positieve trend van de overleving tussen 1985 en 2018. Vergeleken met alle mensen die worden gereanimeerd in het ziekenhuis, zijn patiënten die ECMO krijgen veel meer geselecteerd op bepaalde kenmerken. Vooral patiënten met een hoge kans op een goede uitkomst worden geselecteerd, omdat het bij deze groep patiënten nuttiger is om deze behandeling toe te passen. Mede door deze selectie is de overleving hoger (30%) en is er minder variatie. De neurologische uitkomst na het gebruik van een ECMO is ook erg goed: 84% van de overlevenden herstelt tot een acceptabele neurologische conditie.

In **hoofdstuk 6** vergelijk ik voor reanimatie in ziekenhuizen de structurele kwaliteitsindicatoren van de ziekenhuizen, procedurele kwaliteitsindicatoren van gegeven zorg en de uitkomsten van patiënten tussen zieken-

huizen in Nederland. Er is gemiddelde variatie in uitkomst, wat met name kan worden verklaard door verschillen in de behandelde populatie van patiënten. Alleen de verschillen in de neurologische uitkomst waren niet volledig te verklaren door de verschillen in patiënten. Dat betekent dat als de zorg van deze patiënten beter wordt, er met name een verbetering kan optreden op het gebied van neurologisch uitkomst. Een mogelijke manier om de zorg te verbeteren is om twee keer per jaar een reanimatietraining te geven. Dit leek een positief effect te hebben op de neurologische uitkomst.

16.2 | Deel II – Beste praktijk

In **hoofdstuk 7** bestudeer ik de hoe de uitkomsten van traumatisch hersenletsel worden beïnvloed door intubatie. In landen waar meer geïntubeerd werd, vonden we ook gemiddeld betere functionele uitkomsten. Daarnaast werd het positieve effect van een belangrijke Australische studie niet gevonden in onze studie: een verklaring hiervoor is dat vroeg intuberen op straat in plaats van gecontroleerd in het ziekenhuis misschien meer effectief is in Australië, waar de ziekenhuizen gemiddeld verder weg zijn van het ongeluk dan in Europa. Als laatste vond ik geen effect van intuberen op straat of in het ziekenhuis op functionele uitkomsten. Daarentegen was er wel een positief effect van intuberen in mensen met zwaarder letsel: in mensen met een lager bewustzijn werden er betere uitkomsten gezien na intubatie in het ziekenhuis, en in mensen met zwaarder letsel aan de borst of in de buik werden er betere uitkomsten gezien na intubatie op straat.

In **hoofdstuk 8** wordt het direct vervoeren van patiënten met traumatisch hersenletsel naar gespecialiseerde centra bestudeerd. Ook voor deze interventie werd er veel variatie gezien in Europa, die niet kon worden

verklaard door patiëntkarakteristieken. Ook al hadden patiënten die niet direct werden vervoerd naar een gespecialiseerd centrum vaker ernstige afwijkingen in de hersenen, en ook al kregen ze de behandelingen hiervoor later door de extra vertraging, toch was de uitkomst van de patiënten niet slechter. Sommige patiënten kunnen dus veilig worden gestabiliseerd in een niet gespecialiseerd centrum, om vervolgens alsnog naar zo'n centrum te gaan als dat nodig is.

In **hoofdstuk 9** presenteer ik een kosten-baten analyse voor het gebruik van ECMO tijdens reanimaties in het ziekenhuis. Door de best beschikbare gepubliceerde gegevens te combineren in een besliskundig model, hebben we verschillende scenario's vergeleken. De verschillende scenario's waren:

1. niemand ECMO tijdens reanimatie;
2. iedereen ECMO tijdens reanimatie;
3. alleen patiënten met een grotere kans om te overleven ECMO tijdens reanimatie.

In westerse landen, waar we bereid zijn wel €50.000 – €100.000 per gezond levensjaar te besteden, is iedereen ECMO geven tijdens reanimatie het scenario wat het meest waarschijnlijk kosteneffectief is. Deze strategie implementeren kost €10.818 per extra gezond levensjaar.

16.3 | Deel III - Identificatie van hoog risico patiënten

In **hoofdstuk 10** is een analyse gepresenteerd waarmee ik heb geprobeerd om de classificatie van traumatisch hersenletsel te verbeteren. Door alle

patiënten in de CENTER-TBI studie te groeperen met een algoritme op basis van klinische karakteristieken, zijn 4 groepen gemaakt. De belangrijkste determinanten van deze groepen zijn extra letsel naast hoofdletsel, het type ongeval en de belangrijkste maat voor bewustzijn (GCS). Behalve GCS voorspellen deze determinanten niet goed de uitkomst van patiënten. Wel kan het beschrijven van patiënten op basis van deze determinanten mogelijk een betere verklaring vinden voor verschillen in zorgpaden en ontstaan van afwijkingen.

In **hoofdstuk 11** vergelijk ik hoe goed twee verschillende typen algoritmes zijn in het voorspellen van de gevolgen van een traumatisch hersenletsel. De algoritmes die ik vergelijk zijn zelflerende computer algoritmes en traditionele statistische modellen. Deze algoritmes heten in vaktermen respectievelijk machine learning en regressie. Ongeacht de methode is het belangrijk om het algoritme te trainen om te kunnen voorspellen. Ik beoordeel de verschillende methoden met twee criteria. Ten eerste beoordeel ik hen op hun vermogen om mensen met en zonder negatieve gevolgen van TBI van elkaar te onderscheiden. Ten tweede beoordeel ik hen op het vermogen om waarheidsgetrouwe voorspellingen te doen: Bijvoorbeeld "als er van een groep mensen is gezegd dat ze 10% kans hebben om dood te gaan, gaat er dan ook 10% van die groep dood?". Om dit te doen heb ik alle algoritmes eerst getraind in een selectie van 15 datasets, en daarna getest op een 16e dataset. Het bijzondere aan deze datasets is dat de data een tijdsbestek beslaan van wel 34 jaar. Het bleek dat ieder algoritme hetzelfde scoorde op mijn bovengenoemde twee criteria. Echter, per dataset verschilde de kwaliteit van de voorspellingen wel heel erg. Dit wil zeggen dat de geteste algoritmes op sommige datasets beter werken dan op andere datasets. Daarom heb ik geconcludeerd dat bij het ontwikkelen van een voorspellend algoritme het niet zoveel uitmaakt welk type algoritme je

kiest, omdat ze allemaal even goed lijken. Maar het is wel heel belangrijk om goed na te gaan of het getrainde algoritme goed werkt in de populatie waarin je het wil gaan toepassen.

In **hoofdstuk 12** wordt een andere strategie gebruikt om de gevolgen van traumatisch hersenletsel beter te kunnen voorspellen: om beter te kunnen voorspellen of een patiënt een afwijking op de CT scan heeft, kunnen bepaalde biomarkers worden toegevoegd aan bestaande voorspellende regels. Biomarkers zijn in deze context stofjes die in het bloed kunnen worden gemeten en die iets zeggen over een onderliggend biologisch proces. Ik heb zes verschillende biomarkers geanalyseerd (S100B, NSE, GFAP, UCH-L1, NFL, en t-tau), en vond dat GFAP consistent de voorspellende waarde van de voorspellende regels verbeterde. Helaas was de overeenkomst tussen herhaalde metingen van dezelfde patiënt op het zelfde moment niet erg hoog. Om GFAP dus te kunnen gebruiken in de klinische praktijk, moet de manier waarop het gemeten wordt nog verbeterd worden.

In **hoofdstuk 13** stel ik een 5-stappen plan voor om in onderzoek naar betere beslismodellen, om te gaan met incomplete data. De vijf stappen zijn:

1. beschrijf de patronen in ontbrekende data;
2. kies een manier om te imputeren;
3. imputeer de data;
4. controleer of de imputatie goed is gegaan;
5. analyseer de geïmputeerde data zoals gepland.

Deze tutorial is dus met name gefocust op multipele imputatie, maar we bespreken ook wanneer andere imputatie methoden kunnen worden overwogen, of zelfs nodig zijn.

16.4 | Algemene discussie

Het bestuderen van de huidige klinische praktijk (deel 1) is handig om bepaalde gebieden te vinden die nog verbeterd kunnen worden. De klinische praktijk heb ik bestudeerd aan de hand van behandelingen die patiënten krijgen, maar ook aan de hand van hun uitkomsten. Ik zag veel variatie in de behandelingen die mensen met traumatisch hersenletsel krijgen op straat. Deze situatie is niet ideaal, omdat gelijke mensen gelijk behandeld moeten worden. Deze variatie komt waarschijnlijk deels omdat het bewijs voor effectiviteit van behandelingen niet erg doorslaggevend is, waardoor ieder land zijn eigen nationale richtlijnen maakt. Maar ook kan de variatie van behandelingen in de verschillende Europese landen worden verklaard aan de hand van verschillende geografische contexten (zoals bergen, eilanden) en financiële middelen.

Daarnaast bekeek ik ook de uitkomsten na reanimaties in het ziekenhuis. Hierbij heb ik zowel de gemiddelde uitkomsten bekeken, als de variatie rondom dit gemiddelde. Een interessante bevinding is dat niet alleen betere kwaliteit van zorg voor hogere overlevingskans zorgde, maar ook een strengere selectie van patiënten. Hiermee moet dus rekening worden gehouden als uitkomsten worden onderzocht, aangezien dit niet betekent dat ziekenhuizen betere zorg leveren.

In het tweede deel heb ik bekeken of bepaalde behandelingen voor traumatisch hersenletsel of reanimaties in het ziekenhuis effectief zijn. Zo

ja, heb ik bekeken of de kosten opwegen tegen de te verwachten gezondheidswinst. Om te kunnen bevestigen of een behandeling werkt in observationele data, heb je veel kennis nodig om de data goed te kunnen interpreteren. De ervaring en kennis over de context van het probleem heb je nodig om te weten welke data er verzameld moet worden en waarvoor gecorrigeerd moet worden. Je moet aan de hand van een bepaald causaal model (naar het beeld van wat de onderzoeker van de wereld heeft) dus bepalen hoe je de studie moet vormgeven. Als je dan de effectiviteit van behandelingen hebt bevestigd, is het zaak om de kosten en de baten tegen elkaar af te wegen. Dit heb ik vooral gedaan aan de hand van beslismodelering.

Om hoog risico patiënten beter te kunnen identificeren, is het statistisch niet handig om mensen te clusteren zoals ik in hoofstuk X heb gedaan. Een dergelijke analyse kan wel inzichtelijk maken wat voor "soort" patiënten er bestaan, maar het is niet efficënt om deze groepen dan te gebruiken om de gevolgen te voorspellen na bijvoorbeeld een ongeval. Het is beter om goede prognostische modellen te maken om de gevolgen van een ziekte te voorspellen. Als klinische data met al z'n gebreken wordt gebruikt, kan het beste traditionele statistische methoden worden gebruikt. Vergeleken met machine learning algoritmes, zijn er veel meer hulpmiddelen beschikbaar als traditionele statistische methoden worden gebruikt om de gebreken in de data te addresseren. Wat ook een verbeterstap is, is het toevoegen van sterke, nieuwe voorspellende factoren. Dit hebben we laten zien met GFAP voor het voorspellen van afwijkingen in het brein na traumatisch hersenletsel. Als laatste adviseren we het 5-stappenplan om met incomplete data om te gaan bij het ontwikkelen en testen van prognostische modellen.

Uiteindelijk bieden grote observationele data vele kansen ten opzichte van gerandomiseerde klinische studies: de resultaten zijn makkelijk te ver-

talen naar de “echte” praktijk, het is ethisch en financieel vaak haalbaar, en door de grote aantallen is de kans op een fout-negatieve bevinding laag. Echter, je hebt om causaliteit te kunnen vaststellen uitgebreide kennis nodig van wat je onderzoekt, en de data moet vaak helemaal aangepast worden naar jouw onderzoeksvraag.

17

Publications

Table 17.1: *Manuscripts in this thesis*

Chapter	Status	Journal	Year
2	Published	Prehospital Emergency Care [1]	2020
3	Published	Anaesthesia [2]	2019
4	Published	Resuscitation [3]	2018
5	Published	Critical Care [4]	2020
6	Published	Critical Care [5]	2021
7	Published	British Journal of Anaesthesia [6]	2020
8	Published	Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine [7]	2021
9	Published	Resuscitation [8]	2019
10	Published	Journal of Neurotrauma [9]	2019
11	Published	Journal of Clinical Epidemiology [10]	2020
12	Published	EBioMedicine [11]	2020
13	Published	Journal of Neurotrauma [12]	2021

All personal publications

1. Gravesteijn, B. *et al.* Prehospital management of traumatic brain injury across Europe: a CENTER-TBI study. *Prehospital Emergency Care,* 1–22. ISSN: 1090-3127 (2020).
2. Gravesteijn, B. Y. *et al.* Variation in the practice of tracheal intubation in Europe after traumatic brain injury: a prospective cohort study. *Anaesthesia,* anae.14838. ISSN: 0003-2409. `https://onlinelibrary.wiley.com/doi/abs/10.1111/anae.14838` (2019).
3. Schluep, M., Gravesteijn, B. Y., Stolker, R. J., Endeman, H. & Hoeks, S. E. One-year survival after in-hospital cardiac arrest: A systematic review and meta-analysis. *Resuscitation* **132,** 90–100. ISSN: 1873-1570 (2018).
4. Gravesteijn, B. Y. *et al.* Neurological outcome after extracorporeal cardiopulmonary resuscitation for in-hospital cardiac arrest: a systematic review and meta-analysis. *Critical Care* **24,** 1–12 (2020).
5. Benjamin Y. Gravesteijn *et al.* Between-centre differences in care for in-hospital cardiac arrest: a prospective cohort study. *Critical Care* **25,** 1–12 (2021).
6. Benjamin Y. Gravesteijn *et al.* Tracheal intubation in traumatic brain injury: a multicentre prospective observational study. *British Journal of Anaesthesia* (2020).
7. Sewalt, C. A. *et al.* Primary versus secondary referral to a specialized neurotrauma centre in patients with moderate/severe Traumatic Brain Injury: a CENTER TBI study. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine* (2021).
8. Gravesteijn, B. Y. *et al.* Cost-effectiveness of extracorporeal cardiopulmonary resuscitation after in-hospital cardiac arrest: A Markov decision model. *Resuscitation* **143,** 150–157. ISSN: 03009572 (2019).

9. Benjamin Y. Gravesteijn *et al.* Towards a new multidimensional classification of traumatic brain injury: a CENTER-TBI study. *Journal of Neurotrauma* (2019).

10. Benjamin Y. Gravesteijn *et al.* Machine learning algorithms performed no better than regression models for prognostication in traumatic brain injury. en. *Journal of Clinical Epidemiology*, S0895435619308753. ISSN: 08954356. `https://linkinghub.elsevier.com/retrieve/pii/S0895435619308753` (2020) (Mar. 2020).

11. Czeiter, E. *et al.* Blood biomarkers on admission in acute traumatic brain injury: Relations to severity, CT findings and care path in the CENTER-TBI study. *EBioMedicine* **56,** 102785 (2020).

12. Benjamin Y. Gravesteijn *et al.* Missing Data in Prediction Research: A Five-Step Approach for Multiple Imputation, Illustrated in the CENTER-TBI Study. *Journal of neurotrauma* **38,** 1842–1857 (2021).

13. Benjamin Y. Gravesteijn *et al.* Awake craniotomy versus craniotomy under general anesthesia for the surgical treatment of insular glioma: choices and outcomes. *Neurological research* **40,** 87–96 (2018).

14. Benjamin Y. Gravesteijn *et al.* Minimizing Population Health Loss in Times of Scarce Surgical Capacity During the Coronavirus Disease 2019 Crisis and Beyond: A Modeling Study. *Value in Health* **24,** 648–657 (2021).

15. Benjamin Y. Gravesteijn, van Saase, J., Lingsma, H. & Baatenburg de Jong, R. Rekenmodel helpt bij prioritering ok-tijd. *Medisch Contact* **23** (2020).

16. Voormolen, D. C. *et al.* Post-Concussion Symptoms in Complicated vs. Uncomplicated Mild Traumatic Brain Injury Patients at Three and Six Months Post-Injury: Results from the CENTER-TBI Study. *Journal of Clinical Medicine* **8,** 1921 (2019).

17. Voormolen, D. C. *et al.* Prevalence of post-concussion-like symptoms in the general population in Italy, The Netherlands and the United Kingdom. *Brain Injury* **33.** PMID: 31032649, 1078–1086. eprint: https://doi.org/10.1080/02699052.2019.1607557. https://doi.org/10.1080/02699052.2019.1607557 (2019).

18. Steyerberg, E. W. *et al.* Case-mix, care pathways, and outcomes in patients with traumatic brain injury in CENTER-TBI: a European prospective, multicentre, longitudinal, cohort study. *The Lancet Neurology* **18,** 923–934 (2019).

19. Benjamin Y. Gravesteijn, van Hof, K., Krijkamp, E. & Baatenburg de Jong, R. Rekenmodel helpt bij prioritering ok-tijd. *Nederlands Tijdschrift voor KNO* **27** (2021).

20. Marincowitz, C., Benjamin Y. Gravesteijn, Sheldon, T., Steyerberg, E. & Lecky, F. Performance of the Hull Salford Cambridge Decision Rule (HSC DR) for early discharge of patients with findings on CT scan of the brain: a CENTER-TBI validation study. *Emergency Medicine Journal* (2021).

21. Sewalt, C. A. *et al.* Identifying trauma patients with benefit from direct transportation to Level-1 trauma centers. *BioMed Central: Emergency Medicine* (2021).

18

Portfolio

	Year	Workload (ECTS)
1. PhD training		
Academic training		
BROK	2017	1
Master of Science in Health Sciences, NIHES, Rotterdam	2017-2019	120
Presentations at national and international conferences		
ICP, Leuven, Belgium	2019	1
EuroAnesthesia, Vienna, Austria	2019	1
EuroNeuro, Brussels, Belgium	2019	1
NeuroTrauma, Toronto, Canada	2018	1
EuroNeuro, Barcelona, Spain	2016	1
Seminar and workshops		
Research seminars, Dep. Public Health, Erasmus MC, Rotterdam, the Netherlands	2018-2022	9
Research meetings Medical Decision Making, Dep. Public Health, Erasmus MC, Rotterdam, The Netherlands	2018-2022	1
Pentecilia, research integrity workshop	2018	0.2
Conference ZonMW/NWO: Evolution or Revolution, the Hague	2019	0.2
SMDM 2018 Leiden	2018	0.2

BROK: Basiscursus Regelgeving en Organisatie voor Klinisch onderzoekers; NIHES: National Institute for Health Sciences; SMDM: society for medical decision making

	Year	Workload (ECTS)
2. Teaching activities		
Educational innovation activities		
Master research consultancy project	2019-2020	1.6
CE02 clinical epidemiology (prognosis)	2018-2019	1
CC02 Biostatistics I	2018	1
MOOC population health management	2018-2019	20
Smarter Choices Better Health	2019	0.6
Supervising		
Supervisor medical students minor anesthesiology	2019	1.2
3. Other activities		
Acquisition		
Erasmus trustfonds (applicant: €500.- granted)	2018	0.5
HOKA grant (applicant: €5000.- granted)	2019	0.5
Junior researcher scholarship (applicant €3000.- granted)	2019	0.5
van Walree Beurs (applicant: €1000.- granted)	2019	0.5
	Total	151.6

BROK: Basiscursus Regelgeving en Organisatie voor Klinisch onderzoekers; NIHES: National Institute for Health Sciences; SMDM: society for medical decision making

19

Abbreviations

Abbreviation	Meaning
ACCI	Age-Combined Charlson Comorbidity Index
Adm	Admission
AIS	Abbreviated Injury Scale
ALS	Advanced-Life-Support
ANOVA	Analysis Of Variance
ASA	American Society of Anesthesiologists
AUC	Area under the (Receiver Operating Charachteristic) Curve
BLS	Basic Life Support
BMI	Body Mass Index
CA	Acute Coronary Syndrome
CBS	Statistics Netherlands
CC	Complete Case
CCI	Charlson Comorbidity Index
CCPR	Conventional Cardiopulmonary Resuscitation
CDR	Clinical Decision Rule
CENTER-TBI	Collaborative European NeuroTrauma Effectiveness Research for TBI

Abbreviation	Meaning
CER	Comparative Effectiveness Research
CHEERS	Consolidated Health Economic Evaluation Reporting Standards
CI	Confidence Interval
COSCA	Core Outcome Set for Cardiac Arrest
CPC	Cerebral Performance Category
CPR	Cardiopulmonary Resuscitation
CRASH	Corticosteroid Randomization After Significant Head Injury
CT	Computed Tomography
CV	Cross Validation
DAG	Directed Acyclic graph
Df	Degrees of fredom
DNR	Do-Not- Resuscitate
EALL	ECMO for all
ECLIA	Electrochemiluminescence immunoassay kit
ECMO	Extracorporeal Membrane Oxtgenation
ECPR	Extracorporeal Cardiopulmonary Resuscitation
ED	Emergency Department
EDH	Epidural Haematoma
ELSO	European Life Support Organisation
EMS	Emergency Medical Service
ER	Emergency Room
EU	European Union
EWS	Early Warning Score
FDA	Food and Drug Administration
GBM	Gradient Boosting Machine
GCS	Glasgow Coma Scale
GFAP	Glial Fibrillary Acidic Protein
GFR	Glomerular Filtration Rate
GOS	Glasgow Outcome Scale
GOSE	Glasgow Outcome Scale-Extended
Hb	Hemoglobin
HRQoL	Health Related Quality of Life

Abbreviation	Meaning
Ht	Hematocrit
ICER	Incremental cost-effectiveness ratio
ICH GCP	ICH Harmonized Tripartite Guideline for Good Clinical Practice
ICP	Intracranial Pressure
ICU	Intensive Care Unit
IHCA	In-Hospital Cardiac Arrest
IHI	In-hospital Intubation
IMPACT	International Mission for Prognosis and Analysis of Clinical Trials in TBI
INCF	International Neuroinformatics Coordination Facility
IV	Intravenous
IQR	Interquartile Range
ISS	Injury Severity Score
LMIC	Low and Middle Income Country
Logreg	Logistic regression
LOICUS	Length of ICU Stay
LOS	Length of Stay
LR	Logistic Regression
MAR	Missing At Random
MCAR	Missing Completely At Random
MEI	Major Extracranial Injury
MEWS	Modified Early Warning Score
mGCS	Motor component of the Glasgow Coma Score
MI	Multiple Imputation
MICE	Multiple Imputation with Chained Equations
MID	Muliple-Impute-and-Delete
ML	Machine Learning
MLS	Midline Shift
MNAR	Missing Not At Random
MOR	Median Odds Ratio
MRI	Magnetic Resonance Imaging
MRS	Modified Rankin Scale
NE	ECMO for no one

Abbreviation	Meaning
NEWS	National Early Warning Score
NFL	Neurofilament Protein-Light
NI	Non-intubated
NNs	Neural Networks
NPV	Negative Predictive Value
NSAH	Non Specialist Acute Hospital
NSE	Neuron-Specific Enolase
OHCA	Out-of-Hospital Cardiac Arrest
OR	Odds Ratio
PCA	Principal Component Analysis
PHI	Prehospital intubation
PMM	Predictive Mean Matching
Polr	Proportional Odds Logistic Regression
Polyreg	Polytomous regressiom
PSA	Probabilistic Sensitivity Analysis
QALY	Quality-of-life adjusted life years
RCT	Randomised Controlled Trial
RF	Random Forest
ROC	Reciever Operating Characteristic
ROSC	Return of Spontaneous Circulation
ROUTiNE	Resuscitation Outcomes in the Netherlands
RR	Relative Risks
RRS	Rapid Response Team Warning Score
RTA	Road Traffic Accident
RTI	Road Traffic Incident
RUO	Research-Use Only
S100B	S100 calcium-binding protein B
SD	Standard Deviation
SI	Single Imputation by conditional estimates
SNC	Specialized Neurotrauma Centre
STROBE	Strengthening The Reporting of OBservatiobnal Studies in Epidemiology
SVMs	Support Vector Machines

Abbreviation	**Meaning**
TBI	Traumatic Brain Injury
TRIPOD	Reporting guideline for multivariable prediction models
TSAH	Traumatic Subarachnoid Haemorrhage
UCH-L1	Ubiquitin C-terminal Hydrolase L1
US	United States
VA-ECMO	Veno-Arterial Extracorporeal Membrane Oxygenation
WTP	Willingness to Pay

20

Curriculum Vitae

Benjamin Yaël Gravesteijn was born in 1995, in Capelle aan den IJssel. During the final years of secondary school, he attended the *Junior Med School* at *Erasmus University Medical Center Rotterdam*, where he first came into contact with research. After getting his diploma cum laude at the *Erasmiaans Gymnasium* Rotterdam, he continued by studying medicine at the *Erasmus MC*.

During the first year, he started his scientific endeavors at the department of Anesthesiology, under supervision of dr. Markus Klimek. After getting his bachelor degree, he continued by studying clinical epidemiology at the *Netherlands Institute of Health Sciences*. His research internship was at the department of Public Health, under supervision of both Markus Klimek as well as Hester Lingsma. His research endeavors during the master's degree finally led him to further pursue a PhD degree with prof. dr. Ewout Steyerberg and prof. dr. Hester Lingsma as promotors.

During his time as PhD-student, he visited the department of Anaesthesia at the *University of Cambridge* to collaborate. He was involved in

various educational reform activities, he for example co-developed two Massive Online Open Access Courses and initiated the master consultancy project using his HOKA grant. He presented his work in various international conferences, for which he was invited twice. He was also selected to be an advocate of research for students as one of the *"Faces of Science"* of the KNAW.

Before finishing his PhD, he continued his studies in Medicine. He finished his PhD-thesis during his medical rotations. He will continue by pursuing a career in gynaecology. His goal is to bridge the gap between epidemiologists, data scientists, and clinicians, thereby making healthcare more efficient, value-based, and equal.

21

Dankwoord

Om te beginnen wil ik tegen de mensen zeggen die als allereerst dit deel van mijn proefschrift hebben geopend: Betrapt! De andere negentien hoofdstukken zijn ook zeker het lezen waard. Voor degene die dat hebben gedaan, volgt nu echt het dankwoord, een moment om heel terecht te reflecteren aan wie ik deze prachtige reis allemaal te danken heb.

Graag zou Saskia le Cessie, Marc Sabbe, en Myriam Hunink willen bedanken dat jullie wilden plaatsnemen in de beoordelingscommissie van mijn proefschrift. Het is voor mij een eer om door jullie goed te worden gekeurd als gepromoveerde onderzoeker. Ook de overige leden van mijn commissie wil ik hartelijk bedanken: David Menon, Sanne Hoeks, en Maarten van Smeeden.

Moreover, I would like to thank the unique opportunity to collaborate within the CENTER-TBI family. David Menon, Andrew Maas, as coordinators, you have brought together great minds from all around the continent (and even broader), with the great help of the real boss of CENTER-TBI, Véronique de Keyser. Your efforts have not only produced substan-

tial amounts of beautiful TBI-work, but also stimulated international collaboration and friendships. I might be even more grateful for the latter. About these international collaborators, I would like to thank Ari Ercole, Virginia Newcombe, Sophie Richter, and Dan Whitehouse for the inspiring and stimulating visit to Cambridge. Moreover, I would like to thank Nino Stochetti for his warm welcome in Milan, where I could continue the collaboration with Marco Carbonara. Marco, I would like to thank you personally for the great time we had in Brussels, and the highs and lows we've endured during our -unfinished- collaboration.

Ook het CENTER-TBI Rotterdam team wil ik hartelijk bedankten: Maryse, Jilske, Eveline, Isabel, Victor, Ernest, Ana, ik wil jullie allemaal bedanken voor de samenwerking. De toren, en het eiland, was niet compleet zonder jullie. Ik vind het heel leuk om te zien dat jullie allemaal zulke goede onderzoekers zijn geworden.

Dan wil ik natuurlijk het ROUTiNE-team hartelijk bedanken. Marc, wat heb jij een werk verzet om deze studie op te zetten en succesvol tot een einde te brengen. Ik was blij dat ik je in het begin heb kunnen helpen bij de uitvoering van het onderzoek, maar ook trots dat we uiteindelijk als onderzoekspartners samen vele mooie stukken hebben geschreven. Persoonlijk denk ik dat we ook veel van elkaar hebben geleerd. Daarnaast wil ik Robert-Jan Stolker en Rik Endeman ook nog graag bedanken, voor de productieve samenwerking. Jullie hebben me als relatieve "buitenstaander" ontzettend mooie kansen gegeven. Sanne, jou wil ik nog persoonlijk bedanken voor de persoonlijke en stimulerende begeleiding. Jij hebt mij niet alleen wetenschappelijk-inhoudelijk, maar ook persoonlijk erg geholpen om mezelf tot zelfstandig onderzoeker te ontwikkelen.

Markus, jij stond aan het begin van mijn onderzoekerscarrière, en ik ben jou veel dank verschuldigd voor alle mooie kansen die daaruit zijn

voortgevloeid. Ik ben blij dat jij twee keer in mijn proefschrift staat, en hoop nog veel van je te leren.

Ewout, ik heb wel eens gehoord dat de voornaamste taak van een promotor is om de promovendus te vormen en te inspireren. Als er iets is wat jij hebt gedaan, is het dat wel. Ik heb met jou de allerleukste, scherpste, en leerzaamste discussies gehad door mijn hele traject heen, waar ik ook nog eens schaterlachend vandaan kwam. Ik heb geen enkel gesprek met jou gehad waar ik niet achteraf met meer energie uit ben gekomen. Jij zei altijd over mij dat ik zoveel energie had, wat jij heel leuk vond. Maar in mijn ogen was het juist andersom.

Hester, bedankt voor alle vrijheid die je mij hebt gegeven in mijn PhD-traject. Jij hebt de context gecreëerd waardoor ik me optimaal kon ontwikkelen. Misschien was het soms iets teveel vrijheid, maar achteraf gezien ben ik er denk ik een heel zelfstandig en kritische onderzoeker door geworden. Achteraf bekeken denk ik dat precíes op de juiste momenten gesprekken hebben plaatsgevonden met jou die me hebben geholpen om door te gaan. Door die vrijheid heb ik wel een proefschrift kunnen schrijven waar ik honderd procent achter sta en waar ik heel trots op ben. En omdat we het zo gezellig hadden, ben ik zelfs blijven hangen om met jou nog een geweldig side-project op te starten (het OK triage model). Ik zal nooit vergeten hoe we elkaar met een hartslag van 150 opbelden na het inleveren van onze ZonMW aanvraag, om te vieren dat we het hadden geflikt. Tóch die hardlopersmentaliteit, denk ik zo.

Maaike, toen ik hoorde dat jij weer wilde gaan studeren -onder anderedoordat ik zo enthousiast over onderzoek praatte, was ik zó ontzettend trots. Maar nog trotser nu je "gewoon" journalist bent geworden, na een prachtige persoonlijke reis. Ik hoop dat ik die reis met jou mag blijven meemaken, want je bent stiekem echt een inspiratiebron voor me.

Pap, mam, dit is het dan: ik promoveer! Mam, jij hebt me onderzoek met de paplepel ingegoten. Daarnaast ben ik ongelofelijk blij en dankbaar hoe nauwkeurig en secuur jij mij geholpen hebt de puntjes op de 'i' te zetten voor mijn proefschrift. Pap, ik hoop dat jij jouw rigoreuze administratie- en boekhoudkunsten bij mij terugziet in mijn tabellen en figuren. Ik ben jullie beiden ontzettend dankbaar voor alle kansen die jullie me maar blijven geven. Jullie hebben me geleerd dat kritisch nadenken en discussiëren onderdeel moet zijn van je gehele leven, niet alleen op werk. En ook Thomas en Daniël wil ik bedanken: zonder jullie zou ik niet zijn zoals ik nu ben.

Iris en Jenske, de driehoek, de lijn, mijn paranimfen sinds de middelbare school. Ik weet niet waar ik moet beginnen om jullie te bedanken. Maar ik weet zéker dat dit proefschrift er zonder jullie niet was gekomen. Jullie weten me altijd op te vrolijken, een reflectieve spiegel voor te houden, en ik wil gewoon eigenlijk altijd bij jullie zijn. Ik leer van jullie, ik lach en huil met jullie, en jullie zijn essentieel.

Daphne en Charlie, mijn échte paranimfen. Ik ben jullie beiden zo dankbaar. Dankzij jullie voelde m´n hele PhD traject als iets wat iedereen zou kunnen doen, terwijl ik weet dat het best ingewikkeld is. Wat hebben we gelachen om slaatjes in Toronto, ons verwonderd om struisvogels in de Beekse Bergen, en gehuild om Tony. Charlie, ik kan niet wachten om met jou in een ziekenhuis de boel (on)veilig te maken; en Daphne: ik ga samen met jou nog lekker veel kosten effectiviteits analyses doen (het liefste met zoveel mogelijk discretes choice experiments). Ik ben zo blij dat we hier nog steeds met z´n drieën staan en we al zoveel mee hebben gemaakt. Op wat voor manier dan ook, kunnen we elkaar letterlijk niet meer elkaar uit het oog verliezen. Bij deze beloofd.

Tom, bij deze vereeuwig ik je door te zeggen dat jij iemand bent waar

ik –zonder dat je het misschien zelf door hebt- ontzettend op steun. Ik kan niet wachten om te zien hoe onze reis samen voort zal gaan. Je bent een ideale spiegel, maatje, stressbal, en reisgenoot voor mij en ik houd van jou.