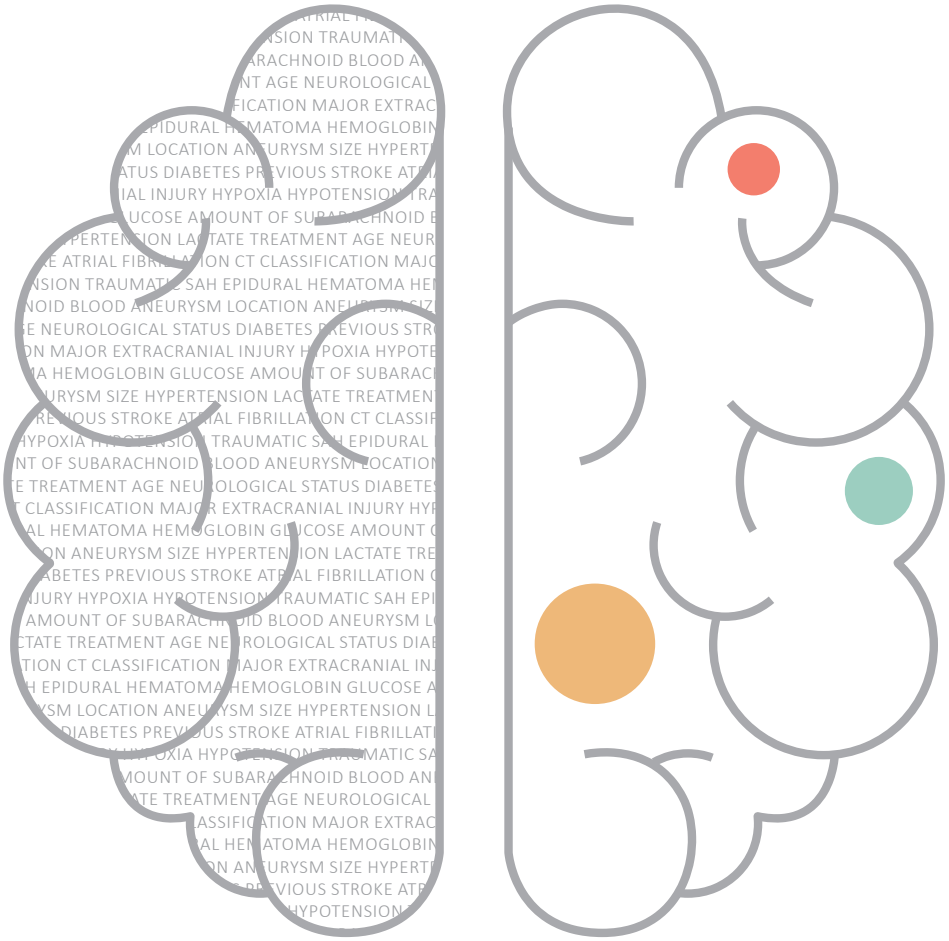
PREDICTION AND OUTCOME ANALYSES
IN ACUTE NEUROLOGICAL DISEASES



Simone A. Dijkland

Prediction and Outcome Analyses in Acute Neurological Diseases

Simone A. Dijkland

Financial support for the publication of this thesis was provided by the Department of Public Health of the Erasmus MC.

Cover design and layout: ProefschriftOntwerp.nl
Printing by: ProefschriftMaken

Prediction and Outcome Analyses in Acute Neurological Diseases

Predictie en uitkomst analyses in acute neurologische ziekten

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus
Prof. Dr. R.C.M.E. Engels

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
15 januari 2020 om 13.30 uur

door

Simone Anna Dijkland

geboren te Utrecht

Erasmus University Rotterdam

Promotiecommissie

Promotoren	Prof. dr. E.W. Steyerberg
	Prof. dr. D.W.J. Dippel
Overige leden	Prof. dr. P.J. Koudstaal
	Prof. dr. S. le Cessie
	Prof. dr. G. Meyfroidt
Copromotoren	Dr. H.F. Lingsma
	Dr. M. van der Jagt

Contents

PART I	**INTRODUCTION**	**9**
Chapter 1	General introduction	11
PART II	**PREDICTION**	**29**
Chapter 2	Early circulating lactate and glucose levels after aneurysmal subarachnoid hemorrhage correlate with poor outcome and delayed cerebral ischemia: A two-center cohort study *Crit Care Med 2016; 44(5): 966–972*	31
Chapter 3	Development and validation of the Dutch Stroke Score for predicting disability and functional outcome after ischemic stroke: A tool to support efficient discharge planning *Eur Stroke J 2018; 3(2): 165–173*	49
Chapter 4	Prognosis in moderate and severe traumatic brain injury: A systematic review of contemporary models and validation studies *J Neurotrauma 2019; Epub ahead of print*	69
Chapter 4.1	Response to Walker et al. (doi: 10.1089/neu.2017.5359): Predicting long-term global outcome after traumatic brain injury *J Neurotrauma 2019; 36(8): 1382-1383*	121
Chapter 5	Prediction of 60-day case fatality after aneurysmal subarachnoid hemorrhage: External validation of a prediction model *Crit Care Med 2016; 44(8): 1523–1529*	127
Chapter 5.1	Letter by Dijkland et al regarding article, "Prediction of outcome after aneurysmal subarachnoid hemorrhage: Development and validation of the SAFIRE grading scale" *Stroke 2019; 50(7): e224*	147
Chapter 6	Outcome prediction after moderate and severe traumatic brain injury: External validation of two established prognostic models in the CENTER-TBI study *Submitted*	153

PART III **OUTCOME ANALYSES** **183**

Chapter 7 Between-center and between-country differences in outcome after aneurysmal subarachnoid hemorrhage in the Subarachnoid Hemorrhage International Trialists (SAHIT) repository 185
J Neurosurg 2019; Epub ahead of print

Chapter 8 Utility-weighted modified Rankin Scale as primary outcome in stroke trials: A simulation study 203
Stroke 2018; 49(4): 965–971

Chapter 8.1 Response by Dijkland et al to letter regarding article, "Utility-weighted modified Rankin Scale as primary outcome in stroke trials: A simulation study" 221
Stroke 2018; 49(12): e338

PART IV **DISCUSSION** **227**

Chapter 9 General discussion 229

APPENDICES **255**

Summary 259

Samenvatting 267

Dankwoord 275

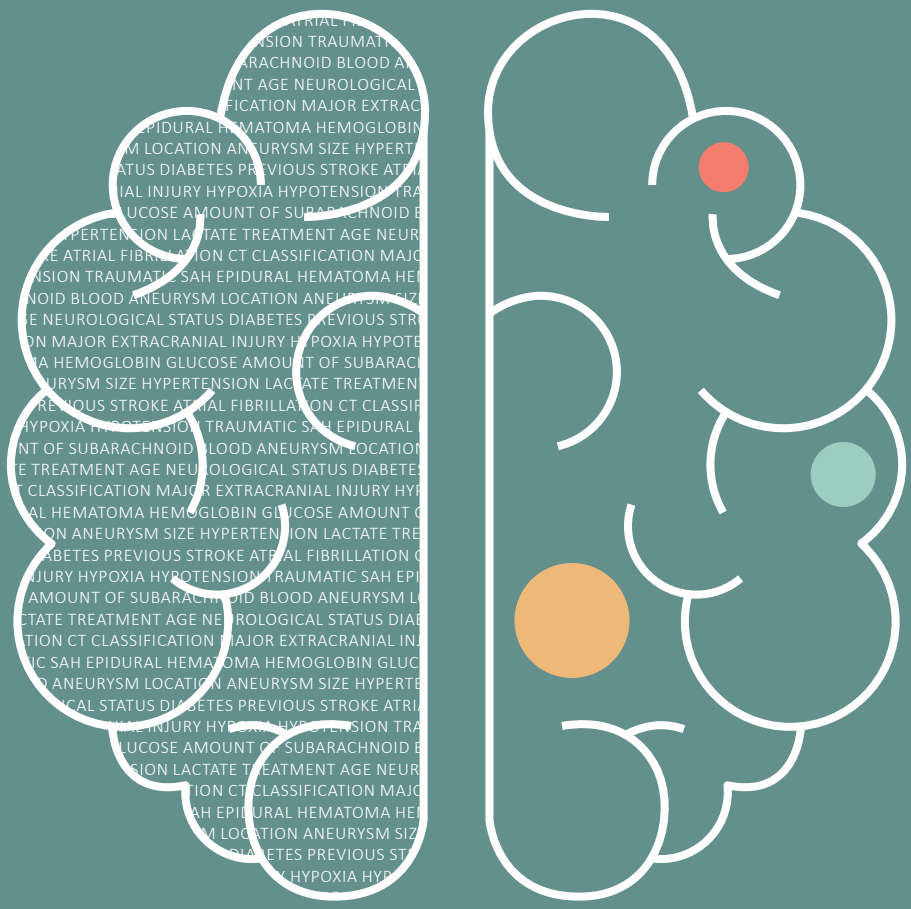
List of publications 281

PhD portfolio 287

About the author 293

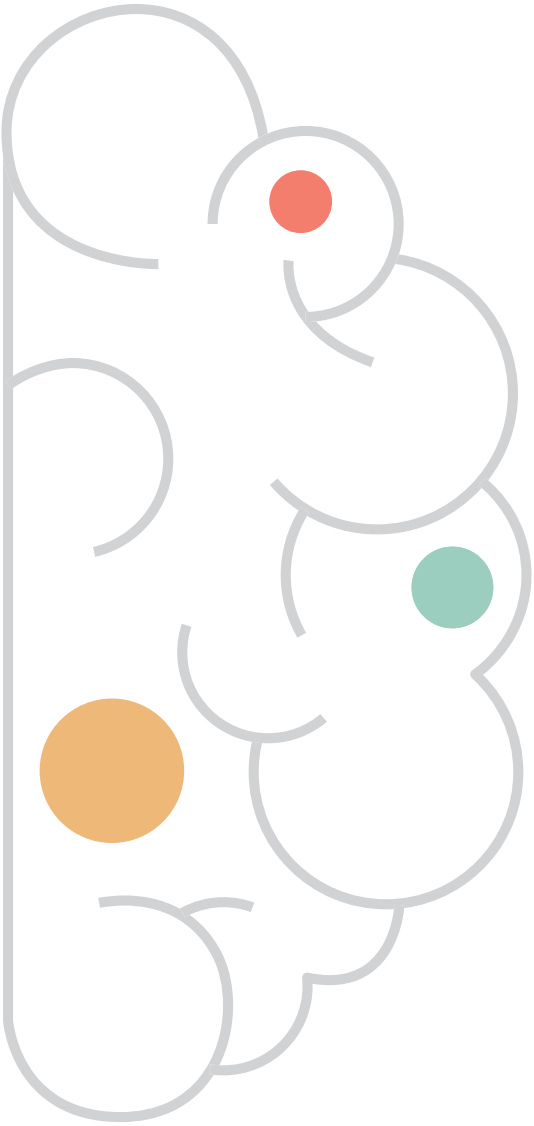
PART I

INTRODUCTION



CHAPTER 1

General introduction



Introduction

Most treatments and interventions in health care are aimed at optimizing clinical outcomes. Clinical outcome refers to the degree to which patients who survived a disease have returned to daily functioning. Clinical outcomes can be measured with different scales and from a variety of perspectives. The spectrum ranges from survival or functional scales focused on activities in daily living scored by a physician[1, 2], to multidimensional questionnaires addressing patient perception regarding physical, mental and emotional wellbeing (quality of life).[3, 4]

Measurement of clinical outcomes may serve different purposes, such as prognostic research and outcomes research. Prognostic research involves estimating the probability of a patient developing a certain clinical outcome over time, based on clinical and other characteristics.[5] Outcomes research refers to the analyses of clinical outcomes related to health care practices and interventions.[6] This includes examining variation in outcomes across different settings and determining the added value of new outcome measures.

This thesis presents the methodology and clinical implications of outcome prediction, assessment of between-hospital variation in clinical outcomes and evaluation of statistical efficiency of new outcome measures. These topics will be studied in the field of acute neurology.

Prediction

Observed or expected improvement or deterioration in patient outcomes is an important driver for changes in clinical management. Early identification of patients at high risk for poor functional outcome in a specific clinical setting may assist clinicians with treatment decisions, inclusion of patients in randomized clinical trials (RCTs) or benchmarking quality of care.[7, 8]

A prognostic factor is any characteristic that is associated with a subsequent clinical outcome.[9] For instance, older age is associated with a higher risk of death (in most diseases as well as in healthy subjects). Multivariable prognostic models combine several prognostic factors to estimate the risk of a specific endpoint for an individual patient.[8] An example is the Corticosteroid Randomisation After Significant Head injury (CRASH) model which estimates the risk of 14-day mortality or 6-month unfavorable outcome (death or severe disability) for patients with traumatic brain injury. The model consists of age, measures for clinical severity, and major extracranial injury (Figure 1.1).[10]

Standards and recommendations for the reporting of studies on multivariable prognostic models have been published.[8, 11] Development of a prognostic model consists of several steps, including selection and coding of predictors and defining the outcome of interest.[12, 13] The validity or quality of a prognostic model should be evaluated in the derivation cohort (internal validation) as well as in a new setting that differs from the derivation cohort (external validation). Several performance measures to determine model validity have been proposed. Prognostic models should adequately distinguish between patients with and without the outcome of interest (= model discrimination). Moreover, good agreement between observed and predicted outcome rates (= model calibration) is required to provide reliable predictions for patients in a specific clinical setting.[11-13] In addition to model discrimination and

calibration, the clinical usefulness of prognostic models should be evaluated, especially for models aiming to support clinical decision making.[11-13]

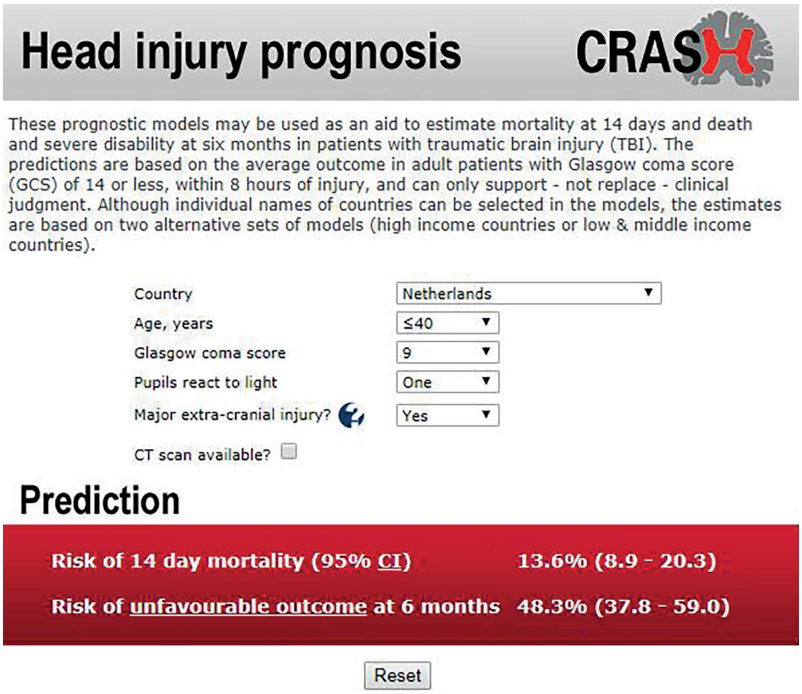


Figure 1.1. Web calculator from the CRASH prognostic model (available from http://www.crash.lshtm.ac.uk/Risk%20calculator/index.html). [10]
CT, computed tomography; CI, confidence interval.

Outcome analyses

Besides outcome prediction, measurement of clinical outcomes is also important to examine outcome variation in clinical outcomes across settings. Differences in clinical outcomes between hospitals and countries are present in many diseases, but are highly undesirable when caused by differences in management. Such differences may reflect poor implementation or even a lack of evidence-based diagnostic and therapeutic policies. Gaining insight in these outcome differences with random effects modeling creates the opportunity to evaluate practice variation.

Further, the introduction of new methods of outcome measurement requires evaluation of their added value in research or practice. Most current functional outcome scales may not be granular enough to detect small changes in clinical status, do not incorporate all aspects that can contribute to the level of disability and exclude patient perception on physical and mental well-being.[14-16] Therefore, a trend exists towards new outcome measures incorporating both functional outcome and quality of life (patient-reported outcome measures [PROMs]).[17] New outcome measures should be statistically

efficient to obtain reliable estimates of treatment effect (i.e. the degree of benefit or harm of an intervention) in clinical trials. Because the true treatment effect is unknown in empirical data, the only valid method to assess statistical efficiency of a new outcome measure is a simulation study.

Random effects modeling

Between-center and between-country differences in patient outcomes are ideally estimated with random effects (multilevel) models. Other than the fixed effects (regression) models that are often used for prognostic modeling, random effects models also take into account the clustering of patients within hospitals and countries.[18] These models facilitate estimation of unexplained outcome differences by enabling adjustment for differences in patient characteristics (i.e. case-mix, at patient level), as well as structure and process characteristics at hospital level. Structure characteristics relate to the organization of care in a hospital, e.g. the number of patients treated. Process characteristics concern treatment in individual patients. A decrease in between-center and between-country differences after correction for case-mix and structure or process characteristics indicates that variation in these factors affects patient outcomes.

Random effects models also account for random variation due to small sample sizes per hospital and country. However, estimates of between-center and between-country differences remain subject to substantial uncertainty. The smaller the sample size per hospital or country, the more uncertain the estimates for differences in clinical outcomes.[19]

Simulations

In short, simulations are computer experiments that involve creating data to reproduce a specific scenario, such as a RCT with a known treatment effect.[20] This simulated dataset can then be used to evaluate the power of the statistical approach required to analyze a new outcome measure, for example ordinal logistic or linear regression. A simulation study also facilitates comparison of new and existing outcome measures and different statistical approaches in the same clinical scenario.[20]

Besides being statistically efficient, new outcome measures should also facilitate interpretation of treatment effects. Treatment effects in clinical trials are currently often expressed on the odds ratio or hazard ratio scale, and researchers and clinicians are used to working with these scales. A new outcome measure should not complicate interpretation of trial results.

Acute neurological diseases

Acute neurological diseases have a heterogeneous disease course and are often associated with poor clinical outcomes, which stimulates measurement of clinical outcomes in terms of prognosis, variation across settings and new assessment methods. In this thesis, outcome prediction and outcome analyses are applied to three acute neurological diseases: ischemic stroke, aneurysmal subarachnoid hemorrhage and traumatic brain injury.

Ischemic stroke

Ischemic stroke occurs when a thrombus is blocking an intracranial artery. This type of stroke accounts for over 80% of all strokes and is a major cause of mortality and disability.[21] In 2017, over 29,000 patients were admitted to hospitals because of ischemic stroke in the Netherlands.[22] Disruption of the blood supply to the brain causes acute neurological deficits, including impaired speech, paresis of arms or legs, facial paralysis, visual loss or even coma. Atherosclerosis and cardioembolism are the main causes of ischemic stroke.[23]

Patients with ischemic stroke should be treated as soon as possible to recover blood flow to the brain (time = brain). Until recently, this could mainly be attempted with intravenous thrombolysis (IVT, administration of intravenous alteplase) within 4.5 hours after stroke onset to dissolve the thrombus blocking the vessel. Over the past five years, acute treatment for ischemic stroke has undergone major change.[24] Intra-arterial treatment (IAT, endovascular removal of the thrombus) within 6 hours after stroke onset has been proven effective for patients with a proximal anterior circulation occlusion in multiple RCTs.[25-30] Recent trials, although conducted in selected groups of patients with ischemic stroke, have shown that IAT is also beneficial within 16 or even 24 hours after "last seen well".[31, 32] However, trials present average treatment effects and benefit of IAT may vary among individual patients with ischemic stroke. This is an example of a clinical scenario where application of a prognostic model estimating individual benefit of IAT may support treatment decisions.[33]

The most widely used primary outcome measure in trials for acute stroke interventions is the modified Rankin Scale (mRS).[34, 35] The mRS is an ordinal scale ranging from 0 (no symptoms) to 6 (death) measuring the degree of disability or dependence in everyday life (Table 1.1).[2] The mRS is often assessed at 3 months after stroke onset, because most improvement in functional outcome is expected to occur within this time window.[34] Although IAT has improved functional outcome after ischemic stroke, many patients experience long-term neurological sequelae in terms of functional, cognitive and behavioral problems that require rehabilitation or nursing home care.[16, 36] Efficient hospital discharge planning is therefore essential.

Table 1.1. Modified Rankin Scale

Category	Interpretation
0	No symptoms at all
1	No significant disability despite symptoms; able to perform all usual activities
2	Slight disability; unable to perform all previous activities, but able to take care of self without assistance
3	Moderate disability; requiring some help, but able to walk without assistance
4	Moderately severe disability; unable to walk without assistance and unable to attend to own bodily needs without assistance
5	Severe disability; requiring constant nursing care and attention
6	Dead

Subarachnoid hemorrhage

Subarachnoid hemorrhage (SAH) is a type of hemorrhagic stroke and accounts for 5% of all strokes. In SAH, blood originating from an intracranial artery accumulates in the subarachnoid space. Of all spontaneous SAHs, 85% is caused by the rupture of an intracranial aneurysm and called an aneurysmal subarachnoid hemorrhage (aSAH).[37] aSAH often occurs in the working population (most patients are <60 years of age) and is associated with poor outcome, with mortality rates around 35%.[38, 39] This makes aSAH a disease with a major individual and economic health impact.[40] The key symptom for aSAH is a sudden-onset headache, described by patients as “the worst headache ever”.

Acute treatment for patients with aSAH consists of occlusion of the aneurysm to prevent rebleeding. This can be achieved by either endovascular coiling or neurosurgical clipping of the aneurysm. Coiling is a less invasive treatment than clipping, and is associated with better short-term outcomes in patients in good clinical condition with a ruptured aneurysm suitable for both interventions.[41] Besides rebleeding, other main complications in the acute phase after aSAH include vasospasm and delayed cerebral ischemia (DCI), and hydrocephalus.[37, 38] The main evidence-based options for medical treatment or prevention of complications after aSAH include administration of oral nimodipine and maintenance of euvolemia to prevent DCI, and drainage of cerebrospinal fluid in patients with hydrocephalus.[42] However, so far, many trials studying interventions to potentially prevent or treat complications after aSAH did not show any additional benefit.[42-45] Because aSAH has a heterogeneous disease course and evidence-based treatment options for complications after aSAH are scarce, it is expected that general management differs between hospitals and countries, which may likely impact on clinical outcomes.

Functional outcome after aSAH is often measured with either the mRS or the Glasgow Outcome Scale (GOS) (Table 1.1 and 1.2).[1, 2] Similar to the mRS, the GOS is an ordinal scale ranging from 1 (death) to 5 (good recovery). Survivors of aSAH often experience deficits on both functional and cognitive domains. Even if patients have made “good” functional recovery, deficits on the cognitive domain (e.g. problems with memory, executive function and language) may cause impaired quality of life for a minimum of 2-3 years after aSAH.[46]

Table 1.2. Glasgow Outcome Scale (Extended)

Category GOS	Category GOSE	Interpretation
1 = Dead	1 = Dead	Dead
2 = Vegetative state	2 = Vegetative state	Unable to interact with the environment, unresponsive
3 = Severe disability	3 = Lower severe disability	Full assistance in activities of daily living
	4 = Upper severe disability	Partial assistance in activities of daily living
4 = Moderate disability	5 = Lower moderate disability	Independent, but cannot resume work, school or all previous activities
	6 = Upper moderate disability	Some disability exists, but can partly resume work or previous activities
5 = Good recovery	7 = Lower good recovery	Minor physical or mental deficits that affect daily life
	8 = Upper good recovery	Full recovery with minor symptoms that do not affect daily life

Traumatic brain injury

Traumatic brain injury (TBI) is a leading cause of injury-related death and disability.[47, 48] In 2016, there were over 27 million new cases of TBI worldwide, with more than 46,000 new cases of TBI in the Netherlands.[47] In short, TBI is defined as an injury to the brain induced by an external force. The epidemiology of TBI has changed substantially over the past years, especially regarding age distribution and injury mechanism. Currently, the main causes of TBI are falls and motor vehicle road accidents.[47, 48]

TBI is a disease with substantial variation in pathophysiology, clinical presentation, and prognosis.[48] Clinical severity of TBI is currently classified according to the Glasgow Coma Scale (GCS). This is a scale for assessment of impaired consciousness based on eye, motor and verbal response ranging from 3 (unresponsive patient) to 15 (fully awake and oriented patient).[49] There are three categories of severity: mild (GCS 13-15), moderate (GCS 9-12) or severe (GCS 3-8) TBI. This thesis focuses mainly on patients with moderate and severe TBI. Age, clinical severity, intracranial abnormalities on brain computed tomography (CT), secondary insults (i.e. hypoxia and hypotension) and laboratory characteristics have been identified as prognostic factors for poor functional outcome in patients with moderate and severe TBI and.[10, 50, 51] Moreover, TBI is often accompanied by extracranial injuries.

Management of the primary injury and secondary brain damage, such as raised intracranial pressure due to swelling of the brain, may include medical or surgical treatment. As for aSAH, knowledge on the best treatment strategies for patients with TBI is scarce, because many trials on potentially effective interventions were inconclusive.[48, 52] Questionnaires among physicians from 71 European centers have shown that substantial between-hospital variation exists in treatment policies and organization of care.[53-58] Moreover, large differences have been observed between hospitals in clinical outcomes of TBI patients, which may be a reflection of the variation in treatment policies.[59]

Functional outcome after TBI is often scored according to the Glasgow Outcome Scale (GOS) ranging from 1 (death) to 5 (complete recovery), or the Glasgow Outcome Scale Extended (GOSE) which is a slightly more granular 8-point scale (Table 1.2).[1] TBI survivors often face a combination of physical,

psychiatric, emotional and cognitive disabilities. The variety in long-term impairments among individual patients requires personalized rehabilitation strategies delivered by a multidisciplinary team.[15, 48]

Data sources

Analyses in this thesis will mainly be based on data from a variety of clinical trials and observational cohort studies in acute neurological diseases (Table 1.3).

Prediction

The following data sources will be used for analyses on outcome prediction:

- Retrospective cohorts of aSAH patients admitted to the intensive care unit from two university hospitals in the Netherlands between 2006 and 2011.
- The Paracetamol (Acetaminophen) In Stroke (PAIS) study, Promoting Acute Thrombolysis in Ischemic StrokE (PRACTISE) study and Preventive Antibiotics in Stroke Study (PASS) conducted between 2003 and 2014. These trials were aimed at improving care for ischemic and/or hemorrhagic stroke patients by evaluating treatment and implementation strategies.[60-62]
- The Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) project. This is a prospective observational cohort study aimed at identifying best clinical care and improving characterization and classification of TBI.[63] Participants for the core study were recruited between December 2014 and December 2017 from 59 neurotrauma centers in 18 countries across Europe and Israel.

Outcome analyses

Random effects analyses regarding outcome differences across hospitals and countries will be based on a selection of data from the Subarachnoid Hemorrhage International Trialists (SAHIT) repository including multiple RCTs and observational studies in patients with aSAH.[64] Data from the Intraoperative Hypothermia during Surgery for Intracranial Aneurysm (IHAST), magnesium sulfate in aneurysmal subarachnoid hemorrhage (MASH) and Tirilazad mesylate in patients with aneurysmal subarachnoid hemorrhage (Tirilazad) trials conducted between 1991 and 2011 will be used.[65-69]

Simulations will be performed on data from the Multicenter Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in the Netherlands (MR CLEAN), designed to evaluate whether acute intra-arterial treatment (within 6 hours of symptom onset) plus usual care would be more effective than usual care alone in patients with ischemic stroke and a proximal arterial occlusion in the anterior cerebral circulation. Patients were recruited from 16 Dutch centers between December 2010 and March 2014.[25]

Table 1.3. Overview of data sources that will be used for analyses

Study	Number of patients used for analysis in this thesis	Design
Ischemic stroke		
PAIS	1227	RCT
PRACTISE	1589	Cluster RCT
PASS	2107	RCT
MR CLEAN	500	RCT
Aneurysmal subarachnoid hemorrhage		
Cohort Erasmus University Medical Center	307	Single-center retrospective observational cohort study
Combined cohort Erasmus University Medical Center and University Medical Center Groningen	285	Multicenter retrospective observational cohort study
Combined cohort based on data from studies in the SAHIT repository	5972	
- IHAST		RCT
- MASH		RCT
- Tirilazad		RCT
Traumatic brain injury		
CENTER-TBI	1742	Multicenter prospective observational cohort study

RCT, randomized clinical trial; PAIS, Paracetamol (Acetaminophen) In Stroke (Netherlands Trial Register, NTR2365); PRACTISE, Promoting Acute Thrombolysis in Ischemic StrokE (International Standard Randomised Controlled Trial Number (ISRCTN) registry ISRCTN20405426); PASS, Preventive Antibiotics in Stroke Study (ISRCTN registry ISRCTN66140176); Multicenter Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in the Netherlands (ISRCTN registry ISRCTN10888758); SAHIT, Subarachnoid Hemorrhage International Trialists; IHAST, Intraoperative Hypothermia during Surgery for Intracranial Aneurysm (NCT00029133); MASH, magnesium sulfate in aneurysmal subarachnoid hemorrhage (ISRCTN68742385 and NTR50); Tirilazad, Tirilazad mesylate in patients with aneurysmal subarachnoid hemorrhage; CENTER-TBI, Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (European Union FP 7th Framework program; grant 602150).

Aims and outline of this thesis

The overall aim of this thesis is to identify patients at high risk for poor outcome after acute neurological diseases and to enhance knowledge on outcome variation and statistical efficiency of new outcome measures.

Specific research questions are:

1. What characteristics are associated with poor outcome after acute neurological diseases?
2. What is the methodological quality of existing prognostic models in acute neurological diseases?
3. Do these models provide reliable predictions for patients in specific clinical settings?
4. What are the differences in clinical outcomes between patients with aSAH in a range of international hospitals, and can these differences be explained by variation in case-mix?
5. What is the statistical efficiency of new outcome measures for acute neurological diseases?

Part II of this thesis investigates different aspects of outcome prediction in acute neurological diseases and answers research questions 1-3. **Chapter 2** describes the association of early serum lactate and glucose levels with delayed cerebral ischemia and functional outcome after aSAH. **Chapter 3** aims to identify prognostic factors for disability and functional outcome early after ischemic stroke and describes the development of a prognostic model to support efficient discharge planning. An overview of contemporary models for prediction of functional outcome in patients with moderate and severe TBI is presented in **Chapter 4**. Related to this topic, **Chapter 4.1** contains a letter discussing the methodological quality of a newly developed model for long-term outcome after TBI. **Chapter 5** describes the external validation of a prognostic model for mortality after aSAH in a specific clinical setting. Additionally, the importance of external validation and updating of a clinical prediction model is shortly discussed in **Chapter 5.1**. **Chapter 6** describes the performance and potential applications of the most widely known prognostic models for functional outcome after moderate and severe TBI in a contemporary European cohort.

Part III focuses on the analyses of clinical outcomes and answers research questions 4 and 5. In **Chapter 7**, random effects modeling is used to assess the presence and magnitude of differences in functional outcome after aSAH between hospitals and countries in a large repository consisting of multiple RCTs and observational studies. In ischemic stroke, a new outcome measure incorporating both functional outcome and quality of life has been proposed called the utility-weighted mRS. **Chapter 8** describes a simulation study evaluating the statistical efficiency of this outcome measure. In response to a discussion initiated by the founders of the UW-mRS, the importance of critically studying the statistical efficiency and interpretability of a new outcome measure is emphasized in **Chapter 8.1**.

Part IV summarizes the main findings of this thesis. **Chapter 9** consists of a discussion of the results of previous chapters and provides recommendations for future studies and clinical practice.

References

1. Wilson JT, Pettigrew LE, Teasdale GM. Structured interviews for the Glasgow Outcome Scale and the Extended Glasgow Gutcome Scale: Guidelines for their use. *J Neurotrauma*. 1998;15:573-585
2. van Swieten JC, Koudstaal PJ, Visser MC, Schouten HJ, van Gijn J. Interobserver agreement for the assessment of handicap in stroke patients. *Stroke*. 1988;19:604-607
3. Euroqol Group. Euroqol – a new facility for the measurement of health-related quality of life. *Health Policy*. 1990;16:199-208
4. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992;30:473-483
5. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: What, why, and how? *BMJ*. 2009;338:b375
6. Jefford M, Stockler MR, Tattersall MHN. Outcomes research: What is it and why does it matter? *Internal Medicine Journal*. 2003;33:110-118
7. Huijben JA, Wiegers EJA, de Keizer NF, Maas AIR, Menon D, Ercole A, et al. Development of a quality indicator set to measure and improve quality of ICU care for patients with traumatic brain injury. *Crit Care*. 2019;23:95-95
8. Steyerberg EW, Moons KGM, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis research strategy (PROGRESS) 3: Prognostic model research. *PLoS Med*. 2013;10:e1001381-e1001381
9. Riley RD, Hayden JA, Steyerberg EW, Moons KGM, Abrams K, Kyzas PA, et al. Prognosis research strategy (PROGRESS) 2: Prognostic factor research. *PLoS Med*. 2013;10:e1001380-e1001380
10. Perel PA, Olldashi F, Muzha I, Filipi N, Lede R, Copertari P, et al. Predicting outcome after traumatic brain injury: Practical prognostic models based on large cohort of international patients. *BMJ*. 2008;336:425-429
11. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. *Ann Int Med*. 2015;162:W1-W73
12. Steyerberg EW. Validation of prediction models. *In: Clinical prediction models: A practical approach to development, validation and updating.* New York: Springer, 2009. P. 299-310.
13. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: Seven steps for development and an ABCD for validation. *European Heart Journal*. 2014;35:1925-1931
14. Berzina G, Sveen U, Paanalahti M, Sunnerhagen KS. Analyzing the modified Rankin Scale using concepts of the international classification of functioning, disability and health. *Eur J Phys Rehabil Med*. 2016;52:203-213
15. Stocchetti N, Zanier ER. Chronic impact of traumatic brain injury on outcome and quality of life: A narrative review. *Crit Care*. 2016;20:148-148
16. Kapoor A, Lanctot KL, Bayley M, Kiss A, Herrmann N, Murray BJ, et al. "Good outcome" isn't good enough: Cognitive impairment, depressive symptoms, and social restrictions in physically recovered stroke patients. *Stroke*. 2017;48:1688-1690
17. Black N. Patient reported outcome measures could help transform healthcare. *BMJ*. 2013;346:f167

18. Merlo J, Chaix B, Ohlsson H, Beckman A, Johnell K, Hjerpe P, et al. A brief conceptual tutorial of multilevel analysis in social epidemiology: Using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *J Epidemiol Community Health*. 2006;60:290-297
19. Lingsma HF, Steyerberg EW, Eijkemans MJC, Dippel DWJ, Scholte Op Reimer WJM, Van Houwelingen HC, et al. Comparing and ranking hospitals based on outcome: Results from the Netherlands Stroke Survey. *QJM*. 2010;103:99-108
20. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. *Stat Med*. 2019;38:2074-2102
21. Valery LF, Bo N, George AM. Global burden of stroke. *Circulation Research*. 2017;120:439-448
22. de Boer AR, van Dis I, Visseren FLJ, Vaartjes I, Bots ML. Ziekte en sterfte aan hart- en vaatziekten. *In: de Boer AR, Bots ML, van Dis I, Vaartjes I, Visseren FLJ. Hart- en vaatziekten in Nederland 2018, cijfers over risicofactoren, hartinterventies, ziekte en sterfte.* Den Haag: Hartstichting, 2018.
23. van der Worp HB, van Gijn J. Acute ischemic stroke. *N Engl J Med*. 2007;357:572-579
24. Zerna C, Thomalla G, Campbell BCV, Rha JH, Hill MD. Current practice and future directions in the diagnosis and acute treatment of ischaemic stroke. *Lancet*. 2018;392:1247-1256
25. Berkhemer OA, Fransen PS, Beumer D, van den Berg LA, Lingsma HF, Yoo AJ, et al. A randomized trial of intraarterial treatment for acute ischemic stroke. *N Engl J Med*. 2015;372:11-20
26. Campbell BCV, Mitchell PJ, Kleinig TJ, Dewey HM, Churilov L, Yassi N, et al. Endovascular therapy for ischemic stroke with perfusion-imaging selection. *N Engl J Med*. 2015;372:1009-1018
27. Goyal M, Demchuk AM, Menon BK, Eesa M, Rempel JL, Thornton J, et al. Randomized assessment of rapid endovascular treatment of ischemic stroke. *N Engl J Med*. 2015;372:1019-1030
28. Jovin TG, Chamorro A, Cobo E, de Miquel MA, Molina CA, Rovira A, et al. Thrombectomy within 8 hours after symptom onset in ischemic stroke. *N Engl J Med*. 2015;372:2296-2306
29. Saver JL, Goyal M, Bonafe A, Diener H-C, Levy EI, Pereira VM, et al. Stent-retriever thrombectomy after intravenous t-PA vs. t-PA alone in stroke. *N Engl J Med*. 2015;372:2285-2295
30. Goyal M, Menon BK, van Zwam WH, Dippel DW, Mitchell PJ, Demchuk AM, et al. Endovascular thrombectomy after large-vessel ischaemic stroke: A meta-analysis of individual patient data from five randomised trials. *Lancet*. 2016;387:1723-1731
31. Albers GW, Marks MP, Kemp S, Christensen S, Tsai JP, Ortega-Gutierrez S, et al. Thrombectomy for stroke at 6 to 16 hours with selection by perfusion imaging. *N Engl J Med*. 2018;378:708-718
32. Nogueira RG, Jadhav AP, Haussen DC, Bonafe A, Budzik RF, Bhuva P, et al. Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct. *N Engl J Med*. 2018;378:11-21
33. Venema E, Mulder MJHL, Roozenbeek B, Broderick JP, Yeatts SD, Khatri P, et al. Selection of patients for intra-arterial treatment for acute ischaemic stroke: Development and validation of a clinical decision tool in two randomised trials. *BMJ*. 2017;357:j1710
34. Lees KR, Bath PM, Schellinger PD, Kerr DM, Fulton R, Hacke W, et al. Contemporary outcome measures in acute stroke research: Choice of primary outcome measure. *Stroke*. 2012;43:1163-1170
35. Lees KR, Khatri P, STAIR IX Collaborators. Stroke treatment academic industry roundtable recommendations for individual data pooling analyses in stroke. *Stroke*. 2016;47:2154-2159

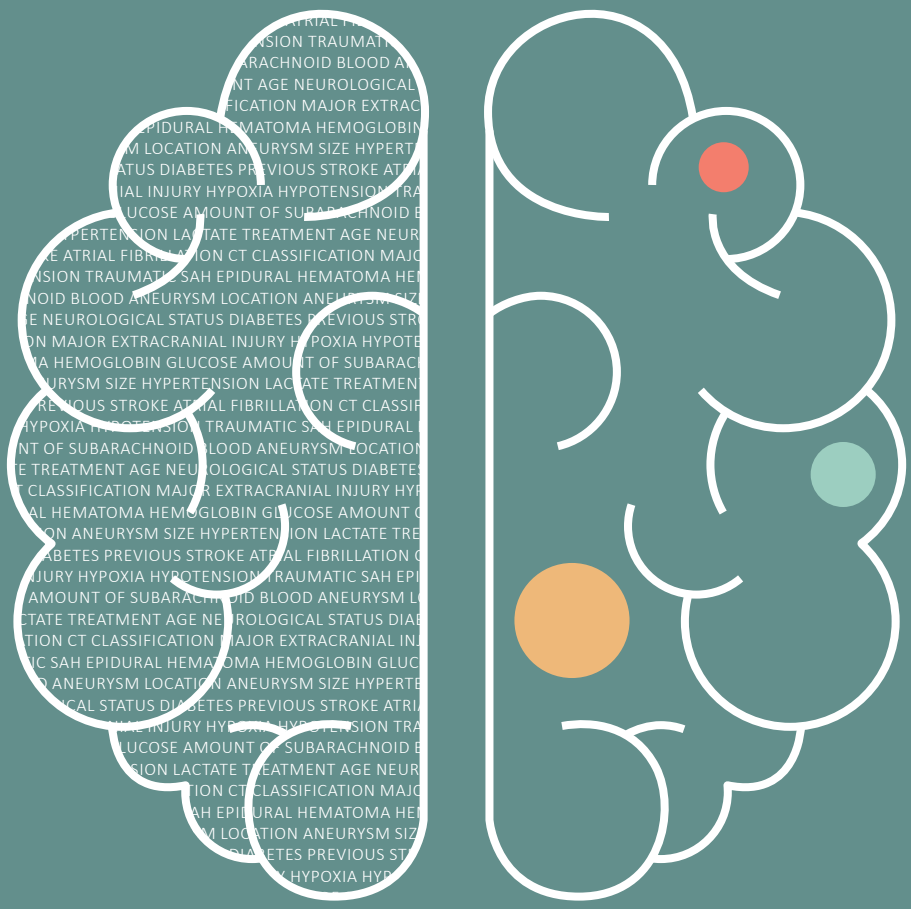
36. Mellon L, Brewer L, Hall P, Horgan F, Williams D, Hickey A, et al. Cognitive impairment six months after ischaemic stroke: A profile from the ASPIRE-S study. *BMC Neurol*. 2015;15:31
37. Macdonald RL, Schweizer TA. Spontaneous subarachnoid haemorrhage. *Lancet*. 2017;389:655-666
38. van Gijn J, Kerr RS, Rinkel GJ. Subarachnoid haemorrhage. *Lancet*. 2007;369:306-318
39. Rinkel GJE, Algra A. Long-term outcomes of patients with aneurysmal subarachnoid haemorrhage. *Lancet Neurol*. 2011;10:349-356
40. Ridwan S, Urbach H, Greschus S, von Hagen J, Esche J, Bostrom A. Health care costs of spontaneous aneurysmal subarachnoid hemorrhage for rehabilitation, home care, and in-hospital treatment for the first year. *World Neurosurg*. 2017;97:495-500
41. Lindgren A, Vergouwen MDI, van der Schaaf IC, Algra A, Wermer MJH, Clarke MJ, et al. Endovascular coiling versus neurosurgical clipping for people with aneurysmal subarachnoid hemorrhage. *Stroke*. 2019;50:e102
42. Connolly ES, Jr., Rabinstein AA, Carhuapoma JR, Derdeyn CP, Dion J, Higashida RT, et al. Guidelines for the management of aneurysmal subarachnoid hemorrhage: A guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*. 2012;43:1711-1737
43. Gathier CS, van den Bergh WM, van der Jagt M, Verweij BH, Dankbaar JW, Muller MC, et al. Induced hypertension for delayed cerebral ischemia after aneurysmal subarachnoid hemorrhage: A randomized clinical trial. *Stroke*. 2018;49:76-83
44. Macdonald RL. Delayed neurological deterioration after subarachnoid haemorrhage. *Nat Rev Neurol*. 2014;10:44-58
45. Steiner T, Juvela S, Unterberg A, Jung C, Forsting M, Rinkel G, et al. European Stroke Organization guidelines for the management of intracranial aneurysms and subarachnoid haemorrhage. *Cerebrovasc Dis*. 2013;35:93-112
46. Timour A-K, Macdonald RL, Tom AS. Cognitive and functional outcome after aneurysmal subarachnoid hemorrhage. *Stroke*. 2010;41:e519-e536
47. Injury GBDTB, Spinal Cord Injury C. Global, regional, and national burden of traumatic brain injury and spinal cord injury, 1990-2016: A systematic analysis for the global burden of disease study 2016. *Lancet Neurol*. 2019;18:56-87
48. Maas AIR, Menon DK, Adelson PD, Andelic N, Bell MJ, Belli A, et al. Traumatic brain injury: Integrated approaches to improve prevention, clinical care, and research. *Lancet Neurol*. 2017;16:987-1048
49. Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. *Lancet*. 1974;2:81-84
50. Lingsma HF, Roozenbeek B, Steyerberg EW, Murray GD, Maas AI. Early prognosis in traumatic brain injury: From prophecies to predictions. *Lancet Neurol*. 2010;9:543-554
51. Steyerberg EW, Mushkudiani N, Perel P, Butcher I, Lu J, McHugh GS, et al. Predicting outcome after traumatic brain injury: Development and international validation of prognostic scores based on admission characteristics. *PLoS Med*. 2008;5:1251-1261
52. Carney N, Totten AM, O'Reilly C, Ullman JS, Hawryluk GW, Bell MJ, et al. Guidelines for the management of severe traumatic brain injury, fourth edition. *Neurosurgery*. 2017;80:6-15
53. Cnossen MC, Polinder S, Lingsma HF, Maas AI, Menon D, Steyerberg EW, 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*. 2016;11:e0161367

54. Huijben JA, van der Jagt M, Cnossen MC, Kruip M, Haitsma IK, Stocchetti N, et al. Variation in blood transfusion and coagulation management in traumatic brain injury at the intensive care unit: A survey in 66 neurotrauma centers participating in the collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury study. *J Neurotrauma*. 2017
55. van Essen TA, den Boogert HF, Cnossen MC, de Ruiter GCW, Haitsma I, Polinder S, et al. Variation in neurosurgical management of traumatic brain injury: A survey in 68 centers participating in the CENTER-TBI study. *Acta Neurochir (Wien)*. 2018
56. Volovici V, Ercole A, Citerio G, Stocchetti N, Haitsma IK, Huijben JA, et al. Intensive care admission criteria for traumatic brain injury patients across Europe. *J Crit Care*. 2019;49:158-161
57. Volovici V, Ercole A, Citerio G, Stocchetti N, Haitsma IK, Huijben JA, et al. Variation in guideline implementation and adherence regarding severe traumatic brain injury treatment: A CENTER-TBI survey study in Europe. *World Neurosurg*. 2019
58. Foks KA, Cnossen MC, Dippel DWJ, Maas A, Menon D, van der Naalt J, 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. *J Neurotrauma*. 2017
59. Lingsma HF, Roozenbeek B, Li B, Lu J, Weir J, Butcher I, 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*. 2011;68:601-607; discussion 607-608
60. den Hertog HM, van der Worp HB, van Gemert HM, Algra A, Kappelle LJ, van Gijn J, et al. The paracetamol (acetaminophen) in stroke (PAIS) trial: A multicentre, randomised, placebo-controlled, phase III trial. *Lancet Neurol*. 2009;8:434-440
61. Dirks M, Niessen LW, van Wijngaarden JD, Koudstaal PJ, Franke CL, van Oostenbrugge RJ, et al. Promoting thrombolysis in acute ischemic stroke. *Stroke*. 2011;42:1325-1330
62. Westendorp WF, Vermeij JD, Zock E, Hooijenga IJ, Kruyt ND, Bosboom HJ, et al. The preventive antibiotics in stroke study (PASS): A pragmatic randomised open-label masked endpoint clinical trial. *Lancet*. 2015;385:1519-1526
63. Maas AI, Menon DK, Steyerberg EW, Citerio G, Lecky F, Manley GT, et al. Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI): A prospective longitudinal observational study. *Neurosurgery*. 2015;76:67-80
64. Jaja BN, Attalla D, Macdonald RL, Schweizer TA, Cusimano MD, Etminan N, et al. The subarachnoid hemorrhage international trialists (SAHIT) repository: Advancing clinical research in subarachnoid hemorrhage. *Neurocrit Care*. 2014;21:551-559
65. Dorhout Mees SM, Algra A, Vandertop WP, van Kooten F, Kuijsten HA, Boiten J, et al. Magnesium for aneurysmal subarachnoid haemorrhage (MASH-2): A randomised placebo-controlled trial. *Lancet*. 2012;380:44-49
66. Haley EC, Jr., Kassell NF, Apperson-Hansen C, Maile MH, Alves WM. A randomized, double-blind, vehicle-controlled trial of tirilazad mesylate in patients with aneurysmal subarachnoid hemorrhage: A cooperative study in North America. *J Neurosurg*. 1997;86:467-474

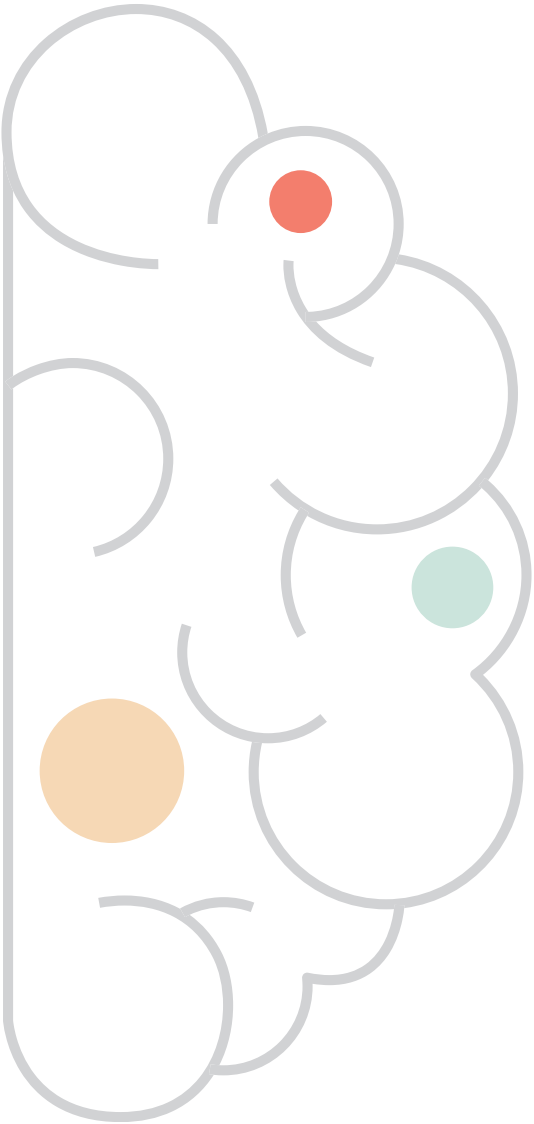
67. Kassell NF, Haley EC, Jr., Apperson-Hansen C, Alves WM. Randomized, double-blind, vehicle-controlled trial of tirilazad mesylate in patients with aneurysmal subarachnoid hemorrhage: A cooperative study in Europe, Australia, and New Zealand. *J Neurosurg*. 1996;84:221-228
68. Todd MM, Hindman BJ, Clarke WR, Torner JC, Intraoperative Hypothermia for Aneurysm Surgery Trial Investigators. Mild intraoperative hypothermia during surgery for intracranial aneurysm. *N Engl J Med*. 2005;352:135-145
69. van den Bergh WM, Algra A, van Kooten F, Dirven CM, van Gijn J, Vermeulen M, et al. Magnesium sulfate in aneurysmal subarachnoid hemorrhage: A randomized controlled trial. *Stroke*. 2005;36:1011-1015

PART II

PREDICTION



CHAPTER 2



Early circulating lactate and glucose levels after aneurysmal subarachnoid hemorrhage correlate with poor outcome and delayed cerebral ischemia: A two-center cohort study

Simone A. Dijkland
Carlina E. van Donkelaar
Walter M. van den Bergh
Jan Bakker
Diederik W.J. Dippel
Maarten W.M. Nijsten
Mathieu van der Jagt

Crit Care Med 2016; 44(5): 966–972

Abstract

Objective: In critically ill patients, elevated blood lactate at admission is associated with poor outcome, but after aneurysmal subarachnoid hemorrhage, this has not been investigated. We studied the association between early circulating lactate and glucose with delayed cerebral ischemia and poor outcome. Lactate and glucose were both studied, hypothesizing that both may be increased due to sympathetic activation after subarachnoid hemorrhage similar to critically ill patients.

Design: Retrospective cohort study.

Setting: ICUs of two academic hospitals in the Netherlands.

Patients: Patients with aneurysmal subarachnoid hemorrhage admitted to the ICU within 24 hours after the bleed surviving beyond 48 hours after ICU admission and who had at least one lactate measurement within 24 hours after admission.

Interventions: None.

Measurements and main results: In 285 patients, maximal lactate and glucose levels within the first 24 hours after admission were determined. Early lactate and glucose were related with delayed cerebral ischemia–related infarction and poor outcome (a modified Rankin Scale score of 4, 5, or death at 3 mo). Delayed cerebral ischemia occurred in 84 patients (29%), and 106 patients (39%) had poor outcome. Multivariable analyses were performed with adjustment of established predictors for delayed cerebral ischemia and outcome: age, sex, World Federation of Neurological Surgeons grade at admission and Hijdra sum scores. Early lactate and glucose were strongly related (Spearman $\rho = 0.55$; $p < 0.001$). Lactate and glucose were both independently associated with delayed cerebral ischemia and poor outcome in multivariable analyses with either lactate or glucose as covariates. When both lactate and glucose were included, only glucose showed an independent association with delayed cerebral ischemia (odds ratio, 1.14; 95% CI, 1.01–1.28) and only lactate showed an independent association with poor outcome (odds ratio, 1.42; 95% CI, 1.11–1.81).

Conclusions: Early lactate and glucose levels after aneurysmal subarachnoid hemorrhage are associated with delayed cerebral ischemia and poor outcome, suggesting that they may be considered in conjunction with other parameters for future prognostic models.

Introduction

Subarachnoid hemorrhage (SAH) caused by a ruptured intracranial aneurysm is a devastating cause of stroke.[1,2] Delayed cerebral ischemia (DCI) occurs in about one third of the patients and is the leading cause of disability and death in patients who survive the first 24 hours.[3] The exact underlying pathophysiological mechanisms of DCI remain obscure, but multifocal cerebral hypoperfusion is considered a final common pathway.[4,5] Prognostic factors for DCI and functional outcome after SAH have been studied, but clinical predictors that are readily available at admission after aneurysmal SAH and are not subject to interobserver variability, such as scoring systems for the amount of subarachnoid blood on CT, are less well established.[6-8] Easily obtainable biomarkers at admission may help early risk assessment of a complicated course and may provide further insights into pathophysiological mechanisms when such factors have a causal link to the outcome.[9]

In critically ill patients, lactate levels are firmly associated with adverse outcomes.[10,11] Although accumulation of cerebral tissue lactate has been associated with poor neurological outcome in patients with SAH and other types of brain injury,[12,13] the prognostic value of blood lactate levels in SAH patients, which are more easily available than brain lactate, has not been investigated. In contrast, several studies have shown that circulating glucose is related with outcome in SAH.[14-16] Lactate and glucose are two key metabolites that are intimately connected: first, because glucose is a direct precursor of lactate; second, because various stress conditions can increase the circulating levels of both lactate and glucose.[17] Indicators of sympathetic stress have been associated with both increased lactate in critically ill patients[18] and DCI and poor outcome after SAH.[19-24]

The objective of this study was to determine whether early increases in circulating lactate and glucose levels are associated with DCI and poor outcome after aneurysmal SAH.

Methods

Study design and population

In this retrospective cohort study, we included adult patients with aneurysmal SAH admitted to the ICUs of two university hospitals in the Netherlands (University Medical Center Groningen and Erasmus Medical Center Rotterdam). Patients with SAH were identified by disease codes as registered in the Dutch National Intensive Care Evaluation or the International Classification of Diseases code retrieved from the hospital's patient registry, indicating SAH in the period between November 2006 and December 2011. Retrieval of subjects was crosschecked with the ICU Patient Data Management System. In the National Intensive Care Evaluation registry database, patient characteristics, presence of chronic disease and comorbidity, reason for admission, disease, ICU course, and outcome characteristics are prospectively collected.[25]

Inclusion criteria were 1) 18 years old or older, 2) admitted to ICU within 24 hours after the initial bleed, 3) at least one lactate and glucose measurement available within 24 hours after admission, 4)

SAH, proven by CT or cerebrospinal fluid spectrophotometry, and 5) ruptured intracranial aneurysm as the presumed cause of spontaneous SAH, preferably demonstrated by digital subtraction angiography or CT angiography.

Patients who met any of the following criteria were not eligible 1) nonaneurysmal (e.g., perimesencephalic or traumatic) SAH, 2) death less than 48 hours after admission, 3) pregnancy, 4) no CT scan on admission available. Patients dying within 48 hours after admission were excluded because these patients frequently had dismal prognosis soon after admission and inclusion in analyses on DCI and outcome was not considered as relevant.

During admission, included patients in both centers were treated according to a standardized protocol that consisted of absolute bed rest until aneurysm treatment, oral doses of nimodipine, cessation of antihypertensive medication, and IV administration of fluid with the aim of normovolemia.

Because this study only involved the anonymized retrospective evaluation of clinical and laboratory parameters acquired during routine clinical care, informed consent was waived as approved by the institutional Medical Ethics Committee of both centers.

Data collection and outcomes

The method of aneurysm treatment (endovascular coiling, neurosurgical clipping, or no treatment) was collected from the electronic patient record at each hospital. The amount of blood at admission CT scans was evaluated using Hijdra sum scores, ranging from 0 to 30 for cisternal amount of blood and from 0 to 12 for ventricular amount of blood.[26] Neurological condition at admission was assessed by the World Federation of Neurological Surgeons (WFNS) grade.[27] Poor neurological condition at admission was defined as WFNS grade 4 or 5.

All blood lactate and glucose levels within the first 24 hours after admission were collected at both hospitals. If more than one measurement was performed during this period, the highest level was used for all analyses and was referred to as “maximum lactate” and “maximum glucose”.[17]

The two main outcomes were DCI defined as a new hypodensity on CT not otherwise explained than by cerebral infarction due to DCI within 30 days after admission, according to earlier proposed definitions,[28] and poor outcome according to the modified Rankin Scale (mRS).

Day of DCI occurrence was the day of brain CT at which the new hypodensity was detected, or the day of clinical symptoms if this obviously occurred the day before a brain CT was performed. The mRS, measuring the degree of dependence or disability in daily activities, was retrieved from the electronic patient record or from the primary care physician and assessed at 3 months after SAH. Poor outcome was defined as an mRS score of 4, 5, or death.

Statistical analysis

Patient baseline characteristics are presented as medians with interquartile range (IQR) for continuous variables and frequencies (percentage) for categorical variables. After testing for normality, continuous variables were analyzed using the unpaired Student t test (normal distribution) or Mann-Whitney U test. Differences between categorical variables were assessed with a chi-square or Fisher exact test.

The association between maximum lactate and glucose levels within the first 24 hours after SAH and DCI was assessed with logistic regression analysis, adjusted for established predictors for DCI-related infarction (age, sex, clinical condition at admission [WFNS grade]), and the amount of subarachnoid blood (cisternal and ventricular Hijdra sum scores). Analysis was similarly performed for poor outcome. Ordinal variables (Hijdra sum scores) were dichotomized at their median, and clinical condition at admission was dichotomized in good (WFNS, 1–3) and poor (WFNS, 4–5) grades, whereas continuous variables were used unaltered for the analyses. Results are presented as odds ratios (ORs) with corresponding 95% CI. Receiver operating characteristic curves with corresponding area under the curve (AUC) and diagnostic test values (sensitivity, specificity, positive predictive value, and negative predictive values [PPV/NPV]) based on the median values of lactate and glucose in all included patients were calculated.

Statistical analyses were performed using SPSS (Statistical Package for Social Sciences, version 22). A p value of less than 0.05 was considered statistically significant for all analyses.

Results

Study population

After exclusion of 240 patients according to exclusion criteria, 285 patients were eligible for analyses (Figure 2.1). Patients with lacking lactate or glucose data (n = 112) more often ($p < 0.001$) had a lower WFNS (corresponding to better neurological status at admission), and less ventricular blood ($p = 0.019$), but did not significantly differ with regard to sex, age, or cisternal amount of blood on initial CT. Data for DCI were complete for all 285 patients. Ten patients had nonretrievable data on mRS. DCI-related infarction after SAH occurred in 84 patients (29%), and 106 patients (39%) had poor outcome. Outcome was assessed at a mean of 3.3 months after admission (sd ± 1.0). Baseline demographic and clinical characteristics are presented in Table 2.1. The medians of the collected maximum values were 1.6 mmol/L (IQR, 1.0–2.7) for lactate and 9.3 mmol/L (IQR, 8.0–11.1) for glucose.

The median number of measurements during the first 24 hours after admission was 4 (IQR, 2–6) for lactate and 5 (IOR, 3–7) for glucose. The median time to occurrence of DCI was 6 days (IQR, 4–11 d) after SAH. Patients who developed DCI had a significantly higher maximum lactate level during the first 24 hours after admission than patients without DCI (2.1 mmol/L [IQR, 1.2–3.1 mmol/L] vs 1.5 mmol/L [IQR, 1.0–2.5 mmol/L]; $p = 0.006$) (Table 2.2). Patients who developed DCI also had a higher maximum glucose level (10.3 mmol/L [IQR, 8.6–11.8 mmol/L] vs 9.1 mmol/L [IQR, 7.8–10.7 mmol/L]; $p = 0.002$) (Table 2.2). Patients with poor outcome had a higher lactate level during the first 24 hours after admission than patients with good outcome (2.2 mmol/L [IQR, 1.3–3.1 mmol/L] vs 1.4 mmol/L [IQR, 0.9–2.3 mmol/L]; $p < 0.001$), which was also seen for glucose (10.4 mmol/L [IQR, 8.7–12.2 mmol/L] vs 8.9 mmol/L [IQR, 7.6–10.1 mmol/L]; $p < 0.001$) (Table 2.2). A substantial correlation existed between lactate and glucose levels (Spearman $\rho = 0.55$; $p < 0.001$).

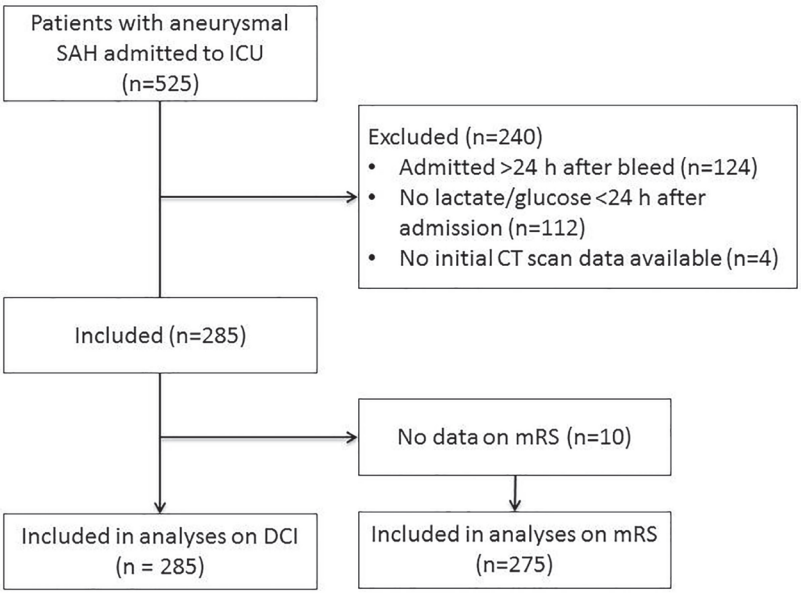


Figure 2.1. Patient flow of included subjects according to inclusion and exclusion criteria.
DCI, infarction caused by delayed cerebral ischemia; mRS, modified Rankin Scale; SAH, subarachnoid hemorrhage.

Table 2.1. Baseline characteristics of study population (n = 285)

Baseline variable	Value (Median or %)
Age, median (IQR)	55 (47-65)
Female sex (%)	189 (66)
Poor clinical condition on admission (World Federation of Neurological Surgeons grade ≥ 4) (%)	141 (49)
Aneurysm treatment	
Endovascular coiling	154 (54)
Neurosurgical clipping	80 (28)
None	51 (19)
Amount of subarachnoid blood, median (IQR)	
Cisternal Hijdra score	21 (12.5-29.0)
Ventricular Hijdra score	3.0 (1.0-6.0)
Maximum lactate[a] within first 24 hr after SAH, median (IQR)	1.6 (1.0-2.7)
Maximum glucose[a] within first 24 hr after ASH, median (IQR)	9.3 (8.0-1.1)

IQR, interquartile range; SAH, subarachnoid hemorrhage.

[a]Unit of measurement mmol/L.

Table 2.2. Medians of maximum lactate and glucose during the first 24 hours after admission related to delayed cerebral ischemia and outcome

Variable	DCI (n = 84; 29%)	No DCI (n = 201; 71%)	p[a]
Maximum lactate,[b] median (IQR)	2.1 (1.2-3.1)	1.5 (1.0-2.5)	0.006
Maximum glucose,[b] median (IQR)	10.3 (8.6-11.8)	9.1 (7.8-10.7)	0.002
Variable	**Poor outcome (n = 106; 39%)**	**Good outcome (n = 169; 61%)**	**p[a]**
Maximum lactate, median (IQR)	2.2 (1.3-3.1)	1.4 (0.9-2.3)	< 0.001
Maximum glucose, median (IQR)	10.4 (8.7-12.2)	8.9 (7.6-10.1)	< 0.001

DCI, delayed cerebral ischemia-related infarction on cerebral CT; IQR, interquartile range.

[a]Mann-Whitney U test.

[b]Unit of measurement: mmol/L.

Main outcomes

Maximum lactate during the first 24 hours after admission was associated with higher risk for DCI (OR, 1.33; 95% CI 1.12–1.58), which persisted after adjustment for known predictors (OR, 1.25; 95% CI, 1.04–1.51). Higher lactate levels were also associated with a higher risk for poor outcome (OR, 1.52; 95% CI, 1.25–1.85 and adjusted OR, 1.56; 95% CI, 1.25–1.94) (Table 2.3). The association between maximum glucose and DCI was significant in both univariable and multivariable analyses (Table 2.3). In multivariable analysis with both glucose and lactate levels as independent variables in the model, only glucose was independently associated with DCI (OR, 1.14; 95% CI, 1.02–1.28). In contrast, in multivariable analysis with both glucose and lactate levels as independent variables in the model for outcome, only lactate was independently associated with poor outcome (OR, 1.42; 95% CI, 1.11–1.81) (Table 2.4). Age was associated with decreased risk of DCI (OR, 0.97; 95% CI, 0.96–0.99) and increased risk for poor outcome (OR, 1.04; 95% CI, 1.02–1.07).

In a sensitivity analysis using the mean values instead of maximum values of lactate and glucose, the associations found did not change (data not shown).

The receiver-operating characteristic curves and corresponding AUCs are shown in Appendix 2.A. AUCs of early lactate and glucose for DCI were 0.60 ($p = 0.006$) and 0.62 ($p = 0.002$), respectively, and for poor outcome 0.68 for both lactate and glucose ($p < 0.001$). For lactate (cutoff value at the median) sensitivity, specificity, PPV and NPV were 58%, 44%, 36%, and 76% for DCI and 64%, 39%, 64%, and 73% for poor outcome; for glucose (cutoff value at the median) sensitivity, specificity, PPV, and NPV were 60%, 44%, 37%, and 77% for DCI and 64%, 38%, 64%, and 73% for poor outcome.

Table 2.3. Univariable and multivariable associations of either lactate or glucose during the first 24 hours after admission with delayed cerebral ischemia and poor outcome

	Odds ratio (95% CI)	
Outcomes and characteristics	**Univariable**	**Multivariable[a]**
Delayed cerebral ischemia-related infarction on cerebral CT (n = 285)		
Maximum lactate (per 1-mmol/L increase)	1.33 (1.12-1.58)	1.25 (1.04-1.51)
Maximum glucose (per 1-mmol/L increase)	1.19 (1.08-1.31)	1.17 (1.05-1.30)
Poor outcome (n = 275)		
Maximum lactate (per 1-mmol/L increase)	1.52 (1.25-1.85)	1.56 (1.25-1.94)
Maximum glucose (per 1-mmol/L increase)	1.25 (1.13-1.39)	1.20 (1.07-1.34)

[a]Ajusted for age, sex, clinical condition at admission, and amount of cisternal and ventricular blood.

Table 2.4. Multivariable analyses with both glucose and lactate within 24 hours after admission as independent variables in the model: association with delayed cerebral ischemia or poor outcome

Outcomes and characteristics	β	Associations, odds ratio (95% CI)	p
Delayed cerebral ischemia (n = 285)			
Age (yr)	-0.036	0.97 (0.94-0.99)	0.003
Sex	0.303	1.35 (0.75-2.42)	0.307
WFNS grade	0.235	1.27 (0.69-2.31)	0.445
Cisternal Hijdra score	0.337	1.40 (0.80-2.45)	0.236
Ventricular Hijdra score	-0.129	0.88 (0.49-1.57)	0.662
Maximum lactate (per 1-mmol/L increase)	0.116	1.12 (0.91-1.38)	0.269
Maximum glucose (per 1-mmol/L increase)	0.131	1.14 (1.01-1.28)	0.027
Outcome (n = 275)			
Age (yr)	0.042	1.04 (1.02-1.07)	<0.001
Sex	0.615	0.54 (0.30-0.99)	0.045
WFNS grade	0.489	1.63 (0.90-2.65)	0.106
Cisternal Hijdra score	0.584	1.79 (1.03-3.14)	0.041
Ventricular Hijdra score	0.402	1.49 (0.84-2.65)	0.171
Maximum lactate (per 1-mmol/L increase)	0.349	1.42 (1.11-1.81)	0.005
Maximum glucose (per 1-mmol/L increase)	0.103	1.11 (0.98-1.26)	0.112

WFNS, World Federation of Neurological Surgeons.

Discussion

The main findings of our study are that maximum lactate and glucose levels early after aneurysmal SAH are associated with both an increased risk of DCI-related cerebral infarction and poor outcome. Lactate and glucose were strongly related. When lactate and glucose were simultaneously entered in the multivariable analysis, only lactate emerged as an independent predictor of poor outcome and only glucose emerged as an independent predictor of DCI. To our knowledge, we are the first to report the association of blood lactate and poor outcome after SAH.

Relationship with previous literature

Because catecholamine levels (epinephrine/norepinephrine) have a prognostic value in patients with SAH,[29] our findings suggest that lactate and glucose levels may rise as a consequence of increased stress. Sympathetic activation in patients in the acute phase of SAH reflects the severity of SAH and is related to the development of DCI and consequently poor outcome.[30] Excessive release of catecholamines has also been suggested to be the principal cause of neurogenic pulmonary edema and cardiac dysfunction after SAH.[19-21] Cardiac dysfunction is a risk factor for poor clinical outcome after SAH, which is partly explained by a higher risk for DCI.[23] Likewise, prolonged elevated heart rate due to sympathetic

activation is associated with major adverse cardiopulmonary events and higher risk of DCI after SAH, whereas lower heart rate has been associated with lower incidence of DCI.[22,24] The excessive release of catecholamines in the acute phase of SAH might thus be a plausible explanation for the increased lactate levels during the first 24 hours after SAH with both DCI and poor neurological outcome.

When lactate levels were considered in the analysis for the association between maximum glucose and outcome, glucose ceased to be independently associated with outcome. This interaction between lactate and glucose has been shown previously in critically ill patients with adrenergic stress.[17] In a recent prospective randomized trial in 497 patients who received either placebo or dexamethasone before cardiac surgery, we have demonstrated that the glucocorticoid component of stress can also induce increases not only in glucose but also in lactate levels.[31] Therefore, our findings may complement the notion that increased serum lactate levels may be related to sympathetic activation. However, for DCI, we found that lactate disappeared as a prognostic factor when glucose was added as an independent variable in the analysis. A possible explanation for this effect in DCI is the proposed mechanism of lactate being preferential fuel for the brain and therefore a glucose-sparing substrate, whereas our finding is also in line with previous studies reporting elevated glucose as a risk factor for cerebral ischemia, which may be mediated by increased cortisol.[12,32]

Although we found elevated serum lactate levels only very slightly above the upper limit of the reference range in our patients with DCI and poor outcome (median, 2.1 and 2.2 mmol/L, respectively), this relative hyperlactatemia has previously been independently associated with an increased hospital mortality rate in critically ill patients.[33,34] Therefore, our findings are not unique in this respect.

Implications of study findings

In SAH patients, prediction of a complicated course remains difficult. Established predictors of DCI and poor outcome are amount of subarachnoid blood, clinical condition at admission, age, and smoking.[6,7] On the basis of results of our study, lactate and glucose are easily available parameters at admission that may be considered for future prognostic models for poor outcome and DCI. It should be noted that in spite of the associations found neither lactate nor glucose values are currently sufficient to predict outcomes with certainty in any individual patient.

An important question that warrants further evaluation is why lactate or glucose levels measured in SAH patients during the first 24 hours after admission can be used to predict development of DCI and poor outcome weeks or months later. This may eventually be helpful to improve individual decision making or even lactate-guided management in these patients.

The proposed mechanism of stress-related increase of lactate levels might have therapeutic consequences. As elevated heart rate and systolic blood pressure are seen during exposure to stress,[35,36] treatment with β-blockers might help in reducing stress-related lactate levels. In previous research, the association between β-blockade and improved outcome after SAH has already been suggested.[22,37] Importantly, lactate levels are easily available in contrast to catecholamine measurements, which renders lactate a much more feasible biomarker for sympathetic activation in routine clinical practice. We cannot entirely exclude that elevated lactates in our patients partly originated from cerebral lactate

release into the systemic circulation due to cerebral anerobic metabolism in an injured brain[38] although the strong association with glucose may argue in favor of the sympathetic hypothesis.

A first step for further research should be confirmation of our findings in prospective studies and, when confirmed, assessment of the pathophysiological relation of increased lactate with physiological derangements related to SAH, such as sympathetic activation, volume status, or cardiac function. In addition, our findings indicate that lactate may hold promise as a variable to be included in future prediction models on outcome. For such prediction models to become useful for every individual patient, they should have good discriminative ability with regard to clinical outcomes. It is important to note that multiple external validations of these findings are necessary before they should be applied outside the setting of this study.

Strengths and weaknesses

An important strength of this study is, first, the completeness of data concerning the DCI endpoint. Second, the inclusion of patients treated at two university hospitals in the Netherlands corroborates the external validity of our findings for similar settings although external validity outside the academic setting and in different countries was not investigated. Importantly, adding treatment center as an independent variable to the analyses (data not shown) did not change our results. Third, we used maximum lactate levels within the first 24 hours after admission, of which the prognostic value has been confirmed in previous research in different settings.[17,39]

Several limitations of our study need to be considered. First, the possibilities for statistical adjustment were limited to variables that were available in the database and we cannot exclude that important variables for adjustment were missing. Further evaluation of the prognostic value of lactate levels in SAH using additional prospectively collected parameters such as catecholamines is therefore required. Second, we only assessed CT-proven DCI. Mild forms of DCI with only clinical symptoms were not included in this study, which underestimates the number of patients with DCI. However, DCI resulting in a cerebral infarction has been shown to be clinically more relevant as a clinical endpoint.[28,40] Third, administration of epinephrine, dobutamine, and/or metformin was not taken into account as a potential confounder. The use of these drugs can affect lactate levels.[41] Fourth, exclusion of a large number of patients without lactate measurements within the first 24 hours after admission may have introduced bias. Because patients with lacking lactate and glucose measurements had better neurological status at admission and less ventricular blood, our results probably apply to patients who were in a somewhat worse condition at admission. Finally, our results only apply to patients who survive the first 48 hours of admission and do not have dismal prognosis very early after admission.

Conclusions

This study shows that maximum early lactate and glucose levels in the acute phase after aneurysmal SAH are associated with an increased risk for DCI-related infarction and poor outcome. These routinely available laboratory measurements may help to improve identification of patients at risk for complications or poor outcome after SAH by studying them in conjunction with other parameters in future prognostic models. Confirmation of the pathophysiological significance of increased lactate and glucose in prospective research seems warranted in SAH, especially with regard to sympathetic activation and its potential adverse consequences.

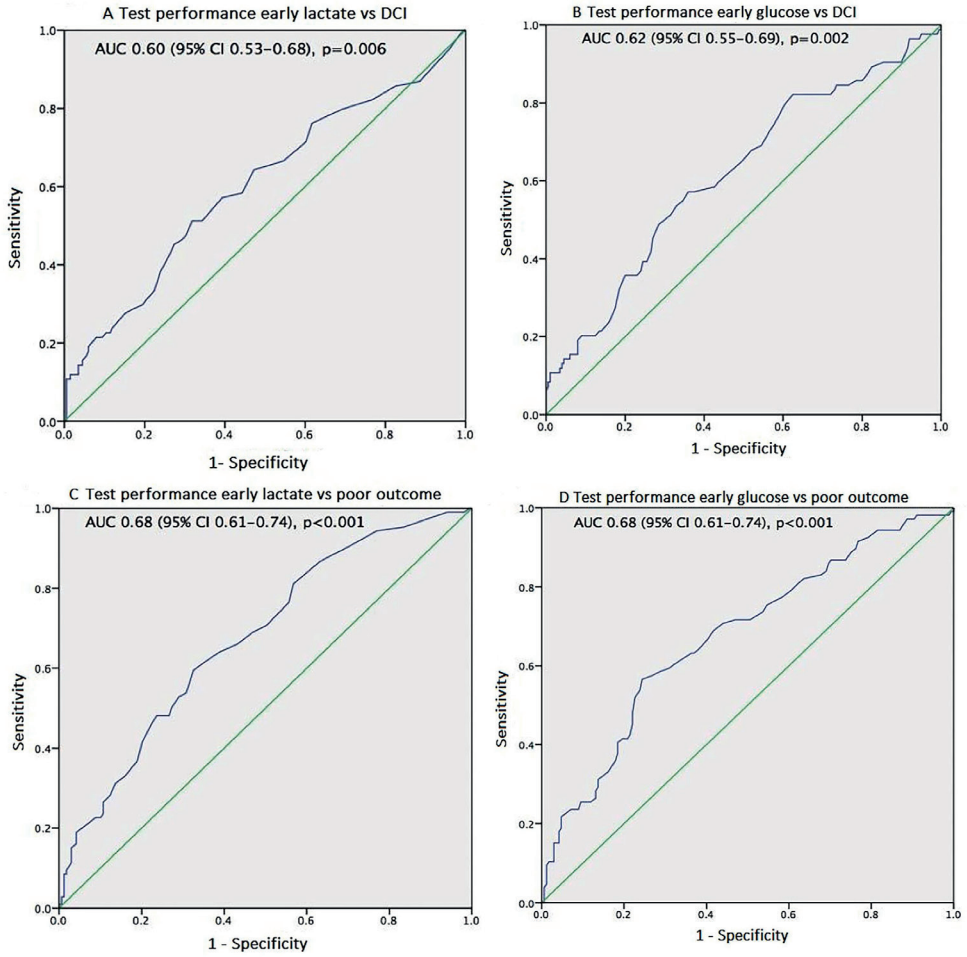
References

1. Nieuwkamp DJ, Setz LE, Algra A, Linn FH, de Rooij NK, Rinkel GJ. Changes in case fatality of aneurysmal subarachnoid haemorrhage over time, according to age, sex, and region: A meta-analysis. *Lancet Neurol.* 2009; 8:635–642
2. Rinkel GJ, Algra A. Long-term outcomes of patients with aneurysmal subarachnoid haemorrhage. *Lancet Neurol.* 2011; 10:349–356
3. Roos YB, de Haan RJ, Beenen LF, Groen RJ, Albrecht KW, Vermeulen M. Complications and outcome in patients with aneurysmal subarachnoid haemorrhage: A prospective hospital based cohort study in the Netherlands. *J Neurol Neurosurg Psychiatry.* 2000; 68:337–341
4. Dankbaar JW, de Rooij NK, Smit EJ, Velthuis BK, Frijns CJ, Rinkel GJ, et al. Changes in cerebral perfusion around the time of delayed cerebral ischemia in subarachnoid hemorrhage patients. *Cerebrovasc Dis.* 2011; 32:133–140
5. Dankbaar JW, de Rooij NK, Velthuis BK, Frijns CJ, Rinkel GJ, van der Schaaf IC. Diagnosing delayed cerebral ischemia with different CT modalities in patients with subarachnoid hemorrhage with clinical deterioration. *Stroke.* 2009; 40:3493–3498
6. de Rooij NK, Greving JP, Rinkel GJ, Frijns CJ. Early prediction of delayed cerebral ischemia after subarachnoid hemorrhage: Development and validation of a practical risk chart. *Stroke.* 2013; 44:1288–1294
7. de Rooij NK, Rinkel GJ, Dankbaar JW, Frijns CJ. Delayed cerebral ischemia after subarachnoid hemorrhage: A systematic review of clinical, laboratory, and radiological predictors. *Stroke.* 2013; 44:43–54
8. Rosengart AJ, Schultheiss KE, Tolentino J, Macdonald RL. Prognostic factors for outcome in patients with aneurysmal subarachnoid hemorrhage. *Stroke.* 2007; 38:2315–2321
9. Riley RD, Hayden JA, Steyerberg EW, Moons KG, Abrams K, Kyzas PA, et al. Prognosis Research Strategy (PROGRESS) 2: Prognostic factor research. *PLoS Med.* 2013; 10:e1001380
10. Jansen TC, van Bommel J, Bakker J. Blood lactate monitoring in critically ill patients: A systematic health technology assessment. *Crit Care Med.* 2009; 37:2827–2839
11. Kruse O, Grunnet N, Barfod C. Blood lactate as a predictor for inhospital mortality in patients admitted acutely to hospital: A systematic review. *Scand J Trauma Resusc Emerg Med.* 2011; 19:74
12. Bouzat P, Oddo M. Lactate and the injured brain: Friend or foe? *Curr Opin Crit Care.* 2014; 20:133–140
13. Oddo M, Levine JM, Frangos S, Maloney-Wilensky E, Carrera E, Daniel RT, et al. Brain lactate metabolism in humans with subarachnoid hemorrhage. *Stroke.* 2012; 43:1418–1421
14. Barletta JF, Figueroa BE, DeShane R, Blau SA, McAllen KJ. High glucose variability increases cerebral infarction in patients with spontaneous subarachnoid hemorrhage. *J Crit Care.* 2013; 28:798–803
15. Frontera JA, Fernandez A, Claassen J, Schmidt M, Schumacher HC, Wartenberg K, et al. Hyperglycemia after SAH: Predictors, associated complications, and impact on outcome. *Stroke.* 2006; 37:199–203
16. Kurtz P, Claassen J, Helbok R, Schmidt J, Fernandez L, Presciutti M, et al. Systemic glucose variability predicts cerebral metabolic distress and mortality after subarachnoid hemorrhage: A retrospective observational study. *Crit Care.* 2014; 18:R89

17. Kaukonen KM, Bailey M, Egi M, Orford N, Glassford NJ, Marik PE, et al. Stress hyperlactatemia modifies the relationship between stress hyperglycemia and outcome: A retrospective observational study. *Crit Care Med.* 2014; 42:1379–1385
18. Griffith FR, Lockwood JE, Emery FE. Adrenalin lactacidemia: Proportionality with dose. *Am J Physiol.* 1939; 127:415–421
19. Fontes RB, Aguiar PH, Zanetti MV, Andrade F, Mandel M, Teixeira MJ. Acute neurogenic pulmonary edema: Case reports and literature review. *J Neurosurg Anesthesiol.* 2003; 15:144–150
20. Inamasu J, Sugimoto K, Yamada Y, Ganaha T, Ito K, Watabe T, et al. The role of catecholamines in the pathogenesis of neurogenic pulmonary edema associated with subarachnoid hemorrhage. *Acta Neurochir (Wien).* 2012; 154:2179–2184
21. Sugimoto K, Inamasu J, Kato Y, Yamada Y, Ganaha T, Oheda M, et al. Association between elevated plasma norepinephrine levels and cardiac wall motion abnormality in poor-grade subarachnoid hemorrhage patients. *Neurosurg Rev.* 2013; 36:259–266
22. Schmidt JM, Crimmins M, Lantigua H, Fernandez A, Zammit C, Falo C, et al. Prolonged elevated heart rate is a risk factor for adverse cardiac events and poor outcome after subarachnoid hemorrhage. *Neurocrit Care.* 2014; 20:390–398
23. van der Bilt I, Hasan D, van den Brink R, Cramer MJ, van der Jagt M, van Kooten F, et al. Cardiac dysfunction after aneurysmal subarachnoid hemorrhage: Relationship with outcome. *Neurology.* 2014; 82:351–358
24. van der Bilt IA, Hasan D, Vandertop WP, Wilde AA, Algra A, Visser FC, et al. Impact of cardiac complications on outcome after aneurysmal subarachnoid hemorrhage: A meta-analysis. *Neurology.* 2009; 72:635–642
25. NICE: Nationale Intensive Care Evaluatie. Available at: www.stichtingnice.nl. Accessed February 24, 2015.
26. Hijdra A, Brouwers PJ, Vermeulen M, van Gijn J. Grading the amount of blood on computed tomograms after subarachnoid hemorrhage. *Stroke.* 1990; 21:1156–1161
27. Report of World Federation of Neurological Surgeons Committee on a Universal Subarachnoid Hemorrhage Grading Scale. *J Neurosurg.* 1988; 68:985–986
28. Vergouwen MD, Vermeulen M, van Gijn J, Rinkel GJ, Wijdicks EF, Muizelaar JP, et al. Definition of delayed cerebral ischemia after aneurysmal subarachnoid hemorrhage as an outcome event in clinical trials and observational studies: Proposal of a multidisciplinary research group. *Stroke.* 2010; 41:2391–2395
29. Benedict CR, Loach AB. Clinical significance of plasma adrenaline and noradrenaline concentrations in patients with subarachnoid haemorrhage. *J Neurol Neurosurg Psychiatry.* 1978; 41:113–117
30. Ogura T, Satoh A, Ooigawa H, Sugiyama T, Takeda R, Fushihara G, et al. Characteristics and prognostic value of acute catecholamine surge in patients with aneurysmal subarachnoid hemorrhage. *Neurol Res.* 2012; 34:484–490
31. Ottens TH, Nijsten MW, Hofland J, Dieleman JM, Hoekstra M, van Dijk D, et al. Effect of high-dose dexamethasone on perioperative lactate levels and glucose control: a randomized controlled trial. *Crit Care.* 2015; 19:41
32. Nichol AD, Egi M, Pettila V, Bellomo R, French C, Hart G, et al. Relative hyperlactatemia and hospital mortality in critically ill patients: A retrospective multi-centre study. *Crit Care.* 2010; 14:R25
33. Rishu AH, Khan R, Al-Dorzi HM, Tamim HM, Al-Qahtani S, Al-Ghamdi G, et al. Even mild hyperlactatemia is associated with increased mortality in critically ill patients. *Crit Care.* 2013; 17:R197

34. Vergouwen MD, van Geloven N, de Haan RJ, Kruyt ND, Vermeulen M, Roos YB. Increased cortisol levels are associated with delayed cerebral ischemia after aneurysmal subarachnoid hemorrhage. *Neurocrit Care.* 2010; 12:342–345
35. Halter JB, Stratton JR, Pfeifer MA. Plasma catecholamines and hemodynamic responses to stress states in man. *Acta Physiol Scand Suppl.* 1984; 527:31–38
36. Stratton JR, Pfeifer MA, Ritchie JL, Halter JB. Hemodynamic effects of epinephrine: Concentration-effect study in humans. *J Appl Physiol.* 1985; 58:1199–1206
37. Neil-Dwyer G, Walter P, Cruickshank JM. Beta-blockade benefits patients following a subarachnoid haemorrhage. *Eur J Clin Pharmacol.* 1985; 28 Suppl:25–29
38. Holbein M, Béchir M, Ludwig S, Sommerfeld J, Cottini SR, Keel M, et al. Differential influence of arterial blood glucose on cerebral metabolism following severe traumatic brain injury. *Crit Care.* 2009; 13:R13
39. Donnino MW, Andersen LW, Giberson T, Gaieski DF, Abella BS, Peberdy MA, et al. Initial lactate and lactate change in postcardiac arrest: A multicenter validation study. *Crit Care Med.* 2014; 42:1804–1811
40. Frontera JA, Fernandez A, Schmidt JM, Claassen J, Wartenberg KE, Badjatia N, et al. Defining vasospasm after subarachnoid hemorrhage: What is the most clinically relevant definition? *Stroke.* 2009; 40:1963–1968
41. Gjedsted J, Buhl M, Nielsen S, Schmitz O, Vestergaard ET, Tønnesen E, et al. Effects of adrenaline on lactate, glucose, lipid and protein metabolism in the placebo controlled bilaterally perfused human leg. *Acta Physiol (Oxf).* 2011; 202:641–648

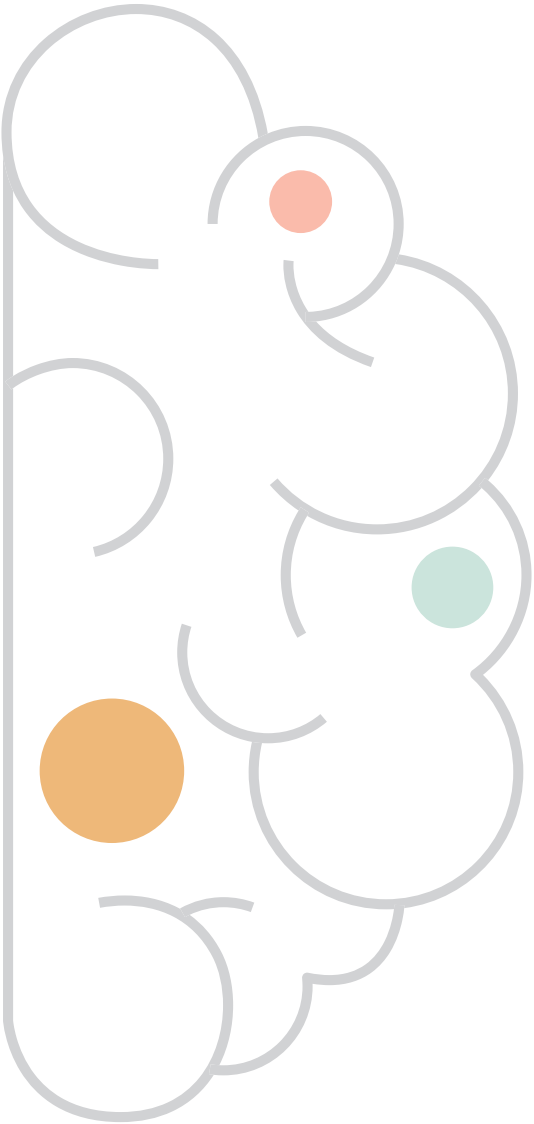
Appendix



Appendix 2.A. ROC plots of early blood lactate and glucose values (test) versus (**A** and **B**) delayed cerebral ischemia or (**C** and **D**) poor outcome ("disease").

ROC, receiver operating characteristic; DCI, delayed cerebral ischemia; AUC, area under the receiver operating characteristic curve; CI, confidence interval

CHAPTER 3



Development and validation of the Dutch Stroke Score for predicting disability and functional outcome after ischemic stroke: A tool to support efficient discharge planning

Simone A. Dijkland
Inger R. de Ridder
Maaike Scheele
Heleen M. den Hertog
Maaike Dirks
Willeke F. Westendorp
Paul J. Nederkoorn
Diederik van de Beek
Gerard M. Ribbers
Ewout W. Steyerberg
Hester F. Lingsma
Diederik W.J. Dippel

Eur Stroke J 2018; 3(2): 165–173

Abstract

Introduction: We aimed to develop and validate a prognostic score for disability at discharge and functional outcome at three months in patients with acute ischemic stroke based on clinical information available on admission.

Patients and methods: The Dutch Stroke Score (DSS) was developed in 1227 patients with ischemic stroke included in the Paracetamol (Acetaminophen) In Stroke study. Predictors for Barthel Index (BI) at discharge ('DSS-discharge') and modified Rankin Scale (mRS) at three months ('DSS-3 months') were identified in multivariable ordinal regression. The models were internally validated with bootstrapping techniques. The DSS-3 months was externally validated in the PRomoting ACute Thrombolysis in Ischemic StrokE study (1589 patients) and the Preventive Antibiotics in Stroke Study (2107 patients). Model performance was assessed in terms of discrimination, expressed by the area under the receiver operating characteristic curve (AUC), and calibration.

Results: At model development, the strongest predictors of Barthel Index at discharge were age per decade over 60 (odds ratio=1.55, 95% confidence interval (CI) 1.41–1.68), National Institutes of Health Stroke Scale (odds ratio=1.24 per point, 95% CI 1.22–1.26) and diabetes (odds ratio=1.62, 95% CI 1.32–1.91). The internally validated AUC was 0.76 (95% CI 0.75–0.79). The DSS-3 months, additionally consisting of previous stroke and atrial fibrillation, performed similarly at internal (AUC 0.75, 95% CI 0.74–0.77) and external validation (AUC 0.74 in PRomoting Acute Thrombolysis in Ischemic StrokE [95% CI 0.72–0.76] and 0.69 in Preventive Antibiotics in Stroke Study [95% CI 0.69–0.72]). Observed outcome was slightly better than predicted.

Discussion: The DSS had satisfactory performance in predicting BI at discharge and mRS at three months in ischemic stroke patients.

Conclusion: If further validated, the DSS may contribute to efficient stroke unit discharge planning alongside patients' contextual factors and therapeutic needs.

Introduction

In 2015, over 26,000 patients were admitted to hospitals because of ischemic stroke in the Netherlands.[1] Most of these patients need rehabilitation to achieve better recovery in the first months after stroke and reduce long-term disability. In the Netherlands, around 8% of all stroke patients is referred to an inpatient rehabilitation centre.[2] Typically, these patients are too disabled to be discharged home, but they are cognitively and physically fit enough to participate in intensive therapy sessions and have sufficient social support to return home within two to four months. Alternatively, patients may be referred to skilled nursing and geriatric rehabilitation facilities. These patients are often elderly, suffer from comorbidities and have a poorer functional prognosis. Still, the majority of stroke patients (60%) is discharged home, mostly with community rehabilitation.[2] Discharge planning may depend on multiple factors such as comorbidities and contextual factors (e.g. the presence of a healthy caregiver and premorbid level of functioning). The importance of the contextual factors increases as the functional prognosis of the stroke decreases. Therefore, early prediction of functional outcome may contribute to efficient discharge planning.

The most widely used functional outcome measure in acute stroke is the modified Rankin Scale (mRS). The mRS measures the degree of disability in daily activities. It is scored on an ordinal scale ranging from 0 (no symptoms) to 6 (death).[3] Another frequently used outcome measure in rehabilitation is the Barthel Index (BI), measuring performance in 10 basic activities of daily living (ADL).[4] BI is associated with duration of hospital stay.[5]

Previous studies identified many prognostic factors for outcome (measured by BI or mRS) after acute stroke.[6] Prognostic factors can be combined in a model to identify patients at risk for poor outcome.[7] Although several prognostic models exist to predict outcome in stroke, very few are adequately validated for use in daily clinical practice.[8] We aimed to develop and validate a prognostic score for disability (BI) at discharge and functional outcome (mRS) at three months after acute ischemic stroke based on clinical information available on admission.

Methods

Derivation cohort

Data from the Paracetamol (Acetaminophen) In Stroke (PAIS) study were used for model development.[9] PAIS was a multicentre, randomised placebo-controlled phase III trial assessing the effect of high dose paracetamol on the functional outcome in patients with acute stroke. In short, patients were eligible for inclusion if they were diagnosed with acute ischemic stroke or intracerebral hemorrhage, had a prestroke mRS<2 and study treatment could be started within 12 h after onset of symptoms. We used data of all patients with ischemic stroke included in PAIS.

Outcome measures

We used the BI at discharge as the outcome measure for short-term disability. The BI is an ordinal scale used to measure performance in ADL. The scale ranges from 0 to 20, with higher scores indicating a greater likelihood of being able to carry out ADL independently.[4] In PAIS, the BI was measured at 14 days after enrolment or at hospital discharge if this occurred earlier (70% of the patients stayed for 3 days).[9] However, choice of the optimal rehabilitation route mostly depends on more than just discharge outcome.[10] Therefore, we additionally evaluated functional outcome at three months with the mRS. The mRS is an ordinal scale used to measure the degree of disability in daily activities and ranges from 0 (no symptoms) to 6, with mRS 5 indicating severe disability and mRS 6 indicating death.[3]

Model development

To identify predictors of disability and functional outcome, we selected variables that were clinically relevant and/or previously reported to predict outcome after stroke in the literature.[6] These variables were sex, age, National Institutes of Health Stroke Scale (NIHSS) score, diabetes, previous stroke, atrial fibrillation and hypertension. All predictors were entered into multivariable ordinal regression with backward selection with $p<0.2$ for inclusion, separately for BI at discharge and mRS at three months. The final associations were presented as a set of odds ratios (ORs) and 95% confidence intervals (CIs) to indicate the individual predictor effects. ORs from an ordinal logistic regression model can be interpreted as a common OR for shifting over the full outcome range.[11]

The resulting models, the Dutch Stroke Score (DSS) for BI at discharge ('DSS-discharge') and mRS at three months ('DSS-3 months'), were internally validated using standard bootstrapping procedures to avoid an optimistic estimate of the model performance, which often occurs when model performance is only evaluated directly in the derivation cohort (apparent validation). In the bootstrap procedure, random samples are drawn from the original sample, each with the same number of patients as the original sample. In each of these samples the modeling steps are repeated and the resulting models are subsequently evaluated on the original sample. The mean model performance in all 500 bootstrap models represents the expected performance of the models in future, similar patients.[12]

Validation cohorts

For external validation, we used data from the PRomoting ACute Thrombolysis in Ischemic StrokE (PRACTISE) study and Preventive Antibiotics in Stroke Study (PASS). PRACTISE was a clusterrandomised trial designed to evaluate an implementation strategy to increase the proportion of patients treated with intravenous thrombolysis.[13] PRACTISE registered adult patients with acute stroke admitted within 24 h after onset of symptoms and had no age restrictions. We used data from ischemic stroke patients admitted within 4 h as in these patients detailed clinical data were available.

PASS was a multicentre, randomised, open-label trial designed to assess whether or not preventive antimicrobial therapy with ceftriaxone improves functional outcome in patients with acute stroke.[14] PASS included adult patients with clinical symptoms of a stroke (ischemic or hemorrhagic) admitted within 24 h after symptom onset. We used data of all patients with ischemic stroke included in PASS.

Model validation

The validity of the DSS-3 months was assessed in terms of discrimination and calibration. The external validation cohorts did not have data on BI at discharge. Discrimination refers to how well the model distinguishes between those who have good outcome (mRS 0–2) vs. those who have poor outcome (mRS 3–6) at three months. Discrimination was assessed by calculating the ordinal area under the curve (AUC) of the receiver operating characteristic (ROC) curve.[15] The AUC ranges from 0.5 for non-informative models to 1.0 for perfect models.[12] Calibration indicates the agreement between predicted and observed probabilities. Calibration was assessed graphically in a calibration graph, and expressed as the calibration slope and an intercept. The calibration slope is ideally equal to 1 and describes the effect of the predictors in the validation cohort versus in the derivation cohort. The intercept indicates whether predictions are systematically too high or too low, and should ideally be zero.[12]

At external validation, the discriminative power of a model may be influenced by differences in predictor effects, but also by differences in distribution of patient characteristics (case-mix) between the derivation and validation cohort.[16] In a more homogeneous population, discrimination between patients with good vs. poor outcome is more difficult than in a heterogeneous population. To take this into account, we calculated the case-mix-corrected AUC. The case-mix-corrected AUC reflects the discriminative power of a model, assuming that the regression coefficients are correct for the validation population. It was calculated by simulating new outcome values for all patients in the validation dataset, based on the predicted risks for each patient.[16]

After external validation, we fitted the DSS-3 months on the combined data of all three trials to get the best estimates for the regression coefficients.[17] The DSS-discharge and DSS-3 months were presented in a score chart, as a score plot simplified to five BI and mRS outcome classes (based on clinically relevant cutoffs), and as formulas to calculate the predicted outcomes.

All statistical analyses were performed using R software, version 3.3.2 (R foundation for statistical computing, Vienna, Austria). The calibration plots were created with an updated version of the *val.prob* function (*rms* library in R). Missing values in the development and validation cohorts were statistically imputed using a multiple imputation method exploiting correlations between predictor variables and between predictor variables and the outcome variables (*mice* function in R). Complete case analyses were done for comparison with the imputed analyses.

Results

Study population

For model development, we included 1227 patients with ischemic stroke from the PAIS trial. Missing data on hypertension (3.1%) were statistically imputed; all other baseline variables and outcomes were complete. For the external validation of the model predicting mRS at three months, we included, 1657 ischemic stroke patients from the PRACTISE study. Sixty-eight patients with missing data on mRS at three months were excluded, resulting in an external validation sample of 1589 patients. Other missing

data (0.6%) were statistically imputed. Additionally, we externally validated the model for functional outcome at three months in, 2125 ischemic stroke patients from the PASS study. Eighteen patients with missing data on the mRS at three months were excluded, resulting in an external validation sample of 2107 patients. Other missing data (0.4%) were statistically imputed.

In all three studies, most patients (55–58%) were male and the mean age was around 70 years (Table 3.1). The three populations are comparable concerning baseline characteristics, except for time from stroke onset to inclusion (PAIS and PRACTISE had a smaller time window compared to PASS), previous stroke (33% in PASS vs. 20% in the other trials) and diabetes (20% in PASS vs. 15–17% in PAIS and PRACTISE). The number of patients with poor outcome (mRS 3–6) was lower in PASS compared to PAIS and PRACTISE (Appendix 3.A). In PAIS, this is reflected in the substantial proportion of patients with favorable outcome on the BI at discharge (Appendix 3.A).

Table 3.1. Baseline characteristics of the included patients from the PAIS, PRACTISE and PASS studies

	PAIS (n = 1227)	PRACTISE (n = 1589)	PASS (n = 2107)
Male sex	675 (55%)	872 (55%)	1212 (58%)
Age in years (mean, sd)	70.1 (13.4)	70.6 (13.4)	71.9 (12.5)
Time from onset to CT in hours (median, IQR)	3.0 (1.8-5.9)	2.0 (1.4-3.0)	NA
NIHSS (median, IQR)	6.0 (3.0-11.0)	5.0 (3.0-12.0)	5.0 (3.0-9.0)
Diabetes mellitus	181 (15%)	266 (17%)	423 (20%)
Previous ischemic stroke	245 (20%)	318 (20%)	698 (33%)
Atrial fibrillation	190 (16%)	290 (18%)	326 (16%)
Hypertension	601 (49%)[a]	811 (51%)	1154 (55%)
Current smoking	380 (31%)	374 (24%)	524 (25%)

NIHSS, National Institutes of Health Stroke Scale; IQR, interquartile range; NA, not available; PRACTISE, Promoting Acute Thrombolysis in Ischemic StrokE; PASS, Preventive Antibiotics in Stroke Study; PAIS, Paracetamol (Acetaminophen) In Stroke.
[a]38 missings.

Model development in PAIS

The relation between age as a continuous variable and the log odds of disability (BI) in the development data was non-linear and intensified when age was above 60 years (Appendix 3.B). Because of this non-linearity, we considered different age effects for patients older vs. younger than 60 years.

Of the variables considered, age per decade above 60, NIHSS per point and diabetes were the strongest predictors of BI at discharge, both in univariable (data not shown) and multivariable analysis (Table 3.2) and were included in the model for disability at discharge. The internally validated ordinal AUC was 0.76 (95%CI 0.75–0.79). Age per decade above 60, NIHSS per point, diabetes, previous stroke and atrial fibrillation were the strongest predictors of mRS at three months, both in univariable (data not shown) and multivariable analysis (Table 3.2) and were included in the final model for mRS at three months. The internally validated ordinal AUC was 0.75 (95%CI 0.74–0.77).

Table 3.2. Associations of predictors in multivariable ordinal regression with lower BI at discharge in PAIS and higher mRS at three months in in PAIS, PRACTISE and PASS

	PAIS (n = 1227) BI at discharge		PAIS (n = 1227) mRS at three months		PRACTISE (n = 1589) mRS at three months		PASS (n = 2107) mRS at three months	
Variable	**OR (95% CI)**	**P value**	**OR (95% CI)**	**P value**	**OR (95% CI)**	**P value**	**OR (95% CI)**	**P value**
Male sex	1.01 (0.79-1.23)	0.923	0.87 (0.71-1.07)	0.189	0.81 (0.67-0.97)	0.022	0.77 (0.61-0.93)	0.002
Age per decade if over 60[a,b]	1.55 (1.41-1.68)	<0.001	1.86 (1.64-2.12)	<0.001	1.80 (1.61-2.01)	<0.001	1.55 (1.41-1.70)	<0.001
Age per decade if under 60	1.07 (0.83-1.30)	0.589	0.93 (0.76-1.15)	0.514	0.89 (0.74-1.07)	0.211	0.70(0.56-0.86)	<0.001
NIHSS per point[a,b]	1.24 (1.22-1.26)	<0.001	1.19 (1.17-1.22)	<0.001	1.19 (1.17-1.21)	<0.001	1.21 (1.19-1.23)	<0.001
Diabetes[a]	1.62 (1.32-1.91)	0.002	1.87 (1.40-2.51)	<0.001	1.70 (1.34-2.17)	<0.001	1.31 (1.11-1.51)	0.007
Previous stroke[b]	1.18 (0.91-1.45)	0.225	1.67 (1.29-2.16)	<0.001	1.59 (1.27-1.99)	<0.001	1.14 (0.98-1.31)	0.111
Atrial fibrillation[b]	1.09 (0.78-1.39)	0.592	1.41 (1.05-1.89)	0.022	1.24 (0.98-1.57)	0.076	1.14 (0.91-1.36)	0.264
Hypertension	1.06 (0.84-1.28)	0.594	1.02 (0.83-1.26)	0.844	1.08 (0.90-1.30)	0.384	0.91 (0.75-1.07)	0.246

BI, Barthel Index; mRS, modified Rankin Scale; OR, odds ratio; CI, confidence interval; NIHSS, National Institutes of Health Stroke Scale; PRACTISE, PRomoting Acute Thrombolysis in Ischemic StrokE; PASS, Preventive Antibiotics in Stroke Study; PAIS, Paracetamol (Acetaminophen) in Stroke.

[a]Parameter included in the final model on BI at discharge.

[b]Parameter included in final model on mRS at three months.

External validation in PRACTISE and PASS

In PRACTISE, the DSS-3 months had an ordinal AUC of 0.74 and an AUC for the cutoff mRS ≥ 3 of 0.81 (95% CI 0.81–0.84) (Appendix 3.C). The model predicted 49.4% poor outcome (mRS ≥ 3); whereas the observed probability of poor functional outcome was 45.2%. The calibration slope was 1.022 and the intercept was -0.238, indicating that the model's predictions of poor outcome were systematically higher than the observed probability of poor outcome (Figure 3.1A).

In PASS, the DSS-3 months had an ordinal AUC of 0.69 and an AUC for the cutoff mRS ≥ 3 of 0.81 (95% CI 0.81–0.83) (Appendix 3.C). The predicted probability of poor outcome was 48.6%, compared to an observed probability of poor functional outcome of 38.5%. The calibration slope was 1.058 and the intercept was -0.555, indicating that the model's predictions of poor outcome were systematically too high (Figure 3.1B). This overestimation was higher than in PRACTISE.

The internal and external validation in the complete cases (PAIS n=1227, PRACTISE n=1581, PASS n=2098) yielded similar results (not shown).

The lower discriminative ability of the DSS-3 months in the external validation cohorts was largely explained by a less heterogeneous case-mix compared to the development cohort. This is illustrated by small differences between the development AUC and casemix-corrected AUCs (Appendix 3.C). The lower discriminative ability in PASS compared to PAIS and PRACTISE was due to both case-mix and differences in predictor effects (relatively large difference between AUC in external validation and case-mix-corrected AUC in PASS).

The final DSS-3 months was developed on the combined data of all three cohorts (n=4923). The model had an ordinal AUC of 0.73 and an AUC for the cutoff mRS ≥ 3 of 0.81 (95% CI 0.81–0.83) (Appendix 3.C).

The final models are presented as the DSS score chart (Table 3.3, and simplified to five outcome classes in Figure 3.2), with higher scores indicating worse outcome. For example, a patient of 70 years with an NIHSS of 13 and a history of previous stroke and diabetes has a DSS-discharge score of 8 and a predicted probability of 17% for BI 19–20 at discharge and a DSS-3 months score of 13 and a predicted probability of 76% for mRS ≥ 3 at three months (Appendix 3.D).

Table 3.3. DSS score chart based on ordinal analysis of the BI and mRS. A higher score indicates a worse outcome (lower predicted BI and higher mRS)

Variable	Points for predicting BI at discharge	Points for predicting mRS score at 3 months
Age		
<60	0	0
60-70	1	2
70-80	2	4
80-90	3	6
90+	4	8
NIHSS		
0	0	0
1-4	1	1
5-15	5	5
16-20	10	10
21-42	15	15
Diabetes	1	2
Previous stroke	-	2
Atrial fibrillation	-	1
Total	0-20	0-28

BI, Barthel Index; mRS, modified Rankin Scale; NIHSS, National Institutes of Health Stroke Scale; DSS, Dutch Stroke Score.

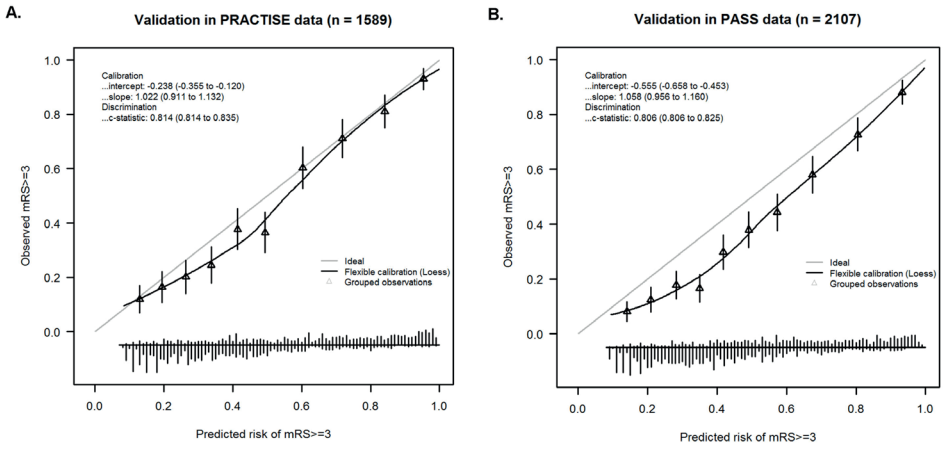


Figure 3.1. Calibration plots of the DSS-3 months in (**A**) PRACTISE and (**B**) PASS.
DSS, Dutch Stroke Score; PRACTISE, PRomoting ACute Thrombolysis in Ischemic StrokE; PASS, Preventive Antibiotics in Stroke Study.

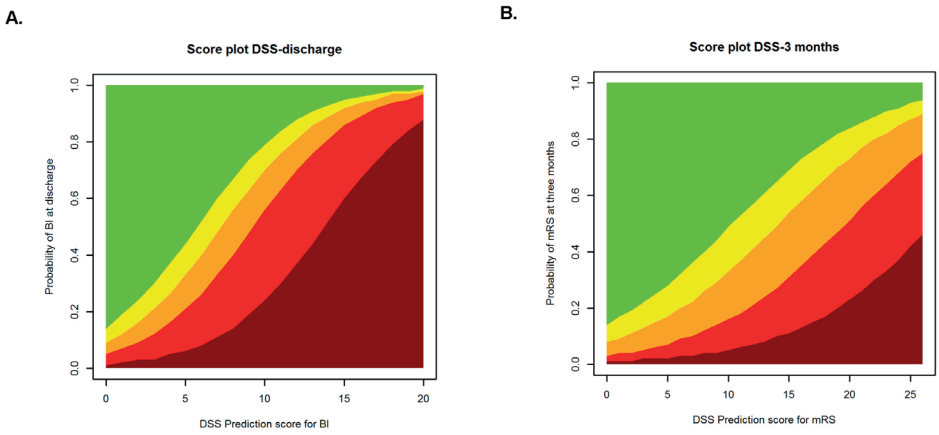


Figure 3.2. DSS score charts simplified to five outcome classes of the (**A**) BI at discharge and (**B**) mRS at three months. Legend of (**A**): Dark red=0, Red=1–9, Orange=10–14, Yellow=15–18, Green=19–20 and legend of (**B**): Dark red=6, Red=4–5, Orange=3, Yellow=2, Green=0–1.
DSS, Dutch Stroke Score; BI, Barthel Index; mRS, modified Rankin Scale.

Discussion

We propose the DSS, consisting of two simple prediction models for disability (BI) at discharge and functional outcome (mRS) at three months after acute ischemic stroke based on clinical information available on admission. The DSS-discharge consists of three variables: age per decade above 60 years, NIHSS per point and diabetes. The DSS-3 months additionally includes previous stroke and atrial fibrillation. Both models showed reasonable performance in internal and external validation.

Relation with previous literature

Previously, several models to estimate the probability of unfavourable outcome after stroke have been developed, with a high variability in endpoints, time between symptom onset and assessment of the variables, and patient populations. Literature reviews have shown that many of these prediction models have methodological shortcomings that limit their use for early discharge planning. For instance, assessment of predictors multiple days after stroke onset[18,19] and the use of a dichotomous outcome such as mortality.[20-26] In addition, previously developed models were not validated, and hence their use in clinical practice is limited.[8,27]

One tool has been developed specifically to predict unfavorable discharge destination from the hospital stroke unit. Functional disability, poor sitting balance, depression, cognitive disability and old age were identified as predictors of poor discharge outcome.[10] However, this model was only applicable for decision-making at 7–10 days post stroke. Moreover, this study had some methodological shortcomings, including dichotomisation of predictors, a small sample size and dichotomisation of the outcome.

Implications of study findings

Prediction models in acute stroke are useful to inform patients and relatives on prognosis and identify patients at risk for poor outcome before treatment decisions are made.[7] On population level, prediction models can be used for adjustment when comparing quality of care for stroke patients across institutions. Additionally, prediction models could be relevant in design and analysis of randomised controlled trials, e.g. for covariate adjustment.[28,29] Further, prediction of functional outcome may contribute to discharge planning. If functional outcome is expected to be poor, contextual factors, such as housing circumstances, financial problems and whether or not a patient is living alone, become more important.

We developed the DSS to be used by stroke unit nurses during the first day after admission. In clinical practice, the NIHSS is mostly scored shortly after the administration of alteplase. Therefore, we did not add treatment with alteplase as a covariable to our analysis. Recently, intra-arterial treatment administered within six hours after stroke onset has been shown beneficial in patients with a proximal intracranial arterial occlusion.[30] However, the majority (90%) of acutely admitted ischemic stroke patients still receives intravenous alteplase as only treatment. Therefore, the DSS is potentially suitable for use in present neurovascular practice. To facilitate discharge planning in endovascular-treated patients, a next step could be to update the models by including treatment (thrombolysis, thrombectomy or both) as a predictor. Moreover, no imaging or laboratory tests are required for clinicians to be able to use the DSS, which allows bedside use of the models early after admission. The DSS score chart can be easily incorporated in clinical practice since it consists of a few readily obtainable clinical variables at admission. Stroke unit nurses will be able to score all variables, including the NIHSS, provided that they are well trained and certified.

The DSS-discharge still needs to be externally validated to give reliable estimates on model performance and study generalisability.

At external validation, the discriminative ability of the DSS-3 months was generally lower than in the development sample. Discrimination was better in PRACTISE compared to PASS, both for the ordinal analysis of the mRS and for three different cutoffs of the mRS (Appendix 3.C). These higher AUCs were partly explained by differences in case-mix, as reflected in the case-mix-corrected AUCs. In addition, the predictor effects were slightly stronger in PRACTISE than in PASS. These differences in regression coefficients were most evident for diabetes and previous stroke, and could be explained by discrepancies in predictor definitions. For instance, in PASS, previous stroke comprised both Transient Ischemic Attack (TIA) and ischemic stroke, while in PRACTISE only ischemic stroke was considered. This implicates that the DSS-3 months is valid, but the definitions of the predictors should be identical to those in the development cohort.

The reasonable discriminative ability of the DSS-3 months was associated with an overall overestimation of the probability of poor outcome. This overestimation was higher in PASS compared to PRACTISE, which might be due to the difference in outcome distribution between these cohorts (lower proportion of patients with poor outcome in PASS). This difference is most likely caused by the exclusion of patients with imminent death and neurological deterioration in PASS. The overestimation

of the probability of poor outcome implies that the DSS-3 months needs updating (e.g. adjustment of the intercept [recalibration]) before it is suitable for individualised predictions in clinical practice.

Strengths and limitations

Strengths of this study are the internal and (partial) external validation of the DSS, and the large size of the development and two independent validation cohorts. Even though many models have been developed for prediction of outcome after stroke, the large sample size and the aim of contributing to efficient discharge planning makes that our study has added value compared to already existing evidence. Also, we predicted outcomes over the whole range from no symptoms to death. Furthermore, we used two well-known and widely implemented outcome measures for functional outcome in our models. The BI is a reliable and valid scale to measure ADL.[31] Since discharge destination (partially) depends on the patient's ability to carry out ADL, the BI is a suitable outcome for our model. Additionally, we selected potential predictors based on the literature and clinical knowledge. This is preferred over selection based on the data as the latter may result in overfitting (model perfect for the development data but performing poor in new patients.[12] The robustness of our approach is represented in the reasonable performance of the models in internal and external validation.

Several limitations of our study need to be considered. We included only hospitalised patients with an ischemic stroke in our analysis. Consequently, our chart does not apply for patients with intracerebral hemorrhage. Further, the development and validation cohorts originated from randomised controlled trials conducted in the Netherlands, potentially limiting the generalisability of the chart. To evaluate the performance of the models beyond the Dutch setting, external validation in observational data from settings with a different healthcare system configuration is necessary. However, the Dutch stroke population is representative for stroke populations in developed countries. Moreover, our external validation cohorts consist of unselected, prospectively included patients, originating from hospitals representative in size, geographic distribution and frequency of stroke treatment procedures. We were able to externally validate the DSS-3 months, but not the DSS-discharge as no data on BI at discharge were available. Also, discharge policy is variable between and within different healthcare systems, which makes it a difficult outcome for prediction purposes. However, these differences in discharge timing resemble the variation in clinical practice. Additionally, in the field of rehabilitation, predicting functional outcome in terms of the mRS has limitations. Important aspects that can contribute to the level of disability and the need for rehabilitation (e.g. pain, communication, cognition) are not entirely covered by the mRS.[32] However, the mRS is a widely used outcome measure in stroke management.

The prognostic performance of the DSS after validation could be classified as satisfactory. This does not disqualify the usefulness of the models for clinical practice, because in general, multivariable prediction models are able to incorporate and accurately weigh more factors than a human mind.[33] Nevertheless, the results should always be regarded as a mere recommendation and should be placed in the context of the personal circumstances, needs and wishes of the patient. Other factors that are worth considering when planning patients' discharge are the presence of social support, cognitive

disability, the therapeutic needs of the patient and the expected future residence destination (e.g. home or nursing facility).

Conclusion

The DSS has satisfactory performance in predicting BI at discharge and mRS at three months in ischemic stroke patients. If further validated, the DSS may contribute to efficient stroke unit discharge planning alongside patients' contextual factors (e.g. social support, housing circumstances and cognitive disability) and therapeutic needs.

Acknowledgements

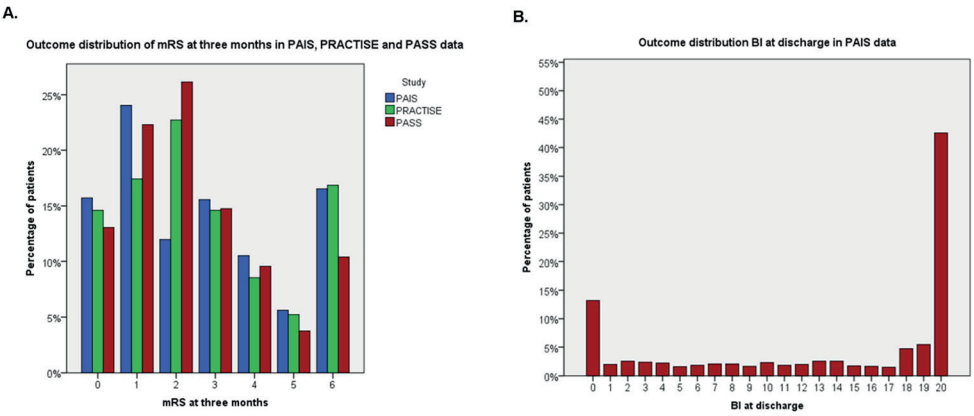
The Dutch Stroke Score was developed by researchers from the Erasmus MC Rotterdam and validated in data provided by researchers from the Erasmus MC Rotterdam and AMC Amsterdam. The authors wish to thank the investigators and patients participating in the PAIS, PRACTISE and PASS trials.

References

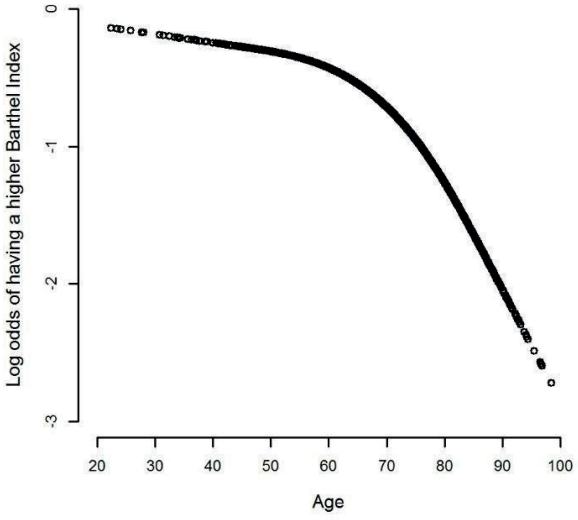
1. Buddeke J, van Dis I, Visseren FLJ, Vaartjes I, Bots ML. Ziekenhuisopnamen wegens hart- en vaatziekten. *In: Bots ML, Buddeke J, van Dis I, Vaartjes I, Visseren FLJ. Hart- en vaatziekten in Nederland 2016, cijfers over prevalentie, ziekte en sterfte.* Den Haag: Hartstichting, 2016.
2. Kennisnetwerk CVA Nederland. Benchmarkindicatoren. *In: Benchmarkresultaten 2014, rapportage en achtergrondinformatie.* 2016, accessed 26 June 2017.
3. van Swieten JC, Koudstaal PJ, Visser MC, Schouten HJ, van Gijn J. Interobserver agreement for the assessment of handicap in stroke patients. *Stroke.* 1988; 19: 604–607.
4. Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. *Md State Med J.* 1965; 14: 61–65.
5. Lees KR, Bath PM, Schellinger PD, Kerr DM, Fulton R, Hacke W, et al. Contemporary outcome measures in acute stroke research: choice of primary outcome measure. *Stroke.* 2012; 43: 1163–1170.
6. Veerbeek JM, Kwakkel G, van Wegen EE, Ket JC, Heymans MW. Early prediction of outcome of activities of daily living after stroke: a systematic review. *Stroke.* 2011; 42: 1482–1488.
7. Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med.* 2013; 10: e1001381.
8. Counsell C, Dennis M. Systematic review of prognostic models in patients with acute stroke. *Cerebrovasc Dis.* 2001; 12: 159–170.
9. den Hertog HM, van der Worp HB, van Gemert HM, Algra A, Kappelle LJ, van Gijn J, et al. The Paracetamol (Acetaminophen) In Stroke (PAIS) trial: a multicentre, randomised, placebocontrolled, phase III trial. *Lancet Neurol.* 2009; 8: 434–440.
10. Meijer R, van Limbeek J, Peusens G, Rulkens M, Dankoor K, Vermeulen M, et al. The Stroke Unit Discharge Guideline, a prognostic framework for the discharge outcome from the hospital stroke unit. A prospective cohort study. *Clin Rehabil.* 2005; 19: 770–778.
11. Senn S, Julious S. Measurement in clinical trials: a neglected issue for statisticians? *Stat Med.* 2009; 28: 3189–3209.
12. Steyerberg EW. Validation of prediction models. *In: Clinical prediction models: a practical approach to development, validation and updating.* New York: Springer, 2009, pp.299–310.
13. Dirks M, Niessen LW, van Wijngaarden JD, Koudstaal PJ, Franke CL, van Oostenbrugge RJ, et al. Promoting thrombolysis in acute ischemic stroke. Stroke 2011; 42: 1325–1330.
14. Westendorp WF, Vermeij JD, Zock E, Hooijenga IJ, Kruyt ND, Bosboom HJ, et al. The Preventive Antibiotics in Stroke Study (PASS): a pragmatic randomised open-label masked endpoint clinical trial. *Lancet.* 2015; 385: 1519–1526.
15. Van Calster B, Van Belle V, Vergouwe Y, Steyerberg EW. Discrimination ability of prediction models for ordinal outcomes: relationships between existing measures and a new measure. *Biom J.* 2012; 54: 674–685.
16. Vergouwe Y, Moons KG, Steyerberg EW. External validity of risk models: Use of benchmark values to disentangle a case-mix effect from incorrect coefficients. *Am J Epidemiol.* 2010; 172: 971–980.
17. Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol.* 2016; 69: 245–247.

18. Counsell C, Dennis M, McDowall M. Predicting functional outcome in acute stroke: comparison of a simple six variable model with other predictive systems and informal clinical prediction. *J Neurol Neurosurg Psychiatry.* 2004; 75: 401–405.
19. Counsell C, Dennis M, McDowall M, Warlow C. Predicting outcome after acute and subacute stroke: development and validation of new prognostic models. *Stroke.* 2002; 33: 1041–1047.
20. Ntaios G, Faouzi M, Ferrari J, Lang W, Vemmos K, Michel P. An integer-based score to predict functional outcome in acute ischemic stroke: the ASTRAL score. *Neurology.* 2012; 78: 1916–1922.
21. Weimar C, Konig IR, Kraywinkel K, Ziegler A, Diener HC; German Stroke Study Collaboration. Age and National Institutes of Health Stroke Scale Score within 6 hours after onset are accurate predictors of outcome after cerebral ischemia: development and external validation of prognostic models. *Stroke.* 2004; 35: 158–162.
22. Myint PK, Clark AB, Kwok CS, Davis J, Durairaj R, Dixit AK, et al. The SOAR (Stroke subtype, Oxford Community Stroke Project classification, Age, prestroke modified Rankin) score strongly predicts early outcomes in acute stroke. *Int J Stroke.* 2014; 9: 278–283.
23. Ayis SA, Coker B, Rudd AG, Dennis MS, Wolfe CD. Predicting independent survival after stroke: a European study for the development and validation of standardised stroke scales and prediction models of outcome. *J Neurol Neurosurg Psychiatry.* 2013; 84: 288–296.
24. O'Donnell MJ, Fang J, D'Uva C, Saposnik G, Gould L, McGrath E, et al. The PLAN score: a bedside prediction rule for death and severe disability following acute ischemic stroke. *Arch Intern Med.* 2012; 172: 1548–1556.
25. Saposnik G, Kapral MK, Liu Y, Hall R, O'Donnell M, Raptis S, et al. IScore: a risk score to predict death early after hospitalization for an acute ischemic stroke. *Circulation.* 2011; 123: 739–749.
26. Smith EE, Shobha N, Dai D, Olson DM, Reeves MJ, Saver JL, et al. Risk score for inhospital ischemic stroke mortality derived and validated within the Get With the Guidelines-Stroke Program. *Circulation.* 2010; 122: 1496–1504.
27. Kwakkel G, Wagenaar RC, Kollen BJ, Lankhorst GJ. Predicting disability in stroke–a critical review of the literature. *Age Ageing.* 1996; 25: 479–489.
28. Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Stat Med.* 2002; 21: 2917–2930.
29. Roozenbeek B, Lingsma HF, Maas AI. New considerations in the design of clinical trials for traumatic brain injury. *Clin Investig Lond.* 2012; 2: 153–162.
30. Goyal M, Menon BK, van Zwam WH, Dippel DW, Mitchell PJ, Demchuk AM, et al. Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. *Lancet.* 2016; 387: 1723–1731.
31. Duffy L, Gajree S, Langhorne P, Stott DJ, Quinn TJ. Reliability (interrater agreement) of the Barthel Index for assessment of stroke survivors: systematic review and meta-analysis. *Stroke.* 2013; 44: 462–468.
32. Berzina G, Sveen U, Paanalahti M, Sunnerhagen KS. Analyzing the modified Rankin scale using concepts of the international classification of functioning, disability and health. *Eur J Phys Rehabil Med.* 2016; 52: 203–213.
33. Tversky A, Kahneman D. The framing of decisions and the psychology of choice. *Science.* 1981; 211: 453–458.

Appendix



Appendix 3.A. Outcome distribution of (**A**) the BI at discharge in PAIS and (**B**) the mRS at three months in PAIS, PRACTISE and PASS.



Appendix 3.B. Non-linear relation between age and the log odds of higher BI at discharge after acute ischemic stroke in PAIS.

Appendix 3.C. Discriminative ability of DSS-3 months at internal and external validation

Internal validation						
Cohort			**Apparent AUC**		**Internally validated AUC**	
PAIS (n = 1227)	mRS at three months		0.75 (0.74-0.77)		0.748 (0.74-0.77)	
External validation						
Cohort	**AUCs for different mRS cutoffs (95%CI)**				**Ordinal AUC (95%CI)**	**Ordinal case-mix-corrected AUC (95%CI)**
	≥ 2	≥ 3	≥ 4	≥ 6		
PRACTISE (n = 1589)	0.77 (0.77-0.79)	0.81 (0.81-0.84)	0.83 (0.83-0.86)	0.83 (0.83-0.86)	0.74 (0.72-0.76)	0.79 (0.76-0.79)
PASS (n = 2107)	0.66 (0.66-0.68)	0.81 (0.81-0.83)	0.84 (0.84-0.86)	0.85 (0.85-0.88)	0.69 (0.69-0.72)	0.77 (0.76-0.78)
Total (n = 4923)	0.73 (0.73-0.74)	0.81 (0.81-0.83)	0.84 (0.84-0.86)	0.83 (0.83-0.85)	0.73 (0.72-0.74)	0.78 (0.77-0.79)

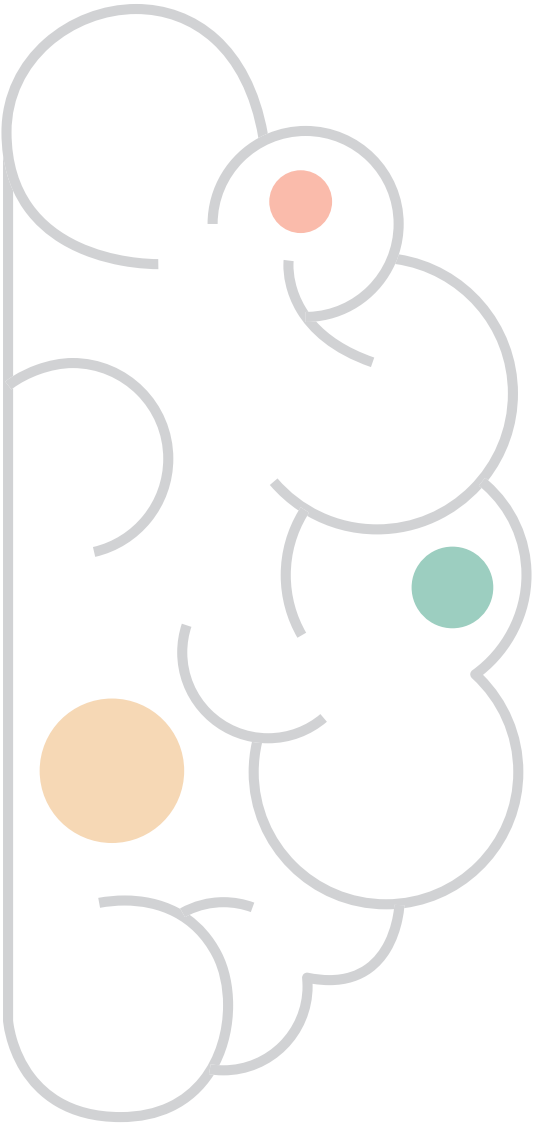
mRS: modified Rankin Scale; AUC: area under the curve.

Appendix 3.D. Details of the (**A**) DSS-discharge and (**B**) DSS-3 months.
The probability of each of the outcome categories is calculated according to the logistic formula: 1/(1 + exp-LP), in which LP stands for linear predictor.

A. To calculate the probability P on each of the five BI categories:
Slope Barthel for age<60 = nihss*-0.213 + 6*-0.468 + diabetes*-0.496
Slope Barthel for age>60 = nihss*-0.213 + age per decade*-0.468 + diabetes*-0.496
LP(Barthel 19-20) = Slope + 4.91
LP(Barthel 15-18) = Slope + 5.45 - Slope + 4.91
LP(Barthel 10-14) = Slope + 6.14 - Slope + 5.45
LP(Barthel 1-9) = Slope + 7.66 - Slope + 6.14
LP(Barthel 0) = 1 - Slope + 7.66

B. To calculate the probability P on each of the five mRS categories:
Slope Rankin for age<60 = nihss*0.182 + 6*0.495 + diabetes*0.410 + previous stroke*0.249 + atrial fibrillation*0.212
Slope Rankin for age>60 = nihss*0.182 + age per decade*0.495 + diabetes*0.410 + previous stroke*0.249 + atrial fibrillation*0.212
LP(Rankin 6) = Slope + -4.68
LP(Rankin 4-5) = Slope + -3.76 - Slope + -4.68
LP(Rankin 3) = Slope + -2.93 - Slope + -3.76
LP(Rankin 2) = Slope + -1.89 - Slope + -2.93
LP(Rankin 0-1) = 1 - Slope + -1.89

CHAPTER 4



Prognosis in moderate and severe traumatic brain injury: A systematic review of contemporary models and validation studies

Simone A. Dijkland
Kelly A. Foks
Suzanne Polinder
Diederik W.J. Dippel
Andrew I.R. Maas
Hester F. Lingsma
Ewout W. Steyerberg

J Neurotrauma 2019; Epub ahead of print

Abstract

Outcome prognostication in traumatic brain injury (TBI) is important but challenging due to heterogeneity of the disease. The aim of this systematic review is to present the current state-of-the-art on prognostic models for outcome after moderate and severe TBI and evidence on their validity. We searched for studies reporting on the development, validation or extension of prognostic models for functional outcome after TBI with Glasgow Coma Scale (GCS) ≤12 published between 2006-2018. Studies with patients aged ≥14 years and evaluating a multivariable prognostic model based on admission characteristics were included. Model discrimination was expressed with the area under the receiver operating characteristic curve (AUC), and model calibration with calibration slope and intercept. We included 58 studies describing 67 different prognostic models, comprising the development of 42 models, 149 external validations of 31 models and 12 model extensions. The most common predictors were GCS (motor) score (n=55), age (n=54) and pupillary reactivity (n=48). Model discrimination varied substantially between studies. The International Mission for Prognosis and Analysis of Clinical Trials (IMPACT) and Corticoid Randomisation After Significant Head injury (CRASH) models were developed on the largest cohorts (8,509 and 10,008 patients, respectively) and were most often externally validated (n=91), yielding AUCs ranging between 0.65-0.90 and 0.66-1.00, respectively. Model calibration was reported with a calibration intercept and slope for 7 models in 53 validations, and was highly variable. In conclusion, the discriminatory validity of the IMPACT and CRASH prognostic models is supported across a range of settings. The variation in calibration, reflecting heterogeneity in reliability of predictions, motivates continuous validation and updating if clinical implementation is pursued.

PROSPERO registry number: CRD42016052100

Introduction

Traumatic brain injury (TBI) is a major cause of injury-related death and disability.[1] It is a disease with a considerable economic impact, often affecting the working population.[2] Patients with TBI show substantial variation in injury mechanism, pathology, clinical severity and prognosis. Due to the heterogeneity of the disease, prediction of functional outcome after TBI is challenging. Outcome prognostication is important to assist clinicians in providing reliable information to patients and relatives, to guide clinical management and trial design, and to give insight in quality of care by comparing observed and expected outcomes.[3] Many prognostic models for functional outcome after moderate and severe TBI have been developed and validated, but their methodological quality was described as poor in reviews performed in 2006 and 2008.[4, 5]

Over the past decade, new prognostic models for moderate and severe TBI have been developed and existing models have been externally validated and extended in new datasets. The question remains whether the quality of the currently available models justifies further implementation in clinical practice. For instance, when informing a relative of a patient with severe TBI in the intensive care unit on prognosis, the physician might want to use a prognostic model to communicate the chance of recovery within the next six months. But can the use of this prognostic model be recommended in this setting and for this patient? The aim of this systematic review is to present the current state-of-the-art on prognostic models for outcome after moderate and severe TBI and to review their performance at internal and external validation.

Methods

This systematic review was conducted and reported in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement.[6] The protocol of this systematic review has been registered on PROSPERO (registration number 2016: CRD42016052100) and can be accessed at: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42016052100.

Literature search

We performed a literature search in Embase, Medline Ovid, Web of Science, Cochrane Central, PsychInfo Ovid and Google Scholar to identify articles published between January 1st 2006 and November 12th 2018 reporting on the development, validation or extension of models predicting outcome after moderate and severe TBI. We used search terms on the following topics: brain or head injury, prediction or prognosis, model, and mortality/survival or recovery (Appendix 4.A). Studies evaluating prognostic models in moderate and severe TBI published before 2006 were already incorporated in previous systematic reviews.[4,5] For comparison of model performance at internal versus external validation, the development studies of models published before 2006 reporting a performance measure were retrieved manually.[7-12]

Eligibility criteria

Studies were eligible if they reported on the development, validation and/or extension of multivariable prognostic models for functional outcome in patients aged ≥ 14 years with moderate and severe TBI. We included original articles that were published in English language between 2006 and 2018. Studies that enrolled both adults and children were included when >80% of the subjects was adult or when adults and children were analyzed and reported separately. Moderate or severe TBI was defined as a Glasgow Coma Scale (GCS) score ≤ 12.[13] When a study only reported inclusion of patients with moderate or severe TBI without defining this in terms of GCS, it was assumed that moderate referred to GCS 9-12 and severe referred to GCS 3-8. In case of a population including TBIs of all severities, the study was included when the data of patients with moderate and severe TBI were incorporated in the analyses (as regards the Corticoid Randomisation After Significant Head injury [CRASH] model) or analyzed separately. Studies that evaluated model performance in specific subgroups of patients (different age groups, patients that underwent neurosurgery) were also included. The predictors used in the models had to be based on patient data obtained in the first 24 hours after injury (on hospital admission), because early outcome prediction is important to provide informed expectations to relatives and to aid early inclusion of patients in clinical trials. Moreover, we wanted to enable comparison between different prognostic models within this review as well as between this study and previous literature.[4] No limitations existed concerning outcome measurement provided that functional outcome was measured between 14 days and 24 months after injury. We excluded reviews and qualitative studies, studies confined to the rehabilitation setting, studies that focused on patients with mild TBI (defined as GCS 13-15) and studies that focused on single predictors instead of a model containing multiple predictors.

One investigator (S.A.D.) carried out the literature search and assessed studies for eligibility on title and abstract, and subsequently on full text. In case of doubt, a second investigator (K.A.F.) was involved. Disagreements were resolved by discussion or by consultation with a third (senior) investigator (H.F.L.).

Data extraction

We used a data extraction form based on the Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modeling Studies (CHARMS) checklist.[14] One investigator (S.A.D.) extracted the data from the included studies, and a random check (20%) was performed by a second investigator (K.A.F.). To ensure consistency of the data extraction, the data extraction form was tested on two studies by both investigators. The random check showed no discrepancies.

For all studies, data on study design, study population and sample size, outcome measure and scale used (e.g. functional outcome according to the Glasgow Outcome Scale [Extended], GOS[E]) and timing of outcome assessment was collected. For each prognostic model described in the included studies, we extracted data on the following topics: type of model (e.g. regression analysis, decision tree), internal or external validation and model performance. Model performance can be expressed in terms of discrimination (ability of the model to distinguish between patients with good and poor outcome) and calibration (agreement between observed and predicted probabilities). A common measure for

discrimination is the area under the receiver operating characteristic curve (AUC or C-statistic). The AUC ranges from 0.5 (no discriminative ability) to 1 (perfect discrimination). Calibration is often tested with the Hosmer-Lemeshow goodness-of-fit test or assessed by a calibration slope and calibration intercept.[15] The calibration slope describes the effect of the predictors in the validation sample and should be equal to 1. The intercept indicates whether predictions are systematically too high or too low, and should ideally be zero.[16]

If one study reported on multiple prognostic models or multiple stages of prognostic modeling (e.g. development and validation), data extraction was performed separately for each model or stage. We classified prognostic models as separate models when they included a different set of prognostic variables. Modifications of existing prognostic models at external validation due to missing predictor data were not defined as separate models, nor were models with identical predictors but for different outcome measures (e.g. mortality and functional outcome) or outcomes measured at different time points. However, when prognostic models consisted of identical predictors but were developed on different cohorts with re-estimation of model parameters, we did consider them as independent models rather than as validation studies.

Model performance in terms of discrimination and calibration was summarized according to AUC, calibration intercept and calibration slope weighted for the square root of study sample size. Analyses were performed with R software version 3.3.1 (R Foundation for Statistical Computing, Vienna, Austria).

Results

The literature search identified 3246 unique studies, of which 3158 were excluded based on title and abstract. Of the 88 full texts screened, 58 studies met the eligibility criteria and were included in this review (Figure 4.1). Data of the 58 studies were collected between 1984 and 2017 (Appendix 4.B). Sample sizes ranged from 41[17] to 10,008 patients.[18] The included studies described the development, validation or extension of 67 different prognostic models (Appendix 4.B). This comprised the development of 42 models, 149 external validations of 31 models and 12 model extensions (Figure 4.1). Half of the studies (n=29, 50%) evaluated multiple models in one study (Appendix 4.B). The most frequently used predictors were GCS (motor) score (n=55), age (n=54) and pupillary reactivity (n=48) (Figure 4.2).

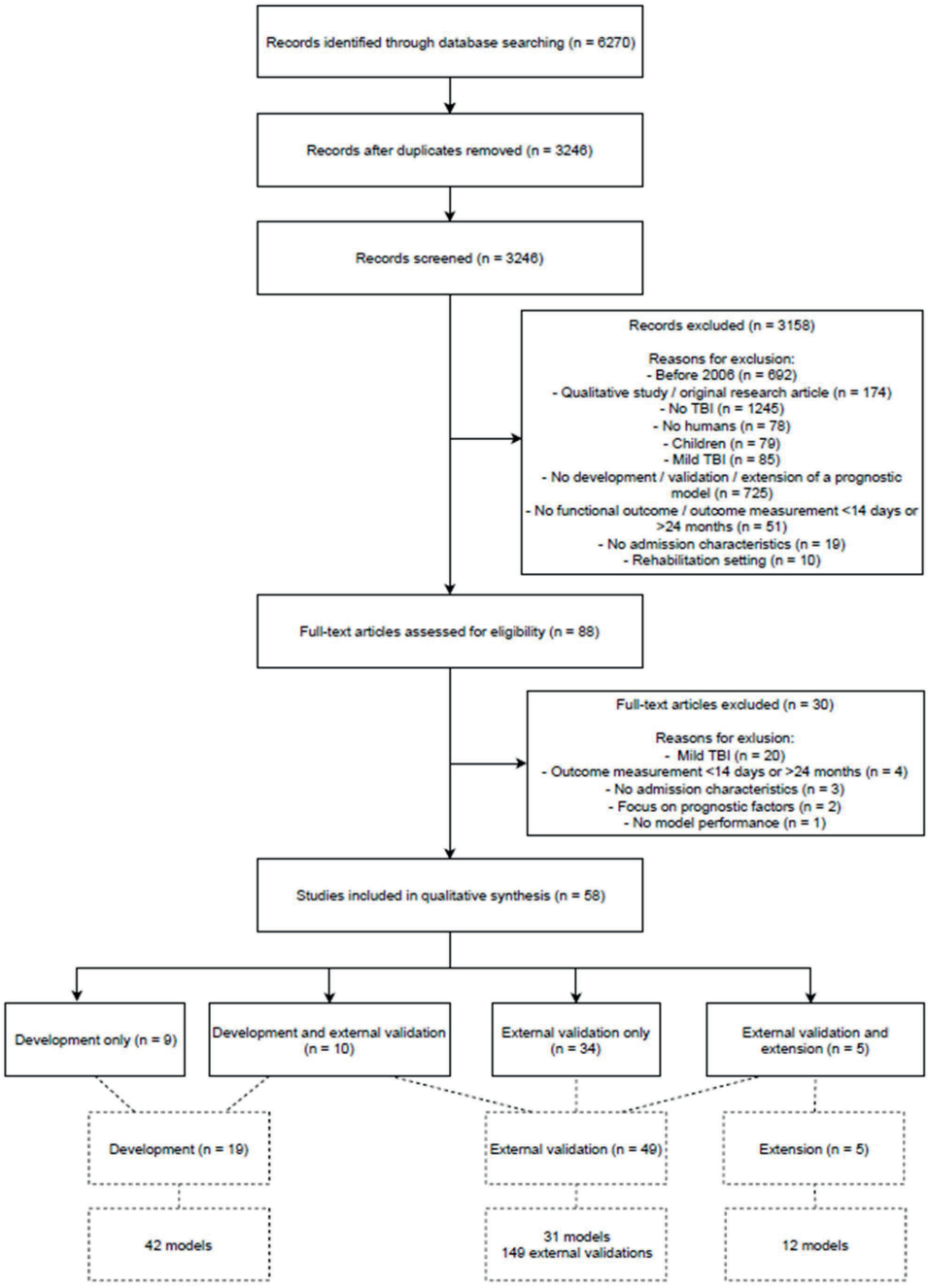


Figure 4.1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram of selected articles.

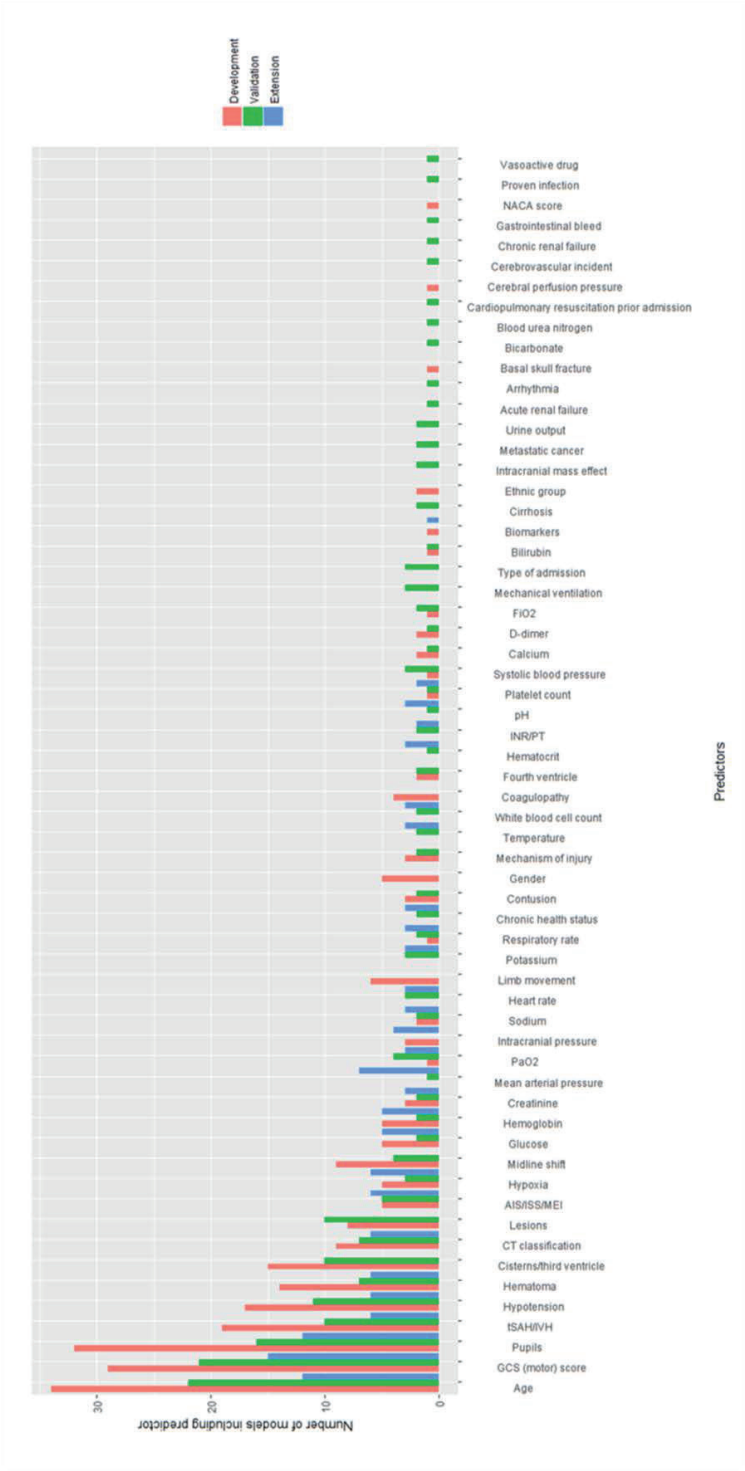


Figure 4.2. Overview of predictors included in 42 models for development, 31 models for external validation and 12 models for extension. GCS, Glasgow Coma Scale; tSAH, traumatic subarachnoid hemorrhage; IVH, intraventricular hemorrhage; CT, computed tomography; AIS, Abbreviated Injury Scale; ISS, Injury Severity Score; MEI, major extracranial injury; PaO2, partial arterial pressure of oxygen; INR, international normalized rate; PT, prothrombin time; pH, potential hydrogen; FiO2, fraction of inspired oxygen; NACA, National Advisory Committee for Aeronautics

Table 4.1. Summary of characteristics of development, validation and extension of models for moderate and severe traumatic brain injury

Characteristics	Development (n=42)	External validation (n=149)	Extension (n=12)
No. of models	42	31	12
Median number of patients (IQR)	700 (381-1466)	409 (290-890)	342 (160-534)
Type of model			
Regression analysis	40 (94)	142 (95)	12 (67)
Classification tree	1 (3)	7 (5)	-
Other[ab]	1 (3)	-	4 (33)
Internal validation		NA	
Apparent	15 (36)		4 (33)
Cross-validation	6 (14)		-
Bootstrapping	11 (26)		8 (67)
Split sample	13 (31)		3 (25)
Performance measures			
Calibration			
Plot	15 (36)	80 (54)	1 (8)
Goodness of fit	36 (86)	77 (52)	10 (83)
Slope	2 (5)	53 (36)	5 (42)
Intercept	2 (5)	53 (36)	5 (42)
Other[c]	2 (5)	7 (5)	3 (25)
Discrimination			
Accuracy rate	1 (8)	6 (4)	-
Sensitivity/specificity	2 (5)	4 (3)	-
ROC/AUC	32 (76)	142 (95)	11 (92)
Other[d]	13 (31)	39 (26)	8 (67)

IQR, interquartile range, ROC, receiver operating characteristic curve, AUC, area under the receiver operating characteristic curve

[a]E.g. Bayesian methods, discriminant analysis, machine learning

[b]One study compared five different statistical approaches on the same cohort: logistic regression, decision tree, neural network, Bayesian methods and discriminant analysis.[27]

[c]E.g. Calibration belt

[d]E.g. Nagelkerke R^2, Brier score

Model development and internal validation

Nineteen studies described the development of 42 prognostic models (1-8 models per study).[18-36] Cohorts for model development were mostly single center and prospective, with a median sample size of 700 patients (Appendix 4.B and Table 4.1). Moderate or severe TBI was defined according to the GCS score in all cohorts. All models had either mortality or unfavorable outcome according to the GOS(E) as outcome measure, assessed between 14 days and one year after trauma (Appendix 4.B). For the vast majority of models, unfavorable outcome was defined as GOS 1-3 or GOSE 1-4 (Appendix 4.B). Age, GCS (motor) score and pupils were the most frequently used predictors (Figure 4.2). Common radiological characteristics were traumatic subarachnoid hemorrhage or intraventricular hemorrhage (19 models), presence of hematoma (14 models), compression of cisterns and third ventricle (15 models) and Marshall or Rotterdam computed tomography (CT) classification (9 models). The most often used physiological predictor was hypotension (17 models). Several laboratory predictors were studied, among which glucose, hemoglobin and coagulopathy (Figure 4.2). Other less frequently used predictors included sex, mechanism of injury, ethnic group, and cerebral perfusion pressure (CPP) (Figure 4.2). Biomarkers, e.g. S100 astroglial calcium-binding protein B (S100B) and glial fibrillary acidic protein (GFAP), were only included in one newly developed model (Figure 4.2). Most models were developed with logistic regression (n=40, 94%) and internally validated with apparent or split-sample validation (Table 4.1). An AUC for internal validation was reported for 32 models (76%). The AUCs for the models for mortality ranged from 0.71 to 0.94, with a mean weighted AUC of 0.84. The models for unfavorable outcome showed AUCs ranging from 0.67 to 0.98 (mean weighted AUC 0.82).

External validation

In 49 studies, 149 external validations of newly developed (n=17) or existing (n=14) prognostic models were described (1-10 models per study).[17-19, 22, 25, 29-31, 33, 34, 36-74] The external validation cohorts had a median sample size of 409 patients, and were often multicenter (n=27, 56%) and prospective (n=37, 77%) (Table 4.1 and Appendix 4.B). The definition of moderate and severe TBI was mostly based on GCS score, but sometimes other criteria were used (e.g. loss of consciousness and Abbreviated Injury Scale ≥ 2) (Appendix 4.B). Five studies only included patients with severe TBI who underwent decompressive craniectomy.[17, 40, 48, 49, 72] The time of outcome assessment according to the GOS(E) was six months in most studies (n=36, 75%), and ranged between hospital discharge and 18 months (Appendix 4.B). The models at external validation included more physiological variables due to validation of several existing Intensive Care severity scores (e.g. Acute Physiology And Chronic Health Evaluation II, Sepsis-related Organ Failure Assessment score) (Figure 4.2). For each external validation, at least one performance measure was reported. Model calibration was most frequently expressed with a calibration plot (54%) or the Hosmer-Lemeshow goodness-of-fit test (52%) (Table 4.1). For 25 external validations, no measure of model calibration was reported. In 95% of the external validations, model discrimination was expressed in terms of an AUC (Table 4.1).

showed substantial variation (Figure 4.3). The AUCs at external validation ranged between 0.61-0.99 (mean weighted AUC 0.80) for the models for mortality, and between 0.66-1.00 (mean

weighted AUC 0.77) for the models for unfavorable outcome. We further focused on models with a reported AUC at internal validation and one or more external validations (n=20). Discriminative ability was slightly poorer at external validation compared to internal validation, with a mean AUC difference of -0.013 (p=0.086 by paired t-test) for prediction of mortality and -0.017 (p=0.031) for unfavorable outcome.

Model calibration, reported with a calibration intercept and slope, was summarized for the models that were externally validated once or more (7 models in 53 validations, Figure 4.4). We observed substantial variation in the agreement between observed and predicted probabilities. The mean weighted calibration intercept was -0.28 (range -3.3-0.93) for the models for mortality, and -0.019 (range -5.7-2.4) for the models for unfavorable outcome. This indicates that both mortality and unfavorable outcome were generally lower than expected. The mean weighted calibration slopes were 1.1 (range 0.42-2.3) and 0.88 (range 0.57-2.5) for mortality and unfavorable outcome respectively. The values at the extremes of the ranges for calibration slope and intercept were mainly due to selection of specific populations with moderate and severe TBI, such as patients who underwent decompressive craniectomy or TBI defined according to the Abbreviated Injury Scale.[48, 67]

The International Mission for Prognosis and Analysis of Clinical Trials (IMPACT) and CRASH models were most frequently externally validated (n=91). The mean weighted AUCs were 0.79 (mortality) and 0.77 (unfavorable outcome) for the IMPACT models (range 0.65-0.90), and 0.82 (mortality) and 0.78 (unfavorable outcome) for the CRASH models (range 0.66-1.00) (Figure 4.3 and Table 4.2). In total 51 external validations reported calibration with an intercept and slope. These 51 validations showed overestimated risks by the IMPACT and CRASH models for mortality and underestimated risks for unfavorable outcome (Figure 4.4 and Table 4.2). The more complex IMPACT and CRASH models, for example including CT characteristics, showed only modest improvement in discriminative ability (Appendix 4.C), and calibration remained highly variable (Figure 4.4). Comparison of the performance of the IMPACT and CRASH models with other models, such as Hukkelhoven and Nijmegen, was not feasible given the limited number of validations of these other models (Table 4.2).

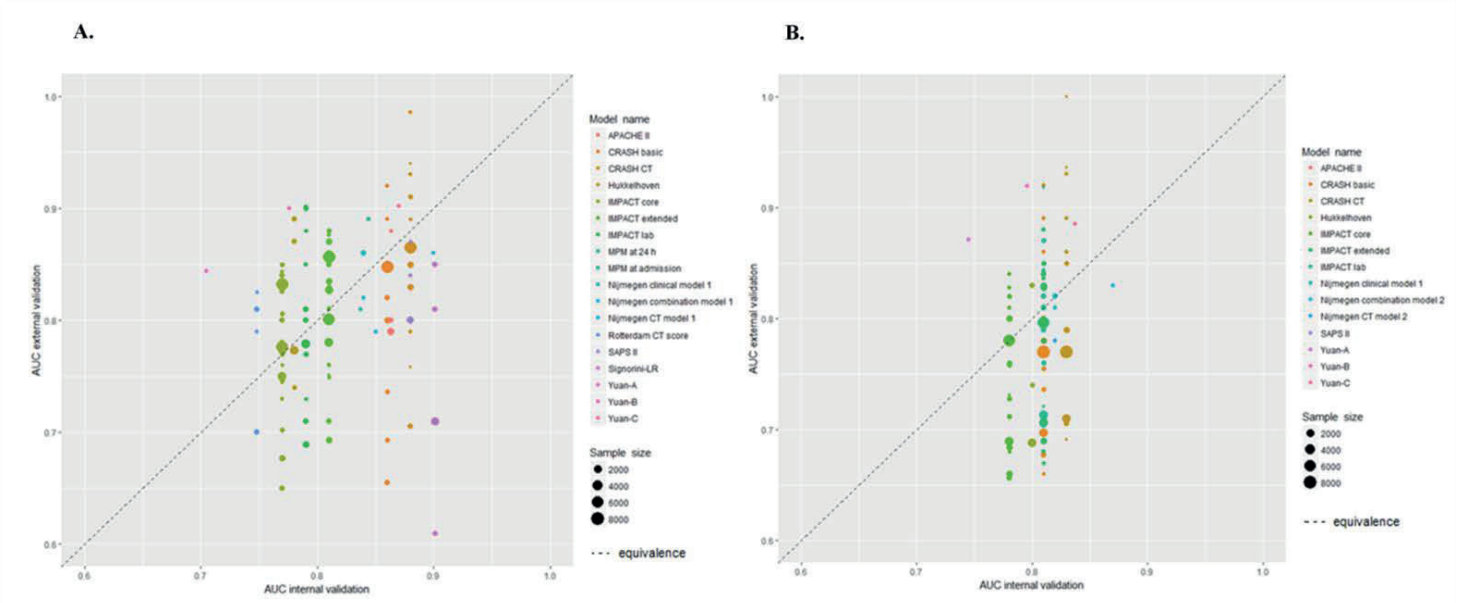


Figure 4.3. Correlation of discriminative ability (area under the receiver operating characteristic curve, AUC) of models for (A) mortality and (B) unfavorable outcome (defined as GOS 1-3 or GOSE 1-4) between internal and external validation. The colors of the dots represent the different prognostic models that have been externally validated once or more, and the dot size refers to the sample size of the different validation cohorts. The diagonal line indicates equivalence between model discrimination at internal and external validation. Above this line, model discrimination at external validation was better than at internal validation. The dots below the line indicate a higher AUC at internal validation compared to external validation.

APACHE, Acute Physiology And Chronic Health Evaluation; CRASH, Corticoid Randomisation After Significant Head injury; CT, computed tomography; IMPACT, International Mission for Prognosis and Analysis of Clinical Trials; MPM, mortality probability models; SAPS, Simplified Acute Physhiology Score; LR, logistic regression; AUC, area under the receiver operating characteristic curve

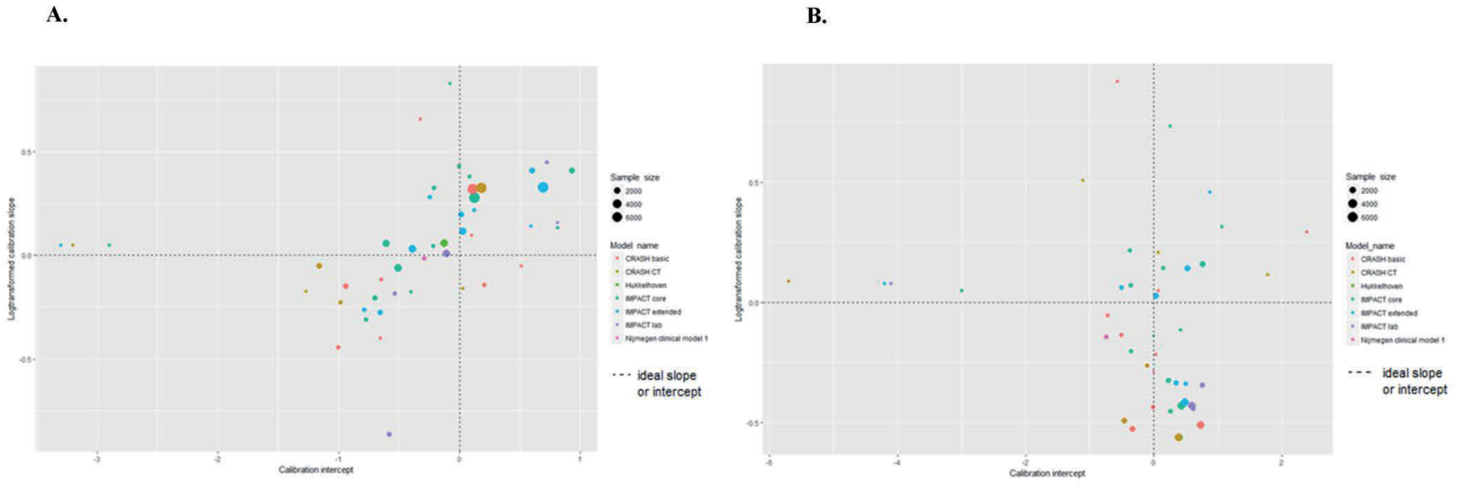


Figure 4.4. Calibration intercept and slope reported for models for (A) mortality and (B) unfavorable outcome (defined as GOS 1-3 or GOSE 1-4) at external validation. The colors of the dots represent the different prognostic models that have been externally validated once or more, and the dot size refers to the sample size of the different validation cohorts. The vertical line indicates the ideal calibration intercept and the horizontal line shows the perfect calibration slope. A calibration intercept > 0 indicates systematic underestimation of mortality or unfavorable outcome, and an intercept < 0 refers to systematic overestimation of outcome risk. A calibration slope below or above the horizontal line indicates that predictions were too extreme: low predictions too low, and high predictions too high. In short, the closer the dots are to the intersection of these lines, the better the ability of the model to provide reliable predictions for patients in that specific setting.

CRASH, Corticoid Randomisation After Significant Head injury; CT, computed tomography; IMPACT, International Mission for Prognosis and Analysis of Clinical Trials

Table 4.2. Summary model performance IMPACT and CRASH models versus other models at external validation

Performance measure	IMPACT models		CRASH models		Nijmegen clinical model 1		Hukkelhoven	
Mortality	Discrimination: 56 validations Calibration: 31 validations		Discrimination: 23 validations Calibration: 16 validations		Discrimination: 2 validations Calibration: 1 validation		Discrimination: 4 validations Calibration: 1 validation	
	Mean[a]	Range	Mean[a]	Range	Mean[a]	Range	Mean[a]	Range
AUC	0.79	0.65-0.90	0.82	0.66-0.99	0.84	0.82-0.86	0.81	0.74-0.89
Calibration slope	1.1	0.42-2.3	1.1	0.64-1.9	0.98	-	1.1	-
Calibration intercept	-0.22	-3.3-0.93	-0.41	-3.2-0.51	-0.29	-	-0.13	-
Unfavorable outcome	Discrimination: 55 validations Calibration: 26 validations		Discrimination: 24 validations Calibration: 17 validations		Discrimination: 2 validations Calibration: 1 validation		Discrimination: 3 validations Calibration: 1 validation	
	Mean[a]	Range	Mean[a]	Range	Mean[a]	Range	Mean[a]	Range
AUC	0.77	0.66-0.92	0.78	0.66-1.00	0.82	0.81-0.82	0.74	0.69-0.83
Calibration slope	0.90	0.63-2.1	0.89	0.57-2.5	0.87	-	0.57	-
Calibration intercept	0.044	-4.2-1.1	-0.13	-5.7-2.4	-0.74	-	0.39	-

[a]Weighted for the square root of sample size

IMPACT, International Mission for Prognosis and Analysis of Clinical Trials; CRASH, Corticoid Randomisation After Significant Head injury; AUC, area under the receiver operating characteristic curve

Model extensions

In five studies, 12 extensions of the IMPACT and CRASH prognostic models were assessed.[41, 44, 51, 57, 58] The median sample size of the extension cohorts was 342 patients (Table 4.1). Moderate and severe TBI patients were selected based on GCS, except for one cohort consisting of consecutive TBI patients requiring intracranial pressure monitoring.[44] Outcomes were assessed between one week and six months (Appendix 4.B). Most studies reported model discrimination with an AUC (n=11, 92%) and calibration with the Hosmer-Lemeshow goodness-of-fit test (n=10, 83%) (Table 4.1). The extensions included several serum and cerebrospinal fluid biomarkers, extracranial injury, coagulation parameters or dynamic predictors containing information on the first 24 hours of the clinical course (Acute Physiology And Chronic Health Evaluation II score, intracranial pressure and mean arterial pressure) (Figure 4.2). Performance of the extended models in terms of both discrimination and calibration improved somewhat compared to the original versions of the models. The mean AUC increase at model extension was 0.013 (p=0.18 by paired t-test) for models for mortality and 0.10 (p=0.026) for models for unfavorable outcome. Calibration was not evaluated or showed no improvement.[41, 44, 51, 57, 58] None of the extended models was externally validated.

Discussion

We systematically reviewed 58 papers describing the development, validation or extension of 67 different multivariable prognostic models for functional outcome in moderate and severe TBI. We identified 149 external validations of prognostic models. The IMPACT and CRASH models currently dominate the field of prognostic modeling in moderate and severe TBI. External validations of these models showed substantial variation in performance: overall moderate to good discrimination, but highly variable calibration.

Strengths and limitations

This systematic review is based on a comprehensive literature search resulting in a large number of prognostic models and validation studies in the field of moderate and severe TBI. A novel feature compared to previous systematic reviews on this topic is that improvements in prognostic research in TBI now permit inclusion of a substantial number of external validation studies. However, some limitations should be considered. We did not consider models for which the outcomes (mortality or unfavorable outcome) were measured at different time points as separate models. Similarly, models with identical predictors but for different outcome measures were not defined as separate models. This may have caused an underestimation of the number of prognostic models for moderate and severe TBI. Another factor that might have unjustly reduced the number of models is the exclusion of studies that were not published in English language. Additionally, most studies in this systematic review were conducted in middle and high income countries. Therefore, our results might not be generalizable to low income countries. Finally, comparing model calibration between different models and settings was difficult due

to variation in, or even absence of, calibration measures. Model calibration was reported in terms of an intercept and slope for only seven models. Our summary of model calibration might therefore not reflect the overall ability of the currently available models to provide predictions in individual patients.

Comparison with previous literature

Previous systematic reviews on prognostic models in moderate and severe TBI mainly focused on their methodological quality. Several recommendations were proposed to improve methodology and reporting of prognostic models.[4, 5] The prognostic models evaluated in the current systematic review showed advancements in reporting and statistical approaches, especially regarding external validation. Models were externally validated in independent cohorts and most validation studies reported appropriate model performance measures in terms of discrimination and calibration.[15] However, measures for discrimination are still more frequently reported than calibration measures. Moreover, although the Hosmer-Lemeshow goodness-of-fit test for model calibration is no longer recommended due to lack of power and interpretability, this was still used in more than half of the validations. The lack of adequate calibration measures is remarkable, since poor calibration implies that the predictions will be misleading when used in clinical practice. This may lead to harmful decision making.[75]

Model development and predictors

After publication of the previous systematic reviews, several new prognostic models for outcome prediction after moderate and severe TBI have been developed. Especially the introductions of the IMPACT and CRASH models have been important to confirm the core predictors for unfavorable outcome after moderate and severe TBI obtained at admission: older age, less responsive pupils and lower GCS (motor) score.[18, 31] Although these baseline predictors included in the IMPACT and CRASH models only explain around 35% of the variance in outcome, more complex models with additional predictors collected within 24 hours may not lead to substantial improvements in model performance.[3] This is supported by our observation that performance of the IMPACT and CRASH models showed only modest improvement in discriminative ability by adding CT characteristics, physiological and laboratory variables obtained within the first 24 hours, both at internal and external validation (Figure 4.3 and Appendix 4.C). However, prognostic estimates will be refined during the course of the disease, as may be considered in dynamic prediction models.[76] Any prognostic model should only be considered an addition to clinical experience.

In line with previous recommendations, other recently developed models introduced several new predictors (e.g. CPP, ethnic group, mechanism of injury, biomarkers).[3] However, many of these predictors were only included in a few models and not yet externally validated (Figure 4.2). Therefore, it remains difficult to assess the added value of these models and predictors. Further research is essential, especially external validation.

External validation

We found a large number of external validations in contemporary series. The IMPACT and CRASH models were externally validated most extensively. Model performance at external validation was on average close to performance at internal validation. Performance at external validation may best reflect the models' discriminative ability when applied in clinical practice.[77] The discriminative ability at external validation was mostly around 0.8, with one very small study even reporting an implausible AUC value of 1 for the CRASH CT model for unfavorable outcome.[74] Calibration varied highly among different models and studies. The variability in discriminative performance and calibration slopes is most likely attributable to differences in measurement of predictors or selection of the validation population.[78] For instance, a few studies investigated model performance in more homogeneous subgroups such as patients with decompressive craniectomy.[17, 40, 48, 49, 72] We also observed a substantial number of variations (i.e. differences in included predictors) on IMPACT and CRASH at external validation (Appendix 4.D), mostly due to discrepancies in predictor definitions or unavailability of predictor data.[18, 31, 38, 61, 62, 64, 68] Further, timing of outcome measurement varied substantially across different studies. Although most models were designed for outcome prediction at six months after injury, model performance was assessed in cohorts with outcome data available up to 18 months after injury.[48] Heterogeneity in baseline risk was noted according to calibration-in-the-large (intercept differences). This variability might be attributed to differences in distributions and effects of unmeasured covariates and is therefore often difficult to explain. The substantial heterogeneity in model performance across different settings indicates that models need to be recalibrated for each new setting before implementation in clinical practice is warranted.

Model extension

Highly variable model performance may be problematic when introducing the models to a specific clinical setting. Several stages have been identified in updating prognostic models, ranging from updating the intercept to addition of predictors.[16] There has been extensive research into the additional prognostic value of baseline biomarkers for TBI.[79] However, extending the IMPACT and CRASH models with markers of coagulation or serum and cerebrospinal fluid biomarkers (S100B and GFAP) barely improved model performance in the few studies that have been performed.[41, 58] Because TBI is a heterogeneous disease with a highly variable clinical course, adding new information as it becomes available over time or including factors that predict treatment response may be more promising to improve outcome prediction.[3] Extending the currently available models with such dynamic predictors has been uncommon so far, and yielded variable improvement in model performance.[44, 57] External validation of these extended models is lacking. Possibilities for updating the IMPACT and CRASH models are currently being evaluated in various studies, including the Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) study, Transforming Research and Clinical Knowledge in TBI (TRACK-TBI) dataset and Collaborative REsearch on ACute Traumatic Brain Injury in intensiVe Care Medicine in Europe (CREACTIVE).[80-82] Given the highly variable calibration, updating of

the baseline risk estimate (the intercept in the regression model) should be considered. Also, machine learning techniques are currently gaining interest and might be helpful for dynamic prediction.

Implementation in clinical practice

The availability of a large number of prognostic models for functional outcome after moderate and severe TBI suggests that outcome prediction is considered relevant for clinical practice. However, despite previous recommendations, none of the available models have been implemented in TBI guidelines. Their use in clinical practice is limited.[3] This might partly be explained by the lack of evidence-based treatment options in TBI,[83] limiting the use of prognostic models to select patients for individualized management. Previous studies evaluating the perceptions of physicians on utilization of the IMPACT calculator in clinical practice showed that approximately half of the clinicians involved in TBI care was aware of its existence. Of those, only 50% occasionally used the model in clinical practice.[84, 85] Factors limiting clinical use of the IMPACT calculator comprised mistrust in the IMPACT development data, utilization for research purposes only, time needed to gather the data required to complete the online tool, and concern about misinterpretation of prognostic estimates by patients and their families.[84, 85] However, the IMPACT calculator was reported to be useful for reducing variability between physicians with different levels of clinical experience.[85]

Model discrimination, although variable, was adequate in most studies. The lack of implementation can therefore not be explained by poor discriminative ability. Moreover, models do not necessarily need high discriminative performance to be accepted in clinical practice. Examples are the Hypertension, Abnormal Renal/Liver Function, Stroke, Bleeding History or Predisposition, Labile INR, Elderly, Drugs/Alcohol Concomitantly (HAS-BLED, AUC 0.65) and Congestive heart failure, Hypertension, Age, Diabetes mellitus, Prior Stroke or TIA or Thromboembolism (CHA2DS2-VASc, AUC 0.61) models that are commonly applied in neurovascular practice, and the extensively used Gail breast cancer models (pooled AUCs between 0.55-0.75).[86-88] Compared to these widely implemented tools, the models for outcome after moderate and severe TBI perform very well (weighted mean AUCs of 0.80 and 0.76 for mortality and unfavorable outcome, respectively).

Model calibration, on the other hand, showed substantial heterogeneity between different settings. The adequate discriminative ability and highly variable calibration may indicate that the models perform well at group level, but caution is required when using them to provide predictions for individual patients in a specific clinical setting.

Based on the main findings of this systematic review, we provided a set of recommendations regarding statistical evaluation and implementation of prognostic models in moderate and severe TBI (Table 4.3).

Table 4.3. Recommendations on (statistical) evaluation and implementation of prognostic models for moderate and severe TBI

- Continuous validation and updating of prognostic models is required to judge generalizability and transportability to other TBI populations.
- Calibration reflects the ability of the prognostic model to provide reliable predictions and should thus be reported at every external validation.
- The currently available prognostic models for moderate and severe TBI discriminate well between low risk and high risk patients.
- Caution is required when providing predictions for patients in a specific clinical setting.
- Prognostic models for moderate and severe TBI may need to be recalibrated for each new setting before implementation in clinical practice is warranted.

Conclusion

The IMPACT and CRASH prognostic models have been developed on the largest datasets and have adequate discriminative ability across a range of settings. The reliability of predictions is highly variable. We recommend implementation of these models in clinical practice, provided that they have been validated or updated for the specific clinical setting.

References

1. Majdan M, Plancikova D, Brazinova A, Rusnak M, Nieboer D, Feigin V, et al. Epidemiology of traumatic brain injuries in Europe: a cross-sectional analysis. *Lancet Public Health.* 2016; 1, e76-e83.
2. Maas AI, Stocchetti N, Bullock R. Moderate and severe traumatic brain injury in adults. *Lancet Neurol.* 2008; 7, 728-741.
3. Lingsma HF, Roozenbeek B, Steyerberg EW, Murray GD, Maas AI. Early prognosis in traumatic brain injury: from prophecies to predictions. *Lancet Neurol.* 2010; 9, 543-554.
4. Mushkudiani NA, Hukkelhoven CW, Hernandez AV, Murray GD, Choi SC, Maas AI et al. A systematic review finds methodological improvements necessary for prognostic models in determining traumatic brain injury outcomes. *J Clin Epidemiol.* 2009; 61, 331-343.
5. Perel P, Edwards P, Wentz R, Roberts I. Systematic review of prognostic models in traumatic brain injury. *BMC Med Inform Decis Mak.* 2006; 6, 38.
6. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol.* 2009; 62, 1006-1012.
7. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: A severity of disease classification system. *Crit Care Med.* 1985; 13, 818-829.
8. Le Gall JR, Lemeshow S, Saulnier, F. A new Simplified Acute Physiology Score (SAPS II) based on a European/ North American multicenter study. *JAMA.* 1993; 270, 2957-2963.
9. Lemeshow S, Teres D, Klar J, Avrunin JS, Gehlbach SH, Rapoport J. Mortality Probability Models (MPM II) based on an international cohort of intensive care unit patients. *JAMA.* 1993; 270, 2478-2486.
10. Signorini DF, Andrews PJ, Jones PA, Wardlaw JM, Miller JD. Predicting survival using simple clinical variables: a case study in traumatic brain injury. *J Neurol Neurosurg Psychiatry.* 1999; 66, 20-25.
11. Hukkelhoven CW, Steyerberg EW, Habbema JD, Farace E, Marmarou A, Murray GD, et al. Predicting outcome after traumatic brain injury: development and validation of a prognostic score based on admission characteristics. *J Neurotrauma.* 2005; 22, 1025-1039.
12. Maas AI, Hukkelhoven CW, Marshall LF, Steyerberg EW. Prediction of outcome in traumatic brain injury with computed tomographic characteristics: a comparison between the computed tomographic classification and combinations of computed tomographic predictors. *Neurosurgery.* 2005; 57, 1173-1182; discussion 1173-1182.
13. Teasdale G, Jennett B Assessment of coma and impaired consciousness. A practical scale. *Lancet.* 1974; 2, 81-84.
14. Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med.* 2014; 11, e1001744.
15. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology.* 2010; 21, 128-138.
16. Steyerberg EW. *In: Clinical Prediction Models: A Practical Approach to Development, Validation and Updating.* New York: Springer, 2009.
17. Honeybul S, Ho KM, Lind CRP, Corcoran T, Gillett GR. The retrospective application of a prediction model to patients who have had a decompressive craniectomy for trauma. *J Neurotrauma*. 2009; 26, 2179-2183.

18. Perel PA, Olldashi F, Muzha I, Filipi N, Lede R, Copertari P, et al. Predicting outcome after traumatic brain injury: Practical prognostic models based on large cohort of international patients. *BMJ.* 2008; 336, 425-429.
19. Cremer OL, Moons KGM, Van Dijk GW, van Balen P, Kalkman CJ. Prognosis following severe head injury: Development and validation of a model for prediction of death, disability, and functional recovery. *J Trauma.* 2006; 61, 1484-1491.
20. Demetriades D, Kuncir E, Brown CVR, Martin M, Salim A., Rhee P, Chan LS. Early prediction of mortality in isolated head injury patients: A new predictive model. *J Trauma.* 2006, 61, 868-872.
21. Fabbri A, Servadei F, Marchesini G, Stein SC, Vandelli A. Early predictors of unfavourable outcome in subjects with moderate head injury in the emergency department. *J Neurol Neurosurg and Psychiatry.* 2008; 79, 567-573.
22. Gradisek P, Osredkar J, Korsic M, Kremzar B. Multiple indicators model of long-term mortality in traumatic brain injury. *Brain Inj.* 2012; 26, 1472-1481.
23. Han JX, See AAQ, Gandhi M, King NKK. Models of mortality and morbidity in severe traumatic brain injury: An analysis of a Singapore neurotrauma database. *World Neurosurg.* 2017; 108, 885-893.e881.
24. Iba J, Tasaki O, Hirao T, Mohri T, Yoshiya K, Hayakawa K, et al. Outcome prediction model for severe traumatic brain injury. *Acute Med Surg.* 2014; 1, 31-36.
25. Jacobs B, Beems T, van der Vliet TM, van Vugt AB, Hoedemaekers C, Horn J, et al. Outcome prediction in moderate and severe traumatic brain injury: A focus on computed tomography variables. *Neurocrit Care.* 2013; 19, 79-89.
26. Kamal VK, Agrawal D, Pandey RM. Prognostic models for prediction of outcomes after traumatic brain injury based on patients admission characteristics. *Brain Inj.* 2016; 30, 393-406.
27. Pang BC, Kuralmani V, Joshi R, Hongli Y, Kah KL, Beng TA, et al. Hybrid outcome prediction model for severe traumatic brain injury. *J Neurotrauma.* 2007; 24, 136-146.
28. Petroni G, Quaglino M, Lujan S, Kovalevski L, Rondina C, Videtta W, et al. Early Prognosis of severe traumatic brain injury in an urban argentinian trauma center. *J Trauma.* 2010; 68, 564-570.
29. Raj R, Skrifvars MB, Bendel S, Selander T, Kivisaari R, Siironen J, et al. Predicting six-month mortality of patients with traumatic brain injury: Usefulness of common intensive care severity scores. *Crit Care.* 2014; 18, R60.
30. Rizoli S, Petersen A, Bulger E, Coimbra R, Kerby JD, Minei J, et al. Early prediction of outcome after severe traumatic brain injury: A simple and practical model. *BMC Emerg Med.* 2016; 16, 32.
31. Steyerberg EW, Mushkudiani N, Perel P, Butcher I, Lu J, McHugh GS, et al. Predicting outcome after traumatic brain injury: Development and international validation of prognostic scores based on admission characteristics. *PLoS Med.* 2008; 5, 1251-1261.
32. Tasaki O, Shiozaki T, Hamasaki T, Kajino K, Nakae H, Tanaka H, et al. Prognostic indicators and outcome prediction model for severe traumatic brain injury. *J Trauma.* 2009; 66, 304-308.
33. Yuan F, Ding J, Chen H, Guo Y, Wang G, Gao WW, et al. Predicting outcomes after traumatic brain injury: The development and validation of prognostic models based on admission characteristics. *J Trauma Acute Care Surg.* 2012; 73, 137-145.
34. Mahadewa TGB, Golden N, Saputra A, Ryalino C. Modified Revised Trauma-Marshall score as a proposed tool in predicting the outcome of moderate and severe traumatic brain injury. *Open Access Emerg Med.* 2018; 10, 135-139.

35. Pannatier M, Delhumeau C, Walder B. Comparison of two prehospital predictive models for mortality and impaired consciousness after severe traumatic brain injury. *Acta Anaesthesiol Scand.* 2018; 63, 74-85.
36. Rached MAKB, Gaudet JG, Delhumeau C, Walder B. Comparison of two simple models for prediction of short term mortality in patients after severe traumatic brain injury. *Injury.* 2018; 50, 65-72.
37. Bonds BW, Dhanda A, Wade C, Massetti J, Diaz C, Stein DM. Prognostication of mortality and long term functional outcomes following traumatic brain injury: Can we do better? *J Neurotrauma.* 2015; Epub ahead of print.
38. Castaño-Leon AM, Lora D, Munarriz PM, Cepeda S, Paredes I, de La Cruz J, et al. Predicting outcomes after severe and moderate traumatic brain injury: An external validation of IMPACT and CRASH prognostic models in a large Spanish cohort. *J Neurotrauma.* 2016; 33, 1598-1606.
39. Charry J, Falla J, Ochoa J, Pinzón M, Tejada J, Henriquez M, et al. External validation of the rotterdam computed tomography score in the prediction of mortality in severe traumatic brain injury. *J Neurosci Rural Pract.* 2017; 8, 23-26.
40. Charry JD, Tejada JH, Pinzon MA, Tejada WA, Ochoa JD, Falla M, et al. Predicted unfavorable neurologic outcome is overestimated by the Marshall computed tomography score, Corticosteroid Randomization After Significant Head Injury (CRASH), and International Mission for Prognosis and Analysis of Clinical Trials in Traumatic Brain Injury (IMPACT) models in patients with severe traumatic brain injury managed with early decompressive craniectomy. *World Neurosurg.* 2017; 101, 554-558.
41. Czeiter E, Mondello S, Kovacs N, Sandor J, Gabrielli A, Schmid K, et al. Brain injury biomarkers may improve the predictive power of the IMPACT outcome calculator. *J Neurotrauma.* 2012; 29, 1770-1778.
42. Egea-Guerrero JJ, Rodríguez-Rodríguez A, Gordillo-Escobar E, Fernández-Delgado E, Martínez-Roldán Á, Roldán-Reina Á, et al. IMPACT score for traumatic brain injury: Validation of the prognostic tool in a Spanish cohort. *J Head Trauma Rehabil.* 2017; 33, E46-E52.
43. Fischler L, Lelais F, Young J, Buchmann B, Pargger H, Kaufmann M. Assessment of three different mortality prediction models in four well-defined critical care patient groups at two points in time: A prospective cohort study. *Eur J Anaesthesiol.* 2007; 24, 676-683.
44. Güiza F, Depreitere B, Piper I, van den Berghe G, Meyfroidt G. Novel methods to predict increased intracranial pressure during intensive care and long-term neurologic outcome after traumatic brain injury: Development and validation in a multicenter dataset. *Crit Care Med.* 2013; 41, 554-564.
45. Han J, King NKK, Neilson SJ, Gandhi MP, Ng I. External validation of the CRASH and IMPACT prognostic models in severe traumatic brain injury. *J Neurotrauma.* 2014; 31, 1146-1152.
46. Harrison DA, Griggs KA, Prabhu G, Gomes M, Lecky FE, Hutchinson PJA, et al. External validation and recalibration of risk prediction models for acute traumatic brain injury among critically ill adult patients in the United Kingdom. *J Neurotrauma.* 2015; 32, 1522-1537.
47. Hashemi B, Amanat M, Baratloo A, Forouzanfar MM, Rahmati F, Motamedi M, et al. Validation of CRASH model in prediction of 14-day mortality and 6-month unfavorable outcome of head trauma patients. *Emergency (Tehran).* 2016; 4, 196-201.
48. Honeybul S, Ho KM. Predicting long-term neurological outcomes after severe traumatic brain injury requiring decompressive craniectomy: A comparison of the CRASH and IMPACT prognostic models. *Injury.* 2016; 47, 1886-1892.

49. Honeybul S, Ho KM, Lind CRP, Gillett GR. Validation of the CRASH model in the prediction of 18-month mortality and unfavorable outcome in severe traumatic brain injury requiring decompressive craniectomy. *J Neurosurg.* 2014; 120, 1131-1137.
50. Hukkelhoven CWPM, Rampen AJJ, Maas AIR, Farace E, Habbema JDF, Marmarou A, et al. Some prognostic models for traumatic brain injury were not valid. *J Clin Epidemiol.* 2006; 59, 132-143.
51. Lingsma H, Andriessen TMJC, Haitsema I, Horn J, van der Naalt J, Franschman G, et al. Prognosis in moderate and severe traumatic brain injury: External validation of the IMPACT models and the role of extracranial injuries. *J Trauma Acute Care Surg.* 2013; 74, 639-646.
52. Majdan M, Brazinova A, Rusnak M, Leitgeb J. Outcome prediction after traumatic brain injury: Comparison of the performance of routinely used severity scores and multivariable prognostic models. J Neurosci Rural Pract. 2017; 8, 20-29.
53. Majdan M, Lingsma HF, Nieboer D, Mauritz W, Rusnak M, Steyerberg EW. Performance of IMPACT, CRASH and Nijmegen models in predicting six month outcome of patients with severe or moderate TBI: an external validation study. *Scand J Trauma Resusc Emerg Med.* 2014; 22, 68.
54. Olivecrona M, Koskinen LOD. The IMPACT prognosis calculator used in patients with severe traumatic brain injury treated with an ICP-targeted therapy. *Acta Neurochirur (Wien).* 2012; 154, 1567-1573.
55. Olivecrona M, Olivecrona Z. Use of the CRASH study prognosis calculator in patients with severe traumatic brain injury treated with an intracranial pressure-targeted therapy. *J Clin Neurosci.* 2013; 20, 996-1001.
56. Panczykowski DM, Puccio AM, Scruggs BJ, Bauer JS, Hricik AJ, Beers SR, et al. Prospective independent validation of IMPACT modeling as a prognostic tool in severe traumatic brain injury. *J Neurotrauma.* 2012; 29, 47-52.
57. Raj R, Siironen J, Kivisaari R, Hernesniemi J, Skrifvars MB. Predicting outcome after traumatic brain injury: Development of prognostic scores based on the IMPACT and the APACHE II. *J Neurotrauma.* 2014; 31, 1721-1732.
58. Raj R, Siironen J, Kivisaari R, Hernesniemi J, Tanskanen P, Handolin L, Skrifvars MB. External validation of the International Mission for Prognosis and Analysis of Clinical Trials model and the role of markers of coagulation. *Neurosurgery.* 2013; 73, 305-311.
59. Røe C, Skandsen T, Manskow U, Ader T, Anke A. Mortality and one-year functional outcome in elderly and very old patients with severe traumatic brain injuries: Observed and predicted. *Behav Neurol.* 2015.
60. Rønning PA, Pedersen T, Skaga NO, Helseth E, Langmoen IA, Stavem K. External validation of a prognostic model for early mortality after traumatic brain injury. *J Trauma.* 2011; 70, E56-E61.
61. Roozenbeek B, Chiu YL, Lingsma HF, Gerber LM, Steyerberg EW, Ghajar J, et al. Predicting 14-day mortality after severe traumatic brain injury: Application of the IMPACT models in the brain trauma foundation TBI-trac® New York state database. *J Neurotrauma.* 2012; 29, 1306-1312.
62. Roozenbeek B, Lingsma HF, Lecky FE, Lu J, Weir J, Butcher I, et al. Prediction of outcome after moderate and severe traumatic brain injury: External validation of the International Mission on Prognosis and Analysis of Clinical Trials (IMPACT) and Corticoid Randomisation after Significant Head injury (CRASH) prognostic models. *Crit Care Med.* 2012; 40, 1609-1617.

63. Staples JA, Wang J, Mills B, Temkin N, Zaros MC, Jurkovich GJ, et al. The application of the CRASH-CT prognostic model for older adults with traumatic brain injury: A population-based observational cohort study. *J Head Trauma Rehabil.* 2016; 31, E8-E14.
64. Staples JA, Wang J, Zaros MC, Jurkovich GJ, Rivara FP. The application of IMPACT prognostic models to elderly adults with traumatic brain injury: A population-based observational cohort study. *Brain Inj.* 2016; 30, 899-907.
65. Sun H, Lingsma HF, Steyerberg EW, Maas AIR. External validation of the International Mission for Prognosis and Analysis of Clinical Trials in traumatic brain injury: Prognostic models for traumatic brain injury on the study of the neuroprotective activity of progesterone in severe traumatic brain injuries trial. *J Neurotrauma.* 2016; 33, 1535-1543.
66. Wan X, Zhao K, Wang S, Zhang H, Zeng L, Wang Y, et al. Is it reliable to predict the outcome of elderly patients with severe traumatic brain injury using the IMPACT prognostic calculator? *World Neurosurg.* 2017; 103, 584-590.
67. Wong GKC, Teoh J, Yeung J, Chan E, Siu E, Woo P, et al. Outcomes of traumatic brain injury in Hong Kong: Validation with the TRISS, CRASH, and IMPACT models. *J Clin Neurosci.* 2013; 20, 1693-1696.
68. Yeoman P, Pattani H, Silcocks P, Owen V, Fuller G. Validation of the IMPACT outcome prediction score using the Nottingham head injury register dataset. *J Trauma.* 2011; 71, 387-392.
69. Kandil A, Kenawi M, Samir A, Hussein K. Traumatic brain injury predictive value of common intensive care severity scores. *Res Opin Anesth Intensive Care.* 2017; 4, 124.
70. Cicuendez M, Castano-Leon A, Ramos A, Hilario A, Gomez PA, Lagares A. The added prognostic value of magnetic resonance imaging in traumatic brain injury: The importance of traumatic axonal injury when performing ordinal logistic regression. *J Neuroradiol.* 2018; Epub ahead of print.
71. Faried A, Satriawan FC, Arifin MZ. Feasibility of online traumatic brain injury prognostic corticosteroids randomisation after significant head injury (CRASH) model as a predictor of mortality. *World Neurosurg.* 2018; 116, e239-e245.
72. Ho KM, Honeybul S, Ambati R. Prognostic significance of magnetic resonance imaging in patients with severe nonpenetrating traumatic brain injury requiring decompressive craniectomy. *World Neurosurg.* 2018; 112, 277-283.
73. Sadaka F, Doctors N, Pearson T, Snyders B, O'Brien J. Does red cell distribution width predict outcome in traumatic brain injury: comparison to corticosteroid randomization after significant head injury. *J Clin Med Res.* 2018; 10, 9.
74. Sadaka F, Jadhav A, Miller M, Saifo A, O'Brien J, Trottier S. Is it possible to recover from traumatic brain injury and a Glasgow coma scale score of 3 at emergency department presentation? *Am J Emerg Med.*2018.
75. Van Calster B, Vickers AJ. Calibration of risk prediction models: impact on decision-analytic performance. *Med Decis Making.* 2015; 35, 162-169.
76. van Houwelingen HC, Putter H. *In: Dynamic Prediction in Clinical Survival Analysis.* Taylor & Francis Inc. 2011.
77. Siontis GC, Tzoulaki I, Castaldi PJ, Ioannidis JP. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *J Clin Epidemiol.* 2015; 68, 25-34.
78. Vergouwe Y, Moons KG, Steyerberg EW. External validity of risk models: Use of benchmark values to disentangle a case-mix effect from incorrect coefficients. *Am J Epidemiol.* 2010; 172, 971-980.

79. Maas AIR, Menon DK, Adelson PD, Andelic N, Bell, MJ, Belli A et al. Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. *Lancet Neurol.* 2017; 16, 987-1048.
80. Maas AI, Menon DK, Steyerberg EW, Citerio G, Lecky F, Manley GT, et al. Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI): a prospective longitudinal observational study. *Neurosurgery.* 2015; 76, 67-80.
81. Yue JK, Vassar MJ, Lingsma HF, Cooper SR, Okonkwo DO, Valadka AB, et al. Transforming research and clinical knowledge in traumatic brain injury pilot: multicenter implementation of the common data elements for traumatic brain injury. *J Neurotrauma.* 2013; 30, 1831-1844.
82. CREACTIVE - Collaborative REsearch on ACute Traumatic Brain Injury in intensiVe Care Medicine in Europe (CREACTIVE). ClinicalTrials.gov. https://clinicaltrials.gov/ct2/show/NCT02004080. Accessed December 3, 2018.
83. Carney N, Totten AM, O'Reilly C, Ullman JS, Hawryluk GW, Bell MJ, et al. Guidelines for the Management of Severe Traumatic Brain Injury, Fourth Edition. *Neurosurgery.* 2017, 80, 6-15.
84. Letsinger J, Rommel C, Hirschi R, Nirula R, Hawryluk GWJ. The aggressiveness of neurotrauma practitioners and the influence of the IMPACT prognostic calculator. *PloS One.* 2017; 12, e0183552.
85. Moskowitz J, Quinn T, Khan MW, Shutter L, Goldberg R, Col N, et al. Should we use the IMPACT-model for the outcome prognostication of TBI patients? A qualitative study assessing physicians' perceptions. *MDM Policy Pract.* 2018; 3, 2381468318757987.
86. Wang X, Huang Y, Li L, Dai H, Song F, Chen K. Assessment of performance of the Gail model for predicting breast cancer risk: a systematic review and meta-analysis with trial sequential analysis. *Breast Cancer Res.* 2018; 20.
87. Lip GY, Frison L, Halperin JL, Lane DA. Comparative validation of a novel risk score for predicting bleeding risk in anticoagulated patients with atrial fibrillation: the HAS-BLED (Hypertension, Abnormal Renal/Liver Function, Stroke, Bleeding History or Predisposition, Labile INR, Elderly, Drugs/Alcohol Concomitantly) score. *J Am Coll Cardiol.* 2011; 57, 173-180.
88. Lip GY, Nieuwlaat R, Pisters R, Lane DA, Crijns HJ. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. *Chest.* 2010; 137, 263-272.

Appendix

Appendix 4. A. Search strategy

Date search: November 12, 2018

Database	Search algorithm
Embase.com	('traumatic brain injury'/exp OR 'brain injury'/de OR 'head injury'/de OR 'acquired brain injury'/de OR ('nervous system injury'/de AND brain/exp) OR (((trauma* OR injur* OR damage*) NEAR/3 (brain* OR cerebral* OR head OR cranial* OR intracranial*)) OR tbi):ab,ti) AND (((model/de OR 'mathematical model'/de OR 'disease model'/de) AND ('prognosis'/de OR 'prediction'/de OR 'mortality'/de OR 'survival'/de OR fatality/de OR 'convalescence'/de OR 'predictive validity'/de)) OR 'nomogram'/de OR (((prognos* OR predict* OR mortal* OR convalescen* OR recover* OR surviv* OR fatal*) NEAR/6 (model*)) OR nomogram*):ab,ti) NOT ([animals]/lim NOT [humans]/lim) NOT ([Conference Abstract]/lim OR [Letter]/lim OR [Note]/lim OR [Editorial]/lim) AND [english]/lim
Medline Ovid	("Brain Injuries"/ OR exp "Brain Hemorrhage, Traumatic"/ OR "Craniocerebral Trauma"/ OR "Head Injuries, Closed"/ OR "Head Injuries, Penetrating"/ OR ("Trauma, Nervous System"/ AND exp brain/) OR (((trauma* OR injur* OR damage*) ADJ3 (brain* OR cerebral* OR head OR cranial* OR intracranial*)) OR tbi).ab,ti.) AND (((exp "Models, Statistical"/ OR exp "Models, Theoretical"/) AND ("prognosis"/ OR exp "mortality"/ OR "mortality".xs. OR survival/ OR "Fatal Outcome"/ OR "Convalescence"/)) OR "Nomograms"/ OR (((prognos* OR predict* OR mortal* OR convalescen* OR recover* OR surviv* OR fatal*) ADJ6 (model*)) OR nomogram*).ab,ti.) NOT (exp animals/ NOT humans/) NOT (letter OR news OR comment OR editorial OR congresses OR abstracts).pt. AND english.la.
PsycINFO Ovid	("Traumatic Brain Injury"/ OR exp "Brain Damage"/ OR "Head Injuries"/ OR (((trauma* OR injur* OR damage*) ADJ3 (brain* OR cerebral* OR head OR cranial* OR intracranial*)) OR tbi).ab,ti.) AND (((("Models"/) AND ("prognosis"/ OR exp "Death and Dying"/ OR "Mortality Rate"/)) OR "Nomograms"/ OR (((prognos* OR predict* OR mortal* OR convalescen* OR recover* OR surviv* OR fatal*) ADJ6 (model*)) OR nomogram*).ab,ti.) NOT (exp animals/ NOT humans/) NOT (letter OR news OR comment OR editorial OR congresses OR abstracts OR books).pt. AND english.la.
Cochrane central	((((trauma* OR injur* OR damage*) NEAR/3 (brain* OR cerebral* OR head OR cranial* OR intracranial*)) OR tbi):ab,ti) AND ((((prognos* OR predict* OR mortal* OR convalescen* OR recover* OR surviv* OR fatal*) NEAR/6 (model*)) OR nomogram*):ab,ti)
Web of science	TS=(((((trauma* OR injur* OR damage*) NEAR/2 (brain* OR cerebral* OR head OR cranial* OR intracranial*)) OR tbi)) AND ((((prognos* OR predict* OR mortal* OR convalescen* OR recover* OR surviv* OR fatal*) NEAR/5 (model*)) OR nomogram*)) NOT ((animal* OR rat OR rats OR mouse OR mice OR murine OR cat OR cats OR feline OR dog OR dogs OR canine OR sheep OR ovine OR cow OR bovine OR cattle OR horse OR equin* OR pig OR swine OR porcine OR monkey* OR primate* OR gerbil* OR rabbit* OR rodent*) NOT (human* OR patient*))) AND DT=(article) AND LA=(english)
Google scholar	"brain|cerebral|head|cranial|intracranial trauma|injury|injuries|damage"|tbi "prognosis|prognostic|predictive|survival model|models"|" model|models*mortality|convalescence|recovery|fatality|fatal"|nomogram|nomograms

Appendix 4.B. Characteristics of 58 studies presenting 67 prediction models for moderate and severe traumatic brain injury

First author	Year	Data collection	Study design	Inclusion criteria	Model	Model no.	Outcome measure	Time of outcome	Number of patients
					Development				
Mahadewa[34]	2018	2017	Prospective observational cohort Single center	Moderate or severe TBI	Modified Revised Trauma-Marshall score	1	Unfavorable outcome (GOS 1-3)	6 months	181
Pannatier[35]	2018	2007-2010	Retrospective observational cohort (based on prospective data) Multicenter	Severe TBI with head AIS > 3	NACA-BM GCS-RM	2 3	Mortality	14 days	677
Rached[36]	2018	2007-2010	Retrospective observational cohort (based on prospective data) Multicenter	Severe TBI with head AIS > 3	HAIS-based model	4	Mortality	14 days	808
Han[23]	2017	2006-2009	Retrospective observational cohort (based on prospective data) Single center	Severe TBI with GCS ≤ 8	Han - NNI - NNI+	 5 6	Mortality and unfavorable outcome (GOS 1-3)	14 days, 6 months	300
Kamal[26]	2016	2010-2012	Retrospective observational cohort Single center	TBI with GCS ≤ 12	Kamal - Model 1 - Model 2 - Model 3 Kamal + gender - Model 1 - Model 2 - Model 3	 7 8 9 10 11 12	Mortality Unfavorable outcome (GOS 1-3)	30 days, 6 months	1466
Rizoli[30]	2016	2006-2009	Retrospective analysis of a RCT Multicenter	Severe TBI with GCS ≤ 8	Decision tree	13	Unfavorable outcome (GOSE 1-4)	6 months	1089

Iba[24]	2014	1997-2009	Retrospective observational cohort (based on prospective data) Multicenter	Severe TBI with GCS ≤ 8	Iba	14	Unfavorable outcome (GOS 1-3)	6 months	253
Raj[29]	2014	2003-2012	Prospective observational cohort Multicenter	TBI with GCS ≤ 13	Adjusted SOFA Reference model	15 16	Mortality	6 months	1625
Jacobs[25]	2013	1998-2006	Prospective observational cohort Single center	TBI with GCS ≤ 12	Nijmegen-clinical - model 1 - model 2 Nijmegen-CT - model 1 - model 2 - model 3 Nijmegen-combination - model 1 - model 2 - model 3	 17 18 19 20 21 22 23 24	Mortality Unfavorable outcome including death (GOSE 1-4) Unfavorable outcome excluding death (GOSE 2-4)	6 months	700
Gradisek[22]	2012	2007-2010	Prospective observational cohort Single center	Severe TBI with GCS ≤ 8	Gradisek – admission	25	Mortality	1 year	84
Yuan[33]	2012	2007-2009	Retrospective observational cohort Single center	TBI with GCS ≤ 12	Yuan - model A - model B - model C - model D	 26 27 28 29	Mortality and unfavorable outcome (GOS 1-3)	30 days, 6 months	1016
Petroni[28]	2010	2000-2003	Prospective observational cohort Single center	Severe TBI with GCS ≤ 8	Petroni - basic - CT	 30 31	Mortality and unfavorable outcome (GOSE 1-4)	6 months	148

Appendix 4.B. continued

Tasaki[32]	2009	1997-2005	Prospective observational cohort Single center	Severe TBI with GCS ≤ 8	Tasaki	32	Unfavorable outcome (GOS 1-3)	6 months	104
Fabbri[21]	2008	1999-2005	Prospective observational cohort Single center	Moderate TBI with GCS 9-13	Fabbri	33	Unfavorable outcome (GOS 1-3)	6 months	309
Perel[18]	2008	1999-2004	Randomized controlled trial Multicenter (MRC CRASH)	TBI with GCS ≤ 14	CRASH - basic - CT	 34 35	Mortality and unfavorable outcome (GOS 1-3)	14 days, 6 months	10008
Steyerberg[31]	2008	1984-1997	8 randomized controlled trials and 3 prospective observational cohorts Multicenter (IMPACT)	TBI with GCS ≤ 12	IMPACT - core - extended - lab	 36 37 38	Mortality and unfavorable outcome (GOS 1-3)	6 months	8509
Pang[27]	2007	1999-2003	Prospective observational cohort Single center	Severe TBI with GCS ≤ 8	Pang-1 (16 predictors) Pang-2 (14 predictors)	39 40	Mortality and ordinal GOS	6 months	Dataset 1: 337 Dataset 2: 513
Cremer[19]	2006	1996-2003	Prospective observational cohort Single center	TBI with GCS ≤ 8	Cremer	41	Ordinal thrichotomized GOSE	1 year	304
Demetriades[20]	2006	1993–2002	Prospective observational cohort Single center	TBI	USC	42	Mortality	Not mentioned	7191

					Validation				
Cicuendez[70]	2018	2000-2014 2014-2016	Retrospective observational cohort Single center	Severe TBI with GCS ≤ 8	IMPACT extended	37	Unfavorable outcome (GOSE 1-4) Unfavorable outcome (GOSE 1-6)	6 months	Dataset 1: 264 Dataset 2: 93
Faried[71]	2018	2016	Retrospective observational cohort Single center	TBI with GCS < 15	CRASH CT	35	Mortality	14 days	266
Ho[72]	2018	2008-2016	Retrospective observational cohort Multicenter	Severe TBI and decompressive craniectomy	IMPACT lab	38	Unfavorable outcome (GOS 1-3)	18 months	56
Mahadewa[34]	2018	2017	Prospective observational cohort Single center	Moderate or severe TBI	Revised Trauma Score Marshall CT score	43 44	Unfavorable outcome (GOS 1-3)	6 months	181
Rached[36]	2018	2007-2010	Retrospective observational cohort (based on prospective data) Multicenter	Severe TBI with head AIS > 3	IMPACT core	36	Mortality	14 days	808
Sadaka[73]	2018	2013-2016	Retrospective observational cohort Single center	TBI with GCS ≤ 14	CRASH CT	35	Mortality Unfavorable outcome (GOS 1-3)	14 days, 6 months	416
Sadaka[74]	2018	2012-2016	Retrospective observational cohort Single center	TBI with GCS 3	CRASH CT	35	Mortality Unfavorable outcome (GOS 1-3)	14 days, 6 months	62
Charry[39]	2017	2014-2015	Retrospective observational cohort Single center	Severe TBI with GCS ≤ 8	Rotterdam CT score	45	Mortality and unfavorable outcome (GOS 1-3)	6 months	127

Appendix 4.B. continued

Charry[40]	2017	2014-2015	Retrospective observational cohort Single center	Severe TBI with GCS ≤ 8 and decompressive craniectomy	Marshall CT score IMPACT - extended CRASH - CT	44 37 35	Mortality and unfavorable outcome (GOS 1-3)	6 months	127
Egea-Guerrero[42]	2017	2011-2014	Prospective observational cohort Multicenter	TBI with GCS ≤ 12	IMPACT - core - extended - lab	 36 37 38	Mortality and unfavorable outcome (GOS 1-3)	6 months	290
Kandil[69]	2017	2014-2015	Prospective observational cohort Multicenter	TBI with GCS ≤ 12	APACHE II SAPS II SOFA	46 47 48	Mortality	6 months	104
Majdan[52]	2017	2002-2005 and 2009-2012	Two prospective observational cohorts (retrospective analysis) Multicenter	TBI with GCS ≤ 8 and TBI with GCS ≤ 12	Marshall CT score Rotterdam CT score IMPACT - extended	44 45 37	Mortality and unfavorable outcome (GOS 1-3)	ICU/ hospital discharge, 6 months	866
Wan[66]	2017	2008-2015	Retrospective observational cohort Single center	Blunt TBI with GCS ≤ 8	IMPACT - core - extended - lab	 36 37 38	Mortality and unfavorable outcome (GOS 1-3)	6 months	137
Castaño-Leon[38]	2016	1993-2013	Prospective observational cohort Single center	TBI with GCS ≤ 12	IMPACT - core - extended CRASH - modified basic - modified CT	 36 37 34-m1 35-m1	Mortality and unfavorable outcome (GOS 1-3)	6 months (if not available GOS at hospital discharge was used)	1301
Hashemi[47]	2016	2012-2014	Prospective observational cohort Multicenter	TBI with GCS ≤ 14	CRASH - basic - CT	 34 35	Mortality and unfavorable outcome (GOS 1-3)	14 days, 6 months	323

Honeybul[48]	2016	2004-2014	Partly retrospective, partly prospective observational cohort Multicenter	Severe TBI and decompressive craniectomy	IMPACT - core - extended - lab CRASH - CT	 36 37 38 35	Mortality and unfavorable outcome (GOS 1-3)	18 months	319
Rizoli[30]	2016	2006-2009	Retrospective analysis of a RCT Multicenter	Severe TBI with GCS ≤ 8	IMPACT - core - extended	 36 37	Unfavorable outcome (GOSE 1-4)	6 months	1089
Staples[63]	2016	2001-2002	Prospective observational cohort Multicenter	TBI with GCS ≤ 14	CRASH - CT	 35	Mortality	14 days	1346
Staples[64]	2016	2001-2002	Prospective observational cohort Multicenter	TBI with GCS ≤ 12	IMPACT - core - lab	 36-m1 38-m1	Mortality and unfavorable outcome (GOS 1-3)	3 months, 6 months	815
Sun[65]	2016	2010-2014	Randomized controlled trial Multicenter	Severe TBI with GCS ≤ 8	IMPACT - core - extended - lab	 36 37 38	Mortality and unfavorable outcome	6 months	1124
Harrison[46]	2015	2009-2011	Prospective observational cohort Multicenter	TBI with GCS ≤ 14	Hukkelhoven IMPACT - core - extended - lab CRASH - basic - CT	49 36 37 38 34 35	Mortality and unfavorable outcome (death or severe disability) Unfavorable outcome	6 months	2975
Røe[59]	2015	2009-2010	Prospective observational cohort Multicenter	Severe TBI with GCS ≤ 8	CRASH - basic - CT	 34 35	Mortality and unfavorable outcome (GOSE 1-4)	14 days, 1 year	97
Bonds[37]	2015	2012-2013	Prospective observational cohort Single center	TBI with GCS ≤ 12 and AIS ≥ 3	IMPACT - lab CRASH - CT	 38 35	Mortality and unfavorable outcome (GOSE 1-4)	14 days, 6 months	86

Appendix 4.B. continued

Han[45]	2014	2006-2009	Prospective observational cohort Single center	Severe TBI with GCS ≤ 8	IMPACT - core - extended - lab CRASH - basic - CT	 36 37 38 34 35	Mortality and unfavorable outcome (GOS 1-3)	14 days, 6 months	300
Honeybul[49]	2014	2004-2012	Prospective and retrospective observational cohort Multicenter	Severe TBI and decompressive craniectomy	CRASH - CT	 35	Mortality and unfavorable outcome (GOS 1-3)	18 months	270
Majdan[53]	2014	2009-2012	Prospective observational cohort Multicenter	TBI with GCS ≤ 12 or AIS > 2	IMPACT - core - extended CRASH - basic Nijmegen-clinical - model 1	 36 37 34 17	Mortality and unfavorable outcome (GOS 1-3)	6 months	778
Raj[57]	2014	2009-2012	Prospective observational cohort Single center	TBI with GCS ≤ 12	IMPACT - core - extended - lab APACHE II	 36 37 38 46	Mortality and unfavorable outcome (GOS 1-3)	6 months	890
Raj[29]	2014	2003-2012	Prospective observational cohort Multicenter	TBI with GCS ≤ 13	APACHE II SAPS II SOFA	46 47 48	Mortality	6 months	1625
Güiza[44]	2013	2003-2005	Prospective observational cohort Multicenter	TBI patients requiring ICP monitoring	IMPACT - core CRASH - basic	 36 34	Poor outcome (GOS 1-2) and unfavorable outcome (GOS 1-3)	6 months	160

Jacobs[25]	2013	1998-2006 1998-2006	Prospective observational cohort Single center Prospective observational cohort Multicenter	TBI with GCS ≤ 12	Nijmegen-clinical - model 1 - model 2 Nijmegen-CT - model 1 - model 2 - model 3 Nijmegen-combination - model 1 - model 2 - model 3 Rotterdam CT score (2x, in dataset 1 and 2) Marshall CT score (2x, in dataset 1 and 2)	 17 18 19 20 21 22 23 24 45 44	Mortality Unfavorable outcome including death (GOSE 1-4) Unfavorable outcome excluding death (GOSE 2-4)	6 months	Dataset 1: 700 Dataset 2: 333
Lingsma[51]	2013	2008-2009	Prospective observational cohort Multicenter	TBI with GCS ≤ 13	IMPACT - core - extended - lab	 36 37 38	Mortality and unfavorable outcome (GOS 1-3)	6 months	415
Olivecrona[55]	2013	2002-2005	Randomized controlled trial Single center	Severe TBI with GCS ≤ 8	CRASH - CT	 35	Mortality and unfavorable outcome (GOS 1-3)	14 days, 6 months	47
Raj[58]	2013	2009-2010	Retrospective observational cohort Single center	TBI with GCS ≤ 12	IMPACT - lab	 38	Mortality and unfavorable outcome (GOS 1-3)	6 months	342
Wong[67]	2013	2006-2008	Prospective observational cohort Single center	TBI with LOC and AIS ≥ 2	IMPACT - core CRASH - basic	 36 34	Mortality and unfavorable outcome (GOS 1-3)	14 days, 6 months	178 (IMPACT) and 310 (CRASH)

Appendix 4.B. continued

Czeiter[41]	2012	Not mentioned	Prospective observational cohort Multicenter	Severe TBI with GCS ≤ 8	IMPACT - core	36	Mortality and unfavorable outcome (GOS 1-3)	1, 2, 3, 4 weeks, 3 months and 6 months	45
Gradisek[22]	2012	2007-2010	Prospective observational cohort Single center	Severe TBI with GCS ≤ 8	IMPACT - extended	37	Mortality	1 year	84
Olivecrona[54]	2012	Not mentioned	Randomized controlled trial Multicenter	Severe TBI with GCS ≤ 8	IMPACT - core - extended - lab	36 37 38	Mortality and unfavorable outcome (GOS 1-3)	6 months	48
Panczykowski[56]	2012	1994-2009	Prospective observational cohort Single center	Severe TBI with GCS ≤ 8	IMPACT - core - extended - lab	36 37 38	Mortality and unfavorable outcome (GOS 1-3)	6 months	587
Roozenbeek[61]	2012	2000-2009	Prospective observational cohort Multicenter	Severe TBI with GCS ≤ 8	IMPACT - core - modified extended	36 37-m1	Mortality	14 days	2513

Roozenbeek[62]	2012	1994-1998	Randomized controlled trial Multicenter (NABIS hypothermia)	Severe TBI	IMPACT - core - modified extended - modified lab CRASH - modified basic - modified CT	 36 37-m2 38-m2 34-m1 35-m2	Mortality and unfavorable outcome	6 months	385
		1996-1997	Randomized controlled trial Multicenter (Cerestat)	Severe TBI with GCS ≤ 8	IMPACT - core - modified extended CRASH - modified basic	 36 37-m3 34-m1	Mortality and unfavorable outcome	6 months	517
		1996-1999	Prospective observational cohort Single center (APOE)	TBI with GCS ≤ 12	IMPACT - core - modified extended	 36 37-m4	Mortality and unfavorable outcome	6 months	404
		2001-2004	Randomized controlled trial Multicenter (Pharmos)	Severe TBI	IMPACT - core - extended - lab CRASH - modified basic - modified extended	 36 37 38 34-m1 35-m1	Mortality and unfavorable outcome	6 months	856
		2001-2009	Prospective observational cohort Multicenter (TARN)	GCS ≤ 12 and AIS ≥ 3	IMPACT - core - extended CRASH - modified basic - modified extended	 36 37 34-m1 35-m1	Mortality	In-hospital	6874
Yuan[33]	2012	2009	Retrospective observational cohort Single center	TBI with GCS ≤ 12	Yuan - model A - model B - model C	 26 27 28	Mortality and unfavorable outcome (GOS 1-3)	30 days, 6 months	203

Appendix 4.B. continued

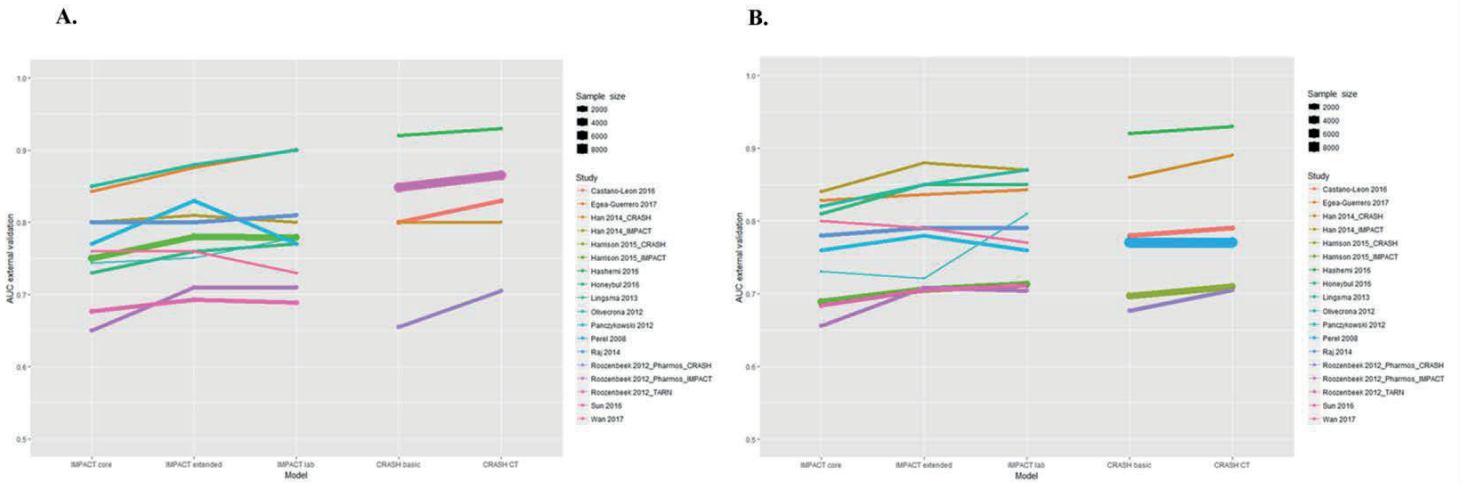
Rønning[60]	2011	2002-2007	Prospective observational cohort Single center	TBI with head AIS ≥ 1 or GCS score <15	USC	42	Mortality	In-hospital, 30 days	3134
Yeoman[68]	2011	1993-2002	Prospective observational cohort Single center	TBI with GCS ≤ 12	IMPACT - modified extended	 37-m5	Mortality and unfavorable outcome (GOS 1-3)	1 year	1276
Honeybul[17]	2009	2006-2007	Retrospective observational cohort Multicenter	Severe TBI and decompressive craniectomy	CRASH - CT	 35	Mortality and unfavorable outcome (GOS 3-5)	6 months, 12 months	41
Perel[18]	2008	1984-1997	8 randomized controlled trials and 3 observational cohorts Multicenter (IMPACT)	TBI with GCS ≤ 12	CRASH - modified basic - modified CT	 34-m2 35-m3	Mortality and unfavorable outcome (GOS 1-3)	14 days, 6 months	8509
Steyerberg[31]	2008	1999-2004	Randomized controlled trial Multicenter (MRC CRASH)	TBI with GCS ≤ 12	IMPACT - core - modified extended	 36 37-m6	Mortality and unfavorable outcome (GOS 1-3)	6 months	6681
Fischler[43]	2007	1995-2000	Prospective observational cohort Single center	Severe TBI with GCS < 8 prehospital or injury needing urgent craniotomy	SAPS II MPM II at admission MPM II at 24 hours ISS	47 50 51 52	Mortality	1 year	299
Cremer[19]	2006	1996-2001	Prospective observational cohort Single center	TBI with GCS ≤ 8	Cremer	41	Ordinal thrichotomized GOSE	1 year	122

Hukkelhoven[50]	2006	1991-1994	Randomized controlled trial Multicenter (Tirilazad)	TBI with GCS ≤ 12	Choi-tree (4x) Signorini-LR (4x) Andrews-tree (3x)	53 54 55	Mortality and unfavorable outcome	6 months	2269
		1994-1995	Randomized controlled trial Multicenter (International Selfotel Trial)	Severe TBI with GCS ≤ 8	Hukkelhoven-LR (3x)	49	Mortality and unfavorable outcome		409
		1995	Prospective observational cohort Multicenter (EBIC survey)	Moderate or severe TBI			Mortality and unfavorable outcome		796 746
		1984-1987	Prospective observational cohort Multicenter (TCDB)	Severe TBI with GCS ≤ 8			Mortality		
					Extension				
Raj[57]	2014	2009-2012	Prospective observational cohort Single center	TBI with GCS ≤ 12	IMPACT-APACHE - core - extended - lab	 56 57 58	Mortality and unfavorable outcome (GOS 1-3)	6 months	890
Güiza[44]	2013	2003-2005	Prospective observational cohort Multicenter	TBI patients requiring ICP monitoring	Güiza-1 - IMPACT core + 19 dynamic - CRASH core + 19 dynamic Güiza-2 - IMPACT core + 11 dynamic - CRASH core + 11 dynamic	 59 60 61 62	Poor outcome (GOS 1-2) and unfavorable outcome (GOS 1-3) Poor outcome (GOS 1-2) Unfavorable outcome (GOS 1-3)	6 months	160

Appendix 4.B. continued

Lingsma[51]	2013	2008-2009	Prospective observational cohort Multicenter	TBI with GCS ≤ 13	IMPACT lab + ISS	39	Mortality and unfavorable outcome (GOS 1-3)	6 months	415
Raj[58]	2013	2009-2010	Retrospective observational cohort Single center	TBI with GCS ≤ 12	IMPACT lab + INR + ISS IMPACT lab + platelets + ISS IMPACT lab + INR + platelets + ISS	64 65 66	Mortality and unfavorable outcome (GOS 1-3)	6 months	342
Czeiter[41]	2012	Not mentioned	Prospective observational cohort Multicenter	Severe TBI with GCS ≤ 8	IMPACT core + biomarkers	67a-e	Mortality and unfavorable outcome (GOS 1-3)	1, 2, 3, 4 weeks, 3 months and 6 months	45

TBI, traumatic brain injury; GCS, Glasgow Coma Scale; NACA-BM, National Advisory Committee for Aeronautics basic model; GCS-RM, Glasgow Coma Scale reference model; HAIS, Head Abbreviated Injury Scale; NNI, National Neuroscience Institute; RCT, randomized clinical trial; GOS(E), Glasgow Outcome Scale (Extended); SOFA, Sequential Organ Failure Assessment; CT, computed tomography; IMPACT, International Mission for Prognosis and Analysis of Clinical Trials; CRASH, Corticoid Randomisation After Significant Head injury; USC, University of Southern California; AIS, Abbreviated Injury Scale; APACHE, Acute Physiology And Chronic Health Evaluation; SAPS, Simplified Acute Physhiology Score; ICP, intracranial pressure; LOC, loss of consciousness; NABIS, The North American Brain Injury Study; APOE, Apolipoproteïne E; TARN, Trauma Audit and Research Network; MPM, Mortality Probability Models; ISS, Injury Severity Score; LR, logistic regression; EBIC, European Brain Injury Consortium; TCDB, Traumatic Coma Data Bank.



Appendix 4.C. Discriminative ability (area under the receiver operating characteristic curve, AUC) of IMPACT and CRASH models for (**A**) mortality and (**B**) unfavorable outcome in studies validating all consecutive IMPACT and CRASH models with increasing complexity. The colors of the dots represent the different validation cohorts, and the dot size refers to the sample size of these cohorts.

IMPACT, International Mission for Prognosis and Analysis of Clinical Trials; CRASH, Corticoid Randomisation After Significant Head injury; TARN, Trauma Audit and Research Network

Appendix 4.D. Overview of predictors included in 67 models for moderate and severe traumatic brain injury (including 12 modifications of IMPACT and CRASH)

Model no.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Demographics															
Age		X	X	X	X	X	X	X	X	X	X	X		X	X
Gender										X	X	X			
Ethnic group															
Clinical															
GCS score	X				X	X									X
GCS motor score			X				X	X	X	X	X	X			
Pupils		X	X	X	X	X	X	X	X	X	X	X	X	X	
AIS/ISS/MEI				X									X		
Limb movement							X	X	X	X	X	X			
Mechanism of injury															
NACA score		X													
Physiological															
Hypoxia						X									
Hypotension								X	X		X	X			X
Mean arterial pressure															
Intracranial pressure														X	
Temperature															
Heart rate															
Respiratory rate	X														
PaO2															X
FiO2															X
Systolic blood pressure	X														
Cerebral perfusion pressure															
Mechanical ventilation															
Urine output															
Arrhythmia															
Radiology															
CT classification													X		
Midline shift	X							X	X		X	X		X	
Cisterns/third ventricle	X					X		X	X		X	X			
tSAH/IVH								X	X		X	X		X	
Hematoma						X		X[a]	X[a]		X[a]	X[a]			
Lesions	X														
Fourth ventricle															
Contusion															
Basal skull fracture															
Intracranial mass effect															

Model no.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Laboratory															
Glucose									X			X			
Hemoglobin									X			X			
Coagulopathy						X									
Sodium									X			X			
Creatinine									X			X			X
Potassium															
White blood cell count															
Hematocrit															
pH															
Bilirubin															X
Platelet count															X
INR/PT															
Biomarkers															
D-dimer															
Calcium															
Blood urea nitrogen															
Bicarbonate															
Other															
Chronic health status															
Type of admission															
Metastatic cancer															
Cirrhosis															
Acute renal failure															
Chronic renal failure															
Cerebrovascular incident															
Cardiopulmonary resuscitation prior admission															
Gastrointestinal bleed															
Proven infection															
Vasoactive drug															
Total	6	3	3	3	3	7	4	10	14	5	11	15	3	5	8

Appendix 4.D. continued

Model no.	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Demographics															
Age	X	X	X				X	X			X	X	X	X	X
Gender															
Ethnic group															
Clinical															
GCS score	X	X													X
GCS motor score											X	X	X	X	
Pupils		X					X	X		X	X	X	X	X	X
AIS/ISS/MEI															
Limb movement															
Mechanism of injury															
NACA score															
Physiological															
Hypoxia															
Hypotension		X	X				X	X	X						X
Mean arterial pressure															
Intracranial pressure														X	
Temperature															
Heart rate															
Respiratory rate															
PaO2															
FiO2															
Systolic blood pressure															
Cerebral perfusion pressure														X	
Mechanical ventilation															
Urine output															
Arrhythmia															
Radiology															
CT classification											X	X	X	X	
Midline shift															
Cisterns/third ventricle				X	X	X	X	X	X	X					
tSAH/IVH				X	X						X	X	X	X	
Hematoma											X	X	X	X	
Lesions				X	X	X		X	X	X					
Fourth ventricle				X			X								
Contusion				X			X								
Basal skull fracture															
Intracranial mass effect															

Model no.	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Laboratory															
Glucose													X	X	
Hemoglobin													X	X	
Coagulopathy															
Sodium															
Creatinine															
Potassium															
White blood cell count															
Hematocrit															
pH															
Bilirubin															
Platelet count															
INR/PT															
Biomarkers										X[b]					
D-dimer													X	X	
Calcium													X	X	
Blood urea nitrogen															
Bicarbonate															
Other															
Chronic health status															
Type of admission															
Metastatic cancer															
Cirrhosis															
Acute renal failure															
Chronic renal failure															
Cerebrovascular incident															
Cardiopulmonary resuscitation prior admission															
Gastrointestinal bleed															
Proven infection															
Vasoactive drug															
Total	2	4	2	5	3	2	6	5	3	5	6	6	10	12	4

Appendix 4.D. continued

Model no.	31	32	33	34	34-m1	34-m2	35	35-m1	35-m2	35-m3	36	36-m1	37	37-m1	37-m2
Demographics															
Age	X	X		X	X	X	X	X	X	X	X		X	X	X
Gender															
Ethnic group															
Clinical															
GCS score	X		X	X		X	X		X	X					
GCS motor score					X			X			X	X	X	X	X
Pupils	X	X		X	X	X	X	X	X	X	X	X	X	X	X
AIS/ISS/MEI				X	X		X	X	X						
Limb movement															
Mechanism of injury															
NACA score															
Physiological															
Hypoxia													X	X	X
Hypotension	X												X	X	X
Mean arterial pressure															
Intracranial pressure		X													
Temperature															
Heart rate															
Respiratory rate															
PaO2															
FiO2															
Systolic blood pressure															
Cerebral perfusion pressure															
Mechanical ventilation															
Urine output															
Arrhythmia															
Radiology															
CT classification			X					X	X				X		X
Midline shift	X	X					X			X				X	
Cisterns/third ventricle	X						X			X				X	
tSAH/IVH	X	X	X				X	X		X			X	X	
Hematoma	X[a]		X				X			X			X		
Lesions							X								
Fourth ventricle															
Contusion	X														
Basal skull fracture			X												
Intracranial mass effect															

Model no.	**31**	**32**	**33**	**34**	**34-m1**	**34-m2**	**35**	**35-m1**	**35-m2**	**35-m3**	**36**	**36-m1**	**37**	**37-m1**	**37-m2**
Laboratory															
Glucose															
Hemoglobin															
Coagulopathy			X												
Sodium															
Creatinine															
Potassium															
White blood cell count															
Hematocrit															
pH															
Bilirubin															
Platelet count															
INR/PT															
Biomarkers															
D-dimer															
Calcium															
Blood urea nitrogen															
Bicarbonate															
Other															
Chronic health status															
Type of admission															
Metastatic cancer															
Cirrhosis															
Acute renal failure															
Chronic renal failure															
Cerebrovascular incident															
Cardiopulmonary resuscitation prior admission															
Gastrointestinal bleed															
Proven infection															
Vasoactive drug															
Total	10	5	6	4	4	3	9	6	5	7	3	2	8	8	6

Appendix 4.D. continued

Model no.	37-m3	37-m4	37-m5	37-m6	38	38-m1	38-m2	39	40	41	42	43	44	45	46
Demographics															
Age	X	X	X	X	X		X	X	X	X	X				X
Gender								X	X						
Ethnic group								X	X						
Clinical															
GCS score								X[d]	X[d]		X	X			X
GCS motor score	X	X	X	X	X	X	X			X					
Pupils	X	X	X	X	X	X	X	X[d]	X[d]	X					
AIS/ISS/MEI											X				
Limb movement															
Mechanism of injury								X	X		X				
NACA score															
Physiological															
Hypoxia	X	X	X		X	X	X	X	X						
Hypotension	X	X	X		X	X	X	X	X	X					
Mean arterial pressure															X
Intracranial pressure															
Temperature															X
Heart rate															X
Respiratory rate												X			X
PaO2															X
FiO2															
Systolic blood pressure												X			
Cerebral perfusion pressure															
Mechanical ventilation															
Urine output															
Arrhythmia															
Radiology															
CT classification			X	X	X	X	X			X					
Midline shift													X	X	
Cisterns/third ventricle													X	X	
tSAH/IVH		X	X	X	X	X		X	X					X	
Hematoma	X	X			X	X								X	
Lesions													X		
Fourth ventricle															
Contusion															
Basal skull fracture															
Intracranial mass effect															

Model no.	37-m3	37-m4	37-m5	37-m6	38	38-m1	38-m2	39	40	41	42	43	44	45	46
Laboratory															
Glucose					X	X	X								
Hemoglobin					X	X	X								
Coagulopathy								X	X						
Sodium															X
Creatinine															X
Potassium															X
White blood cell count															X
Hematocrit															X
pH															X
Bilirubin															
Platelet count															
INR/PT															
Biomarkers															
D-dimer															
Calcium															
Blood urea nitrogen															
Bicarbonate															
Other															
Chronic health status															X
Type of admission															
Metastatic cancer															
Cirrhosis															
Acute renal failure															
Chronic renal failure															
Cerebrovascular incident															
Cardiopulmonary resuscitation prior admission															
Gastrointestinal bleed															
Proven infection															
Vasoactive drug															
Total	6	7	7	5	10	9	8	12	12	5	4	3	3	4	14

Appendix 4.D. continued

Model no.	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61
Demographics															
Age	X		X	X	X		X	X	X	X	X	X	X	X	X
Gender															
Ethnic group															
Clinical															
GCS score	X	X		X	X			X	X	X	X	X		X	
GCS motor score			X				X			X	X	X	X		X
Pupils			X				X	X	X	X	X	X	X	X	X
AIS/ISS/MEI						X[d]		X						X	
Limb movement															
Mechanism of injury									X						
NACA score															
Physiological															
Hypoxia			X								X	X			
Hypotension		X	X								X	X			
Mean arterial pressure										X	X	X	X[e]	X[e]	X[f]
Intracranial pressure													X[e]	X[e]	X[f]
Temperature	X									X	X	X			
Heart rate	X			X						X	X	X			
Respiratory rate										X	X	X			
PaO2	X	X			X					X	X	X			
FiO2	X	X													
Systolic blood pressure	X			X											
Cerebral perfusion pressure															
Mechanical ventilation	X			X	X										
Urine output	X				X										
Arrhythmia				X											
Radiology															
CT classification			X								X	X			
Midline shift															
Cisterns/third ventricle															
tSAH/IVH			X								X	X			
Hematoma											X	X			
Lesions							X	X							
Fourth ventricle															
Contusion															
Basal skull fracture															
Intracranial mass effect				X	X										

Model no.	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61
Laboratory															
Glucose												X			
Hemoglobin												X			
Coagulopathy															
Sodium	X									X	X	X			
Creatinine		X			X					X	X	X			
Potassium	X									X	X	X			
White blood cell count	X									X	X	X			
Hematocrit										X	X	X			
pH										X	X	X			
Bilirubin	X	X													
Platelet count		X													
INR/PT					X										
Biomarkers															
D-dimer															
Calcium															
Blood urea nitrogen	X														
Bicarbonate	X														
Other															
Chronic health status	X									X	X	X			
Type of admission	X			X	X										
Metastatic cancer				X	X										
Cirrhosis				X	X										
Acute renal failure				X											
Chronic renal failure				X											
Cerebrovascular incident				X											
Cardiopulmonary resuscitation prior admission				X											
Gastrointestinal bleed				X	X										
Proven infection					X										
Vasoactive drug															
Total	17	7	7	15	13	6	4	5	4	16	21	23	22	23	14

Appendix 4.D. continued

Model no.	62	63	64	65	66	67 a-e	Total n (%)
Demographics							
Age	X	X	X	X	X	X	54 (82)
Gender							5 (8)
Ethnic group							2 (3)
Clinical							
GCS score	X						28 (42)
GCS motor score		X	X	X	X	X	27 (42)
Pupils	X	X	X	X	X	X	48 (73)
AIS/ISS/MEI	X	X	X	X	X		13 (19)
Limb movement							6 (10)
Mechanism of injury							4 (6)
NACA score							1 (1)
Physiological							
Hypoxia		X	X	X	X		12 (19)
Hypotension		X	X	X	X		25 (40)
Mean arterial pressure	X[f]						8 (13)
Intracranial pressure	X[f]						7 (11)
Temperature							5 (8)
Heart rate							6 (10)
Respiratory rate							6 (6)
PaO2							8 (13)
FiO2							3 (5)
Systolic blood pressure							4 (3)
Cerebral perfusion pressure							1 (2)
Mechanical ventilation							3 (5)
Urine output							2 (3)
Arrhythmia							1 (2)
Radiology							
CT classification		X	X	X	X		16 (26)
Midline shift							11 (16)
Cisterns/third ventricle							17 (26)
tSAH/IVH		X	X	X	X		27 (44)
Hematoma		X	X	X	X		21 (34)
Lesions							11 (16)
Fourth ventricle							2 (3)
Contusion							3 (5)
Basal skull fracture							1 (2)
Intracranial mass effect							2 (3)

Model no.	62	63	64	65	66	67 a-e	Total n (%)
Laboratory							
Glucose		X	X	X	X		10 (16)
Hemoglobin		X	X	X	X		10 (16)
Coagulopathy							4 (6)
Sodium							7 (11)
Creatinine							9 (15)
Potassium							5 (8)
White blood cell count							5 (8)
Hematocrit							4 (6)
pH							4 (6)
Bilirubin							3 (5)
Platelet count				X	X		4 (6)
INR/PT			X		X		3 (5)
Biomarkers						X[g]	2 (3)
D-dimer							2 (3)
Calcium							2 (3)
Blood urea nitrogen							1 (2)
Bicarbonate							1 (2)
Other							
Chronic health status							5 (8)
Type of admission							3 (5)
Metastatic cancer							2 (3)
Cirrhosis							2 (3)
Acute renal failure							1 (2)
Chronic renal failure							1 (2)
Cerebrovascular incident							1 (2)
Cardiopulmonary resuscitation prior admission							1 (2)
Gastrointestinal bleed							
Proven infection							1 (2)
Vasoactive drug							1 (2)
							1 (2)
Total	15	11	12	12	13	6	

GCS, Glasgow Coma Scale; tSAH, traumatic subarachnoid hemorrhage; IVH, intraventricular hemorrhage; CT, computed tomography; AIS, Abbreviated Injury Scale; ISS, Injury Severity Score; MEI, major extracranial injury; PaO2, partial arterial pressure of oxygen; INR, international normalized rate; PT, prothrombin time; pH, potential hydrogen; FiO2, fraction of inspired oxygen; NACA, National Advisory Committee for Aeronautics

[a]EDH and SDH

[b]GFAP and S-100B peak concentration

[c]Pre- and post-resuscitation

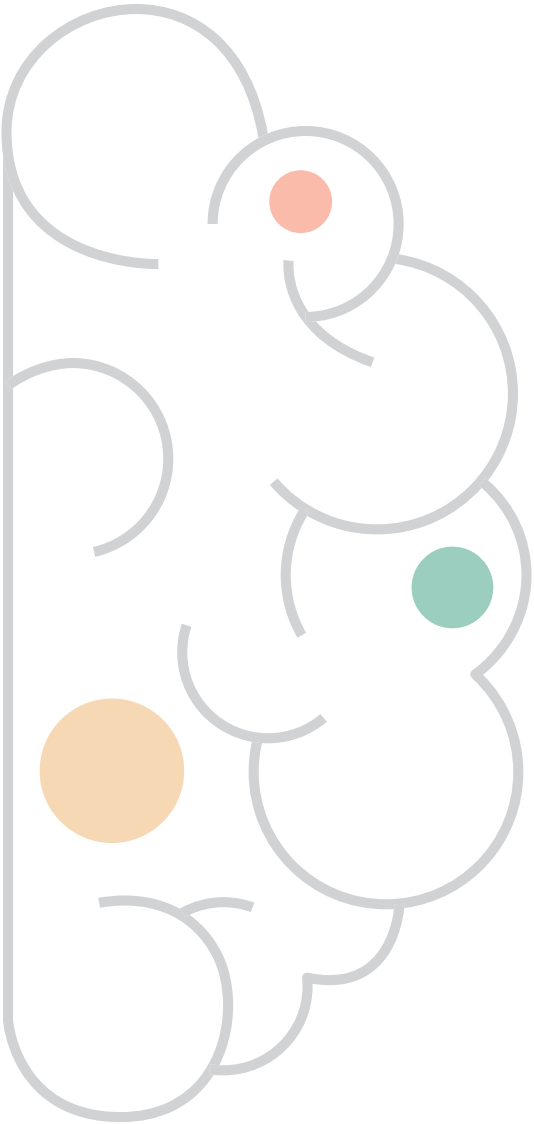
[d]Six individual predictors within ISS

[e]19 dynamic predictors related to ICP and MAP

[f]11 dynamic predictors related to ICP and MAP

[g]Combinations of 3 different biomarkers

CHAPTER 4.1



Response to Walker et al. (doi: 10.1089 neu.2017.5359): Predicting long-term global outcome after traumatic brain injury

Kelly A. Foks
Simone A. Dijkland
Ewout W. Steyerberg

J Neurotrauma 2019; 36(8): 1382-1383

Dear Editor:

With great interest we read the recent study by Walker and colleagues on the development of a prognostic model to predict long-term functional outcomes for adult patients with moderate and severe traumatic brain injury (TBI).[1] The authors used a large prospective multi-center cohort of patients with TBI receiving inpatient rehabilitation, representative for clinical practice in the United States. Prognostic modeling for outcome after TBI in the rehabilitation setting could help set expectations and plan treatments in those patients who are in inpatient rehabilitation after sustaining a TBI. We noted, however, several methodological shortcomings that necessitate a cautious interpretation of findings from this study.

First, the authors seem to have excluded or removed patients with the outcomes death or vegetative state from the analysis, arguing that including these would not have added much significant information. Obviously, leaving out these patients introduces some bias toward better outcome. Moreover, it is unknown in advance which patients will die or remain vegetative, and hence use of the model in clinical practice is impossible.

Second, the authors performed a complete case analysis by removing all patients with missing Glasgow Outcome Scale scores or a missing covariate from the analysis. Systematic differences between patients with missing data and patients with complete data could cause bias. A solution for this problem that is now widely implemented in clinical research is a multiple imputation procedure, where missing values are substituted with plausible values based on correlations with covariates and with outcome variables.[2]

Third, the authors claim that a decision tree model is the best method to define a prognostic model in this context. Thorough methodological research has shown quite suboptimal performance of decision trees for modeling prognosis in TBI and other medical domains, however.[3,4] Studies comparing different modeling strategies concluded that logistic regression analysis is the preferred method to develop a prognostic model for outcomes of TBI.[3] A key prognostic characteristic such as age is then dealt with in a natural, continuous way rather than creating artificial groupings.

Fourth, the authors state that they demonstrated a reasonable predictive accuracy of themodel. Indeed, a randomsplit sample is an independent test for the model, but cannot be considered as external validation. To assess generalizability of the model, validation is required with meaningful geographic or temporal splitting.[5]

Remarkably, the authors cite a systematic review that includes all the above mentioned recommendations for improvement of methodological quality in prognostic models in TBI.[6] Moreover, promising prognostic models for functional outcome after moderate and severe TBI have been developed over the last decade, including the International Mission for Prognosis and Analysis of Clinical Trials (IMPACT) and Corticoid Randomisation After Significant Head injury (CRASH) models.[7,8] Relevant admission characteristics included in these models, such as pupillary reactivity and extracranial injury, unfortunately were not incorporated in the current analyses.

In conclusion, we observe multiple methodological shortcomings in both development and validation of the proposed prognostic tool. In addition, important advances in prognostic modeling in

TBI over the last decade should be considered. Application of the proposed model in patients with TBI in inpatient rehabilitation can only be recommended after satisfactory performance is shown in fully independent external validation studies with adequate design.

References

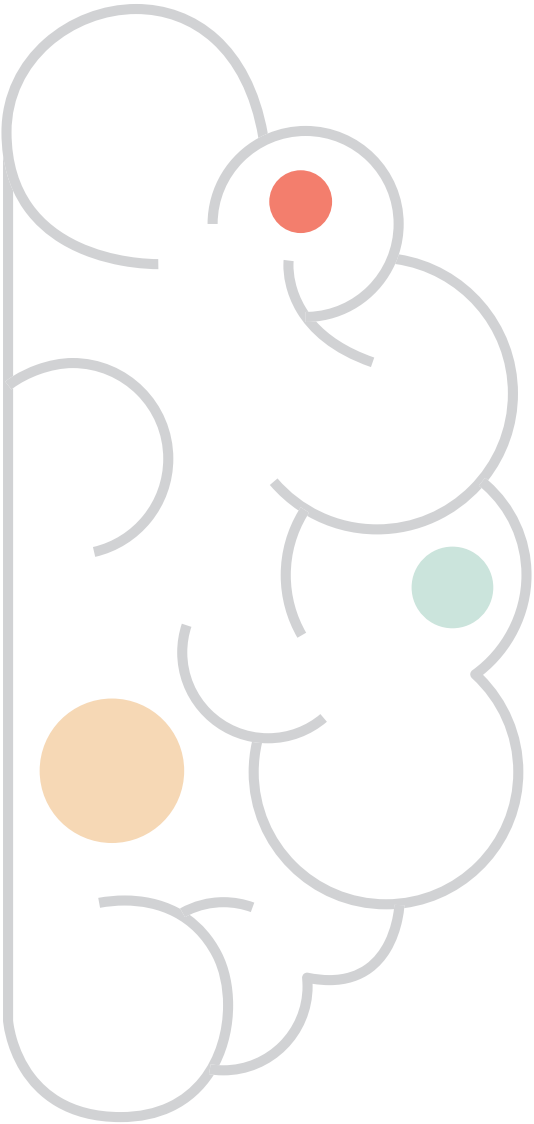
1. Walker WC, Stromberg KA, Marwitz JH, Sima AP, Agyemang AA, Graham KM, et al. Predicting long-term global outcome after traumatic brain injury: development of a practical prognostic tool using the traumatic brain injury model systems national database. *J Neurotrauma.* 2018, 35, 1587–1595.
2. Steyerberg EW. *In: Clinical Prediction Models: A Practical Approach to Development, Validation and Updating.* New York: Springer, 2009, p. 115–129.
3. van der Ploeg T, Smits M, Dippel DW, Hunink M, Steyerberg EW. Prediction of intracranial findings on CT-scans by alternative modelling techniques. *BMC Med Res Methodol.* 2011, 11, 143.
4. Austin PC. A comparison of regression trees, logistic regression, generalized additive models, and multivariate adaptive regression splines for predicting AMI mortality. *Stat Med.* 2007,26, 2937–2957.
5. Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol.* 2016, 69, 245–247.
6. Mushkudiani NA, Hukkelhoven CW, Hernandez AV, Murray GD, Choi SC, Maas AI, et al. A systematic review finds methodological improvements necessary for prognostic models in determining traumatic brain injury outcomes. *J Clin Epidemiol.* 2008, 61, 331–343.
7. Steyerberg EW, Mushkudiani N, Perel P, Butcher I, Lu J, McHugh GS, et al. Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. *PLoS Med.* 2008, 5, e165.
8. MRC CRASH Trial Collaborators, Perel P, Arango M, Clayton T, Edwards P, Komolafe E, et al. Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients. *BMJ.* 2008, 336, 425–429.

Reference to response letter by Walker et al:

Walker WC, Sima AP, Hoffman JM, Harrsion-Felix C, Agyemang AA, Stromberg KA, et al. Response to Foks et al. (doi: 10.1089/neu.2018.5979): Why Our Long-Term Functional Prognosis Tools are a Valuable Contribution to the Traumatic Brain Injury Outcome Literature. *J Neurotrauma.* 2019, 36:8, 1384-1385.

CHAPTER 5

Prediction of 60-day case fatality after aneurysmal subarachnoid hemorrhage: External validation of a prediction model



Simone A. Dijkland
Bob Roozenbeek
Patrick A. Brouwer
Hester F. Lingsma
Diederik W.J. Dippel
Leonie J. Vergouw
Mervyn D.I. Vergouwen
Mathieu van der Jagt

Crit Care Med 2016; 44(8): 1523–1529

Abstract

Objective: External validation of prognostic models is crucial but rarely done. Our aim was to externally validate a prognostic model to predict 60-day case fatality after aneurysmal subarachnoid hemorrhage developed from the International Subarachnoid Aneurysm Trial in a retrospective unselected cohort of subarachnoid hemorrhage patients.

Design: The model's predictors were age, aneurysm size, Fisher grade, and World Federation of Neurological Surgeons grade. Two versions of the model were validated: one with World Federation of Neurological Surgeons grade scored at admission and the other with World Federation of Neurological Surgeons grade at treatment decision. The outcome was 60-day case fatality. Performance of the model was assessed by studying discrimination, expressed by the area under the receiver operating characteristic curve, and calibration.

Setting: University hospital.

Patients: We analyzed data from 307 consecutive aneurysmal subarachnoid hemorrhage patients admitted between 2007 and 2011 (validation cohort).

Interventions: None.

Measurements and main results: The observed 60-day case fatality rate was 30.6%. Discrimination was good, and differed between the model with World Federation of Neurological Surgeons grade at treatment decision (area under the receiver operating characteristic curve, 0.89) and at admission (area under the receiver operating characteristic curve, 0.82). Mean predicted probabilities were lower than observed: 17.0% (model with World Federation of Neurological Surgeons grade at admission) and 17.7% (model with World Federation of Neurological Surgeons grade at treatment decision).

Conclusions: The model discriminated well between patients who died or survived within 60 days. In addition, we found that using World Federation of Neurological Surgeons grade at moment of treatment decision of the ruptured aneurysm improved model performance. However, since predicted probabilities were much lower than observed probabilities, the International Subarachnoid Aneurysm Trial prediction model needs to be adapted to be used in clinical practice.

Introduction

Aneurysmal subarachnoid hemorrhage (aSAH) is associated with significant morbidity and mortality.[1,2] Although the case fatality rate has decreased, mortality is still around 35%.[2,3]

Reliable prediction of short-term mortality risk is useful to inform patients and relatives on prognosis and select patients at risk for poor outcome before therapeutic decisions are made, both in clinical practice and in intervention studies.[4-6] In addition, outcome prediction may be important in benchmarking quality of care. Several prognostic factors can be combined in a prognostic model to calculate the risk of a specific endpoint for an individual patient.[7] For aSAH, various prognostic models have been developed.[4,5] However, to date no prognostic model for aSAH has found its way into clinical practice. This might be explained by methodological problems with the development of these models; typically too many predictors are tested for the number of outcome events in datasets, leading to overfitted models with limited generalizability and overoptimistic estimates of model performance.[4] To reveal the performance of prediction models in new datasets, external validation is a crucial step, but is rarely done.[4-8] Finally, application in clinical practice is further hampered by the necessity to adapt prediction models to specific clinical settings that in addition may rapidly change over time, for instance, when new therapies are introduced.[9]

A recent prognostic model, predicting 60-day case fatality after aSAH using data from the International Subarachnoid Aneurysm Trial (ISAT),[9] has shown reasonable performance.[4,5] The objective of this study is to externally validate the ISAT prediction model in an independent cohort.

Methods

Study design and population

In this retrospective cohort study, we included consecutively admitted aSAH patients to the Erasmus MC, University Medical Center, Rotterdam, The Netherlands. Patients were identified through a hospital registry and were admitted between October 2007 and October 2011. All aSAH patients were routinely managed at an ICU.

Inclusion criteria were 1) 18 years or older, 2) admitted to hospital less than or equal to 28 days after ictus, 3) SAH, proven by CT or cerebrospinal fluid spectrophotometry, and 4) ruptured intracranial aneurysm as the presumed cause. Exclusion criteria were 1) explicit objection by the subject to view the medical data, and 2) missing data on 60-day case fatality. The study protocol was approved by the local medical ethics committee.

Derivation cohort

The prediction model was based on patients included in the International Subarachnoid Aneurysm Trial (ISAT; n = 2,143), which compared the safety and efficacy of endovascular coiling with neurosurgical clipping.[9] The ISAT prediction model included the predictors age, maximum lumen size of the

ruptured aneurysm, Fisher grade, and World Federation of Neurological Surgeons (WFNS) grade at randomization.[5] The model aimed to predict case fatality at 60 days. The model performed reasonably at internal validation with an area under the curve (AUC) of 0.70.[5]

Validation cohort: data collection and outcome

The following data were collected: age, sex, Fisher grade, WFNS grade at admission, location of ruptured aneurysm, and aneurysm treatment mode. Assessment of Fisher grade[10] and maximum lumen size (in millimeters, on CT angiography or digital subtraction angiography) of the ruptured aneurysm was done by an interventional neuroradiologist (P.A.B.). Because the ISAT prediction model used WFNS grade[11] at time of randomization for coiling or clipping, we additionally assessed WFNS grade at treatment decision regarding suitability for coiling or clipping, which was deemed a proxy for the moment of randomization in ISAT.

The outcome was 60-day case fatality, which was collected from our electronic patient record. When these data were not available, a letter to the general practitioner was sent for retrieval of this information.

Discrimination and calibration

The external validity of the ISAT prediction model was assessed in terms of discrimination and calibration. Discrimination refers to how well the model distinguishes between those who die within 60 days and those who survive. Discrimination was assessed by calculating the AUC of the receiver operating curve (ROC). The discriminative power of a model may be influenced by differences in case-mix between the derivation and validation cohort.[12] In a population with a prognostically homogeneous case-mix, it will be more difficult to distinguish between patients with good or poor outcome than in a heterogeneous population. To take this into account, we calculated the case-mix-corrected AUC. The case-mix-corrected AUC indicates the discriminative power of a model, under the assumption that the predictor effects are fully correct for the validation population. It was calculated by simulating new outcome values for all patients in the validation dataset, based on the predicted risks for each patient calculated by the prognostic model.[12]

Calibration refers to the agreement between predicted and observed probabilities. Calibration was assessed graphically in a calibration graph, and expressed as the calibration slope and an intercept. The calibration slope describes the effect of the predictors in the validation sample versus in the derivation sample. Ideally, the calibration slope is equal to 1. The intercept indicates whether predictions are systematically too high or too low, and should ideally be zero.

Statistical analyses

Patient baseline characteristics are presented as medians (interquartile range [IQR]) or frequencies (percentage). The association of the predictors with 60-day case fatality was assessed with univariable and multivariable logistic regression analyses. Associations were expressed as odds ratios (ORs) and 95% CIs. For adequate comparison of the prognostic effects for Fisher grade, we converted the ISAT reference category for this variable from grade 1 to grade 4 by recalculating the ORs and the 95% CIs.

Two versions of the model were validated: one with WFNS grade at admission and the other with WFNS grade at treatment decision.

The main analysis was performed on the entire cohort. A sensitivity analysis was performed by excluding patients in whom either coiling or clipping had not been performed, patients who died within 48 hours after admission, and patients who had emergency decompressive craniotomy because of impending herniation due to intracerebral hematoma. The remaining patients were considered to approximate the original ISAT population.

The statistical analyses were performed using SPSS (Statistical Package for Social Sciences, version 22; IBM Corporation, Armonk, NY) and R software (R Foundation for Statistical Computing, Vienna, Austria). Missing values in the validation cohort were statistically imputed using a multiple imputation method with the *AregImpute* function in R statistical software. Complete case analyses were done for comparison with the imputed analyses. The calibration plots were created with an adapted version of the *val.prob* function from the *rms* library in the R package.

Results

Study population

We retrieved 410 patients with aSAH. Reasons for exclusion are shown in Appendix 5.A. Thirteen patients with missing data on case fatality were excluded. The main reason for loss to follow-up was transfer of patients to another hospital. We performed analyses on 307 patients (96%), of whom 94 patients (30.6%) died within 60 days. In the dataset with four independent variables (i.e., age, aneurysm lumen size, Fisher grade, and WFNS grade) and one outcome variable per patient, 47 of 1,228 data points (3.8%) were missing and statistically imputed in the validation sample. The highest percentage of missings was in the variables lumen size (11.7%) and Fisher grade (3.3%).

The distribution of demographic data and prognostic variables of both the validation cohort and the ISAT derivation population are shown in Table 5.1. In total, 93 patients (30%) did not receive aneurysm treatment. Among these, 41 died less than 48 hours, 32 died between 48 hours and 60 days, and 20 survived more than 60 days. The decision whether or not to treat the aneurysm was based on local clinical guidelines (Appendix 5.B). The Fisher grades and WFNS grades were significantly higher in the validation cohort than the ISAT sample.

The median time between SAH and randomization in the ISAT derivation cohort was 2 days (coiling: IQR, 1–4; range, 0–26 and clipping: IQR, 1–5; range, 0–28). In the validation cohort, the median time between SAH and treatment decision was 1 day (IQR, 0–3; range, 0–25). The median interval between the moment of assessment of WFNS grade at admission and WFNS grade at time of treatment decision was 1 day (IQR, 0–2; range, 0–22). There was no statistically significant difference between WFNS grade at admission and WFNS grade at time of treatment decision in the unselected cohort (n = 307; p = 0.69, Wilcoxon signed rank test); in the cohort of patients who met the original ISAT criteria

(i.e., were clipped or coiled), WFNS grade at admission differed from WFNS grade at treatment decision (n = 211; p = 0.04, Wilcoxon signed rank test) (Appendix 5.C).

Prognostic effects

In the validation cohort, the strongest univariable predictor of case fatality was WFNS grade at time of treatment decision (WFNS grade 4: OR, 6.95; 95% CI, 2.30–21.01 and WFNS grade 5: OR, 299.20; 95% CI, 83.53–1071.74) (Table 5.2). WFNS grade was also the strongest predictor in the ISAT population. Associations of both age and lumen size with 60-day case fatality were similar in derivation and validation cohort. No patients with Fisher grade 1 and 2 died within 60 days in the validation cohort. The prognostic effects of WFNS grade in the multivariable analysis showed the same trend as in the univariable analysis (Table 5.2).

Table 5.1. Baseline characteristics of patients in the International Subarachnoid Aneurysm Trial development cohort and in the Rotterdam validation cohort

Characteristics	Measure or category	International Subarachnoid Aneurysm Trial derivation cohort (n=2,143)	Rotterdam validation cohort (n=307)	p[a]
Age[b] (yr)		52 (44-60)	56 (47-66)	<0.001
Maximum lumen size aneurysm (mm)	Total available	2128 (100%) 5.0 (4.0-7.0)	271 (88%) 6.0 (4.8-8.1)	<0.001
Fisher grade, n (%)	Total available 1 2 3 4	2128 (100) 114 (6) 360 (17) 902 (42) 752 (35)	297 (97) 7 (2) 7 (2) 62 (21) 221 (75)	<0.001
World Federation of Neurological Surgeons grade, n (%)	Total available 1 2 3 4 5 6 (not assessable)	2128 (100) 1324 (62) 546 (26) 133 (6) 74 (4) 20 (1) 31 (1)	306 (99) 115 (38) 62 (20) 6 (2) 50 (16) 73 (24) NA	<0.001
Sex, n (%)	Total available Female Male	2128 (100) 1339 (63) 789 (37)	307 (100) 200 (65) 107 (35)	0.450
Treatment, n (%)	Total available Coil Clip None	2128 (100) 1062 (50) 1066 (50) NA	307 (100) 153 (50) 61 (20) 93 (30)	<0.001
Location ruptured aneurysm, n (%)	Total available Anterior circulation Posterior circulation None	2128 (100) 2070 (97) 58 (3) NA	307 (100) 210 (68) 77 (25) 20 (7)	<0.001

NA, not applicable.
[a]p were calculated by Mann-Whitney U test, chi-square test, or Fisher exact test.
[b]Median (interquartile range).

Table 5.2. Univariable and multivariable association of predictors with 60-day case fatality

	ISAT derivation cohort (n=2,128), OR (95% CI)		Rotterdam validation cohort (n=307), OR (95% CI)	
Predictor	**Univariable**	**Multivariable**	**Univariable**	**Multivariable**
Age (10 yr)	1.43 (1.23–1.66)	1.32 (1.13–1.55)	1.26 (1.04–1.52)	1.50 (1.12–2.02)
Maximum lumen size aneurysm (mm)	1.10 (1.05–1.15)	1.08 (1.03–1.13)	1.12 (1.06–1.19)	1.04 (0.95–1.14)
Fisher grade				
1	0.15 (0.04–0.63)	0.36 (0.09–1.49)	0.00 (—)	0.00 (—)
2	0.24 (0.12–0.48)	0.52 (0.27–1.02)	0.00 (—)	0.00 (—)
3	0.63 (0.44–0.89)	0.97 (0.69–1.37)	0.21 (0.09–0.48)	0.93 (0.31–2.81)
4	Reference	Reference	Reference	Reference
World Federation of Neurological Surgeons grade				
1	Reference	Reference	Reference	Reference
2	2.42 (1.63–3.60)	1.87 (1.23–2.83)	3.20 (1.00–10.24)	2.56 (0.78–8.42)
3	2.55 (1.35–4.80)	1.70 (0.87–3.32)	4.40 (0.43–45.07)	4.45 (0.39–50.61)
4	8.12 (4.51–14.63)	4.87 (2.60–9.14)	6.95 (2.30–21.01)	5.71 (1.79–18.24)
5	12.66 (4.86–33.02)	7.00 (2.54–19.28)	299.20 (83.53–1071.74)	272.82 (68.97–1079.24)
6 (not assessable)	9.62 (4.23–21.89)	5.75 (2.41–13.73)	NA	NA

OR, odds ratio, NA, not applicable.

Model performance

Assessment of discriminative ability of the model in all patients (n = 307) including WFNS grade at time of treatment decision showed an AUC of 0.89 (Figure 5.1A). When WFNS grade at admission was used, the AUC was 0.82 (Figure 5.1B), indicating less optimal discrimination.

The model with WFNS grade at admission predicted 17.0% 60-day case fatality, and the model with WFNS grade at time of treatment decision 17.7%, whereas the observed case fatality was 30.6%. The calibration slopes were 1.417 for the model with WFNS grade at admission, and 1.959 for WFNS grade at time of treatment decision. The intercepts were 1.502 and 2.248, respectively, indicating that the model's predictions of case fatality were systematically lower than observed case fatality. When WFNS grade at time of treatment decision was used as a predictor, this overall underestimation increased. In patients with low observed case fatality risk (≤ 20%) (Figure 5.2A), the calibration plot shows adequate agreement between predicted and observed 60-day case fatality. The model was also tested in the nonimputed dataset (only complete cases, n = 266), which showed similar results (not shown).

Sensitivity analysis in patients similar to the original ISAT population (n = 211) showed reasonable calibration and discrimination in the model with WFNS grade at time of treatment decision (AUC = 0.72; calibration slope, 0.916) (Figure 5.2). The model with WFNS grade at admission showed lower discriminative ability between survivors and nonsurvivors (AUC = 0.65; calibration slope, 0.599). Intercepts were –0.423 for the model with WFNS grade at time of treatment decision and –1.031 for WNFS at admission. This indicates an overall overestimation of case fatality in this selection of patients, which decreased when using WFNS grade at time of treatment decision.

For both versions of the prediction model, discrimination was better in the unselected validation cohort, which was largely explained by a more heterogeneous case-mix, compared with the cohort of patients who met the original ISAT selection criteria. This is indicated by a small difference between case-mix–corrected AUCs of the two cohorts: the case-mix–corrected AUCs in the unselected cohort (n = 307) were 0.77 for WFNS grade at treatment decision and 0.76 for WFNS grade at admission, versus 0.74 (WFNS grade at treatment decision) and 0.73 (WFNS grade at admission) in the cohort that met the original ISAT selection criteria.

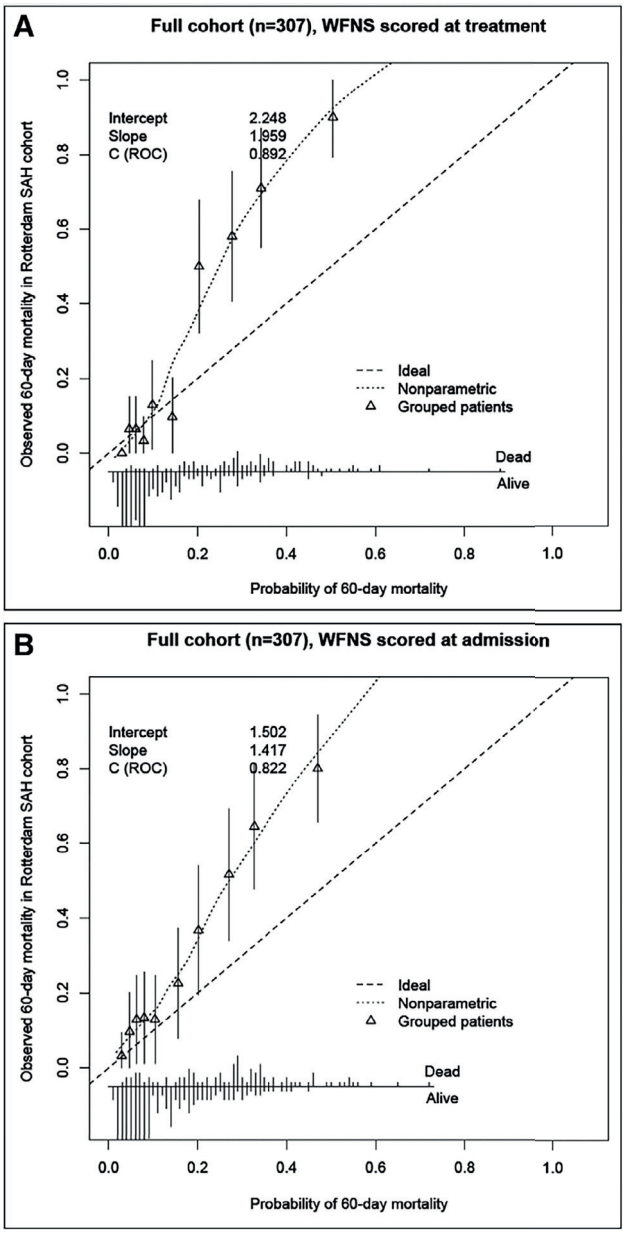


Figure 5.1. Calibration plots of the model with (**A**) World Federation of Neurological Surgeons (WFNS) grade at time of treatment decision and (**B**) WFNS grade at admission in the unselected Rotterdam validation cohort (n = 307). C, area under the receiver operating characteristic (ROC) curve, SAH, subarachnoid hemorrhage.

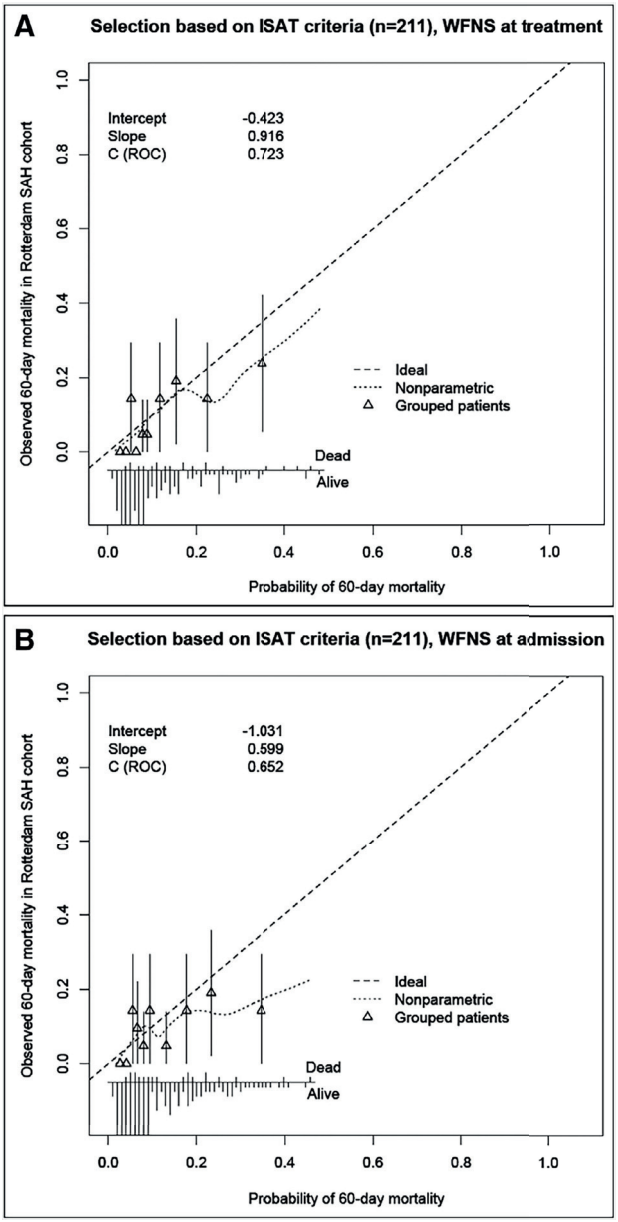


Figure 5.2. Calibration plots of the model with patients clipped or coiled after consideration for both treatment modalities (similar to original International Subarachnoid Aneurysm Trial [ISAT] inclusion criteria, n = 211) with (**A**) World Federation of Neurological Surgeons (WFNS) grade at time of treatment decision and (**B**) WFNS grade at admission.

C, area under the receiver operating characteristic (ROC) curve, SAH, subarachnoid hemorrhage.

Discussion

This study is the first to externally validate a prognostic model for SAH based on data from ISAT to predict 60-day case fatality. External validation yielded a discriminative performance superior to the derivation setting, suggesting generalizability. However, predicted probabilities were lower than observed 60-day case fatality, implicating relatively poor calibration. An important secondary finding was that timing of WFNS grade assessment influenced model performance.

A recent systematic review showed that the ISAT prediction model has reasonable performance and good design compared with other SAH prediction models.[4] The most commonly used predictors in this review were age, WFNS grade, Fisher grade, and aneurysm size.[4] These variables are easily obtainable at admission, facilitating a prognostic estimate early in the disease course. However, we found greater predictive ability of the model with WFNS grade at time of treatment decision compared with WFNS grade at admission. This finding is in line with a previous study demonstrating that WFNS grade obtained at admission was inferior to WFNS grade after admission.[13] This indicates that including a change over time may help to improve model performance.[13,14] This seems especially true for aSAH, which is characterized by variable clinical course.[14-16] However, the ideal timing of obtaining predictors does depend not only on model performance but also on the timing of various treatments, based on such predictors (e.g., immediately after admission or later).

Assessing the performance of a prognostic model and interpreting its clinical relevance is complex.[8] We aimed to investigate the generalizability of the ISAT model by testing it in an unselected cohort within a different setting (observational cohort vs randomized controlled trial) and case-mix (more severely affected patients). In principle, a model is generalizable to populations comparable to the development data, based on the data (i.e., comparability of mean age, severity) or on clinical judgment (are populations expected to be comparable between center A and B). However, generalizability is not by definition limited to populations comparable to the development setting; model estimates may also be valid in broader populations. External validation is useful to see whether the model can be used in different settings. Thus the differences between derivation and validation cohorts are more an advantage than a limitation of our study.

We found higher AUCs in the heterogeneous validation population than described in both the derivation cohort and the sensitivity analysis in patients who were clipped or coiled. The higher AUCs reflect the less restrictive enrollment criteria: the greater the heterogeneity, the better the model can distinguish patients with or without the outcome of interest. Discriminative ability of the ISAT prediction model in our cohort remained adequate (well over 0.70) after correcting for the more heterogeneous case-mix. This finding suggests that the model might be applicable for prognostic classification of future aSAH patient populations.

The good discriminative ability of the model is accompanied by an overall underestimation of case fatality, especially in patients with high case fatality risk. This finding is partly explained by the fact that patients who died early are included in our cohort (13% of our patients died < 48 hr), but not in ISAT. Furthermore, the modest performance of the model in patients with a case fatality risk

greater than 20% is in line with the higher proportion of poor-grade patients in the validation cohort. It is indeed common that RCTs—as ISAT—typically include lower risk patients. When risk estimates are used for clinical decision making, reliable absolute risk estimates are needed. Therefore, a model with good agreement between observed and predicted probabilities (calibration) is required.[17] For clinical practice, this implicates that the ISAT prediction model needs to be updated in more recent data and for specific settings. Specifically, we would recommend adjustment of the intercept of the model such that the overall mean predicted probability is equal to the observed overall outcome frequency (recalibration). A second step in updating the model would be reestimation of the regression coefficients of the predictors in the model. Whether such updating is needed should be decided based on external validation results, the comparability of the development and validation setting based on clinical knowledge, and the number of patients in the development setting. For example, one would not decide to completely refit a model based on a small validation set when the development sample was very large. A general message is that existing prognostic models should always be considered and validated instead of developing new models.

Strengths of this study are external validation in an unselected population of aSAH patients, reflecting real-life clinical practice and replication of predictors and outcome.[5] Several limitations of our study need to be considered. First, since the validation cohort consisted of ICU managed aSAH patients, model performance may not apply to non-ICU patients. Second, this study is a single-center study and external validity of the ISAT model needs to be confirmed. Third, there was a small number of missing outcomes in the unselected cohort, but sensitivity analyses accounting for missing outcomes did not differ (results not shown). Finally, we only had case fatality at 60 days and not functional status. Although 60-day case fatality is a very robust outcome, long-term disability is a more relevant outcome for patients and should be included in future prognostic studies.

Outcome prediction in individual aSAH patients remains difficult due to the variable clinical course and multiple treatment options.[18] Our findings might indicate that acute phase variables are not ideal predictors in diseases with variable clinical course, in contrast to neurologic diseases with a less variable course.[19] To improve outcome prediction, including dynamic variables over time in future models may benefit performance. Additionally, we could focus on predictors with a higher prognostic value. Since Fisher grade has suboptimal interobserver variability,[20-23] the use of other grading scales for blood on CT may further improve model performance.[20,24,25] Importantly, future prognostic models on mortality should include data on causes of death and withdrawal of care practices to further scrutinize external validity of such models.

Conclusions

Validation of existing models should always be taken as a starting point in prognostic model development. This external validation study confirms generalizability of the ISAT prognostic model in terms of discrimination, in an independent unselected cohort of more severely affected aSAH

patients. In addition, WFNS grade at treatment decision for the ruptured aneurysm benefitted model performance. However, predicted probabilities were lower than observed case fatality, illustrating the need for continuous external validation and updating over time and to specific settings before implementation in clinical practice.

Acknowledgments

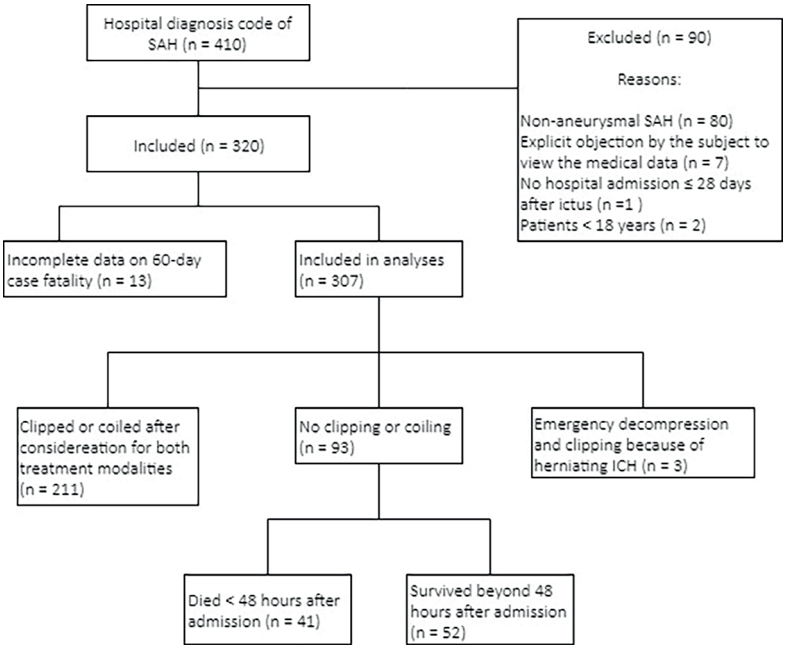
We thank Prof A. J. Molyneux (International Subarachnoid Aneurysm Trial [ISAT] investigator) and Dr. R. Risselada for providing the summary statistics of the ISAT baseline data as presented in Table 5.1.

References

1. van Gijn J, Kerr RS, Rinkel GJ. Subarachnoid haemorrhage. *Lancet.* 2007; 369:306–318
2. Rinkel GJ, Algra A. Long-term outcomes of patients with aneurysmal subarachnoid haemorrhage. *Lancet Neurol.* 2011; 10:349–356
3. Nieuwkamp DJ Vaartjes I Algra A, Bots ML, Rinkel GJ. Age- and gender-specific time trend in risk of death of patients admitted with aneurysmal subarachnoid hemorrhage in the Netherlands. *Int J Stroke.* 2013; 8(SupplA100):90–94
4. Jaja BN, Cusimano MD, Etminan N, Hanggi D, Hasan D, Ilodigwe D, et al. Clinical prediction models for aneurysmal subarachnoid hemorrhage: A systematic review. *Neurocrit Care.* 2013; 18:143–153
5. Risselada R, Lingsma HF, Bauer-Mehren A, Friedrich CM, Molyneux AJ, Kerr RS, et al. Prediction of 60 day case-fatality after aneurysmal subarachnoid haemorrhage: Results from the International Subarachnoid Aneurysm Trial (ISAT). Eur J Epidemiol 2010; 25:261–266
6. Riley RD, Hayden JA, Steyerberg EW, Moons KG, Abrams K, Kyzas PA, et al. Prognosis Research Strategy (PROGRESS) 2: Prognostic factor research. *PLoS Med.* 2013; 10:e1001380
7. Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: Prognostic model research. *PLoS Med.* 2013; 10:e1001381
8. Siontis GC, Tzoulaki I, Castaldi PJ, Ioannidis JP. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *J Clin Epidemiol.* 2015; 68:25–34
9. Molyneux A, Kerr R, Stratton I, Sandercock P, Clarke M, Shrimpton J, et al. International Subarachnoid Aneurysm Trial (ISAT) Collaborative Group: International Subarachnoid Aneurysm Trial (ISAT) of neurosurgical clipping versus endovascular coiling in 2143 patients with ruptured intracranial aneurysms: A randomized trial. *Lancet.* 2002; 360:1267–1274
10. Fisher CM, Kistler JP, Davis JM. Relation of cerebral vasospasm to subarachnoid hemorrhage visualized by computerized tomographic scanning. *Neurosurgery.* 1980; 6:1–9
11. Drake CG, Hunt WE, Sano K, Kassell N, Teasdale G, Pertuiset B, et al. Report of World Federation of Neurological Surgeons Committee on a Universal Subarachnoid Hemorrhage Grading Scale. *J Neurosurg.* 1988; 68:985–986
12. Vergouwe Y, Moons KG, Steyerberg EW. External validity of risk models: Use of benchmark values to disentangle a case-mix effect from incorrect coefficients. *Am J Epidemiol.* 2010; 172:971–980
13. Giraldo EA, Mandrekar JN, Rubin MN, Dupont SA, Zhang Y, Lanzino G, et al. Timing of clinical grade assessment and poor outcome in patients with aneurysmal subarachnoid hemorrhage. *J Neurosurg.* 2012; 117:15–19
14. Ehlenbach WJ, Cooke CR. Making ICU prognostication patient centered: Is there a role for dynamic information? *Crit Care Med.* 2013; 41:1136–1138
15. Helbok R, Kurtz P, Vibbert M, Schmidt MJ, Fernandez L, Lantigua H, et al. Early neurological deterioration after subarachnoid haemorrhage: Risk factors and impact on outcome. *J Neurol Neurosurg Psychiatry.* 2013; 84:266–270
16. Sasaki T, Sato M, Oinuma M, Sakuma J, Suzuki K, Matsumoto M, et al. Management of poor-grade patients with aneurysmal subarachnoid hemorrhage in the acute stage: Importance of close monitoring for neurological grade changes. *Surg Neurol.* 2004; 62:531–535

17. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med.* 1999; 130:515–524
18. Suarez JI, Tarr RW, Selman WR. Aneurysmal subarachnoid hemorrhage. *N Engl J Med.* 2006; 354:387–396
19. van Koningsveld R, Steyerberg EW, Hughes RA, Swan AV, van Doorn PA, Jacobs BC. A clinical prognostic scoring system for Guillain-Barré syndrome. *Lancet Neurol.* 2007; 6:589–594
20. Smith ML, Abrahams JM, Chandela S, Smith MJ, Hurst RW, Le Roux PD. Subarachnoid hemorrhage on computed tomography scanning and the development of cerebral vasospasm: The Fisher grade revisited. *Surg Neurol.* 2005; 63:229–234
21. Lindvall P, Runnerstam M, Birgander R, Koskinen LO. The Fisher grading correlated to outcome in patients with subarachnoid haemorrhage. *Br J Neurosurg.* 2009; 23:188–192
22. van Norden AG, van Dijk GW, van Huizen MD, Algra A, Rinkel GJ. Interobserver agreement and predictive value for outcome of two rating scales for the amount of extravasated blood after aneurysmal subarachnoid haemorrhage. *J Neurol.* 2006; 253:1217–1220
23. Kapapa T, Tjahjadi M, König R, Wirtz CR, Woischneck D. Which clinical variable influences health-related quality of life the most after spontaneous subarachnoid hemorrhage? Hunt and Hess scale, Fisher score, World Federation of Neurosurgeons score, Brussels coma score, and Glasgow coma score compared. *World Neurosurg.* 2013; 80:853–858
24. Hijdra A, Brouwers PJ, Vermeulen M, van Gijn J. Grading the amount of blood on computed tomograms after subarachnoid hemorrhage. *Stroke.* 1990; 21:1156–1161
25. Wilson DA, Nakaji P, Abla AA, Uschold TD, Fusco DJ, Oppenlander ME, et al. A simple and quantitative method to predict symptomatic vasospasm after subarachnoid hemorrhage based on computed tomography: Beyond the Fisher scale. *Neurosurgery.* 2012; 71:869–875

Appendix



Appendix 5.A. Flowchart clarifying patient flow according to in- and exclusion criteria

Appendix 5.B. Management guideline with regard to acutely treat or not (yet) treat ruptured intracranial aneurysm in Erasmus Medical Center during period of study (validation cohort):

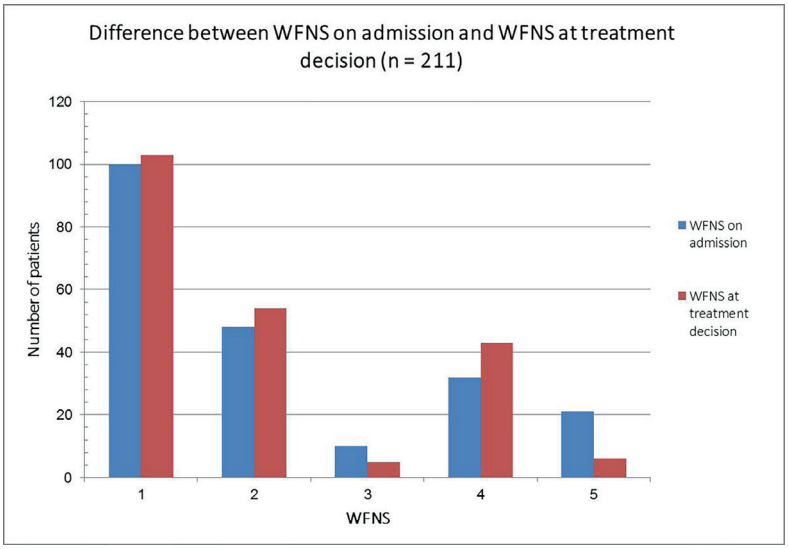
The decision whether or not to treat the aneurysm of individual patients was made based on the local multidisciplinary clinical guidelines in our university hospital. These guidelines state that no endovascular or neurosurgical aneurysm treatment immediately after SAH is considered in patients with WFNS grade 5 who do not improve after resuscitation within the first 24 hours and/or CSF drainage in case of hydrocephalus (excluding those who have a space occupying ICH with impending herniation necessitating emergency decompression craniotomy).

The combination of early deaths and the adherence to these local clinical guidelines explain why 30% of our cohort did not receive aneurysm treatment.

With regard to poor grade patients on admission:

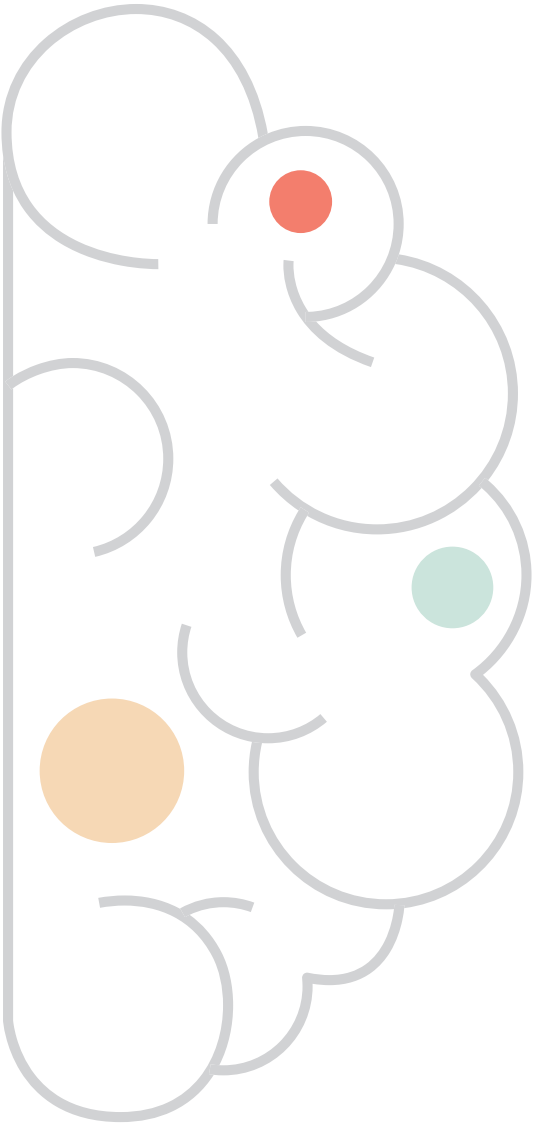
In severely affected patients (i.e. those with WFNS 5 or even those with (partially) absent brainstem reflexes after resuscitation), we adhere to a policy of treatment at the ICU (including CSF drainage in case of hydrocephalus) of at least 24 hours after the ictus during which time the course of the neurological examination and level of consciousness will guide our multi-disciplinary decision to stay on active treatment or to stop treatment because of infaust prognosis or consider organ donation in potential organ donors.

We acknowledge the fact that patients after SAH may eventually improve after successful resuscitation even when brain stem reflexes are initially absent.



Appendix 5.C. Histogram showing the significantly different distribution of WNFS on admission and WFNS at time of treatment decision among patients clipped or coiled after consideration for both treatment modalities (original ISAT selection criteria). Wilcoxon Signed Ranks test, p = 0.04.

CHAPTER 5.1



Letter by Dijkland et al regarding article, “Prediction of outcome after aneurysmal subarachnoid hemorrhage: Development and validation of the SAFIRE grading scale”

Simone A. Dijkland
Mathieu van der Jagt
Hester F. Lingsma

Stroke 2019; 50(7): e224

To the Editor:

With great interest, we read the study by van Donkelaar et al,[1] which describes the development and validation of the SAFIRE (size of the aneurysm, age, Fisher grade, World Federation of Neurosurgical Societies after resuscitation) grading scale to predict functional outcome after aneurysmal subarachnoid hemorrhage (aSAH). Indeed, early identification of aSAH patients at risk for poor functional outcome is important for clinical decision making. However, is the development of a new prognostic model the most logical approach, given the contemporary evidence on outcome prediction in aSAH?

Several cross-validated or externally validated prognostic models for mortality and functional status after aSAH exist.[2–5] Although the generalizability and transportability of prognostic models to other populations can only be established after a continuous process of model validation and updating, existing models should always be taken into account to prevent development of multiple models with unknown generalizability. For instance, poor calibration of the ISAT (International Subarachnoid Aneurysm Trial) model in a single-center aSAH population implied that the model should be updated,[3] not discarded. According to van Donkelaar et al,[1] the currently available prognostic models lack accuracy and generalizability, but this was not fully tested in their own data. Instead of developing a new model, validation and updating of available prognostic models for aSAH would have been preferred.[3]

Additionally, it is not evident that this study provides novel insights for clinicians and researchers in the field of aSAH. The authors state that the SAFIRE grading scale excels in simplicity.[1] However, the final predictors of this prognostic model (age, World Federation of Neurological Surgeons grade, Fisher grade, and aneurysm size) are identical to those in the ISAT model.[2] The potential limitation of the ISAT model that World Federation of Neurological Surgeons grade was assessed at randomization, which is not clinically applicable, has been addressed in a previous external validation.[3] The SAHIT (Subarachnoid Hemorrhage International Trialists) prognostic models that were developed on >10 000 patients from multiple randomized clinical trials and observational studies,[4] were deemed by the authors to be complex and confusing for use in clinical practice.[1] But these SAHIT models with increasing complexity (core: age, hypertension and World Federation of Neurological Surgeons grade; neuroimaging: core+Fisher grade, aneurysm location, and size; full: neuroimaging+aneurysm treatment)[4] facilitate insight in the added value of new predictors and allow clinicians to predict outcome depending on the clinical situation (eg, before or after imaging). Moreover, model simplicity is merely a matter of model presentation: when the regression equation of a prognostic model is available, a risk score or nomogram can easily be developed.[4]

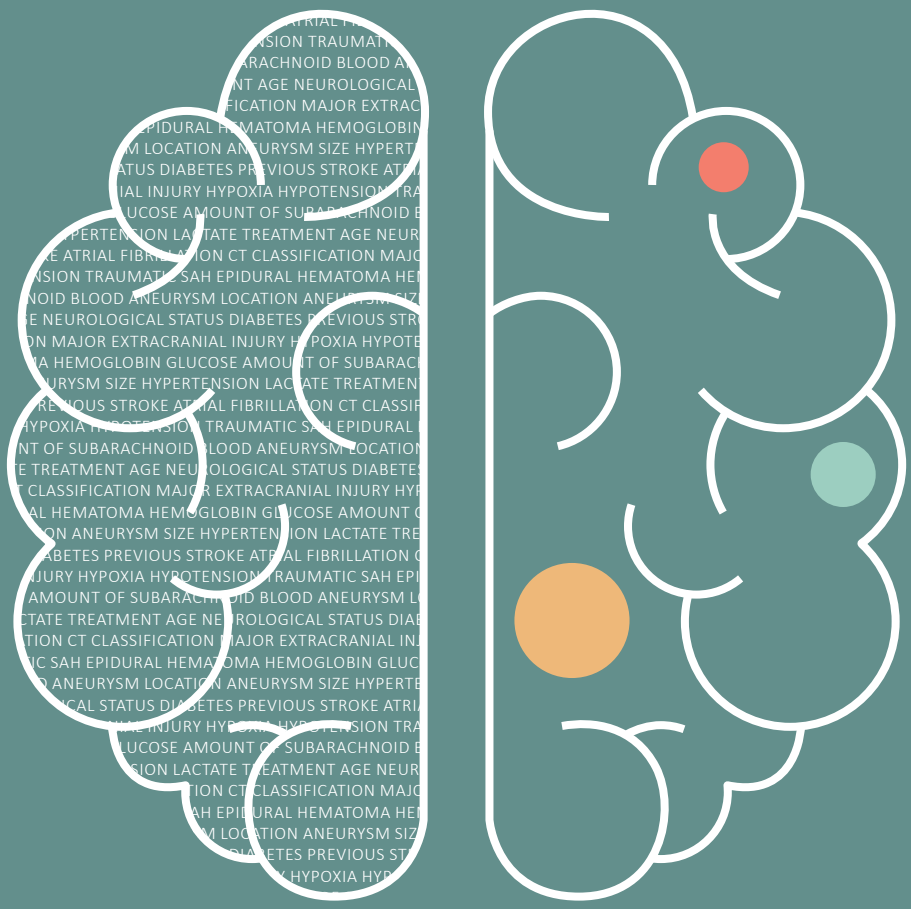
In conclusion, the proposed SAFIRE grading scale resembles existing prognostic models for clinical outcome after aSAH in terms of predictors, performance, and simplicity and does, therefore, not seem to contribute to current knowledge. External validation and updating of existing prognostic models should always be considered before development of a new model. This is especially relevant in a disease like aSAH for which a variety of neurological and imaging grading scales are being used worldwide, while core predictors of clinical outcome have been established.

References

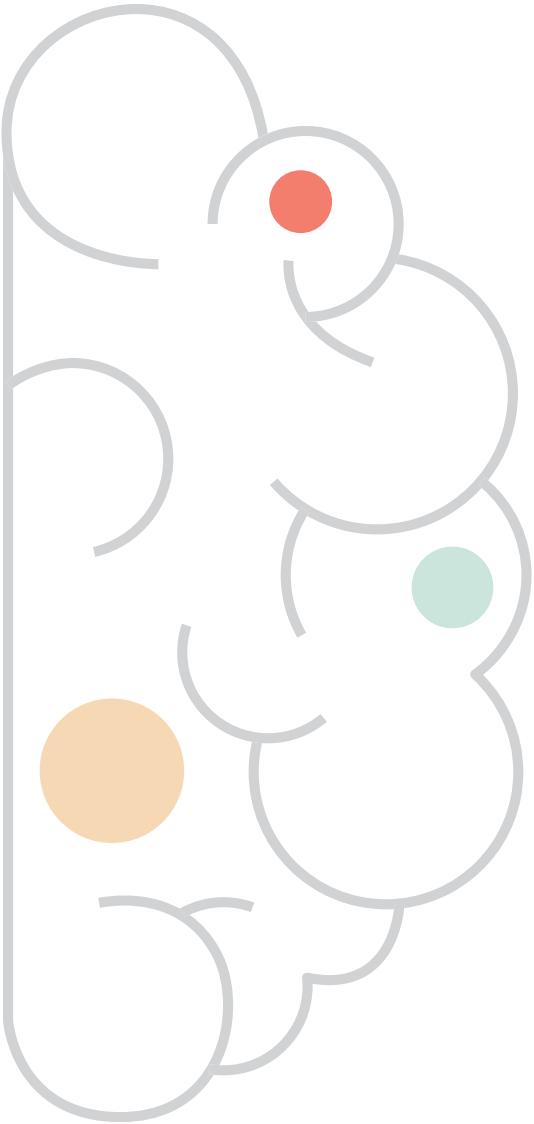
1. van Donkelaar CE, Bakker NA, Birks J, Veeger NJGM, Metzemaekers JDM, Molyneux AJ, et al. Prediction of outcome after aneurysmal subarachnoid hemorrhage: development and validation of the SAFIRE grading scale. *Stroke.* 2019;50:837–844.
2. Risselada R, Lingsma HF, Bauer-Mehren A, Friedrich CM, Molyneux AJ, Kerr RS, et al. Prediction of 60 day case-fatality after aneurysmal subarachnoid haemorrhage: results from the International Subarachnoid Aneurysm Trial (ISAT). *Eur J Epidemiol.* 2010;25:261–266.
3. Dijkland SA, Roozenbeek B, Brouwer PA, Lingsma HF, Dippel DW, Vergouw LJ, et al. Prediction of 60-day case fatality after aneurysmal subarachnoid hemorrhage: external validation of a prediction model. *Crit Care Med.* 2016;44:1523–1529.
4. Jaja BNR, Saposnik G, Lingsma HF, Macdonald E, Thorpe KE, Mamdani M, et al.; SAHIT collaboration. Development and validation of outcome prediction models for aneurysmal subarachnoid haemorrhage: the SAHIT multinational cohort study. *BMJ.* 2018;360:j5745.
5. Mascitelli JR, Cole T, Yoon S, Nakaji P, Albuquerque FC, McDougall CG, et al. External validation of the Subarachnoid Hemorrhage International Trialists (SAHIT) predictive model using the Barrow-Ruptured Aneurysm Trial (BRAT) cohort. [published online December 19, 2018]. *Neurosurgery.* 2018. https://doi.org/10.1093/neuros/nyy600. Accessed March 20, 2019.

PART III

OUTCOME ANALYSES



CHAPTER 7



Between-center and between-country differences in outcome after aneurysmal subarachnoid hemorrhage in the Subarachnoid Hemorrhage International Trialists (SAHIT) repository

Simone A. Dijkland
Blessing N.R. Jaja
Mathieu van der Jagt
Bob Roozenbeek
Mervyn D.I. Vergouwen
Jose I. Suarez
James C. Torner
Michael M. Todd
Walter M. van den Bergh
Gustavo Saposnik
Daniel W. Zumofen
Michael D. Cusimano
Stephan A. Mayer
Benjamin W.Y. Lo
Ewout W. Steyerberg
Diederik W.J. Dippel
Tom A. Schweizer
R. Loch Macdonald
Hester F. Lingsma
on behalf of the SAHIT collaboration

J Neurosurg 2019; Epub ahead of print

Abstract

Object: Differences in clinical outcomes between centers and countries may reflect variation in patient characteristics, diagnostic and therapeutic policies or quality of care. The purpose of this study was to investigate the presence and magnitude of between-center and between-country differences in outcome after aneurysmal subarachnoid hemorrhage (aSAH).

Methods: We analyzed data from 5972 aSAH patients enrolled in randomized clinical trials of 3 different treatments from the Subarachnoid Hemorrhage International Trialists (SAHIT) repository including data from 179 centers and 20 countries. We used random effects logistic regression adjusted for patient characteristics and timing of aneurysm treatment to estimate between-center and between-country differences in unfavorable outcome, defined as Glasgow Outcome Scale score of 1-3 (severe disability, vegetative state or death) or modified Rankin Scale score of 4-6 (moderately severe disability, severe disability or death) at three months. Between-center and between-country differences were quantified with the median odds ratio (MOR), which can be interpreted as the ratio of odds of unfavorable outcome between a typical high-risk and a typical low-risk center or country.

Results: The proportion of patients with unfavorable outcome was 27% (n=1599). We found substantial between-center differences (MOR = 1.26, 95% CI 1.16-1.52), which could not be explained by patient characteristics and timing of aneurysm treatment (adjusted MOR = 1.21, 95% CI 1.11-1.44). We observed no between-country differences (adjusted MOR = 1.13, 95% CI 1.00-1.40).

Conclusions: Clinical outcomes after aSAH differ between centers. These differences could not be explained by patient characteristics or timing of aneurysm treatment. Further research is needed to confirm the presence of differences in outcome after aSAH between hospitals in more recent data and to investigate potential causes.

Introduction

Despite advances in treatment, functional outcome after aneurysmal subarachnoid hemorrhage (aSAH) remains poor.[1, 2] The combination of a relatively young age of onset and poor clinical outcomes makes aSAH a disease with major individual and economic impact.[3] The main evidence-based treatment recommendations in aSAH include endovascular coil embolization in patients with a ruptured aneurysm eligible for both endovascular coiling and neurosurgical clipping, administration of oral nimodipine and maintenance of euvolemia to prevent delayed cerebral ischemia (DCI), and drainage of cerebrospinal fluid in patients with hydrocephalus.[4] However, many other interventions to prevent or treat complications in aSAH are less evidence-based.[4, 5] Also, discrepancies have been found between centers regarding clinical practice and adherence to guidelines for aSAH,[6, 7] suggesting differences in diagnostic and therapeutic policies between centers and countries that may contribute to variations in observed case-fatality rates across regions.[1]

Between-center and between-country differences in outcome can be caused by random variation or by center-, country- or patient-related factors (e.g. differences in country economic status or severity of aSAH), but they may also reflect differences in processes of care including diagnostic and therapeutic policies and adherence to guidelines (quality of care). Insight into between-center or between-country differences in outcome may facilitate research evaluating the comparative effectiveness of structures and processes of care in aSAH (e.g. organizational structures, individual treatment interventions), and may consequently contribute to improvement in quality of care. We aimed to investigate the presence and magnitude of between-center and between-country differences in clinical outcome after aSAH.

Methods

Study population

The Subarachnoid Hemorrhage International Trialists (SAHIT) repository contains data on more than 15,000 SAH patients from 10 randomized clinical trials (RCTs) and 11 observational studies or registries. For the present study, we used data from multicenter studies of 3 different treatments: the Intraoperative Hypothermia for Aneurysm Surgery Trial (IHAST), Magnesium Sulfate in Aneurysmal Subarachnoid Hemorrhage (MASH I and II) trials, and trials of tirilazad mesylate in patients with aneurysmal subarachnoid hemorrhage (tirilazad trials),[8-12] including a total of 6036 patients. The other studies in the SAHIT database could not contribute to the estimation of between-center and between-country differences, either because they were single-center studies (and therefore no distinction could be made between study effect and center or country effect) or because no information on center or country was available in the SAHIT database. Details on the development of the SAHIT repository and the included studies have been reported previously.[13] The SAHIT database was approved by the research ethics board at St Michael's Hospital, Toronto, Canada. Patients previously consented to the use of their

data for future related studies, and all data for the current study were anonymized. Therefore, neither approval from an institutional review board nor informed consent was required.

Primary outcome measure

The RCTs used either the Glasgow Outcome Scale (GOS)[8-10] or modified Rankin Scale (mRS) score[11, 12] at 3 months for functional outcome. We therefore defined our primary outcome measure as functional outcome according to the GOS or mRS score at 3 months, combined into a composite endpoint by dichotomizing both outcomes into favorable (GOS score 4-5 or mRS score 0-3) versus unfavorable (GOS score 1-3 or mRS score 4-6).

Between-center and between-country differences

We used random effects (multilevel) logistic regression to estimate differences in functional outcome after aSAH between centers and countries in order to be able to account for random variation due to small sample sizes per center or country and for differences in patient characteristics and process measures. In a random effects model, fixed effects are estimated for patient and process characteristics, and random effects are estimated for the effect of center and country. The random effects model assumes a normal distribution of the random effects. The variance of the random effects (T^2) estimated in the random effects logistic regression model is a measure for the unexplained between-center or between-country differences, independent of both random variation (chance) and patient and process characteristics as included in the model. Since between-center and between-country differences may influence each other, we used one random effects logistic regression model with both center and country as random effects (Appendix 7.A).

To facilitate interpretation of the between-center or between-country differences and allow for a direct comparison with the effect size (odds ratios) of patient characteristics, we calculated the median odds ratio (MOR) with 95% confidence interval (CI).[14, 15] For each pair of patients from different centers or countries, an odds ratio was computed between a patient from the center or country with the highest risk for unfavorable outcome and a patient from the center or country with the lowest risk for unfavorable outcome. The MOR represents the median value of the distribution of these odds ratios for unfavorable outcome for all pairs of patients in our dataset. The MOR is calculated based on the T^2 estimated in the random effects model, using the following formula: $MOR = \exp(\sqrt{[2 \times T^2]} \times \Phi^{-1}[0.75])$, where Φ corresponds to the cumulative distribution function of the normal distribution with mean 0 and variance 1. Hence, $\Phi^{-1}(0.75)$ is the 75th percentile.[14, 15] If there are no unexplained between-center or between-country differences, $T^2 = 0$ and MOR = 1.

The random effects logistic regression model was considered for both unadjusted between-center and between-country differences, and for between-center and between-country differences adjusted for differences in patient and process characteristics (fixed effects) between centers and countries. To enable comparison between the variance components of the unadjusted and adjusted models, we rescaled the variance of the adjusted models according to previously proposed methods.[16] The patient characteristics included in the model were age, history of hypertension, World Federation

of Neurosurgical Societies (WFNS) grade, Fisher grade, aneurysm location (anterior cerebral artery aneurysms [including anterior communicating artery aneurysms], internal cerebral artery aneurysms [including posterior communicating artery aneurysms], middle cerebral artery aneurysms or posterior circulation aneurysms [including vertebral and basilar artery aneurysms]), aneurysm size (≤12 mm, 13-24 mm or ≥25 mm)[17] and aneurysm treatment (clipping, coiling or none). These variables are known predictors of poor outcome after aSAH.[17-20] Because recommendations on the timing of aneurysm treatment differ between American and European guidelines, we additionally adjusted for the process measure "time from aSAH to aneurysm treatment".[4, 21] All analyses were also adjusted for study as a fixed effect because the overall outcome may vary across studies. Centers that participated in multiple studies were given the same center code across studies. We performed sensitivity analyses in the centers that included more than 10 patients to evaluate the robustness of our results.

Because the MOR is an overall measure for between-center and between-country differences, we also compared the effect estimates for the individual centers and countries to identify the hospitals or countries with the highest and lowest risk of unfavorable outcome. The estimated random effects (betas) for unfavorable outcome of the individual centers and countries were presented graphically by plotting them with a 95% CI.

Statistical analyses were performed with R software version 3.3.1 (R Foundation for Statistical Computing, Vienna, Austria). Missing data were statistically imputed using single imputation (*mice* package R). The CIs around the MOR were computed with the *confint.merMod* function (*lme4* package R).

Results

Study population

We analyzed data from 5972 aSAH patients from 179 centers in 20 different countries, after excluding patients with missing data on functional outcome (n=54) or unknown center (n=10). Missing data on history of hypertension (22%), Fisher grade (22%), aneurysm location (18%), aneurysm size (23%) and timing of aneurysm treatment (8%) were imputed. Unfavorable outcome at 3 months occurred in 1599 patients (27%), and 872 patients (15%) died. The patients' median age was 53 years (interquartile range [IQR] 44-62). A total of 1132 patients (19%) had a poor WFNS grade (4 or 5) at admission (Table 7.1). The number of included patients per center ranged from 1 to 846 (Figure 7.1A). The majority of patients were from the US (n=1765, 30%) or from one of 14 countries in Europe (n=3155, 53%). Other participating countries were Canada (n=536), Australia (n=344), New Zealand (n=142), Chile (n=21) and Mexico (n=9) (Figure 7.1B). The centers located in the US participated in the IHAST and tirilazad studies. The United Kingdom was the only country that contributed to the studies of all 3 treatments (Appendix 7.B). Patient characteristics, such as age, history of hypertension and poor WFNS or Fisher grade at admission, were predictive of unfavorable outcome (Appendix 7.C).

Table 7.1. Descriptive statistics of the studies in the SAHIT repository used for analysis of between-center and between-country differences

	IHAST	MASH I & II	Tirilazad
Study period	2000-2003	2000-2011	1991-1997
Original publication	Todd et al (2005)[10]	Van den Bergh et al 2005[11] Dorhout Mees et al 2012[12]	Kassell et al (1996)[9] Haley et al (1997)[8]
Patients, n	**1000**	**1484**	**3488**
Centers, n	30	9	148
Countries, n	7	3	19
Continents	Europe North America Oceania	Europe South America	Europe North America Oceania
Age in years, median (IQR)	52 (43-60)	56 (48-65)	51 (42-62)
History of hypertension, n (%)[a]	398 (40)	57 (4)	1124 (33)
Initial WFNS grade, n (%)			
1	660 (66)	728 (49)	1265 (36)
2	289 (29)	346 (23)	1028 (29)
3	51 (5)	64 (4)	408 (12)
4	0 (0)	218 (15)	346 (10)
5	0 (0)	127 (8)	441 (13)
Fisher grade, n (%)[b]			
1	54 (5)	1 (0)	330 (9)
2	342 (34)	22 (1)	451 (13)
3	474 (47)	43 (3)	2271 (66)
4	130 (13)	141 (10)	414 (12)
Aneurysm location, n (%)[c]			
ACA/ACoA	391 (39)	190 (13)	1243 (36)
ICA/PCoA	318 (32)	117 (8)	1019 (29)
MCA	206 (21)	89 (6)	695 (20)
Pst circ (incl BA & VA)	84 (8)	61 (4)	469 (13)
Aneurysm size, n (%)[d]			
≤12 mm	878 (88)	143 (10)	2549 (73)
13-24 mm	94 (9)	14 (1)	785 (23)
≥25 mm	24 (3)	2 (1)	126 (4)
Aneurysm treatment			
Clipping	1000 (100)	551 (37)	3151 (90)
Coiling	0 (0)	735 (50)	0 (0)
None	0 (0)	198 (13)	337 (10)
Time from aSAH to aneurysm treatment in days, median (IQR)	2.0 (1.0-4.0)	1.0 (1.0-2.0)	1.4 (1.0-1.8)
Outcome at 3 mos, n (%)[e]			
Unfavorable	144 (14)	398 (27)	1057 (30)
Mortality	61 (6)	234 (16)	577 (17)

ACA, anterior cerebral artery; ACoA, anterior communicating artery; BA, basilar artery; circ, circulation; ICA, internal cerebral artery; MCA, middle cerebral artery; PCoA, posterior communicating artery; pst, posterior; VA, vertebral artery.

[a]MASH 1276 missing

[b]MASH 1277 missing. In the MASH trials, the Hijdra score was used to measure the amount of subarachnoid blood.

[c]MASH 1027 missing

[d]MASH 1325 missing

[e]Outcome was based on 3-month GOS scores for IHAST and the tirilazad studies and 3-month mRS scores for the MASH trials.

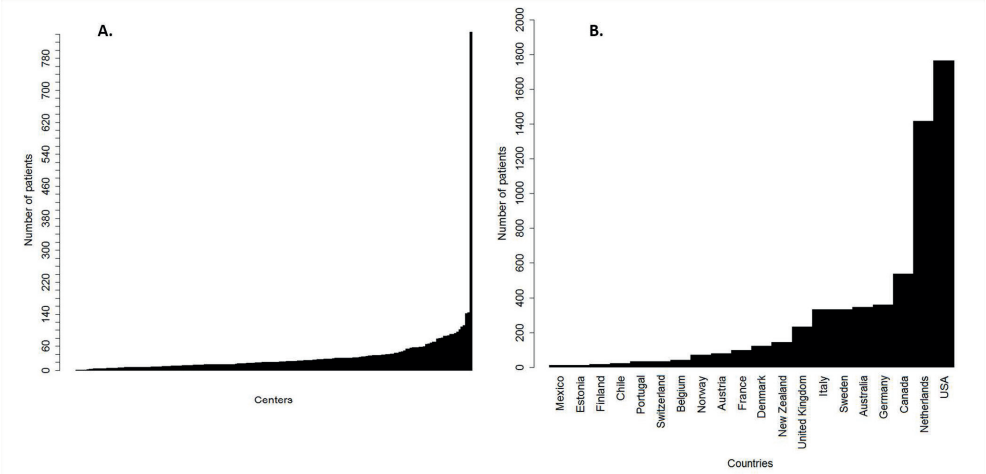


Figure 7.1. Observed number of patients (**A**) per center in one of 179 centers, with numbers varying from 1 to 846 (median 20; IQR 11-37) and (**B**) per country in one of 20 countries, with numbers varying from 9 to 1765 (median 109; IQR 31-334).

Between-center differences

We found between-center differences in functional outcome, both before and after adjustment for patient characteristics and time to aneurysm treatment (MOR = 1.26, 95% CI 1.16-1.52, and adjusted MOR = 1.21, 95% CI 1.11-1.44, respectively) (Table 7.2). The MOR of 1.21 implies a median increase of 21% in odds of unfavorable outcome if a patient was treated in a hospital with higher risk of unfavorable outcome. This order of magnitude is comparable to the effect of hypertension or aneurysm size larger than 12 mm (Appendix 7.C). While between-center differences were substantial in the tirilazad trials (adjusted MOR = 1.22, 95% CI 1.10-1.46), we found no between-center differences beyond random variation, patient characteristics and timing of aneurysm treatment in the IHAST (adjusted MOR = 1.00, 95% CI 1.00-1.02) and MASH studies (adjusted MOR = 1.00, 95% CI 1.00-1.50) (Table 7.2).

The effect estimates for unfavorable outcome in individual centers were subject to substantial uncertainty (Figure 7.2A), making it difficult to identify individual centers that perform better or worse than others.

Between-country differences

No between-country differences were observed in the unadjusted (MOR = 1.14, 95% CI 1.00-1.43) and adjusted (adjusted MOR = 1.13, 95% CI 1.00-1.40) analyses (Table 7.2 and Figure 7.2B). Between-country differences beyond random variation, patient characteristics and timing of treatment were absent in the IHAST (adjusted MOR = 1.00, 95% CI 1.00-1.02) and the MASH studies (adjusted MOR = 1.00, 95% CI 1.00-1.38) and nonsignificant in the tirilazad trials (adjusted MOR = 1.14, 95% CI 1.00-1.46) (Table 7.2).

Sensitivity analyses with only centers that included 10 or more patients yielded similar between-center and between-country differences (Appendix 7.D).

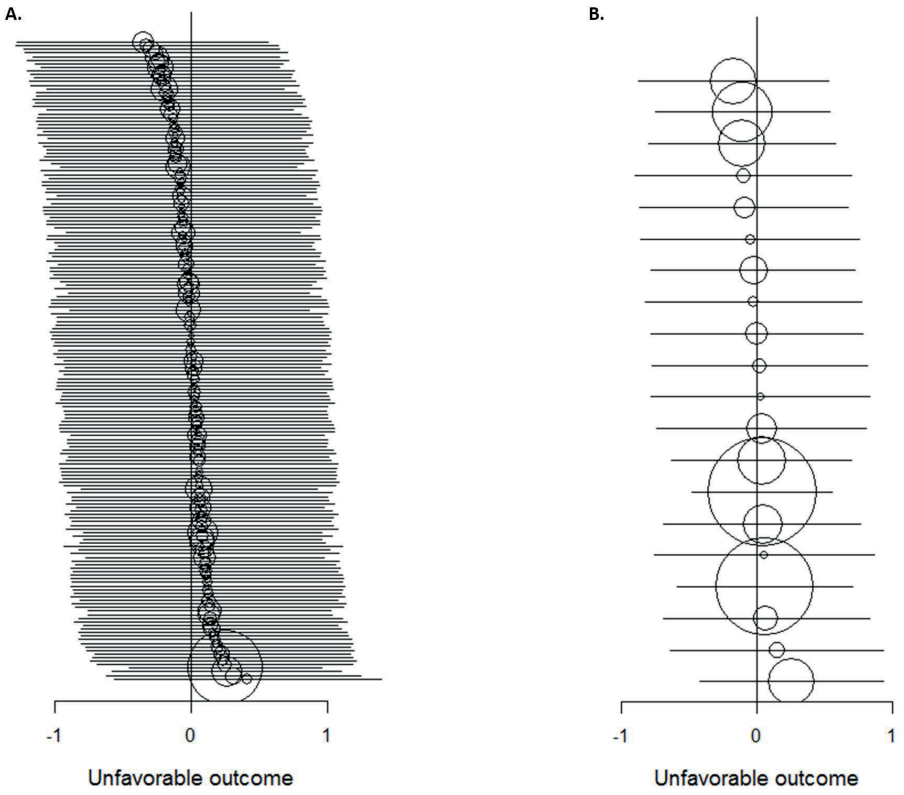


Figure 7.2. Differences between (**A**) centers and (**B**) countries in unfavorable outcome, adjusted for age, history of hypertension, WFNS, Fisher grade, aneurysm location, aneurysm size and time from SAH to aneurysm treatment in a random effects model. The circles indicate the random effects for the individual centers (betas), and the size of the circle refers to the number of patients in each center. The lines reflect the 95% confidence interval.

Table 7.2. Between-center and between-country differences in the total database (n=5972) and within studies.

	Unfavorable outcome n (%)	Unadjusted		Adjusted[a]	
Between-center differences[b]		T^2	MOR (95% CI)	T^2	MOR (95% CI)
Total[c] (n=5972)	1599 (27)	0.062	1.26 (1.16-1.52)	0.045	1.21 (1.11-1.44)
IHAST (n=1000)	144 (14)	0.000	1.00 (1.00-1.53)	0.000	1.00 (1.00-1.02)
MASH (n=1484)	398 (27)	0.050	1.23 (1.00-1.85)	0.000	1.00 (1.00-1.50)
Tirilazad (n=3488)	1057 (30)	0.074	1.28 (1.15-1.60)	0.047	1.22 (1.10-1.46)
Between-country differences[d]					
Total[c] (n=5972)	1599 (27)	0.021	1.14 (1.00-1.43)	0.016	1.13 (1.00-1.40)
IHAST (n=1000)	144 (14)	0.000	1.00 (1.00-1.69)	0.000	1.00 (1.00-1.02)
MASH (n=1484)	398 (27)	0.000	1.00 (1.00-1.70)	0.000	1.00 (1.00-1.38)
Tirilazad (n=3488)	1057 (30)	0.038	1.20 (1.05-1.58)	0.020	1.14 (1.00-1.46)

[a]Adjusted for age, hypertension, WFNS grade, Fisher grade, aneurysm location, aneurysm size, aneurysm treatment and time from aSAH to aneurysm treatment.

[b]Adjusted for country as a random effect.

[c]Models in the total database were adjusted for study.

[d]Adjusted for center as a random effect.

Discussion

We analyzed data from a large international repository of aSAH patients and observed substantial between-center differences in functional outcome that could not be explained by random variation, differences in patient characteristics or timing of aneurysm treatment. We observed no statistically significant between-country differences.

Previous studies have reported substantial between-center differences in other neurological diseases. Large between-center differences in outcome were found in a study in traumatic brain injury (TBI), based on more than 15,000 patients from both RCTs and observational studies.[22] The between-center differences in our study were similar to those reported in TBI (comparable variances).[22] Another example is the considerable between-center variability in functional outcome that was observed in patients enrolled in the Tinzaparin in Acute Ischemic Stroke Trial (TAIST).[23] In aSAH, only a few studies have reported on between-center or between-country differences in outcome.[24, 25] Moreover, studies that evaluated between-center and between-country variability generally used fixed effect models, while random effects logistic regression is preferred to better take into account clustering of patients, especially with a small number of patients per center or country.[26] The present study confirms the previously reported absence of between-center differences in outcome after aSAH within the IHAST study, but contradicts prior analyses by showing that between-center differences in outcome do exist

within the Tirilazad trials.[24,25] Our results were based on a large repository and we used advanced statistical methods accounting for differences due to random variation and patient or process characteristics.

Between-center differences in clinical outcomes after aSAH persisted after adjustment for patient characteristics and timing of aneurysm treatment. Other factors that might explain between-center differences are residual confounding and registration bias. However, these factors are unlikely to account for our results. We adjusted for known prognostic factors for outcome after aSAH as well as for time from aSAH to aneurysm treatment. This reduced the risk for residual confounding, although we acknowledge that data on several other factors that might influence outcome (e.g. withdrawal of life-sustaining measures or severity of underlying systemic illness) were unavailable. Also, our analyses were performed on multiple RCTs with high-quality data. Altogether, differences in unfavorable outcome between centers might be best explained by differences in diagnostic and therapeutic policies or quality of care. We observed no statistically significant between-country differences, suggesting that hospitals with similar patient outcomes are not clustered within one country.

Differences in outcome after aSAH between centers due to different treatment policies or quality of care are undesirable. However, because of limited evidence regarding treatment strategies and differences in adherence to guidelines,[4-6] it is expected that diagnostic and therapeutic policies for aSAH vary between centers and countries. This has been confirmed in previous studies.[27-29] In our study, the causality between variation in treatment policies or quality of care (other than timing of aneurysm treatment) and observed outcome differences could not be verified. We are therefore unable to present recommendations for current clinical practice. However, gaining insight into outcome differences between centers and countries is an important first step to evaluate practice variation and eventually improve clinical outcomes after aSAH. Our results provide the opportunity to perform comparative effectiveness research relating differences in structures and processes of care in aSAH between centers to differences in outcome. In TBI, such comparative effectiveness research is currently being conducted in a large prospective observational study.[30]

Assessing the performance of individual hospitals and countries is challenging since the estimates for specific centers and countries are subject to substantial uncertainty. Because the effect of chance increases with a decrease in the number of treated patients or outcomes,[31] a recommendation for future comparative effectiveness research is to focus on sufficient numbers of patients per center or country.

We found that between-center differences were substantial in the tirilazad trials, but were absent in the more recent IHAST and MASH trials. The tirilazad trials included more centers than the IHAST and MASH trials (Appendix 7.B), which increases the statistical power to identify differences in outcome. Moreover, progress has been made in diagnostic and therapeutic management since publication of the tirilazad trials and prognosis after aSAH may therefore have improved. For instance, the tirilazad studies and IHAST were (largely) conducted before publication of the International Subarachnoid Aneurysm Trial, so only 12% of the patients in our dataset underwent coil embolization. This and other factors related to the relatively old data limit the generalizability of our results to the contemporary aSAH population. Unfortunately, the more recent observational studies in the SAHIT repository could

not contribute to the estimation of between-center and between-country differences, because they were conducted in a single center or information on center or country was not available in the SAHIT database.[13] Given the evidence in aSAH and from related disease fields,[7, 22, 32] we consider it unlikely that between-center differences in clinical outcomes after aSAH are no longer present in current clinical practice. Our results should however be confirmed in a multicenter prospective cohort study.

Some other limitations should be acknowledged. Our data are based on RCTs with strict inclusion criteria. This created a relatively homogeneous study population, which might have caused an underestimation of the between-center and between-country differences. Further, the varying inclusion criteria (e.g. neurological condition on admission, time from onset of aSAH to inclusion) across the studies[8-11] made it impossible to assess the previously studied effect of center-volume on outcome.[33, 34] Information on other center- and country-specific aspects could not be retrieved due to the historic nature of the data, and the current center- and country-specific characteristics would not be applicable to the time when the data were collected for these studies. For example, the presence of neurocritical care teams has been associated with improved outcomes and inclusion of this factor in future observational studies would be very important.[35-37] Finally, we were unable to assess the effect of time on outcome differences, because the inclusion periods of the trials were relatively short, and only analyses on within-study time trends could be performed, since adjustment for study is required to distinguish between time effect and study effect.

Conclusions

Clinical outcomes after aSAH differ between centers. These differences could not be explained by random variation, patient characteristics or timing of aneurysm treatment. Further research is needed to confirm the presence of differences between hospitals with respect to outcome after aSAH between hospitals in more recent data and to investigate potential causes, such as variation in diagnostic and therapeutic policies or quality of care, in order to identify best practices and inform guidelines.

References

1. Nieuwkamp DJ, Setz LE, Algra A, Linn FH, de Rooij NK, Rinkel GJ. Changes in case fatality of aneurysmal subarachnoid haemorrhage over time, according to age, sex, and region: A meta-analysis. *The Lancet. Neurology*. 2009;8:635-642
2. Rinkel GJ, Algra A. Long-term outcomes of patients with aneurysmal subarachnoid haemorrhage. *The Lancet. Neurology*. 2011;10:349-356
3. Ridwan S, Urbach H, Greschus S, von Hagen J, Esche J, Bostrom A. Health care costs of spontaneous aneurysmal subarachnoid hemorrhage for rehabilitation, home care, and in-hospital treatment for the first year. *World Neurosurg*. 2017;97:495-500
4. Connolly ES, Jr., Rabinstein AA, Carhuapoma JR, Derdeyn CP, Dion J, Higashida RT, et al. Guidelines for the management of aneurysmal subarachnoid hemorrhage: A guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*. 2012;43:1711-1737
5. Macdonald RL. Delayed neurological deterioration after subarachnoid haemorrhage. *Nat Rev Neurol*. 2014;10:44-58
6. Gritti P, Akeju O, Lorini FL, Lanterna LA, Brembilla C, Bilotta F. A narrative review of adherence to subarachnoid hemorrhage guidelines. *J Neurosurg Anesthesiol*. 2017
7. Citerio G, Gaini SM, Tomei G, Stocchetti N. Management of 350 aneurysmal subarachnoid hemorrhages in 22 Italian neurosurgical centers. *Intensive Care Med*. 2007;33:1580-1586
8. Haley EC, Jr., Kassell NF, Apperson-Hansen C, Maile MH, Alves WM. A randomized, double-blind, vehicle-controlled trial of tirilazad mesylate in patients with aneurysmal subarachnoid hemorrhage: A cooperative study in North America. *J Neurosurg*. 1997;86:467-474
9. Kassell NF, Haley EC, Jr., Apperson-Hansen C, Alves WM. Randomized, double-blind, vehicle-controlled trial of tirilazad mesylate in patients with aneurysmal subarachnoid hemorrhage: A cooperative study in Europe, Australia, and New Zealand. *J Neurosurg*. 1996;84:221-228
10. Todd MM, Hindman BJ, Clarke WR, Torner JC, Intraoperative Hypothermia for Aneurysm Surgery Trial I. Mild intraoperative hypothermia during surgery for intracranial aneurysm. *N Engl J Med*. 2005;352:135-145
11. van den Bergh WM, Algra A, van Kooten F, Dirven CM, van Gijn J, Vermeulen M, et al. Magnesium sulfate in aneurysmal subarachnoid hemorrhage: A randomized controlled trial. *Stroke*. 2005;36:1011-1015
12. Dorhout Mees SM, Algra A, Vandertop WP, van Kooten F, Kuijsten HA, Boiten J, et al. Magnesium for aneurysmal subarachnoid haemorrhage (MASH-2): A randomised placebo-controlled trial. *Lancet*. 2012;380:44-49
13. Jaja BN, Attalla D, Macdonald RL, Schweizer TA, Cusimano MD, Etminan N, et al. The subarachnoid Hemorrhage International Trialists (SAHIT) repository: Advancing clinical research in subarachnoid hemorrhage. *Neurocrit Care*. 2014;21:551-559
14. Larsen K, Merlo J. Appropriate assessment of neighborhood effects on individual health: Integrating random and fixed effects in multilevel logistic regression. *Am J Epidemiol*. 2005;161:81-88
15. Merlo J, Chaix B, Ohlsson H, Beckman A, Johnell K, Hjerpe P, et al. A brief conceptual tutorial of multilevel analysis in social epidemiology: Using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *J Epidemiol Community Health*. 2006;60:290-297

16. Austin PC, Merlo J. Intermediate and advanced topics in multilevel logistic regression analysis. *Stat Med*. 2017;36:3257-3277
17. Jaja BNR, Saposnik G, Lingsma HF, Macdonald E, Thorpe KE, Mamdani M, et al. Development and validation of outcome prediction models for aneurysmal subarachnoid haemorrhage: The SAHIT multinational cohort study. *BMJ*. 2018;360:j5745
18. Jaja BN, Cusimano MD, Etminan N, Hanggi D, Hasan D, Ilodigwe D, et al. Clinical prediction models for aneurysmal subarachnoid hemorrhage: A systematic review. *Neurocrit Care*. 2013;18:143-153
19. Jaja BN, Lingsma H, Schweizer TA, Thorpe KE, Steyerberg EW, Macdonald RL, et al. Prognostic value of premorbid hypertension and neurological status in aneurysmal subarachnoid hemorrhage: Pooled analyses of individual patient data in the SAHIT repository. *J Neurosurg*. 2015;122:644-652
20. Dijkland SA, Roozenbeek B, Brouwer PA, Lingsma HF, Dippel DW, Vergouw LJ, et al. Prediction of 60-day case fatality after aneurysmal subarachnoid hemorrhage: External validation of a prediction model. *Crit Care Med*. 2016;44:1523-1529
21. Steiner T, Juvela S, Unterberg A, Jung C, Forsting M, Rinkel G, et al. European Stroke Organization guidelines for the management of intracranial aneurysms and subarachnoid haemorrhage. *Cerebrovasc Dis*. 2013;35:93-112
22. Lingsma HF, Roozenbeek B, Li B, Lu J, Weir J, Butcher I, 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*. 2011;68:601-607; discussion 607-608
23. Gray LJ, Sprigg N, Bath PM, Sorensen P, Lindenstrom E, Boysen G, et al. Significant variation in mortality and functional outcome after acute ischaemic stroke between western countries: Data from the Tinzaparin in Acute Ischaemic Stroke Trial (TAIST). *J Neurol Neurosurg Psychiatry*. 2006;77:327-333
24. Lipsman N, Tolentino J, Macdonald RL. Effect of country or continent of treatment on outcome after aneurysmal subarachnoid hemorrhage. *J Neurosurg*. 2009;111:67-74
25. Bayman EO, Chaloner KM, Hindman BJ, Todd MM, IHAST Investigators. Bayesian methods to determine performance differences and to quantify variability among centers in multi-center trials: The IHAST trial. *BMC Med Res Methodol*. 2013;13:5
26. Guo G, Zhao H. Multilevel modeling for binary data. *Annual Review of Sociology*. 2000;26:441–462
27. Fargen KM, Soriano-Baron HE, Rushing JT, Mack W, Mocco J, Albuquerque F, et al. A survey of intracranial aneurysm treatment practices among United States physicians. *J Neurointerv Surg*. 2018;10:44-49
28. Hollingworth M, Chen PR, Goddard AJ, Coulthard A, Soderman M, Bulsara KR. Results of an international survey on the investigation and endovascular management of cerebral vasospasm and delayed cerebral ischemia. *World Neurosurg*. 2015;83:1120-1126 e1121
29. Velly LJ, Bilotta F, Fabregas N, Soehle M, Bruder NJ, Nathanson MH, et al. Anaesthetic and ICU management of aneurysmal subarachnoid haemorrhage: A survey of European practice. *Eur J Anaesthesiol*. 2015;32:168-176
30. Maas AI, Menon DK, Steyerberg EW, Citerio G, Lecky F, Manley GT, et al. Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI): A prospective longitudinal observational study. *Neurosurgery*. 2015;76:67-80

31. Lingsma HF, Steyerberg EW, Eijkemans MJ, Dippel DW, Scholte Op Reimer WJ, Van Houwelingen HC, et al. Comparing and ranking hospitals based on outcome: Results from the Netherlands stroke survey. *QJM*. 2010;103:99-108
32. van Essen TA, den Boogert HF, Cnossen MC, de Ruiter GCW, Haitsma I, Polinder S, et al. Variation in neurosurgical management of traumatic brain injury: A survey in 68 centers participating in the CENTER-TBI study. *Acta Neurochir (Wien)*. 2018
33. Boogaarts HD, van Amerongen MJ, de Vries J, Westert GP, Verbeek AL, Grotenhuis JA, et al. Caseload as a factor for outcome in aneurysmal subarachnoid hemorrhage: A systematic review and meta-analysis. *J Neurosurg*. 2014;120:605-611
34. Pandey AS, Gemmete JJ, Wilson TJ, Chaudhary N, Thompson BG, Morgenstern LB, et al. High subarachnoid hemorrhage patient volume associated with lower mortality and better outcomes. *Neurosurgery*. 2015;77:462-470; discussion 470
35. Egawa S, Hifumi T, Kawakita K, Okauchi M, Shindo A, Kawanishi M, et al. Impact of neurointensivist-managed intensive care unit implementation on patient outcomes after aneurysmal subarachnoid hemorrhage. *J Crit Care*. 2016;32:52-55
36. Harrison DA, Prabhu G, Grieve R, Harvey SE, Sadique MZ, Gomes M, 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 Technol Assess*. 2013;17:vii-viii, 1-350
37. Suarez JI, Zaidat OO, Suri MF, Feen ES, Lynch G, Hickman J, et al. Length of stay and mortality in neurocritically ill patients: Impact of a specialized neurocritical care team. *Crit Care Med*. 2004;32:2311-2317

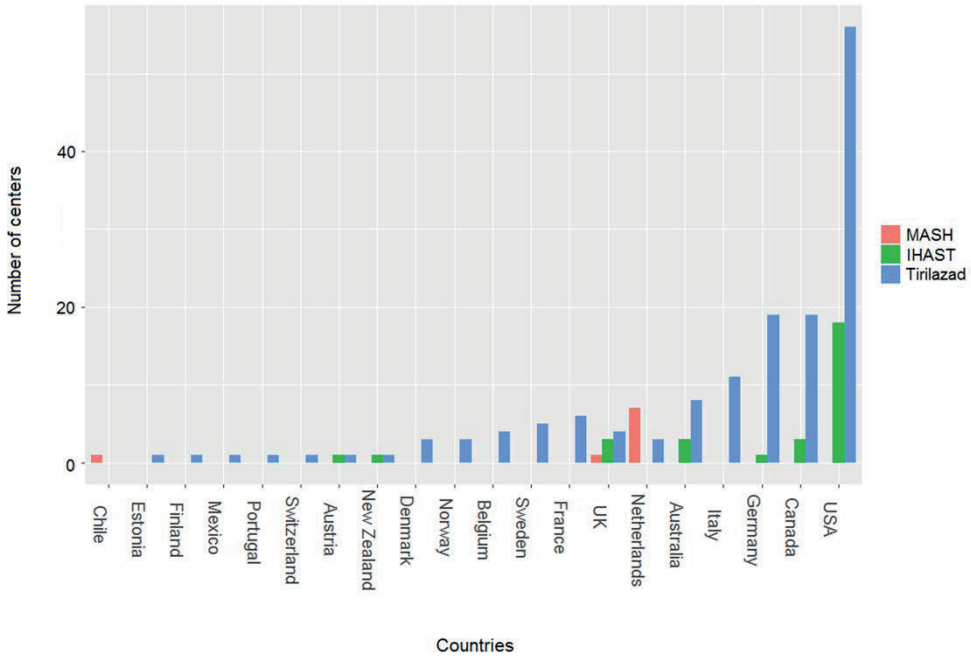
Appendix

Appendix 7.A. Random effects logistic regression model for between-center and between-country differences

Random effect logistic regression with random intercepts for center and country

$$\text{Logit}\ (p(Y_{ij} = 1)) = \beta_0 + \beta_1 + \beta_2 + (u_{0j} + u_{0k} + e_{0ijk})$$

With Y_{ij} the outcome for patient i in center j, β_0 the intercept, β_1 the patient and process characteristics, β_2 the study, u_{0j} the random intercept for center, u_{0k} the random intercept for the country, and e_{0ijk} the residuals. The random intercepts are assumed to be normally distributed with $\tau^2_{0j} = \text{var}(u_{0j})$ and $\tau^2_{0kj} = \text{var}(u_{0k})$.



Appendix 7.B. Number of centers per country within each of the trials

Appendix 7.C. Predictor effects for unfavorable outcome after aSAH in the multivariable logistic regression model ('fixed effects model')

Predictor	OR (95% CI)
Age per decade	1.45 (1.37-1.54)
Hypertension	1.52 (1.29-1.78)
WFNS grade	
1	1.0 (reference)
2	1.83 (1.54-2.18)
3	4.58 (3.65-5.73)
4	5.98 (4.80-7.46)
5	12.73 (10.11-16.03)
Fisher grade	
1	1.0 (reference)
2	1.27 (0.82-1.98)
3	2.01 (1.38-2.95)
4	1.97 (1.24-3.13)
Aneurysm location	
ACA/ACoA	1.0 (reference)
ICA/PCoA	0.84 (0.70-1.01)
MCA	0.68 (0.56-0.83)
Pst circ (incl BA & VA)	1.04 (0.81-1.33)
Aneurysm size	
≤ 12 mm	1.0 (reference)
13-24 mm	1.33 (1.10-1.60)
≥ 25 mm	1.54 (0.94-2.52)
Aneurysm treatment	
Clipping	1.0 (reference)
Coiling	0.69 (0.53-0.89)
None	3.35 (2.66-4.22)
Time from aSAH to aneurysm treatment in days	1.01 (0.99-1.04)

ACA, anterior cerebral artery; ACoA, anterior communicating artery; BA, basilar artery; circ, circulation; ICA, internal cerebral artery; MCA, middle cerebral artery; PCoA, posterior communicating artery; pst, posterior; VA, vertebral artery.

Appendix 7.D. Sensitivity analysis of between-center and between-country differences in centers with more than ten patients

	Unfavorable outcome n (%)	Unadjusted		Adjusted[a]	
Between-center differences[b]		**T^2**	**MOR (95% CI)**	**T^2**	**MOR (95% CI)**
Total[c] (n=5757)	1537 (27)	0.064	1.26 (1.17-1.52)	0.042	1.21 (1.09-1.43)
IHAST (n=971)	137 (14)	0.000	1.00 (1.00-1.56)	0.000	1.00 (1.00-1.02)
MASH (n=1484)	398 (27)	0.050	1.23 (1.00-1.85)	0.000	1.00 (1.00-1.50)
Tirilazad (n=3302)	1002 (30)	0.076	1.29 (1.16-1.61)	0.020	1.14 (1.06-1.29)
Between-country differences[d]					
Total[c] (n=5757)	1537 (27)	0.023	1.15 (1.00-1.44)	0.020	1.14 (1.00-1.42)
IHAST (n=971)	137 (14)	0.000	1.00 (1.00-1.71)	0.000	1.00 (1.00-1.02)
MASH (n=1484)	398 (27)	0.000	1.00 (1.00-1.70)	0.000	1.00 (1.00-1.38)
Tirilazad (n=3302)	1002 (30)	0.041	1.21 (1.06-1.64)	0.012	1.11 (1.00-1.32)

[a]Adjusted for age, hypertension, WFNS grade, Fisher grade, aneurysm location, aneurysm size, aneurysm treatment and time from aSAH to aneurysm treatment.

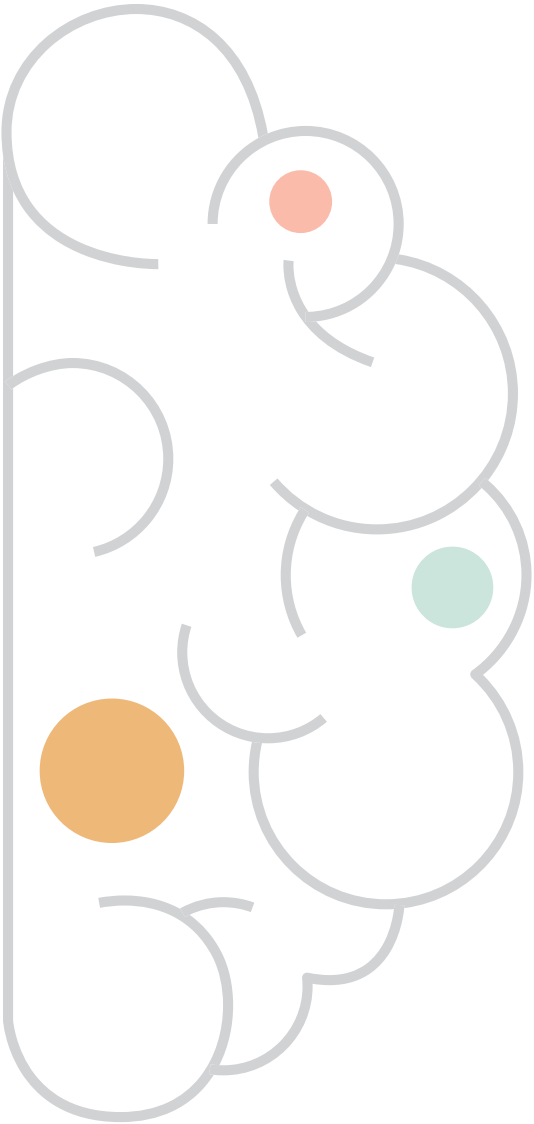
[b]Adjusted for country as a random effect.

[c]Models in the total database were adjusted for study.

dAdjusted for center as a random effect.

CHAPTER 8

Utility-weighted modified Rankin Scale as primary outcome in stroke trials: A simulation study



Simone A. Dijkland
Daphne C. Voormolen
Esmee Venema
Bob Roozenbeek
Suzanne Polinder
Juanita A. Haagsma
Daan Nieboer
Vicky Chalos
Albert J. Yoo
Jennifer Schreuders
Aad van der Lugt
Charles B.L.M. Majoie
Yvo B.W.E.M. Roos
Wim H. van Zwam
Robert J. van Oostenbrugge
Ewout W. Steyerberg
Diederik W.J. Dippel
Hester F. Lingsma
on behalf of the MR CLEAN investigators

Stroke 2018; 49(4): 965–971

Abstract

Background and purpose: The utility-weighted modified Rankin Scale (UW-mRS) has been proposed as a new patient-centered primary outcome in stroke trials. We aimed to describe utility weights for the mRS health states and to evaluate the statistical efficiency of the UW-mRS to detect treatment effects in stroke intervention trials.

Methods: We used data of the 500 patients enrolled in the MR CLEAN (Multicenter Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in the Netherlands). Utility values were elicited from the EuroQol Group 5-Dimension Self-Report Questionnaire assessed at 90 days after inclusion, simultaneously with the mRS. Utility weights were determined by averaging the utilities of all patients within each mRS category. We performed simulations to evaluate statistical efficiency. The simulated treatment effect was an odds ratio of 1.65 in favor of the treatment arm, similar for all mRS cutoffs. This treatment effect was analyzed using 3 approaches: linear regression with the UW-mRS as outcome, binary logistic regression with a dichotomized mRS (0–1/2–6, 0–2/3–6, and 0–4/5–6), and proportional odds logistic regression with the ordinal mRS. The statistical power of the 3 approaches was expressed as the proportion of 10,000 simulations that resulted in a statistically significant treatment effect ($p \leq 0.05$).

Results: The mean utility values (SD) for mRS categories 0 to 6 were: 0.95 (0.08), 0.93 (0.13), 0.83 (0.21), 0.62 (0.27), 0.42 (0.28), 0.11 (0.28), and 0 (0), respectively, but varied substantially between individual patients within each category. The UW-mRS approach was more efficient than the dichotomous approach (power 85% versus 71%) but less efficient than the ordinal approach (power 85% versus 87%).

Conclusions: The UW-mRS as primary outcome does not capture individual variation in utility values and may reduce the statistical power of a randomized trial.

Introduction

The modified Rankin Scale (mRS) is the most widely used primary outcome measure in trials for acute stroke interventions.[1,2] The mRS is an ordinal scale ranging from 0 (no symptoms) to 6 (death) measuring the degree of disability or dependence in everyday life.[3] Previously, dichotomizing the mRS into dead or dependent (mRS, 3–6) versus independent (mRS, 0–2) was common, but this results in a reduction in statistical power to detect relevant treatment effects.[4] Therefore, statistical approaches preserving the ordinal nature of outcome measures, such as proportional odds logistic regression, have been recommended for stroke and other neurological disorders.[1,5-8]

Currently, the importance of incorporating quality of life (QoL) in outcome analysis in stroke trials is increasingly recognized.[9-11] For the mRS to reflect both treatment effect and patient perception, the utility-weighted mRS (UW-mRS) has been proposed and used as primary end point.[2,12,13] In the UW-mRS, utilities based on the EuroQol Group 5-Dimension Self-Report Questionnaire (EQ-5D-3L) values are assigned to the mRS health states. Two prior studies reported utility weights for the mRS health states: 1 representing the values of patients and 1 representing the values of clinicians. The utility weights that were proposed for the UW-mRS are based on these 2 studies.[12] Compared with the ordinal mRS, the UW-mRS showed similar statistical power to detect treatment effects in empirical data in a wide range of stroke trials.[12] However, because in empirical data, the true treatment effect is unknown, the only valid method to assess statistical power is simulation.

We aimed to describe utility weights for the mRS health states and to evaluate the statistical efficiency of the UW-mRS to detect treatment effects in stroke trials.

Methods

Study population

We used individual patient data of the 500 patients enrolled in the MR CLEAN (Multicenter Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in the Netherlands). MR CLEAN was a phase III, multicenter randomized clinical trial, designed to evaluate whether intra-arterial treatment (within 6 hours of symptom onset) plus usual care would be more effective than usual care alone in patients with acute ischemic stroke and a proximal arterial occlusion in the anterior cerebral circulation. The primary outcome was the mRS at 90 days, and the secondary outcome was the EQ-5D-3L at 90 days. In MR CLEAN, ethics approval was obtained from the local institutional review boards of the participating centers, and written informed consent was obtained from patients or legal representatives before randomization.[14]

Modified Rankin Scale

The mRS is a measure of functional outcome after stroke, evaluating the degree of disability or dependence in daily life. The scale is derived from clinical assessment by a trained nurse or a physician

and consists of 7 grades ranging from 0 (no symptoms) to 6, with 5 indicating severe disability and 6 indicating death. A score of ≤2 indicates functional independence.[4]

Utilities

Utilities represent preferences for mRS health states and range from 0 (death) to 1 (perfect health). Utility values of poor outcome categories might even be negative, indicating that they are valued worse than death.[15] In MR CLEAN, utility values were elicited using the EQ-5D-3L responses of patient, proxy, or healthcare provider assessed at 90 days after inclusion, simultaneously with the mRS. The EQ-5D-3L consists of 5 dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) with 3 levels each (no problems, some problems, and extreme problems), thus defining 243 (3^5) distinct health states.[16] Converting the EQ-5D-3L responses into utility values was done according to the Dutch tariff—a countryspecific value set established based on the time trade-off method.[17] Patients who died before the follow-up interviews at 90 days received a utility value of zero. The utility values ranged from −0.33 to 1.00. We determined utility weights for each mRS category by averaging the derived utilities (including the negative values) of all patients within each mRS health state (eg, the utility weight for mRS=1 is the average of the utilities of all patients with mRS=1). Additionally, we matched the utility values proposed by Chaisinanunkul et al,[12] who collapsed mRS 5 to 6 by assigning a utility weight of zero to both categories, to our mRS values.

Simulations for statistical efficiency

Statistical efficiency was evaluated based on simulations that utilized the MR CLEAN database. For a single simulation, 500 patients were sampled at random with replacement. For each patient, the predicted probability of each possible outcome on the 7-point ordinal mRS was modeled as a function of the baseline covariates. These covariates were identical to those in MR CLEAN and included age, stroke severity (National Institutes of Health Stroke Scale) at baseline, time from stroke onset to randomization, status with respect to previous stroke, atrial fibrillation, diabetes mellitus, and occlusion of the internal carotid artery terminus (yes/no).[14]

Using these estimated probabilities, an actual outcome in terms of an mRS or UW-mRS was simulated. Treatment (yes/no) was randomly assigned, and the simulated treatment effect was an odds ratio (OR) of 1.65 (β=0.5) in favor of the treatment arm, similar for all mRS cutoffs. We also evaluated a scenario with no treatment effect, by simulating a treatment effect of OR=1.0 (β=0). During this process, samples of 500 subjects were generated representing 250 patients from the control group and 250 from the intervention group, with a known treatment effect. This was then repeated 10,000×.

The data were analyzed by 3 different statistical approaches. First, we dichotomized the 90-day mRS in 3 different ways of favorable versus unfavorable outcome: 0 to 1 versus 2 to 6, 0 to 2 versus 3 to 6, and 0 to 4 versus 5 to 6. The treatment effect on the dichotomized mRS was determined using binary logistic regression. Second, we used proportional odds logistic regression for analysis of the treatment effect on the ordinal mRS. We fitted a proportional odds logistic regression model with the 7-point ordinal mRS scale as outcome. The proportional odds model estimates a common OR over all

health state transitions within the mRS. According to the proportional odds assumption, the common OR is an accurate reflection of the overall treatment effect if the ORs are the same for each health state transition. If there is agreement regarding the ordinality of the mRS, the common OR can be interpreted as a summary measure of treatment effect even if the proportional odds assumption is violated.[18] Third, treatment effect on the UW-mRS was analyzed using linear regression, as proposed by Chaisinanunkul et al.[12]

Each of the 3 approaches yielded either a significant (P≤0.05) or a nonsignificant treatment effect (p >0.05, 2 sided). The power (or type 1 error in case of no treatment effect) of each statistical approach was estimated as the proportion of the 10,000 analyses, which resulted in a statistically significant treatment effect.

Associations were expressed as ORs or β with 95% confidence intervals (CIs), averaged over all simulations. All analyses were performed unadjusted and adjusted for the prespecified covariates identical to those mentioned above. Statistical analyses were performed with R software, version 3.3.2 (R Foundation for Statistical Computing, Vienna, Austria). Missing data on time from stroke to randomization (0.4%) and level of vessel occlusion (0.2%) was statistically imputed using simple imputation (replacement by mean or mode, as applicable).

Results

Study population

All 500 participants from the MR CLEAN trial were included in our analysis. The mRS at 90 days was available for all patients. The EQ-5D-3L assessments, and consequently the utility values, were available in 457 patients (including 108 patients who died before follow-up). In 43 patients (8.6%), mRS assessment could not be followed by an EQ-5D-3L assessment. In 192 patients (38%), the EQ-5D-3L was completed by a proxy.

The total study population had a mean age of 65 years (SD, 14 years), and most patients (58%) were men (Table 8.1). The intervention and control groups were similar in terms of baseline and treatment characteristics. The number of patients with poor outcome (mRS, 3–6) at 90 days was lower in the intervention group than in the control group (Figure 8.1).

Utility weights

The mean utility values (SD) for mRS categories 0 to 6 were: 0.95 (0.08), 0.93 (0.13), 0.83 (0.21), 0.62 (0.27), 0.42 (0.28), 0.11 (0.28), and 0 (0), respectively (Table 8.2). We observed substantial variation in utility values within each mRS category (Figure 8.2). Within MR CLEAN, the mean UW-mRS for the intervention group was significantly higher when compared with the control group (Table 8.2).

Outcome analysis in MR CLEAN

Ordinal analysis of the mRS showed improved functional outcomes in favor of the intervention, consistent throughout all categories of the mRS except for death (adjusted common OR, 1.67; 95% CI, 1.21–2.30) (Figure 8.1). The dichotomous approach led to slightly stronger treatment effects for cutoffs mRS 0 to 1 and 0 to 2 (adjusted OR, 2.07 [95% CI, 1.07–4.02] and 2.16 [95% CI, 1.39–3.38], respectively). The fact that the ORs were not equal for the different cutoffs might imply that the proportional odds assumption did not hold perfectly in the empirical data. Linear analysis of the UW-mRS resulted in an adjusted β of 0.086 (95% CI, 0.033–0.131).

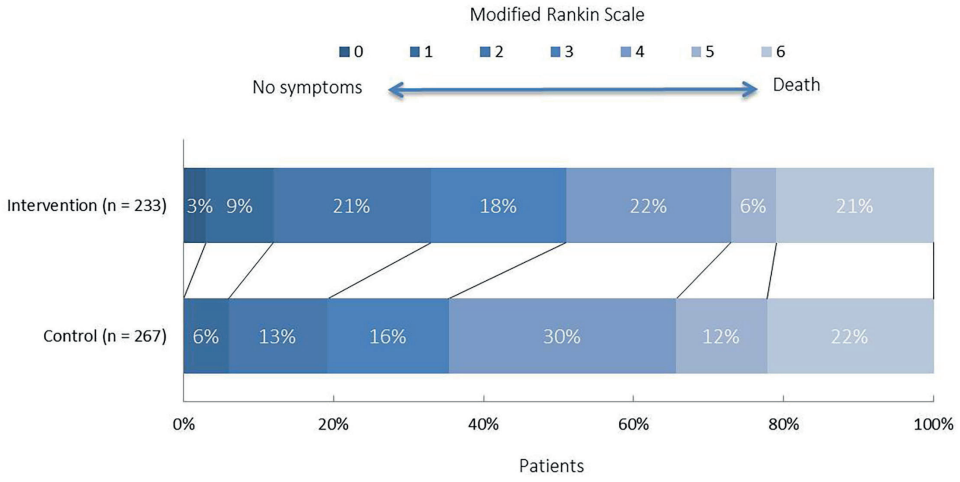


Figure 8.1. Distribution of the modified Rankin Scale at 90 days among intervention and control groups

Table 8.1. Baseline characteristics of the 500 patients in the MR CLEAN trial

	Intervention (n = 233)	Control (n = 267)
Baseline variable	**Intra-arterial treatment plus usual care**	**Usual care**
Age, y; median (IQR)	65.8 (54.5-76.0)	65.7 (55.5-76.4)
Male sex	135 (58%)	157 (59%)
NIHSS score, median (IQR)	17 (14-21)	18 (14-22)
Previous ischemic stroke	29 (12%)	25 (9%)
Atrial fibrillation	66 (28%)	69 (26%)
Diabetes mellitus	34 (15%)	34 (13%)
Prestroke mRS		
0	190 (82%)	214 (80%)
1	21 (9%)	29 (11%)
2	12 (5%)	13 (5%)
>2	10 (4%)	11 (4%)
Treatment with IV alteplase	203 (87%)	242 (91%)
Time from stroke onset to start of IV alteplase, min; median (IQR)	85 (67-110)	87 (65-116)
Occlusion of the internal carotid artery terminus[a]	59 (25%)	75 (28%)
Time from stroke onset to randomization, min; median (IQR)[b]	204 (152-251)	196 (149-266)
Time from stroke onset to groin punction, min; median (IQR)	260 (210-313)	NA

IQR interquartile range; IV, intravenous; mRS, modified Rankin Scale; NA, not applicable; NIHSS, National Institutes of Health Stroke Scale.

[a]No vessel imaging in 1 patient in the control group.

[b]Data were missing for 2 patients in the intervention group.

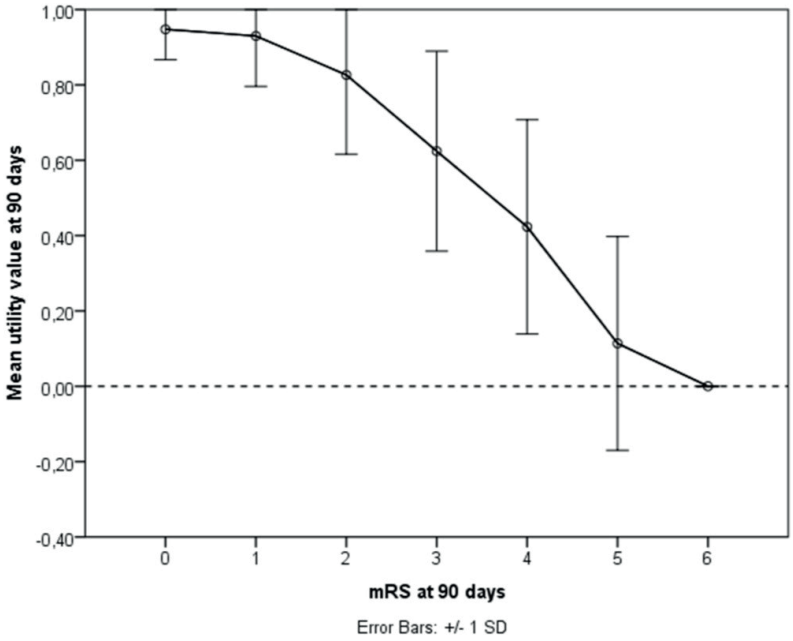


Figure 8.2. Mean EuroQol Group 5-Dimension Self-Report Questionnaire (EQ-5D-3L) utility values per modified Rankin Scale (mRS) category in MR CLEAN (Multicenter Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in the Netherlands)

Simulations

For all 3 prespecified mRS dichotomizations, intra-arterial treatment was positively associated with better outcomes (adjusted OR, 1.66–1.68) (Table 8.3). The estimated treatment effects were similar to the simulated (true) treatment effect of 1.65. When comparing the 3 different mRS cutoffs, the statistical efficiency for the cutoff of mRS 0 to 2 versus 3 to 6 was highest (power 71% versus 62% for mRS 0–1 and 35% for mRS 0–4). This could be explained by an almost equal distribution of patients among both categories for this cutoff (Table 8.3).

Ordinal analysis of the mRS estimated an adjusted treatment effect of common OR=1.66 (95% CI, 1.41–1.95) (Table 8.3), similar to the dichotomous approach. However, the ordinal approach was statistically more efficient (power 87% versus 71%).

Linear regression analysis of the UW-mRS estimated an adjusted beneficial treatment effect of β=0.075 (95% CI, 0.027–0.125) (Table 8.3). The UW-mRS approach was statistically less efficient in detecting treatment effects compared with the ordinal approach (power 85% versus 87%). Matching the utilities of Chaisinanunkul et al to the mRS values in MR CLEAN led to similar results (Table 8.2 and Table 8.3). However, the assumptions of the linear model were not met because there was non-normality of the residuals (Appendix 8.A).

In the simulations without a treatment effect, a proportion of false-positives (type 1 error) of around 5% was estimated for all 3 statistical approaches (data not shown).

Table 8.2. Mean utility values per mRS category and mean UW-mRS in MR CLEAN and the study by Chaisinanunkul et al

	No. of patients MR CLEAN	Mean (SD)	Chaisinanunkul et al[12], mean utility values
mRS			
0	7	0.95 (0.08)	1.00
1	36	0.93 (0.13)	0.91
2	84	0.83 (0.21)	0.76
3	87	0.62 (0.27)	0.65
4	133	0.42 (0.29)	0.33
5	45	0.11 (0.28)	0.00
6	108	0.00	0.00
UW-mRS			
Overall	500	0.45 (0.32)	0.40
Intervention group	233	0.50 (0.33)[a]	0.46
Control group	267	0.41 (0.31)	0.36

MR CLEAN, Multicenter Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in the Netherlands; mRS, modified Rankin Scale; UW, utility weighted.

[a]Mean utility for the intervention group vs control group within MR CLEAN: P=0.002 (Mann-Whitney *U* test).

Table 8.3. Univariable and multivariable estimated treatment effects in simulations (n=500)

		SE	Power		SE	Power
Binary logistic regression	Univariable OR (95% CI)[a]			Multivariable OR (95% CI)[a b]		
0-1 vs 2-6		0.205	56%		0.226	62%
0-1 (n = 146)	1.54 (1.29-1.83)			1.67 (1.08-2.61)		
2-6 (n = 354)	Reference			Reference		
0-2 vs 3-6		0.181	63%		0.203	71%
0-2 (n = 272)	1.51 (1.30-1.74)			1.66 (1.12-2.48)		
3-6 (n = 228)	Reference			Reference		
0-4 vs 5-6		0.303	32%		0.326	35%
0-4 (n = 448)	1.58 (1.21-2.07)			1.68 (0.89-3.19)		
5-6 (n = 52)	Reference			Reference		
Proportional odds logistic regression	Univariable OR (95% CI[a]			Multivariable OR (95% CI)[a b]		
mRS at 90 d	1.53 (1.34-1.75)	0.159	76%	1.66 (1.41-1.95)	0.163	87%
Linear regression	Univariable β (95% CI)[a]			Multivariable β (95% CI)[a b]		
UW-mRS with MR CLEAN utilities	0.075 (0.020-0.131)	0.028	76%	0.075 (0.027-0.125)	0.025	85%
UW-mRS with utilities from Chaisinanunkul et al	0.076 (0.020-0.133)	0.029	75%	0.077 (0.026-0.128)	0.026	84%

CI, confidence interval; MR CLEAN, Multicenter Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in the Netherlands; mRS, modified Rankin Scale; NIHSS, National Institutes of Health Stroke Scale; OR, odds ratio; UW, utility weighted.

[a]Simulated treatment effect β=0.5 (OR, 1.65).

[b]Adjusted for age, NIHSS at baseline, time from stroke onset to randomization, status with respect to previous stroke, atrial fibrillation, diabetes mellitus, and occlusion of the internal carotid artery terminus (yes/no).

Discussion

We evaluated the UW-mRS—a recently proposed patient-centered outcome measure in stroke. Our study, based on a Dutch stroke intervention trial, showed that the UW-mRS does not capture the individual variation in utility values within each mRS category. Moreover, our simulations revealed that the UW-mRS approach was more efficient in detecting treatment effects than dichotomous analysis of the mRS but less efficient than the ordinal approach.

Widely used functional outcome measures in stroke intervention trials, such as the mRS, have been extensively studied concerning their feasibility in measuring disability after stroke.[19,20] Nevertheless, more attention has recently been aimed at incorporating patient-reported QoL in stroke outcome measures.[10,11]

As part of this trend, the UW-mRS has been proposed as a new primary patient-centered outcome measure in acute stroke intervention trials. In empirical data, the UW-mRS was equally statistically efficient in detecting treatment effects compared with ordinal analysis of the mRS.[12] Based on that study, the UW-mRS was recently used as the primary end point in the DAWN trial (Diffusion-Weighted Imaging or Computerized Tomography Perfusion Assessment With Clinical Mismatch in the Triage of Wake Up and Late Presenting Strokes Undergoing Neurointervention With Trevo),[13] and it is expected that more trials will follow. However, the study by Chaisinanunkul et al was only based on analyses of empirical sets of data. Because the true treatment effect in empirical data is unknown and different treatment effects on different outcome measures could be caused by random variation, the only valid method to assess the power of a statistical approach is a simulation study, as we performed.

Intuitively, patient-centered outcomes, such as the UW-mRS, are clinically useful because they concern patientreported measures combined with the perception of the general public. These outcomes reflect patient perception and respect the nonequality of health state transitions on an ordinal scale. Nevertheless, averaging utility values for each mRS category does not reflect individual valuation of these health states: all patients within 1 mRS category receive the same utility weight, irrespective of their own valuation of this health state (Figure 8.2). So, the UW-mRS is in fact a revaluation of the mRS. Moreover, the utility distribution with mRS=5 being worse than death for some patients does not support collapsing mRS categories 5 to 6 as proposed by Chaisinanunkul et al. To reflect true individual valuation of health states, QoL instruments should rather be used as outcome. However, utility values derived from the EuroQol Group 5-Dimension Self-Report Questionnaire may not cover the full range of limitations relevant to patients with stroke[21] and may, therefore, overestimate QoL in this group. An alternative would be to use utility values derived from QoL instruments designed specifically for patients with neurological disorders, such as Neuro-QoL.[22] Nevertheless, because QoL depends on many external factors, it might introduce noise, making it less suitable as a primary outcome measure.[23,24]

Our simulations revealed that the UW-mRS is not as statistically efficient as ordinal analysis of the mRS and may, therefore, cause a reduction in statistical power when used in randomized trials. Chaisinanunkul et al[12] analyzed the UW-mRS with a t test, implying a continuous outcome variable. We used linear regression, which is a comparable approach but allows for multivariable analysis. In theory,

linear analysis is expected to be more efficient than ordinal analysis when the assumptions of the linear model are met. A linear model assumes that the errors between observed and predicted values, that is, the residuals of the regression, are normally distributed. In our analyses, however, we found non-normality of the residuals of the linear model for the UW-mRS. Because the UW-mRS remains a scale with 7 outcome categories, the assumption of normally distributed residuals can never be met. Non-normality of the residuals might cause bias because of underestimation of the standard error.

Therefore, the actual power of the UW-mRS approach will be even <85%. Ordinal analysis also makes an assumption (the proportional odds assumption), but it should be noted that the assumption of a normal distribution of the residuals in a linear model is more difficult to fulfill than the assumption of ordinality in proportional odds analyses. In line with theoretical expectations, the UW-mRS showed to be exactly as efficient as the mRS when it was analyzed with a proportional odds model (data not shown).

Defining a beneficial treatment effect in terms of the UW-mRS, and, therefore, clinical interpretability, might be difficult. Treatment effect on the UW-mRS scale is expressed as a difference in mean UW-mRS between treatment and control groups.[12] This difference can be converted into quality-adjusted life-years (QALYs) gained or lost by a certain treatment.[12,25] The QALY measure assumes that a year of life lived in perfect health is worth 1 QALY, and a year of life lived in a state less than perfect health is worth <1 QALY, proportional to its utility value (QALY=years of life×utility). QALYs can be used to calculate cost-effectiveness to select a certain intervention for funding.[26] Also, the QALY measure has been argued to be more intuitive to patients (healthy life-years gained) and, therefore, to improve communication of treatment effects.[12,25] However, when not converted into QALYs, treatment effects expressed as utility differences remain difficult to interpret. Moreover, clinicians and researchers are now used to working with the (common) OR.

Ordinal outcome scales are also used in other neurological disorders besides stroke. Examples are the Glasgow Outcome Scale in traumatic brain injury and the Guillain-Barre syndrome disability score in Guillain-Barre syndrome.[6,7,27] These ordinal outcomes could be transformed to patient-centered outcomes using utility values, similar to the UW-mRS. For
randomized trials in patients with other neurological diseases, such as traumatic brain injury and Guillain-Barre syndrome, our study might, therefore, also implicate that ordinal analysis
should remain the gold standard.

Our study has several strengths. The simulation study was based on data from the MR CLEAN trial, with relatively broad inclusion criteria.[14] As such, our findings should be generalizable to future stroke trials. Furthermore, simulation is the most adequate method to evaluate statistical power. Also, we used utility values derived using the recommended time trade-off method, which should be less prone to bias compared with other elicitation methods.[24]

Some limitations should also be acknowledged. As with all simulation studies, we do not know how far our findings may be extrapolated beyond the modeled situations. For instance, we only simulated a model with a uniform treatment effect across all mRS health state transitions, which, therefore, adheres perfectly to the proportional odds assumption. However, if the proportional odds assumption

would be violated, and treatment effect would not be uniform across the different outcome categories, ordinal analysis would still be the most efficient (6). Nevertheless, further validation of our results is required. Finally, we used the EuroQol Group 5-Dimension Self-Report Questionnaire assessed at 90 days after inclusion, which reflects neither short-term QoL nor the final health state. A better reflection of patient perception could be achieved by calculating QALYs based on multiple QoL measurements in 1 patient. Nevertheless, the aim of this study is not to describe QoL but to evaluate efficiency in detecting treatment effects.

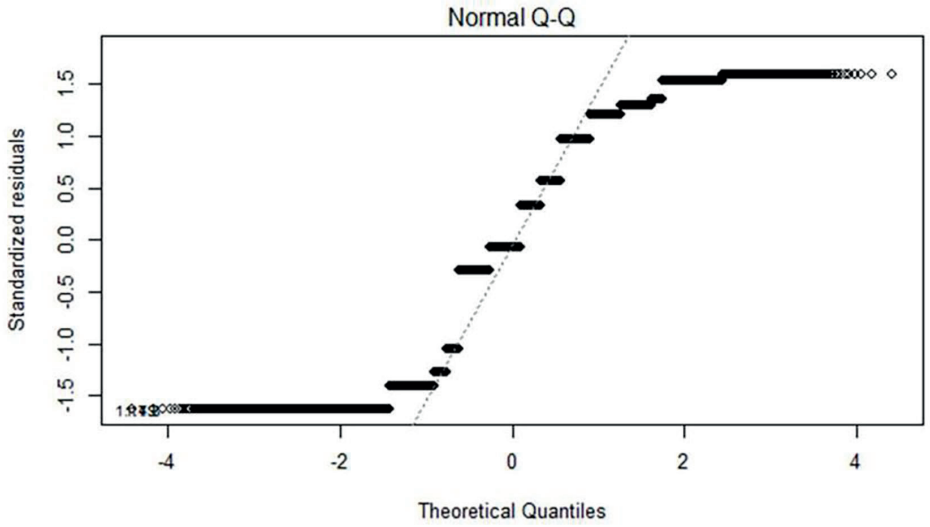
In conclusion, the UW-mRS has been received as a promising new patient-centered outcome in stroke research. However, the UW-mRS does not capture individual variation in utilities within each mRS health state. Also, interpretation of treatment effect on the UW-mRS scale might be more challenging than was first suggested. Finally, clinicians and researchers should be aware of the reduction in power compared with ordinal analysis of the mRS when they use the UW-mRS as outcome measure in acute stroke intervention trials. More thorough evaluation of the UW-mRS in terms of its added value, analytic approach, and interpretation is required.

References

1. Lees KR, Bath PM, Schellinger PD, Kerr DM, Fulton R, Hacke W, et al; European Stroke Organization Outcomes Working Group. Contemporary outcome measures in acute stroke research: choice of primary outcome measure. *Stroke.* 2012;43:1163–1170.
2. Lees KR, Khatri P; STAIR IX Collaborators. Stroke Treatment Academic Industry Roundtable recommendations for individual data pooling analyses in stroke. *Stroke.* 2016;47:2154–2159.
3. van Swieten JC, Koudstaal PJ, Visser MC, Schouten HJ, van Gijn J. Interobserver agreement for the assessment of handicap in stroke patients. *Stroke.* 1988;19:604–607.
4. Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ.* 2006;332:1080.
5. Bath PM, Gray LJ, Collier T, Pocock S, Carpenter J; Optimising Analysis of Stroke Trials (OAST) Collaboration. Can we improve the statistical analysis of stroke trials? Statistical reanalysis of functional outcomes in stroke trials. *Stroke.* 2007;38:1911–1915.
6. McHugh GS, Butcher I, Steyerberg EW, Marmarou A, Lu J, Lingsma HF, et al. A simulation study evaluating approaches to the analysis of ordinal outcome data in randomized controlled trials in traumatic brain injury: results from the IMPACT Project. *Clin Trials.* 2010;7:44–57.
7. Roozenbeek B, Lingsma HF, Perel P, Edwards P, Roberts I, Murray GD, et al; IMPACT (International Mission on Prognosis and Clinical Trial Design in Traumatic Brain Injury) Study Group; CRASH (Corticosteroid Randomisation After Significant Head Injury) Trial Collaborators. The added value of ordinal analysis in clinical trials: an example in traumatic brain injury. *Crit Care.* 2011;15:R127.
8. Saver JL. Optimal end points for acute stroke therapy trials: best ways to measure treatment effects of drugs and devices. *Stroke.* 2011;42:2356– 2362.
9. Ali M, Fulton R, Quinn T, Brady M; VISTA Collaboration. How well do standard stroke outcome measures reflect quality of life? A retrospective analysis of clinical trial data. *Stroke.* 2013;44:3161–3165.
10. Carod-Artal FJ, Egido JA. Quality of life after stroke: the importance of a good recovery. *Cerebrovasc Dis.* 2009;27(suppl 1):204–214.
11. Kapoor A, Lanctôt KL, Bayley M, Kiss A, Herrmann N, Murray BJ, et al. "Good Outcome" isn't good enough: cognitive impairment, depressive symptoms, and social restrictions in physically recovered stroke patients. *Stroke.* 2017;48:1688–1690.
12. Chaisinanunkul N, Adeoye O, Lewis RJ, Grotta JC, Broderick J, Jovin TG, et al; DAWN Trial and MOST Trial Steering Committees; Additional Contributors From DAWN Trial Steering Committee. Adopting a patient-centered approach to primary outcome analysis of acute stroke trials using a utility-weighted modified Rankin Scale. *Stroke.* 2015;46:2238–2243.
13. Nogueira RG, Jadhav AP, Haussen DC, Bonafe A, Budzik RF, Bhuva P, et al. Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct. *N Engl J Med.* 2018;378:11–21.
14. Berkhemer OA, Fransen PS, Beumer D, van den Berg LA, Lingsma HF, Yoo AJ, et al; MR CLEAN Investigators. A randomized trial of intraarterial treatment for acute ischemic stroke. *N Engl J Med.* 2015;372:11–20.
15. Patrick DL, Starks HE, Cain KC, Uhlmann RF, Pearlman RA. Measuring preferences for health states worse than death. *Med Decis Making.* 1994;14:9–18.

16. EuroQol Group. EuroQol–a new facility for the measurement of healthrelated quality of life. *Health Policy.* 1990;16:199–208.
17. Lamers LM, McDonnell J, Stalmeier PF, Krabbe PF, Busschbach JJ. The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. *Health Econ.* 2006;15:1121–1132.
18. Senn S, Julious S. Measurement in clinical trials: a neglected issue for statisticians? *Stat Med.* 2009;28:3189–3209.
19. Harrison JK, McArthur KS, Quinn TJ. Assessment scales in stroke: clinimetric and clinical considerations. *Clin Interv Aging.* 2013;8:201–211.
20. Quinn TJ, Dawson J, Walters MR, Lees KR. Reliability of the modified Rankin Scale: a systematic review. *Stroke.* 2009;40:3393–3395.
21. Schreuders J, van den Berg LA, Fransen PS, Berkhemer OA, Beumer D, Lingsma HF, et al; MR CLEAN Investigators. Quality of life after intraarterial treatment for acute ischemic stroke in the MR CLEAN trial- Update. *Int J Stroke.* 2017;12:708–712.
22. Gershon RC, Lai JS, Bode R, Choi S, Moy C, Bleck T, et al. Neuro- QOL: quality of life item banks for adults with neurological disorders: item development and calibrations based upon clinical and general population testing. *Qual Life Res.* 2012;21:475–486.
23. Cnossen MC, Polinder S, Vos PE, Lingsma HF, Steyerberg EW, Sun Y, et al. Comparing health-related quality of life of Dutch and Chinese patients with traumatic brain injury: do cultural differences play a role? *Health Qual Life Outcomes.* 2017;15:72.
24. Post PN, Stiggelbout AM, Wakker PP. The utility of health states after stroke: a systematic review of the literature. *Stroke.* 2001;32:1425–1429.
25. Broderick JP, Adeoye O, Elm J. Evolution of the modified Rankin Scale and its use in future stroke trials. *Stroke.* 2017;48:2007–2012.
26. Prieto L, Sacristán JA. Problems and solutions in calculating qualityadjusted life years (QALYs). *Health Qual Life Outcomes.* 2003;1:80.
27. Hughes RA, Newsom-Davis JM, Perkin GD, Pierce JM. Controlled trial prednisolone in acute polyneuropathy. *Lancet.* 1978;2:750–753.

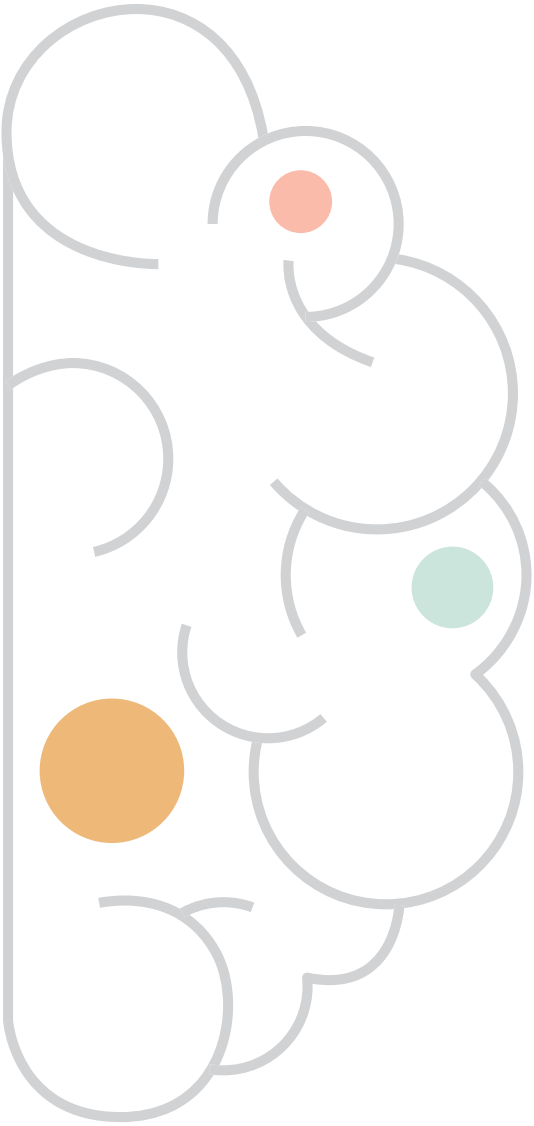
Appendix



Appendix 8.A. Q-Q plot to test normality of the residuals of the UW-mRS in simulations

Legend: Univariable linear model with UW-mRS as outcome and treatment effect as variable. (Standardized) residuals are the errors between observed and predicted values in a model. Theoretical quantiles are the residuals as theoretically expected when they are normally distributed. In a Q-Q plot, the residuals are normally distributed when they fall on the dashed line.

CHAPTER 8.1



Response by Dijkland et al to letter regarding article, "Utility-weighted modified Rankin Scale as primary outcome in stroke trials: A simulation study"

Simone A. Dijkland
Hester F. Lingsma
Diederik W.J. Dippel

Stroke 2018; 49(12): e338

In Response:

We thank Berry et al for starting this interesting discussion to critically assess the utility-weighted modified Rankin Scale (UW-mRS) as outcome measure in stroke intervention trials.[1] Before responding to their comments, we want to point out that we were surprised by the description of our analyses as misleading because it suggests deliberate tampering with results. We are grateful for the opportunity to counter the concerns raised and expect that our arguments will convince the readership of stroke, and hopefully Berry et al, that this qualification is entirely inappropriate.

Berry et al based their conclusion about the advantage of the UW-mRS over the ordinal mRS on a slight gain in statistical power with multinomial analysis. This statistical approach is fundamentally flawed because it ignores the ordering of the mRS categories. The corresponding test statistic is the $\chi 2$ and its P value tests differences in distributions between mRS categories, independent of how these categories are valued. Therefore, the category death might as well be renamed blue, and the additional utility weights are useless. In addition, multinomial regression of the UW-mRS yields 1 odds ratio for each category, highly limiting interpretation of the overall treatment effect. Although Berry et al promote this multinomial approach in their letter, in the DAWN trial (Diffusion-Weighted Imaging or Computerized Tomography Perfusion Assessment With Clinical Mismatch in the Triage of Wake Up and Late Presenting Strokes Undergoing Neurointervention With Trevo), they used an entirely different approach to analyze the UW-mRS: a (Bayesian normal dynamic) linear model.[2]

The mRS analyzed with proportional odds (PO) logistic regression does facilitate interpretation of the overall treatment effect. The PO model assumes a similar treatment effect across all cutoffs of the scale. However, the dependence of the PO model on this proportionality assumption should not be aggravated. As stated in our article, if there is agreement on ordinality of the mRS, the common odds ratio can be interpreted as a summary

measure of treatment effect even if the PO assumption is violated.[1,3] Therefore, testing for the PO assumption is redundant.

We strongly disagree that assigning health values to the different mRS categories is a feature of the UW-mRS. As clearly substantiated in our article, the UW-mRS does not capture the individual variation in utilities within each mRS health state and does not add new information: it still consists of 7 ordered categories.[1] Measuring quality of life in stroke trials is very important but should be done at individual level accounting for variation between patients. In contrast with the remark by Berry et al, a treatment effect was observed on the EuroQol Group 5-Dimension Self-Report Questionnaire (EQ-5D) in the MR CLEAN trial

(Multicenter Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in the Netherlands) after including the deceased patients.[1,4]

In conclusion, after the success of the DAWN trial, it might seem appealing to use the UW-mRS as primary outcome in future stroke trials and Berry et al are clearly advocating their approach. However, we should not refrain from critically studying its added value in terms of statistical accuracy and interpretability. As this added value appears to be absent, we still recommend analyzing the mRS

with PO logistic regression as a primary outcome measure in stroke trials. Individual variation in quality of life should be measured as a secondary outcome using the EQ-5D or disease-specific instruments.

Acknowledgments

We thank E.W. Steyerberg and D. Nieboer for their helpful comments.

References

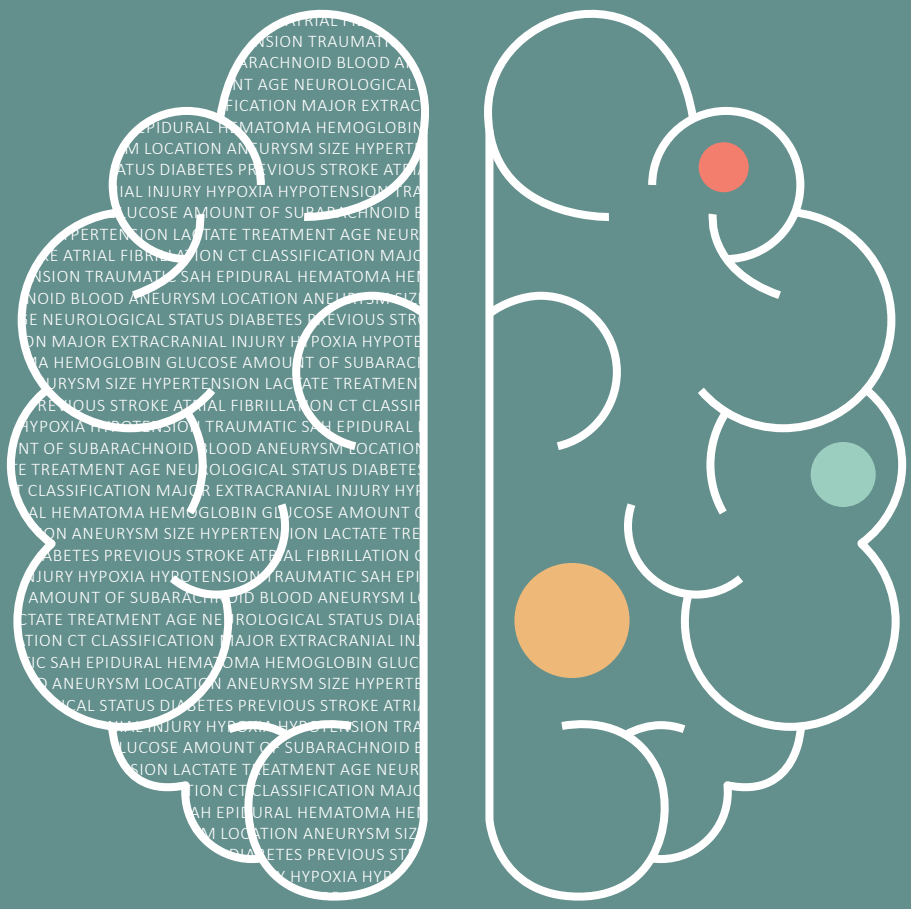
1. Dijkland SA, Voormolen DC, Venema E, Roozenbeek B, Polinder S, Haagsma JA, et al; MR CLEAN Investigators. Utility-weighted modified Rankin scale as primary outcome in stroke trials: a Simulation Study. *Stroke.* 2018;49:965–971.
2. Nogueira RG, Jadhav AP, Haussen DC, Bonafe A, Budzik RF, Bhuva P, et al; DAWN Trial Investigators. Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct. *N Engl J Med.* 2018;378:11–21.
3. Senn S, Julious S. Measurement in clinical trials: a neglected issue for statisticians? *Stat Med.* 2009;28:3189–3209.
4. Schreuders J, van den Berg LA, Fransen PS, Berkhemer OA, Beumer D, Lingsma HF, et al; MR CLEAN Investigators. Quality of life after intraarterial treatment for acute ischemic stroke in the MR CLEAN trial-Update. *Int J Stroke.* 2017;12:708–712.

Reference to letter by Berry et al:

Berry S, Nogueira RG, Saver JL. Letter by Berry et al. Regarding Article, "Utility-Weighted Modified Rankin Scale as Primary Outcome in Stroke Trials". *Stroke.* 2018;49(12):e337.

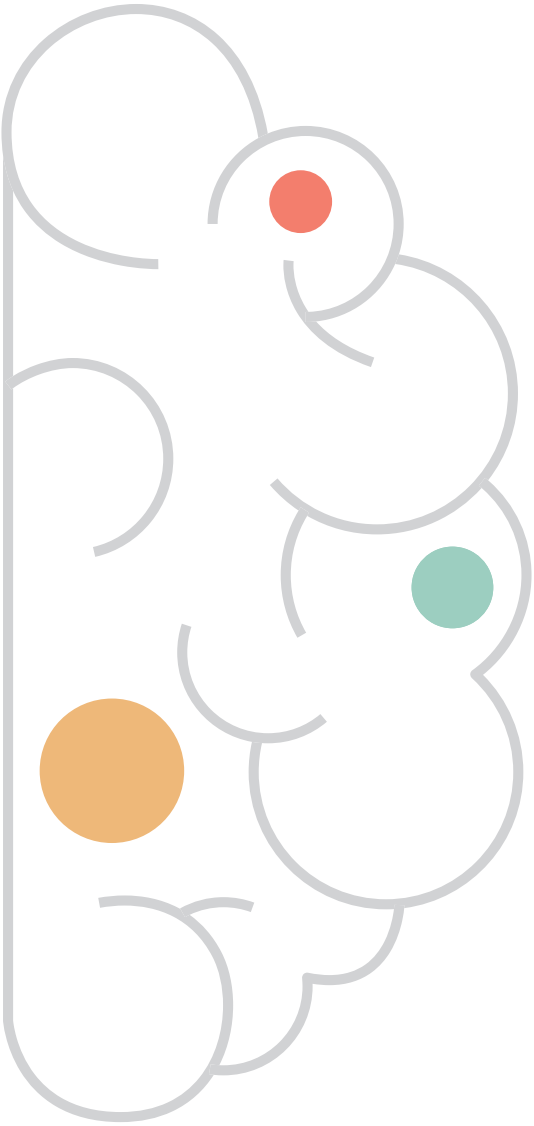
PART IV

DISCUSSION



CHAPTER 9

General discussion



General discussion

The main objective of this thesis was to identify patients at high risk for poor outcome after acute neurological diseases and to enhance knowledge on outcome variation and statistical efficiency of new outcome measures. An overview of the main findings for the five specific research questions posed in Chapter 1 can be found in Box 9.1. In this chapter, the main findings will be discussed separately for outcome prediction and outcome analyses, followed by implications for clinical practice and policy, and recommendations for future research.

Box 9.1. Overview of main findings per research question.

1. What characteristics are associated with poor outcome after acute neurological diseases?
Similar to previous studies, we observed that the main characteristics that are independently associated with poor outcome after acute neurological diseases are age and neurological status at hospital admission.

2. What is the methodological quality of existing prognostic models in acute neurological diseases?
We identified a large number of external validation studies of prognostic models in moderate and severe traumatic brain injury. However, there are still opportunities for improvement of the methodological quality of existing prognostic models for functional outcome after acute neurological diseases. For instance, bootstrapping techniques were infrequently used at internal validation and the importance of model calibration is often underestimated.

3. Do these models provide reliable predictions for patients in specific clinical settings?
Providing reliable predictions for patients with acute neurological diseases in a specific clinical setting remains challenging, and model performance across different settings is highly variable. This may be problematic when intending to apply prognostic models in clinical practice.

4. What are the differences in clinical outcomes between patients with aSAH in a range of international hospitals, and can differences be explained by variation in case-mix?
We observed between-hospital variation in clinical outcomes after aneurysmal subarachnoid hemorrhage. Random effects analyses revealed that between-hospital differences could not be explained by random variation, patient characteristics and timing of aneurysm treatment.

5. What is the statistical efficiency of new outcome measures for acute neurological diseases?
Simulations showed that the utility-weighted modified Rankin Scale (UW-mRS), a recently proposed patient-centered outcome measure for ischemic stroke, may reduce the power of clinical trials in detecting treatment effects. Further, the UW-mRS could complicate interpretation of trial results.

Prediction in acute neurological diseases

Characteristics associated with poor outcome

A first step in prognostic research is to identify characteristics associated with the outcome of interest. To facilitate early classification of disease severity and inclusion of patients in clinical trials, prognostic factor research is often focused on characteristics that can be obtained early in the disease course, e.g. at hospital admission.

In line with previous literature,[1, 2] we observed that age and the National Institutes of Health Stroke Scale (NIHSS) are the main drivers of prognosis after ischemic stroke (Chapter 3). In aneurysmal subarachnoid hemorrhage (aSAH), age and the World Federation of Neurological Surgeons (WFNS) grade had the strongest association with mortality (Chapter 5). Age and WFNS are often included in prognostic models for functional outcome after aSAH.[3, 4] The most frequently included predictors in prognostic models for functional outcome after moderate and severe traumatic brain injury (TBI) were age, the full Glasgow Coma Scale (GCS) or its motor component, and pupillary reactivity (Chapter 4). These characteristics combined explain 35% of the variance in outcomes after moderate and severe TBI.[5] These findings indicate that similarities seem to exist between ischemic stroke, aSAH and TBI in terms of prognosis. Age and neurological status at hospital admission are essential for adequate identification of patients at high risk for poor clinical outcomes after acute neurological diseases.

Several other admission characteristics were associated with poor clinical status after acute neurological diseases (Figure 9.1). Brain computed tomography (CT) characteristics related to the severity of intracranial lesions, such as the amount of subarachnoid blood in aSAH or presence of subdural or epidural hematoma in TBI, are also predictive of poor outcome (Chapters 4 and 5). In ischemic stroke, radiological characteristics such as the Alberta Stroke Program Early CT score (ASPECTS) and collateral score on CT angiography predict clinical outcome and have been included in prognostic models.[6-8] In aSAH, we observed that elevated serum lactate and glucose levels within the first 24 hours after ictus were associated with delayed cerebral ischemia (DCI) related infarction and poor functional outcome (Chapter 2). Elevated glucose levels and other characteristics related to critical illness are also relevant for prognosis in TBI (Figure 9.1). However, only little prognostic information was added when combining these CT and laboratory characteristics with age and neurological status in prognostic models (Chapters 4 and 6). Leaving out additional admission characteristics in prognostic models may, on the other hand, affect individual patient classification.

The additional value of major extracranial injury for models predicting outcome in patients with moderate and severe TBI seems very limited (Chapter 6).[9] This may be explained by an inverse relation of major extracranial injury with TBI severity: the more severe the brain injury, the smaller appears the effect of extracranial injuries on clinical outcome. Also, the association between major extracranial injury and functional outcome may be influenced by patient selection. Major extracranial injury has more prognostic value in studies considering all trauma patients from time of injury, than in studies selecting patients based on presence of TBI who survived the early stage.[10]

In our study on the associations between early lactate and glucose levels and poor outcome or delayed cerebral ischemia (DCI) after aSAH, lactate and glucose were associated with each of the two

outcome measures after adjustment for patient and imaging characteristics. However, when including both lactate and glucose in the multivariable model, only glucose was independently associated with DCI and only lactate was associated with poor outcome (Chapter 2). This finding emphasizes the importance of adjusting for all relevant prognostic factors when analyzing potential prognostic variables and outcomes.

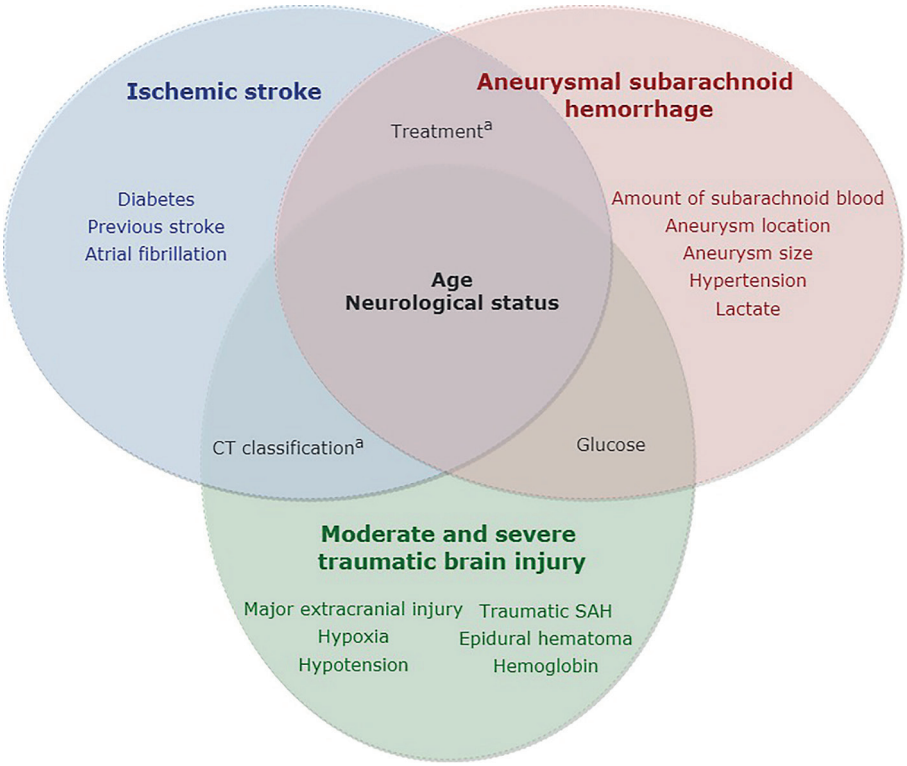


Figure 9.1. Overview of admission characteristics that are independently associated with poor clinical outcome after acute neurological diseases. NIHSS, National Institutes of Health Stroke Scale; WFNS, World Federation of Neurological Surgeons; GCS, Glasgow Coma Scale; SAH, subarachnoid hemorrhage; CT, computed tomography. [a]Not evaluated in this thesis for ischemic stroke.

Timing of predictor and outcome assessment

Variation exists in timing of predictor assessment, which may affect associations between predictors and outcomes. Predictors obtained at hospital admission do not account for changes during the clinical course, such as neurological deterioration due to rebleeding of the aneurysm in aSAH. Assessment of prognostic factors at a later stage may improve outcome prediction. For instance, in line with results from a previous study,[11] we observed that assessment WFNS at time of treatment decision improved discriminative ability of the International Subarachnoid Aneurysm Trial (ISAT) prognostic model for

mortality after aSAH, when compared to assessment of WFNS at admission (Chapter 5). Characteristics obtained during the clinical course are gaining attention, but yielded variable improvement in performance of prognostic models (Chapter 4). Variables obtained at hospital admission facilitate early prediction of clinical outcomes, which may, for instance, facilitate starting the process of referring patients to rehabilitation facilities or nursing homes soon after hospital admission (Chapter 3).

Timing of clinical outcome measurement is also highly variable across different studies evaluating predictors or prognostic models (Chapter 4).[1, 12] Moreover, especially in aSAH, different scales are used to measure functional outcome (e.g. modified Rankin Scale [mRS] or Glasgow Outcome Scale [GOS]) and differences exist in cutoffs for favorable versus unfavorable outcome.[13] Variation in timing of predictor and outcome measurement may cause heterogeneity in predictor effects and performance of prognostic models.

Methodological quality of prognostic models

Although guidelines have been proposed to improve development and reporting of prognostic models, a majority of the published models is not thoroughly developed or validated.[14-17] Several systematic reviews demonstrated opportunities for improvement of methodological quality of prognostic models for functional outcome after ischemic stroke, aSAH and moderate and severe TBI.[1, 3, 12, 18] Some main concerns were the small and selected cohorts used for model development, complete approach to handling of missing data, limited use of bootstrapping techniques for internal validation, and the lack of external validation studies.

Our systematic review on prognostic models in moderate and severe TBI demonstrated a good trend towards external validation of existing prognostic models. Within one decade, 31 prognostic models were externally validated 149 times (Chapter 4). Also, regression analyses were most frequently used for development of new models, which is in principle the preferred method for outcome prediction in TBI (Chapter 4.1). However, methodological quality of prognostic models could still be improved. For instance, bootstrapping techniques for internal validation were only applied in 25% of the developed models. Additionally, model calibration at external validation (i.e. agreement between observed and predicted outcome rates) was only assessed graphically in half of the validations (54%) (Chapter 4). Poor methodological quality of prognostic models may reduce reliability of predictions for patients in specific clinical and research settings. Therefore, recommendations on model development and validation remain current and relevant for future studies.

We provided examples on the development and validation of prognostic models for outcomes after acute neurological diseases. An overview of the main do's and don'ts in prognostic modeling resulting from this thesis is provided in Table 9.1. At model development, the specification and coding of predictors for the model is preferably based on literature and expert opinion (as done in Chapter 3).[12, 14, 19] Reliable estimation of model parameters requires sufficient sample size and is ideally performed with logistic regression analyses (Chapters 4 and 4.1).[19, 20] Concerning the handling of missing data, complete case analysis is still regularly performed, although multiple imputation has been advocated for prognostic research (Chapter 4.1). Finally, dichotomization of predictor and outcome variables

causes loss of information.[21] Continuous predictors should therefore rather be included in the model as such, and ordinal or continuous outcome measures should be analyzed with proportional odds logistic regression or linear regression, respectively.[20]

Before application in clinical practice, prognostic models should be internally and externally validated.[22] Evaluating model performance directly in the derivation cohort (i.e. apparent validation) may cause optimistic estimates of model performance. Random splitting of the original sample into a derivation and validation cohort (i.e. split-sample validation) is an inefficient approach (Chapter 4.1).[14, 23] Therefore, recommended methods for internal validation are cross-validation or bootstrap resampling. With cross-validation, a prognostic model is developed on a part of the derivation cohort and validated on the remaining patients. This process is repeated until all patients have been used for model validation, and model performance is estimated over all validations.[19] A 10-fold cross-validation uses 90% of the derivation sample for development with validation at 10%; repeated 10 times.[14] In the bootstrap procedure, random samples with replacement are drawn from the derivation cohort, with sample size equal to that of the original cohort. The modeling steps are repeated in each of the bootstrap samples, and performance of the constructed models is additionally evaluated in the original cohort. The difference in performance (i.e. optimism) is subtracted from the apparent performance to indicate the expected model performance for future patients similar to the derivation cohort (Chapter 3).[14, 19]

External validation is important to judge the generalizability and transportability of prognostic models to new populations, based on model discrimination and calibration.[14, 24] The area under the receiver operating characteristic curve (AUC) is almost always used to report discrimination between patients with and without the outcome of interest (Chapter 4). The AUC ranges between 0.5 (no discrimination) to 1 (perfect discrimination). Calibration is ideally assessed with a calibration graph, in which a 45-degree line with calibration slope 1 and intercept 0 indicates perfect agreement between observed and predicted outcome rates. In the current literature on prognostic models, the importance of model calibration is often underestimated (Chapter 4). Adequate model calibration is however crucial for adequately informing patients about their risks, and for decision support.[14, 25] Ideally, a prognostic model should be refitted on development and validation cohort combined to obtain the best estimates of the regression coefficients (Chapter 3).[20]

Table 9.1 .Overview of do's and don'ts for development and validation of prognostic models for acute neurological diseases resulting from this thesis[a]

DON'T	DO
Model development	
Use small cohorts	Use a dataset with at least 100 events, or at least 10 patients with the outcome of interest for each candidate predictor (10 events per variable)
Perform complete case analysis	Multiple imputation
Dichotomize predictors or outcomes	Include continuous predictors as such (e.g. age); analyze ordinal or continuous outcomes with proportional odds logistic regression or linear regression, respectively
Use decision trees	Logistic regression
Model validation - internal	
Use apparent or split-sample approaches	Use bootstrapping techniques or cross-validation
Model validation - external	
Stop after internal validation	External validation whenever possible
Forget model calibration	Assess model calibration graphically (with intercept and slope) in addition to model discrimination

[a]Statements on all relevant recommendations for conducting and reporting prognostic research have been published[15, 16]

Validation and updating of promising existing models is preferred over development of new models.[26] Especially in the field of acute neurological diseases, where the main predictors of clinical outcome have been confirmed (Chapter 5.1). In our systematic review on prognostic models for functional outcome after moderate and severe TBI, we found that model discrimination at external validation is often good, but providing reliable predictions for individual patients (i.e. model calibration) remains challenging (Chapter 4). This was also observed in external validation studies of the ISAT, International Mission on Prognosis and Analysis of Clinical Trials (IMPACT) and Corticoid Randomisation After Significant Head injury (CRASH) prognostic models included in this thesis (Chapters 5 and 6) (Table 9.2).[27-29] Model performance at external validation can be affected by several factors, including patient selection and definition and measurement of predictors and outcomes.[30] In the external validation studies included in this thesis, the ISAT and IMPACT prognostic models showed improved model discrimination in broader, less selected cohorts (Chapters 5 and 6). Also, performance of prognostic models for moderate and severe TBI is highly variable across different settings (Chapter 4). These findings underscore the need for model validation and updating before implementation in research or clinical practice.

For models aimed at clinical decision making, a decision analysis is required beyond discrimination and calibration.[14] A decision analysis evaluates the consequences of applying the prognostic model in a specific setting, by balancing the relative importance of the benefits (true positives) and harms (false positives) of a clinical decision based on the model.[31] If a decision analysis shows potential, the final step is to perform an impact study. This includes evaluating whether care provided based on the model is

better than usual care, and determining the applicability in daily routine according to clinicians. These evaluation steps have not been performed for the prognostic models evaluated in this thesis, but are important to clarify whether or not a prognostic model can be used in clinical practice.[26]

Table 9.2 .Overview of external validity of the ISAT model for aneurysmal subarachnoid hemorrhage and the IMPACT and CRASH models for traumatic brain injury evaluated in this thesis (Chapters 5 and 6, broadest patient selections).

Prognostic model	Outcome	Discrimination AUC (95% CI)	Calibration Observed versus predicted outcome rates
Aneurysmal SAH (n = 307)			
ISAT[a]	Mortality at 60 days	0.82 (-)	30.6% vs. 17.7%
Traumatic brain injury (GCS 3-14, n = 1742)			
IMPACT core	Mortality at 6 months	0.85 (0.83 to 0.87)	15% vs. 37%
	Unfavorable outcome at 6 months (GOSE 1-4)	0.80 (0.78 to 0.82)	43% vs. 46%
IMPACT extended	Mortality at 6 months	0.88 (0.86 to 0.90)	15% vs. 34%
	Unfavorable outcome at 6 months (GOSE 1-4)	0.84 (0.82 to 0.86)	43% vs. 47%
IMPACT lab	Mortality at 6 months	0.88 (0.86 to 0.90)	15% vs. 29%
	Unfavorable outcome at 6 months (GOSE 1-4)	0.84 (0.82 to 0.86)	43% vs. 44%
CRASH basic	Mortality at 14 days	0.86 (0.83 to 0.88)	15% vs. 15%
	Unfavorable outcome at 6 months (GOSE 1-4)	0.82 (0.80 to 0.84)	43% vs. 43%
CRASH CT	Mortality at 14 days	0.88 (0.86 to 0.90)	15% vs. 33%
	Unfavorable outcome at 6 months (GOSE 1-4)	0.84 (0.82 to 0.86)	43% vs. 56%

[a]Original model with World Federation of Neurological Surgeons grade at hospital admission.

AUC, area under the receiver operating characteristic curve; CI, confidence interval; SAH, subarachnoid hemorrhage; GCS, Glasgow Coma Scale; IMPACT, International Mission on Prognosis and Analysis of Clinical Trials; GOSE, Glasgow Outcome Scale Extended; CRASH, Corticoid Randomisation After Significant Head injury; CT, computed tomography.

Applications of prognostic models

Some main (potential) applications for prognostic models in research settings and clinical practice can be distinguished (Table 9.3).

Table 9.3. Overview of potential applications of prognostic models for acute neurological diseases

Applications	Example in acute neurological diseases
Research	
Inform clinical trial design	Prognostic targeting and covariate adjustment based on established predictors for clinical outcome for TBI could reduce required sample size for clinical trials[32, 33]
Provide insight in possible (and modifiable) causes for clinical outcomes	Markers of critical illness (e.g. elevated glucose levels, hypoxia, hypotension) are included in prognostic models for moderate and severe TBI, giving insight in the systemic consequences of the brain injury
Clinical	
Assist clinicians with communication regarding the disease course of individual patients	Use of predicted probabilities provided by the IMPACT model to inform a relative of a patient with severe TBI in the intensive care unit on the chance of recovery within the next 6 months
Guide therapeutic decisions for individual patients	Prognostic model to select patients with atrial fibrillation at high risk for ischemic stroke for preventive treatment with anticoagulants[34]
Reduction of heterogeneity in prognostic estimates across physicians	Use of a prognostic score influenced the prognostic estimates made by physicians in patients with intracerebral hemorrhage (e.g. more optimistic prognostic estimates when the score indicated better prognosis)[35]
Improving quality of care	Use established prognostic models for case-mix adjustment in analyses on variation in clinical outcomes after aSAH and TBI across hospitals and countries (Chapter 7)[36, 37]

TBI, traumatic brain injury; aSAH, aneurysmal subarachnoid hemorrhage; IMPACT, International Mission on Prognosis and Analysis of Clinical Trials; CRASH, Corticoid Randomisation After Significant Head injury.

In the field of acute neurological diseases, clinicians are involved in many publications on prognostic models. This implies that outcome prediction is considered relevant for clinical practice. Most studies on clinical prognostic models claim that the model predictions can assist clinicians with risk communications concerning the disease course for (the relatives of) individual patients.[38] Predictions of clinical outcomes for individual patients with TBI based on clinical expertise alone may be too pessimistic.[39] Prognostic models can incorporate a broad range of characteristics relevant for the subsequent clinical outcome, and may antagonize these views. Moreover, clinicians have indicated that the use of prognostic models may assist in reducing heterogeneity in prognostic estimates across physicians.[40, 41]

However, in spite of the vast body of clinical research on prognostic modeling, large-scale implementation of prognostic models in care for individual patients is not established. Multiple barriers for implementation of prognostic models in clinical practice, in general as well as specifically for acute neurological diseases, can be identified (Box 9.2):

- An important barrier is the lack of knowledge about existence of prognostic models among clinicians. For the IMPACT prognostic models, questionnaires revealed that only 50% of clinicians knew about their existence.[40, 42]
- Clinicians also indicated that they considered prognostic models to be research tools, designed for populations and not for clinical decision making in individual patients.[40] Further, prognostic estimates are difficult to interpret and are therefore seen as misleading for individual patients and relatives.[26, 40] The mistrust in prognostic models does however not seem to apply to all models and is not related to model performance, because several models with only moderate performance are frequently used in clinical practice.[34, 43-45]
- Factors related to usability of a prognostic model may also limit application in clinical practice.[26] Prognostic models often require computer support to calculate predicted probabilities of clinical outcomes.[38] Moreover, in routine clinical practice, characteristics are considered that are not included in prognostic models (e.g. comorbidities). Also, given the variety in neurological and imaging grading scales in acute neurological diseases, measurement of predictors may differ across clinical settings.
- Changes in clinical practice, e.g. availability of new treatments or innovations in imaging techniques may change prognosis of individual patients.[26] For instance, the Dutch Stroke Score, was developed on data that was collected before the introduction of IAT (Chapter 3). Another example is the historic nature of the IMPACT development data. However, an effect of changes in clinical practice is not always observed at external validation, as shown by the adequate performance of the IMPACT and CRASH models throughout the past decade (Chapters 4 and 6).
- Finally, heterogeneity of the disease course and/or lack of evidence-based treatment options may complicate application of a prognostic model in clinical practice. In acute neurological diseases, outcomes may be different for patients with similar clinical and radiological characteristics.[46] Additionally, limited evidence exists on treatment for complications related to aSAH and TBI that occur in the acute phase (e.g. DCI or raised intracranial pressure). A survey among clinicians revealed that the prognostic estimates provided by the IMPACT calculator have only little impact on (aggressiveness of) care of patients with TBI.[42] Further, a RCT showed that documenting prognosis in the intensive care setting had only limited impact on treatment decisions.[47] Prognostic models that do affect diagnostic or therapeutic decisions are more likely to be implemented in guidelines and/or clinical practice.[34, 43-45] The prognostic models evaluated in this thesis (ISAT, IMPACT and CRASH) are not (yet) recommended for clinical decision making.

Box 9.2. Overview of barriers for application of prognostic models in clinical practice and potential solutions

Clinician related

- Lack of awareness regarding availability of prognostic models

  Solution: improve 'marketing', e.g. by designing and promoting online tools and apps

- Mistrust in prognostic estimates for individual patients:
  - Prognostic models are considered research tools designed for populations
  - Interpretation of prognostic estimates for individual patients is challenging

  Solution: provide clear guidance on the intended use of prognostic models and their risk estimates in clinical practice

Model related

- Limitations concerning usability:
  - Computer support required to calculate predicted probabilities of outcome
  - Characteristics that are considered in routine practice are not included in prognostic models
  - Ambiguous predictors or differences in predictor measurement across hospitals

  Solution: focus on simple models with easily obtainable characteristics that are preferably not subject to measurement variation across different settings

- Changes in clinical practice over time are not accounted for

  Solution: externally validate prognostic models in more recent data

Disease related

- Heterogeneous disease course
- Lack of evidence-based treatment options

Because of these barriers, prognostic models are currently more seen as tools that may support clinicians to increase their confidence in outcome prognostication, than as crucial for changing prognostic estimates for individual patients based on clinical experience.[41] Addressing the barriers from both research and clinical practice perspectives could enhance application of prognostic models in clinical practice.

Research should focus on simple models with easily obtainable characteristics. Additionally, external validation and updating of prognostic models in recent data is important to address changes in clinical practice and provide reliable predictions for specific settings. In this thesis, examples have been provided for external validation of the ISAT model in a cohort of aSAH patients admitted to the intensive care unit of our hospital, and the IMPACT and CRASH models in a large contemporary cohort of TBI patients across Europe (Chapters 5 and 6). Finally, decision analyses and impact studies should be performed to evaluate feasibility of implementation in clinical practice.

From a clinical perspective, the most fundamental aspect is to create awareness among clinicians regarding availability of prognostic models, and provide clear guidance on the intended use of prognostic models and their risk estimates in clinical practice (as done in Chapter 6). It should be evident that application of prognostic models is merely meant to complement clinical judgement, not

to replace it. Further, when a prognostic model has been externally validated extensively, variation in model performance is commonly observed across different settings. Therefore, validated prognostic models should be implemented only if the model is expected to be applicable to the specific setting and patient.[14]

Outcome analyses in acute neurological diseases

Differences in clinical outcomes between hospitals and countries

Based on random effects modeling, we observed between-center differences in clinical outcomes after aSAH that could not be explained by random variation, patient characteristics, and timing of aneurysm treatment (Chapter 7). Similar differences in clinical outcomes beyond case-mix have also been observed for patients with moderate and severe TBI.[36] Also, other studies have identified differences between hospitals and countries in clinical outcomes after ischemic stroke and aSAH, but used other methodology.[48-50]

After establishing differences in clinical outcomes, the next step would be to relate this variation in clinical outcomes to variation in diagnostic and treatment policies. For TBI, variation in treatment policies was observed based on questionnaires among physicians from multiple centers across Europe.[51-56] One way to evaluate the effect of differences in clinical practice on outcomes is to adjust for structure and process characteristics in random effects models. A decrease in between-center differences in clinical outcomes after adjustment may indicate that these factors affect clinical outcomes. Further, comparative effectiveness research (CER) can be performed to generate evidence on the benefits and harms of health care interventions (e.g. concerning therapeutic policies or organization of care). By providing evidence-based recommendations for best clinical practice at individual and population level, CER has the potential to improve the quality and outcomes of care for patients with acute neurological diseases.[57, 58]

However, the available data should facilitate evaluation of potential causes for variation in clinical outcomes. So far, random effects modeling for between-hospital variation in clinical outcomes was performed on data from the Subarachnoid Hemorrhage International Trialists (SAHIT) repository and the IMPACT database (Chapter 7).[36] Both data sources consist of a combination of multiple RCTs, observational studies and hospital registries.[59, 60] Data collection for the studies included in these repositories was not uniform, making it impossible to combine all data points for the included studies. Moreover, the SAHIT repository and IMPACT database were mainly designed for prognostic research. In this way, meta-analyses on all studies combined are restricted due to loss of valuable data. For instance, we were unable to evaluate whether variation in structure and process characteristics explained some of the between-center differences in clinical outcomes after aSAH. Also, to pool clinical outcomes, we had to dichotomize the outcome scales into favorable versus unfavorable because either the GOS or the mRS was used (Chapter 7).

The National Institute of Health (NIH) and the National Institute of Neurological Disorders and Stroke (NINDS) aim to stimulate more uniform collection, coding and definition of data points for clinical trials in acute neurological diseases through establishing common data elements (CDEs).[61-64]

Standardizing names and definitions of variables and agreement on methods for data framing may facilitate pooling and comparing data from different studies. This may also reduce the variation in measurement of predictors and outcomes that exists in the field of acute neurological diseases.[13, 65] Additionally, large observational cohort studies are required to confirm presence of differences in clinical outcomes in more recent data, and to investigate potential causes with CER. An example is provided by the Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) study.[66] This observational cohort of contemporary TBI patients facilitates adjustment for structure and process characteristics at hospital level in random effects models, and aims to provide recommendations for best clinical practice in TBI based on CER.

Statistical efficiency of new outcome measures

Introduction of new outcome measures is common, with a current emphasis on measures other than functional outcome, such as patient-reported outcome measures (PROMs). PROMs reflect patients' views on their symptoms, functional status and quality of life.[67] New (patient-centered) outcome measures have to meet several requirements that should be evaluated before application in clinical practice and research. One of these requirements concerns validity: the degree to which a new outcome scale measures what we intend to measure.[68] A specific aspect of validity is the statistical efficiency of a new outcome measure to detect treatment effects.

We provided an example on a simulation study evaluating the statistical efficiency of a newly developed PROM for ischemic stroke: the utility-weighted mRS (UW-mRS). Before proper evaluation of statistical efficiency, the UW-mRS was used as a (co-)primary outcome in a clinical trial on IAT in patients with ischemic stroke presenting more than 6 hours after stroke onset.[69] However, our simulation study revealed that the UW-mRS approach was less efficient in detecting treatment effects than ordinal analysis of the mRS (Chapter 8). This finding underscores the importance of evaluating statistical efficiency and interpretability of a new outcome measure before implementation in research or clinical practice.

By deriving mean utility weights for each mRS category, the UW-mRS remains an ordinal scale with 7 categories and does not add new information. Moreover, this approach does not account for individual variation in utilities within each health state of the mRS (Chapters 8 and 8.1). A recent study confirmed that substantial variation exists in utility values between and within mRS categories and over time post-stroke, which is not accounted for by the UW-mRS. Moreover, differences in methods used to derive utility values also cause variability in UW-mRS values,[70] which complicates evaluation and interpretation of treatment effects.

The UW-mRS has been described as an "imperfect solution to an important problem."[70] In TBI, efforts have been made to determine health state preference weights for the GOS. Only few preference weights with highly variable magnitude have been estimated for the different GOS categories. Several factors, such as age and comorbidities, affected the mean EuroQol Group 5-Dimension Self-Report Questionnaire (EQ-5D) utility values per GOS category.[71]

The development of alternative measures that capture both functional outcome and quality of life is difficult,[58] and it remains uncertain whether implementation of these outcome scales in research or clinical practice is achievable given the challenges associated with their validity.

Table 9.4. Pros (+) and cons (-) of using patient-reported versus functional outcome measures

Patient-reported outcome measure		Functional outcome measure	
+	Incorporates patient perception on physical and mental well-being	+	Objective evaluation of treatment effects
+	Reduces floor and ceiling effects	+	Clinicians are used to interpretation of treatment effects on odds or hazard ratio scale
+	Allows for individual variation in clinical outcomes	+	Extensively studied and proven useful
-	May reduce statistical power to detect treatment effects	-	May not be sensitive to subtle changes in clinical status (floor and ceiling effects)
-	May result in false-negative or false-positive treatment effects	-	Does not include all domains relevant for the level of disability
-	Complicates interpretation of treatment effects	-	Subject to interobserver variability
-	Introduces noise because it is affected by external factors		

Use of outcome measures in clinical practice and research

How should we primarily measure outcomes after acute neurological diseases in clinical practice and research settings? Functional outcome measures and PROMs each have their advantages and disadvantages (Table 9.4).

For acute neurological diseases, functional outcome measures (e.g. mRS and GOS) are widely implemented in research settings and clinical practice. Functional outcome measures are simple, have been extensively studied and proven useful in detecting disability after acute neurological diseases.[72, 73] They also facilitate objective evaluation of treatment effects on the odds or hazard ratio scale, which is currently common practice for clinicians and researchers.

Criticism on functional outcome scales is that they are not granular enough to detect subtle but relevant changes in clinical status and do not include all domains that are important for the level of disability.[58] Therefore, these outcome measures may be subject to floor and ceiling effects, meaning that patients may score at the extreme ends of the distribution despite relevant changes in clinical status.[72] Slightly more detailed measures of functional outcome exist, such as the Barthel Index and the Glasgow Outcome Scale Extended (GOSE).[73, 74] Functional outcome scales have also been criticized for interobserver variability.[75]

PROMs are increasingly popular and many studies regarding development of new PROMs or evaluation of their clinical relevance have been published. The main reason that PROMs are strongly advocated, is that they may have the capacity to narrow the gap between patient and physician. They allow for individual variation in clinical outcomes, and are therefore less sensitive to floor and

ceiling effects than functional outcome measures. PROMs have many potential applications, aimed at improving quality of care for individual patients as well as for healthcare systems.[67, 76]

Some important limitations of generic and disease-specific PROMs should however be considered:

- As shown for the UW-mRS, PROMs may reduce statistical power to detect treatment effects in clinical trials (Chapter 8). Lack of statistical efficiency of an outcome measure could cause unnecessary patient inclusion in RCTs, which might cause delay in release of new treatments for acute neurological diseases.
- Further, evaluation of treatment benefit on PROMs may result in false-negative or false-positive treatment effects. For instance, the Multicenter Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in the Netherlands (MR CLEAN) showed a clear benefit of IAT on the ordinal mRS in patients with ischemic stroke based on a proximal arterial occlusion in the anterior cerebral circulation.[77] However, IAT only had a limited effect on quality of life measured with the EQ-5D.[78] Relying on a PROM as primary outcome measure could therefore wrongfully affect clinical management: when quality of life does only marginally improve, why would we still perform IAT in ischemic stroke?
- Also, PROMs may impede interpretation of treatment effects. Concerning the UW-mRS, interpretation of treatment effects expressed as differences in utility values or quality-adjusted life years (QALYs) is quite complex in research and clinical practice (Chapter 8).
- Finally, the advantage of considering multiple domains of daily functioning also comes with a downside: most PROMs depend on external factors (i.e. not related to the specific disease or treatment). This 'noise' introduced by PROMs makes them less appropriate as efficacy measures for treatment (Chapter 8). Disease-specific PROMs may provide a solution in this regard, but a major disadvantage is that these PROMs cannot be compared with other disease groups or population norms.

Therefore, PROMs are quite complex and there is currently not enough evidence to implement them in research and clinical practice as primary outcome measures. Two generic PROMs have been validated for many diseases: the EQ-5D and the Short-Form (SF-36).[79, 80] In aSAH, there is however limited evidence for selection of suitable generic or disease-specific PROMs. None of the available PROMs complied with the standards for validity, and only one PROM was specifically developed for aSAH.[81] For ischemic stroke, there is somewhat more evidence for both generic and stroke-specific PROMs, but the implementation in research settings and clinical practice is still lacking.[82] The Quality of Life after Brain Injury (QOLIBRI) is a promising disease-specific PROM for TBI.[83] Nevertheless, more should be done to evaluate the role of multidimensional outcome measures in TBI research and clinical practice.[58, 84]

In acute neurological diseases, PROMs are relevant since many patients experience impairments that affect both functional status and quality of life. However, PROMs have some important limitations and their impact on clinical practice still needs to be established. Therefore, PROMs are currently not sensitive enough to be used as primary outcome measure in clinical trials or routine clinical practice.[67] For now, treatment effects in clinical trials are still recommended to be analyzed with functional scales

as primary outcome measures (Chapter 8).[58, 85] Individual variation in quality of life can be measured as a secondary outcome with validated generic or disease-specific scales (Chapters 8 and 8.1). Patients' impairments on domains other than functional outcome should also be considered in clinical practice.

Limitations

The development data of the majority of prognostic models evaluated in this thesis mainly originated from clinical trials (Chapter 3-6). The evaluation of between-hospital variation in clinical outcomes after aSAH was also based on data from multicenter RCTs (Chapter 6). For prognostic and outcomes research, observational cohorts are preferred. Moreover, these RCTs were conducted before the introduction of some highly beneficial interventions in ischemic stroke and aSAH (e.g. IAT and aneurysm coiling). These factors limit the generalizability of our results to contemporary patients with ischemic stroke or aSAH. Our findings should therefore be interpreted with caution and should be validated in current settings. However, most ischemic stroke patients receive intravenous alteplase as only treatment (Chapter 3). Further, given the lack of evidence-based treatment options and variation in guidelines for treatment of aSAH, it is expected that between-center differences in clinical outcomes still exist in current clinical practice (Chapter 6).

The prognostic models presented in this thesis consist of characteristics obtained within 24 hours after hospital admission, but do not take into account changes in the clinical course. Therefore, characteristics obtained during the disease course or variables that predict treatment response may also be relevant. In routine clinical practice, variables such as improvement in neurological status and medical comorbidities are usually considered by clinicians, even if no model currently includes such variables. However, prognostic models based on admission characteristics enable early disease classification and timely inclusion of patients in clinical trials. Further, model extension with dynamic predictors has not been widely investigated and yielded variable improvement in model performance (Chapter 4).

Next steps in research and clinical practice

Based on our main findings regarding prediction and outcome analyses in acute neurological diseases and their interpretation, specific recommendations on the next steps in future research and clinical practice can be summarized.

Prediction

- Perform decision-analytic evaluations and impact studies to get an impression of the clinical applicability of existing prognostic models
- Attempt updating of promising existing prognostic models to enhance reliability of predictions for patients in a specific clinical setting

Outcome analyses

- Reduce heterogeneity in definitions and measurement of clinical characteristics and outcomes in acute neurological diseases by standardizing data collection through CDE
- Relate between-hospital differences in clinical outcomes to variation in clinical practice with CER in large datasets that allow for sufficient sample size per hospital to provide recommendations for best clinical practice
- Use functional outcome scales to evaluate treatment effects in research and clinical practice and assess individual variation in quality of life separately with validated scales (e.g. EQ-5D or SF-36)
- Pursue development of multidimensional outcome measures for acute neurological diseases, provided that their statistical efficiency and interpretability are ensured

Overall conclusions

The aim of this thesis was to identify patients at high risk for poor outcome after acute neurological diseases and to enhance knowledge on outcome variation and statistical efficiency of new outcome measures. The core predictors to identify patients with acute neurological diseases at high risk for poor outcome are age and neurological status at hospital admission. Prognostic models are increasingly externally validated, which is a crucial step before we start implementation in clinical practice. Providing reliable predictions for individual patients with acute neurological diseases remains challenging, so validated prognostic models should be applied in addition to clinical experience and only if the model is expected to be applicable to the specific setting and patient.

Variation between hospitals in clinical outcomes after aSAH exists and could not be explained by random variation, case-mix and timing of aneurysm treatment. Our results need to be confirmed in more recent data, ideally a large observational cohort study. Differences in clinical outcomes should be related to practice variation in future (CER) studies, to provide evidence-based recommendations for best clinical practice.

Patient-centered outcome measures may reduce the power of clinical trials in detecting treatment effects, and may complicate interpretation of trial results. An example was provided for the UW-mRS in ischemic stroke. This finding underscores the importance of evaluating statistical efficiency and interpretability of a new outcome measure before using it.

References

1. Veerbeek JM, Kwakkel G, van Wegen EE, Ket JC, Heymans MW. Early prediction of outcome of activities of daily living after stroke: A systematic review. *Stroke*. 2011;42:1482-1488
2. Weimar C, Konig IR, Kraywinkel K, Ziegler A, Diener HC, German Stroke Study Collaborators. Age and national institutes of health stroke scale score within 6 hours after onset are accurate predictors of outcome after cerebral ischemia: Development and external validation of prognostic models. *Stroke*. 2004;35:158-162
3. Jaja BNR, Cusimano MD, Etminan N, Hanggi D, Hasan D, Ilodigwe D, et al. Clinical prediction models for aneurysmal subarachnoid hemorrhage: A systematic review. *Neurocrit Care*. 2013;18:143-153
4. Jaja BNR, Saposnik G, Lingsma HF, Macdonald E, Thorpe KE, Mamdani M, et al. Development and validation of outcome prediction models for aneurysmal subarachnoid haemorrhage: The SAHIT multinational cohort study. *BMJ*. 2018;360:j5745
5. Lingsma HF, Roozenbeek B, Steyerberg EW, Murray GD, Maas AI. Early prognosis in traumatic brain injury: From prophecies to predictions. *Lancet Neurol*. 2010;9:543-554
6. Barber PA, Demchuk AM, Zhang J, Buchan AM. Validity and reliability of a quantitative computed tomography score in predicting outcome of hyperacute stroke before thrombolytic therapy. Aspects study group. Alberta Stroke Programme Early CT Score. *Lancet*. 2000;355:1670-1674
7. Tan IYL, Demchuk AM, Hopyan J, Zhang L, Gladstone D, Wong K, et al. CT angiography clot burden score and collateral score: Correlation with clinical and radiologic outcomes in acute middle cerebral artery infarct. *Am J Neuroradiol*. 2009;30:525
8. Venema E, Mulder MJHL, Roozenbeek B, Broderick JP, Yeatts SD, Khatri P, et al. Selection of patients for intra-arterial treatment for acute ischaemic stroke: Development and validation of a clinical decision tool in two randomised trials. *BMJ*. 2017;357:j1710
9. Roozenbeek B, Lingsma HF, Lecky FE, Lu J, Weir J, Butcher I, et al. Prediction of outcome after moderate and severe traumatic brain injury: External validation of the International Mission on Prognosis and Analysis of Clinical Trials (IMPACT) and Corticoid Randomisation After Significant Head injury (CRASH) prognostic models. *Crit Care Med*. 2012;40:1609-1617
10. van Leeuwen N, Lingsma HF, Perel P, Lecky F, Roozenbeek B, Lu J, et al. Prognostic value of major extracranial injury in traumatic brain injury: An individual patient data meta-analysis in 39,274 patients. *Neurosurgery*. 2012;70:811-818; discussion 818
11. Giraldo EA, Mandrekar JN, Rubin MN, Dupont SA, Zhang Y, Lanzino G, et al. Timing of clinical grade assessment and poor outcome in patients with aneurysmal subarachnoid hemorrhage. *J Neurosurg*. 2012;117:15-19
12. Mushkudiani NA, Hukkelhoven CW, Hernandez AV, Murray GD, Choi SC, Maas AI, et al. A systematic review finds methodological improvements necessary for prognostic models in determining traumatic brain injury outcomes. *J Clin Epidemiol*. 2008;61:331-343
13. Andersen CR, Fitzgerald E, Delaney A, Finfer S. A systematic review of outcome measures employed in aneurysmal subarachnoid hemorrhage (aSAH) clinical research. *Neurocrit Care*. 2019;30:534-541
14. Dijkland SA, Retel Helmrich IRA, Steyerberg EW. Validation of prognostic models: Challenges and opportunities. *J Emerg Crit Care Med*. 2018;2

15. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. *Ann Intern Med*. 2015;162:W1-W73
16. Steyerberg EW, Moons KGM, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis research strategy (PROGRESS) 3: Prognostic model research. *PLoS Med*. 2013;10:e1001381-e1001381
17. Collins GS, de Groot JA, Dutton S, Omar O, Shanyinde M, Tajar A, et al. External validation of multivariable prediction models: A systematic review of methodological conduct and reporting. *BMC Med Res Methodol*. 2014;14:40-40
18. Perel P, Edwards P, Wentz R, Roberts I. Systematic review of prognostic models in traumatic brain injury. *BMC Med Inform Decis Mak*. 2006;6:38
19. Steyerberg EW. Validation of prediction models. *In: Clinical prediction models: A practical approach to development, validation and updating*. New York: Springer, 2009. P. 299-310.
20. Dijkland SA, Dippel DWJ, Lingsma HF. Letter by Dijkland et al regarding article, "Development and validation of a predictive model for functional outcome after stroke rehabilitation: The Maugeri model". *Stroke*. 2018;49:e133
21. Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ*. 2006;332:1080
22. Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol*. 2016;69:245-247
23. Steyerberg EW, Harrell FE, Jr., Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol*. 2001;54:774-781
24. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: Validating a prognostic model. *BMJ*. 2009;338:b605
25. Van Calster B, Vickers AJ. Calibration of risk prediction models: Impact on decision-analytic performance. *Med Decis Making*. 2015;35:162-169
26. Moons KGM, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: Application and impact of prognostic models in clinical practice. *BMJ*. 2009;338:b606
27. Risselada R, Lingsma HF, Bauer-Mehren A, Friedrich CM, Molyneux AJ, Kerr RSC, et al. Prediction of 60 day case-fatality after aneurysmal subarachnoid haemorrhage: Results from the International Subarachnoid Aneurysm Trial (ISAT). *Eur J Epidemiol*. 2010;25:261-266
28. Perel PA, Olldashi F, Muzha I, Filipi N, Lede R, Copertari P, et al. Predicting outcome after traumatic brain injury: Practical prognostic models based on large cohort of international patients. *BMJ*. 2008;336:425-429
29. Steyerberg EW, Mushkudiani N, Perel P, Butcher I, Lu J, McHugh GS, et al. Predicting outcome after traumatic brain injury: Development and international validation of prognostic scores based on admission characteristics. *PLoS Med*. 2008;5:e165; discussion e165
30. Damen J, Debray TPA, Pajouheshnia R, Reitsma JB, Scholten R, Moons KGM, et al. Empirical evidence of the impact of study characteristics on the performance of prediction models: A meta-epidemiological study. *BMJ Open*. 2019;9:e026160
31. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ*. 2016;352:i6

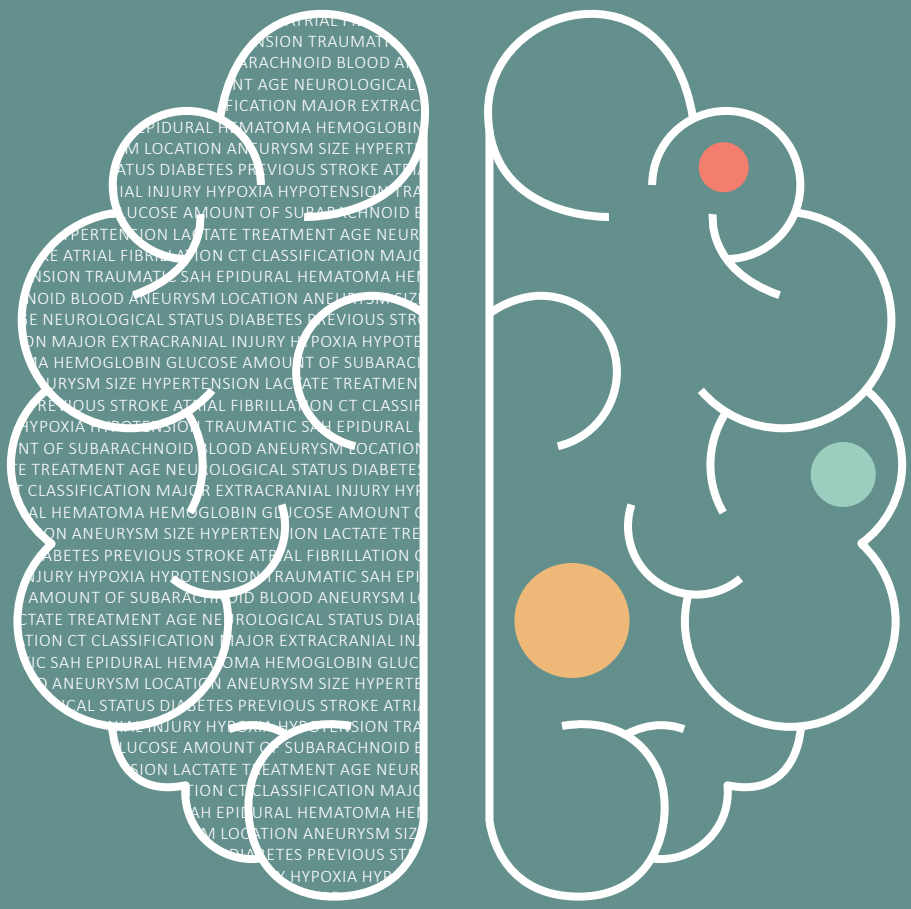
32. Roozenbeek B, Maas AI, Lingsma HF, Butcher I, Lu J, Marmarou A, et al. Baseline characteristics and statistical power in randomized controlled trials: Selection, prognostic targeting, or covariate adjustment? *Crit Care Med*. 2009;37:2683-2690
33. Hernandez AV, Steyerberg EW, Butcher I, Mushkudiani N, Taylor GS, Murray GD, et al. Adjustment for strong predictors of outcome in traumatic brain injury trials: 25% reduction in sample size requirements in the IMPACT study. *J Neurotrauma*. 2006;23:1295-1303
34. Lip GYH, Nieuwlaat R, Pisters R, Lane DA, Crijns HJGM. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: The Euro heart survey on atrial fibrillation. *Chest*. 2010;137:263-272
35. Zahuranec DB, Fagerlin A, Sánchez BN, Roney ME, Thompson BB, Fuhrel-Forbis A, et al. Variability in physician prognosis and recommendations after intracerebral hemorrhage. *Neurology*. 2016;86:1864-1871
36. Lingsma HF, Roozenbeek B, Li B, Lu J, Weir J, Butcher I, 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*. 2011;68:601-608
37. Huijben JA, Wiegers EJA, de Keizer NF, Maas AIR, Menon D, Ercole A, et al. Development of a quality indicator set to measure and improve quality of ICU care for patients with traumatic brain injury. *Crit Care*. 2019;23:95-95
38. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: What, why, and how? *BMJ*. 2009;338:b375
39. Harvey D, Butler J, Groves J, Manara A, Menon D, Thomas E, et al. Management of perceived devastating brain injury after hospital admission: A consensus statement from stakeholder professional organizations. *Br J Anaesth*. 2018;120:138-145
40. Moskowitz J, Quinn T, Khan MW, Shutter L, Goldberg R, Col N, et al. Should we use the IMPACT-model for the outcome prognostication of TBI patients? A qualitative study assessing physicians' perceptions. *MDM Policy Pract*. 2018;3:2381468318757987
41. Hallen SAM, Hootsmans NAM, Blaisdell L, Gutheil CM, Han PKJ. Physicians' perceptions of the value of prognostic models: The benefits and risks of prognostic confidence. *Health Expect*. 2015;18:2266-2277
42. Letsinger J, Rommel C, Hirschi R, Nirula R, Hawryluk GWJ. The aggressiveness of neurotrauma practitioners and the influence of the IMPACT prognostic calculator. *PLoS One*. 2017;12:e0183552-e0183552
43. Lip GYH, Frison L, Halperin JL, Lane DA. Comparative validation of a novel risk score for predicting bleeding risk in anticoagulated patients with atrial fibrillation: The HAS-BLED (Hypertension, Abnormal renal/liver function, Stroke, Bleeding history or predisposition, Labile INR, Elderly, Drugs/alcohol concomitantly) score. *J Am Coll Cardiol*. 2011;57:173-180
44. Wang X, Huang Y, Li L, Dai H, Song F, Chen K. Assessment of performance of the GAIL model for predicting breast cancer risk: A systematic review and meta-analysis with trial sequential analysis. *Breast Cancer Res*. 2018;20:18-18
45. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation*. 1998;97:1837-1847

46. Carter EL, Hutchinson PJ, Kolias AG, Menon DK. Predicting the outcome for individual patients with traumatic brain injury: A case-based review. *Br J Neurosurg*. 2016;30:227-232
47. Turnbull AE, Hayes MM, Brower RG, Colantuoni E, Basyal PS, White DB, et al. Effect of documenting prognosis on the information provided to ICU proxies: A randomized trial. *Crit Care Med*. 2019;47:757-764
48. Bayman EO, Chaloner KM, Hindman BJ, Todd MM, IHAST Investigators. Bayesian methods to determine performance differences and to quantify variability among centers in multi-center trials: The IHAST trial. *BMC Med Res Methodol*. 2013;13:5
49. Gray LJ, Sprigg N, Bath PM, Sorensen P, Lindenstrom E, Boysen G, et al. Significant variation in mortality and functional outcome after acute ischaemic stroke between western countries: Data from the Tinzaparin in Acute Ischaemic Stroke Trial (TAIST). *J Neurol Neurosurg Psychiatry*. 2006;77:327-333
50. Lipsman N, Tolentino J, Macdonald RL. Effect of country or continent of treatment on outcome after aneurysmal subarachnoid hemorrhage. *J Neurosurg*. 2009;111:67-74
51. Cnossen MC, Polinder S, Lingsma HF, Maas AI, Menon D, Steyerberg EW, 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*. 2016;11:e0161367
52. Foks KA, Cnossen MC, Dippel DWJ, Maas A, Menon D, van der Naalt J, 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. *J Neurotrauma*. 2017
53. Huijben JA, van der Jagt M, Cnossen MC, Kruip M, Haitsma IK, Stocchetti N, et al. Variation in blood transfusion and coagulation management in traumatic brain injury at the intensive care unit: A survey in 66 neurotrauma centers participating in the Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury study. *J Neurotrauma*. 2017
54. van Essen TA, den Boogert HF, Cnossen MC, de Ruiter GCW, Haitsma I, Polinder S, et al. Variation in neurosurgical management of traumatic brain injury: A survey in 68 centers participating in the CENTER-TBI study. *Acta Neurochir (Wien)*. 2018
55. Volovici V, Ercole A, Citerio G, Stocchetti N, Haitsma IK, Huijben JA, et al. Intensive care admission criteria for traumatic brain injury patients across Europe. *J Crit Care*. 2019;49:158-161
56. Volovici V, Ercole A, Citerio G, Stocchetti N, Haitsma IK, Huijben JA, et al. Variation in guideline implementation and adherence regarding severe traumatic brain injury treatment: A CENTER-TBI survey study in Europe. *World Neurosurg*. 2019
57. Federal Coordinating Council for Comparative Effectiveness Research. Report to the president and congress. Washington, DC: Department of Health and Human Services. 2009. https://osp.od.nih.gov/wp-content/uploads/FCCCER-Report-to-the-President-and-Congress-2009.pdf. Accessed June 20, 2019.
58. Maas AIR, Menon DK, Adelson PD, Andelic N, Bell MJ, Belli A, et al. Traumatic brain injury: Integrated approaches to improve prevention, clinical care, and research. *Lancet Neurol*. 2017;16:987-1048
59. Jaja BN, Attalla D, Macdonald RL, Schweizer TA, Cusimano MD, Etminan N, et al. The Subarachnoid Hemorrhage International Trialists (SAHIT) repository: Advancing clinical research in subarachnoid hemorrhage. *Neurocrit Care*. 2014;21:551-559

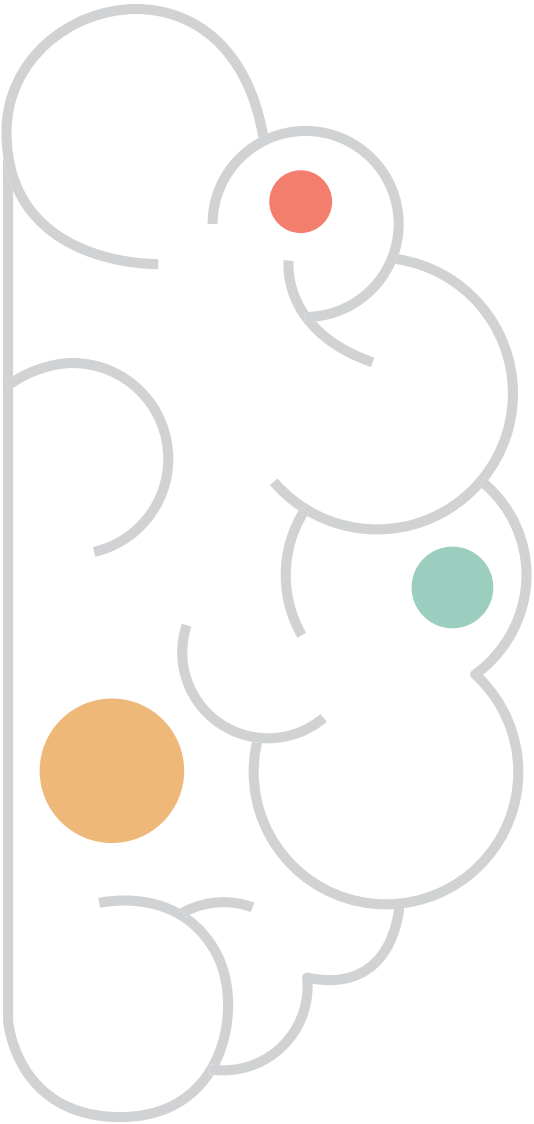
60. Marmarou A, Lu J, Butcher I, McHugh GS, Mushkudiani NA, Murray GD, et al. IMPACT database of traumatic brain injury: Design and description. *J Neurotrauma*. 2007;24:239-250
61. Maas AI, Harrison-Felix CL, Menon D, Adelson PD, Balkin T, Bullock R, et al. Common data elements for traumatic brain injury: Recommendations from the interagency working group on demographics and clinical assessment. *Arch Phys Med Rehabil*. 2010;91:1641-1649
62. Yue JK, Vassar MJ, Lingsma HF, Cooper SR, Okonkwo DO, Valadka AB, et al. Transforming research and clinical knowledge in traumatic brain injury pilot: Multicenter implementation of the common data elements for traumatic brain injury. *J Neurotrauma*. 2013;30:1831-1844
63. Suarez JI, Sheikh MK, Macdonald RL, Amin-Hanjani S, Brown RD, de Oliveira Manoel AL, et al. Common data elements for unruptured intracranial aneurysms and subarachnoid hemorrhage clinical research: A National Institute for Neurological Disorders and Stroke and National Library of Medicine project. *Neurocrit Care*. 2019
64. Saver JL, Warach S, Janis S, Odenkirchen J, Becker K, Benavente O, et al. Standardizing the structure of stroke clinical and epidemiologic research data: The National Institute of Neurological Disorders and Stroke (NINDS) stroke common data element (CDE) project. *Stroke*. 2012;43:967-973
65. Horton L, Rhodes J, Wilson L. Randomized controlled trials in adult traumatic brain injury: A systematic review on the use and reporting of clinical outcome assessments. *J Neurotrauma*. 2018;35:2005-2014
66. Maas AI, Menon DK, Steyerberg EW, Citerio G, Lecky F, Manley GT, et al. Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI): A prospective longitudinal observational study. *Neurosurgery*. 2015;76:67-80
67. Black N. Patient reported outcome measures could help transform healthcare. *BMJ*. 2013;346:f167
68. Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res*. 2013;22:1889-1905
69. Nogueira RG, Jadhav AP, Haussen DC, Bonafe A, Budzik RF, Bhuva P, et al. Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct. *N Engl J Med*. 2017
70. Rethnam V, Bernhardt J, Dewey H, Moodie M, Johns H, Gao L, et al. Utility-weighted modified rankin scale: Still too crude to be a truly patient-centric primary outcome measure? *Int J Stroke*. 2019:1747493019830583
71. Ward Fuller G, Hernandez M, Pallot D, Lecky F, Stevenson M, Gabbe B. Health state preference weights for the Glasgow Outcome Scale following traumatic brain injury: A systematic review and mapping study. *Value Health*. 2017;20:141-151
72. Harrison JK, McArthur KS, Quinn TJ. Assessment scales in stroke: Clinimetric and clinical considerations. *Clin Interv Aging*. 2013;8:201-211
73. Wilson JT, Pettigrew LE, Teasdale GM. Structured interviews for the Glasgow Outcome Scale and the Extended Glasgow Outcome Scale: Guidelines for their use. *J Neurotrauma*. 1998;15:573-585
74. Mahoney FI, Barthel DW. Functional evaluation: The Barthel Index. *Md State Med J*. 1965;14:61-65
75. Quinn TJ, Dawson J, Walters MR, Lees KR. Reliability of the modified Rankin Scale: A systematic review. *Stroke*. 2009;40:3393-3395
76. Nelson EC, Eftimovska E, Lind C, Hager A, Wasson JH, Lindblad S. Patient reported outcome measures in practice. *BMJ*. 2015;350:g7818

77. Berkhemer OA, Fransen PS, Beumer D, van den Berg LA, Lingsma HF, Yoo AJ, et al. A randomized trial of intraarterial treatment for acute ischemic stroke. *N Engl J Med*. 2015;372:11-20
78. Schreuders J, van den Berg LA, Fransen PS, Berkhemer OA, Beumer D, Lingsma HF, et al. Quality of life after intra-arterial treatment for acute ischemic stroke in the MR CLEAN trial – Update. *Int J Stroke*. 2017;12:708-712
79. Euroqol Group – a new facility for the measurement of health-related quality of life. *Health Policy*. 1990;16:199-208
80. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992;30:473-483
81. Nobels-Janssen E, van der Wees PJ, Verhagen WIM, Westert GP, Bartels RHMA, Boogaarts JD. Patient-reported outcome measures in subarachnoid hemorrhage. *Neurology*. 2019;92:1096
82. Mathew R, Lynda L, Linda W, Irene K, Moira K, Anne D, et al. Patient-reported outcome measures (PROMs) for acute stroke: Rationale, methods and future directions. *Stroke*. 2018;49:1549-1556
83. von Steinbuchel N, Wilson L, Gibbons H, Hawthorne G, Hofer S, Schmidt S, et al. Quality of life after brain injury (QOLIBRI): Scale validity and correlates of quality of life. *J Neurotrauma*. 2010;27:1157-1165
84. Nelson Lindsay D, Ranson J, Ferguson Adam R, Giacino J, Okonkwo David O, Valadka Alex B, et al. Validating multi-dimensional outcome assessment using the traumatic brain injury common data elements: An analysis of the track-tbi pilot study sample. *J Neurotrauma*. 2017;34:3158-3172
85. Lees KR, Bath PM, Schellinger PD, Kerr DM, Fulton R, Hacke W, et al. Contemporary outcome measures in acute stroke research: Choice of primary outcome measure. *Stroke*. 2012;43:1163-1170

APPENDICES



APPENDICES



Summary
Samenvatting
Dankwoord
List of publications
PhD portfolio
About the author

Summary

Summary

Part I Introduction

Most treatments and interventions in health care are aimed at optimizing clinical outcomes. Measurement of clinical outcomes may serve different purposes, such as outcome prediction and outcome analyses. Clinical outcome prediction involves the use of prognostic factors or a prognostic model for early identification of patients at high risk for poor functional outcome in a specific clinical setting. This may assist clinicians with treatment decisions, inclusion of patients in randomized clinical trials or benchmarking quality of care. Outcome analyses include examining variation in outcomes across different settings and determining the added value of new outcome measures. Variation in clinical outcomes between hospitals and countries is present in many diseases, but is highly undesirable when caused by differences in management. Gaining insight in outcome differences across settings with random effects modeling creates the opportunity to evaluate practice variation. Further, a trend exists towards new outcome measures incorporating both functional outcome and quality of life (patient-reported outcome measures [PROMs]). New outcome measures should be statistically efficient to obtain reliable estimates of treatment effects in clinical trials, and should also facilitate interpretation of treatment effects.

Ischemic stroke, aneurysmal subarachnoid hemorrhage (aSAH) and traumatic brain injury (TBI) are acute neurological diseases with a heterogeneous disease course that are often associated with poor functional outcomes and reduced quality of life. This stimulates measurement of clinical outcomes in terms of prognosis, variation across settings and new assessment methods.

The overall aim of this thesis was to identify patients at high risk for poor outcome after acute neurological diseases (**Part II**) and to enhance knowledge on outcome variation and statistical efficiency of new outcome measures (**Part III**).

Specific research questions were:

1. What characteristics are associated with poor outcome after acute neurological diseases?
2. What is the methodological quality of existing prognostic models in acute neurological diseases?
3. Do these models provide reliable predictions for patients in specific clinical settings?
4. What are the differences in clinical outcomes between patients with aSAH in a range of international hospitals, and can these differences be explained by variation in case-mix?
5. What is the statistical efficiency of new outcome measures for acute neurological diseases?

Part II Prediction

In **Chapter 2**, we performed a two-center retrospective cohort study in 285 aSAH patients to evaluate the associations between maximum serum lactate and glucose levels measured within the first 24 hours after onset of aSAH, and delayed cerebral ischemia (DCI) or poor functional outcome (modified Rankin Scale [mRS] 4-6). After adjustment for patient and imaging characteristics, lactate and glucose

were independently associated with DCI and poor outcome with odds ratios between 1.17 and 1.56. Lactate and glucose were strongly related, and after inclusion of both parameters in the multivariable model, only glucose was independently associated with DCI and only lactate was associated with poor outcome. The role of early glucose and lactate in prognostic models for outcome after aSAH and the associated pathophysiological mechanisms (e.g. relation with sympathetic stress) should be evaluated in future studies.

Chapter 3 presents the development and validation of prognostic models for the Barthel Index (BI) at hospital discharge (Dutch Stroke Score [DSS]-discharge) and mRS at three months (DSS-3 months) after ischemic stroke. We analyzed individual patient data from three clinical trials, of which one served as development cohort (n=1227) and two as external validation cohorts (n=1589 and n=2107). The DSS-discharge included age, National Institutes of Health Stroke Scale (NIHSS) and diabetes as predictors, and showed reasonable discrimination at internal validation (area under the receiver operating characteristic curve [AUC] of 0.76). The DSS-3 months consisted of age, NIHSS, diabetes, previous stroke and atrial fibrillation, and yielded ordinal AUCs around 0.70 at internal and external validation. However, model calibration showed that the DSS-3 months overestimated the proportion of poor outcome (mRS 3-6) in the validation cohorts. If further validated, the DSS may assist clinicians with efficient stroke unit discharge planning.

Chapter 4 provides a systematic overview of contemporary prognostic models for functional outcome after moderate and severe TBI. We included 58 studies describing 67 unique prognostic models. The most frequently included predictors were age, the full Glasgow Coma Scale or its motor component, and pupillary reactivity. We observed that existing prognostic models are increasingly externally validated (149 external validations of 31 models). However, methodological quality of prognostic models for moderate and severe TBI could still be improved. For instance, model calibration was reported graphically in only half of the validations (54%). The International Mission on Prognosis and Analysis of Clinical Trials (IMPACT) and Corticoid Randomisation After Significant Head injury (CRASH) models were externally validated most extensively (n=91) and showed adequate discriminative ability across different settings (mean weighted AUCs 0.77-0.82). However, the reliability of predictions was highly variable. This illustrates the need for continuous external validation and updating of prognostic models over time and to specific clinical settings.

External validation studies of prognostic models for functional outcome after aSAH and moderate and severe TBI were presented in **Chapters 5 and 6**, respectively. The International Subarachnoid Aneurysm Trial (ISAT) model for 60-day mortality after aSAH consists of age, World Federation of Neurological Surgeons grade, Fisher grade and aneurysm size as predictors. The ISAT model showed good discriminative ability in a retrospective cohort of 307 aSAH patients admitted to the intensive care unit of the Erasmus University Medical Center (AUC 0.82). The IMPACT and CRASH models for mortality and unfavorable outcome (Glasgow Outcome Scale Extended 1-4) showed good discrimination in a prospective observational cohort of 1742 moderate and severe TBI patients across Europe (AUCs 0.80-0.88). However, providing predictions for patients with aSAH and TBI in specific clinical settings remains difficult.

Chapters 4.1 and 5.1 addressed some core methodological concepts of model development and validation in acute neurological diseases. Adequate reporting of prognostic research, taking into consideration the available evidence in the field, is crucial for the reliability and reproducibility of prognostic models. Importantly, the core clinical predictors of functional outcome after acute neurological diseases have been established. Therefore, validation and updating of promising existing prognostic models is preferred over the development of new models.

Part III Outcome analyses

In **Chapter 7**, random effects modeling was performed to evaluate between-hospital and between-country variation in clinical outcomes after aSAH. We analyzed data from 5972 aSAH patients treated at 179 centers in 20 countries included in a large international repository. We found substantial between-hospital variation, that could not be explained by random variation, patient characteristics or timing of aneurysm treatment (adjusted median odds ratio 1.21, 95% confidence interval 1.11-1.44). There were no statistically significant between-country differences. Identifying individual hospitals that performed better or worse than others was difficult, because the individual random effect estimates were subject to substantial uncertainty. The data were relatively old, and we were unable to evaluate the causality between observed outcome differences and variation in treatment policies (other than timing of aneurysm treatment) or quality of care. Therefore, we could not provide recommendations for current clinical practice.

A simulation study evaluating the statistical efficiency of the utility-weighted mRS (UW-mRS), a recently proposed patient-centered outcome measure in ischemic stroke, was presented in **Chapter 8**. The simulations were based on individual patient data of 500 patients enrolled in a multicenter clinical trial evaluating the effectiveness of intra-arterial treatment in ischemic stroke. Linear analysis of the UW-mRS was less efficient in detecting treatment effects than ordinal analysis of the mRS (power 85% versus 87%). Moreover, the UW-mRS does not capture individual variation in utilities within each mRS category, and may impede interpretation of treatment effects. These findings underscore the importance of studying the statistical efficiency and interpretability of new patient-centered outcome measures, as outlined in **Chapter 8.1**.

Part IV Discussion

The main objective of this thesis was to identify patients at high risk for poor outcome after acute neurological diseases and to enhance knowledge on outcome variation and statistical efficiency of new outcome measures. We found that age and neurological status on admission are the main characteristics associated with poor clinical outcome after ischemic stroke, aSAH and moderate and severe TBI. Prognostic models are increasingly externally validated, which is a crucial step before we start implementation in clinical practice. However, providing reliable predictions for individual patients with acute neurological diseases remains challenging. Further, we found substantial variation between hospitals in clinical outcomes after aSAH, which could not be explained by random variation, case-mix and timing of aneurysm treatment. Finally, a simulation study evaluating the statistical efficiency of the

UW-mRS showed that patient-centered outcome measures may reduce the power of clinical trials in detecting treatment effects, and may complicate interpretation of trial results.

Although the vast body of prognostic research in acute neurological diseases implies that outcome prediction is considered relevant for clinical practice, prognostic models for ischemic stroke, aSAH and TBI are rarely implemented. Several barriers for use of prognostic models in clinical practice can be identified, including the lack of knowledge among clinicians on existence and use of prognostic models, and the few evidence-based treatment recommendations in acute neurological diseases. The clinical applicability of existing prognostic models should be examined with decision-analytic evaluations and impact studies.

Differences in clinical outcomes between hospitals should, as a next step, be related to variation in clinical practice with comparative effectiveness research to provide recommendations for best clinical practice, preferably based on large observational cohort study. To enable this, heterogeneity in definitions and measurement of clinical characteristics and outcomes in acute neurological diseases needs to be reduced by standardizing data collection through common data elements.

Functional outcome measures have been widely implemented and provide objective evaluation of treatment effects. PROMs, on the other hand, incorporate patient perception on physical and mental well-being and allow for individual variation in clinical outcomes. However, the development of PROMs is difficult, and it remains uncertain whether implementation of these outcome scales in research or clinical practice is achievable given the challenges associated with their validity. Therefore, we recommend to use functional outcome scales to evaluate treatment effects and to assess individual variation in quality of life separately with validated scales.

Samenvatting

Samenvatting

Deel I Introductie

De meeste behandelingen en interventies in de gezondheidszorg zijn gericht op het verbeteren van klinische uitkomsten. Het meten van klinische uitkomsten is belangrijk voor diverse doeleinden, waaronder predictie en analyse van uitkomsten. Predictie van klinische uitkomsten omvat het gebruik van prognostische factoren of een prognostisch model om patiënten met een hoog risico op ongunstige uitkomst in een specifieke klinische context vroeg te herkennen. Dit kan clinici ondersteunen bij beslissingen rondom behandeling, inclusie van patiënten in gerandomiseerde studies en het benchmarken van kwaliteit van zorg. Analyses van uitkomsten omvatten onder andere het evalueren van verschillen in uitkomsten tussen ziekenhuizen en het beoordelen van de toegevoegde waarde van nieuwe uitkomstmaten. Variatie in klinische uitkomsten tussen ziekenhuizen en landen komt voor bij veel ziektebeelden, maar is zeer ongewenst wanneer dit wordt veroorzaakt door verschillen in management. Het verkrijgen van inzicht in verschillen in uitkomsten tussen ziekenhuizen faciliteert evaluatie van variatie in behandeling van individuele patiënten en kwaliteit van zorg. Verder is er een trend richting het ontwikkelen en gebruiken van uitkomsten waarin zowel functionele uitkomst als kwaliteit van leven verenigd zijn, de zogenaamde "patient-reported outcome measures" (PROMs). Nieuwe uitkomstmaten moeten statistisch efficiënt zijn om betrouwbare schattingen van behandeleffecten te verkrijgen en moeten daarnaast voorzien in eenvoudige interpretatie van behandeleffecten.

Het herseninfarct, de aneurysmatische subarachnoïdale bloeding (aSAB) en traumatisch hersenletsel zijn acute neurologische ziekten met een heterogeen beloop, en zijn vaak geassocieerd met ongunstige functionele uitkomsten en verminderde kwaliteit van leven. Dit stimuleert het meten van uitkomsten in termen van prognose, variatie tussen ziekenhuizen en nieuwe meetmethoden.

Het doel van dit proefschrift was het identificeren van patiënten met acute neurologische ziekten met een hoog risico op ongunstige uitkomst (**Deel II**) en het vergroten van onze kennis ten aanzien van variatie in uitkomsten en statistische efficiëntie van nieuwe uitkomstmaten (**Deel III**).
Specifieke onderzoeksvragen waren:

1. Welke kenmerken zijn geassocieerd met ongunstige uitkomst na acute neurologische ziekten?
2. Wat is de methodologische kwaliteit van prognostische modellen voor acute neurologische ziekten?
3. Kunnen deze modellen betrouwbare voorspellingen genereren voor patiënten in een specifieke klinische context?
4. Wat zijn de verschillen in klinische uitkomsten tussen patiënten met aSAB in een reeks internationale ziekenhuizen, en kunnen deze verschillen verklaard worden door variatie in patiëntkarakteristieken?
5. Wat is de statistische efficiëntie van nieuwe uitkomstmaten voor acute neurologische ziekten?

Deel II Predictie

In **Hoofdstuk 2** hebben we een retrospectieve cohortstudie verricht in 285 aSAB patiënten voor het onderzoeken van de relatie tussen maximaal serum lactaat en glucose gemeten binnen 24 uur na ontstaan van de aSAB en het optreden van secundaire cerebrale ischemie of ongunstige functionele uitkomst (modified Rankin Scale [mRS] 4-6). Na correctie voor patiënt- en radiologische karakteristieken waren lactaat en glucose onafhankelijk geassocieerd met cerebrale ischemie en ongunstige uitkomst met odds ratios tussen 1.17 en 1.56. Lactaat en glucose waren sterk gecorreleerd, en na inclusie van beide parameters in het multivariabele model was alleen glucose geassocieerd met cerebrale ischemie en alleen lactaat geassocieerd met ongunstige uitkomst. De rol van serum glucose en lactaat in prognostische modellen voor uitkomst na aSAB en de gerelateerde pathofysiologische mechanismen (bijv. de relatie met activatie van het sympathische systeem bij stress) moet verder bestudeerd worden.

Hoofdstuk 3 beschreef de ontwikkeling en validatie van prognostische modellen voor de Barthel Index (BI) bij ontslag uit het ziekenhuis (Dutch Stroke Score [DSS]-ontslag) en de mRS op 3 maanden (DSS-3 maanden) na een herseninfarct. We hebben data geanalyseerd van patiënten uit drie gerandomiseerde studies, waarvan één als ontwikkelcohort (n=1227) en twee als validatiecohort fungeerden (n=1589 en n=2107). De DSS-ontslag bevatte leeftijd, National Institutes of Health Stroke Scale (NIHSS) en diabetes als prognostische factoren, en liet redelijke discriminatie zien bij interne validatie (area under the receiver operating characteristic curve [AUC] van 0.76). The DSS-3 maanden bestond uit leeftijd, NIHSS, diabetes, eerder herseninfarct en atriumfibrilleren, en resulteerde in ordinale AUCs rond 0.70 bij interne en externe validatie. Kalibratie liet echter zien dat de DSS-3 maanden het aantal patiënten met ongunstige uitkomst (mRS 3-6) overschatte. De DSS kan clinici ondersteunen bij het efficiënt plannen van ontslag vanaf de stroke unit, mits verder gevalideerd.

Hoofdstuk 4 bestond uit een systematisch overzicht van beschikbare prognostische modellen voor functionele uitkomst na matig ernstig en ernstig traumatisch hersenletsel. We hebben 58 studies geïncludeerd waarin 67 verschillende modellen werden beschreven. De meest voorkomende prognostische factoren waren leeftijd, de volledige Glasgow Coma Scale of de motor component, en pupilreacties. We vonden dat bestaande prognostische modellen steeds vaker extern worden gevalideerd (149 externe validaties van 31 modellen). De methodologische kwaliteit van prognostische modellen voor matig ernstig en ernstig traumatisch hersenletsel kan echter nog steeds worden verbeterd. Kalibratie werd bijvoorbeeld slechts in de helft van de validaties (54%) grafisch gerapporteerd. De International Mission on Prognosis and Analysis of Clinical Trials (IMPACT) en Corticoid Randomisation After Significant Head injury (CRASH) modellen waren het meest vaak extern gevalideerd (n=91) en hadden goede discriminatie in verschillende populaties (gemiddelde gewogen AUCs 0.77-0.82). De betrouwbaarheid van de voorspellingen was echter zeer variabel. Dit illustreert het belang van continue externe validatie en updaten van prognostische modellen over de tijd en voor iedere specifieke klinische context.

In **Hoofdstuk 5 en 6** werden externe validatie studies van prognostische modellen gepresenteerd voor respectievelijk aSAB en matig ernstig en ernstig traumatisch hersenletsel. Het International Subarachnoid Aneurysm Trial (ISAT) model voor mortaliteit 60 dagen na aSAB bevat leeftijd, World

Federation of Neurological Surgeons score, Fisher score en grootte van het aneurysma als prognostische factoren. Het ISAT model liet goede discriminatie zien in een retrospectief cohort van 307 aSAB patiënten die opgenomen waren op de intensive care van het Erasmus Universitair Medisch Centrum (AUC 0.82). De IMPACT en CRASH modellen voor mortaliteit en ongunstige uitkomst (Glasgow Outcome Scale Extended 1-4) na matig ernstig en ernstig traumatisch hersenletsel hadden goede discriminatie in een prospectief observationeel cohort van 1742 Europese patiënten (AUCs 0.80-0.88). Het genereren van betrouwbare voorspellingen voor patiënten met aSAB en traumatisch hersenletsel in een specifieke klinische context blijft echter lastig.

Hoofdstukken 4.1 en 5.1 behandelden een aantal methodologische concepten met betrekking tot het ontwikkelen en valideren van prognostische modellen in acute neurologische ziekten. Adequate rapportage van prognostisch onderzoek met aandacht voor de beschikbare literatuur is cruciaal voor de betrouwbaarheid en reproduceerbaarheid van prognostische modellen. Voor acute neurologische ziekten zijn de belangrijkste prognostische factoren voor functionele uitkomst bevestigd. Daarom heeft validatie en updaten van bestaande prognostische modellen de voorkeur boven het ontwikkelen van nieuwe modellen.

Deel III Uitkomst analyses

In **Hoofdstuk 7** hebben we gekeken naar verschillen in klinische uitkomsten na aSAB tussen ziekenhuizen en landen. We hebben data geanalyseerd van 5972 aSAB patiënten uit een grote internationale database die waren behandeld in 179 ziekenhuizen uit 20 landen. We vonden aanzienlijke variatie tussen ziekenhuizen, welke niet verklaard kon worden door random variatie, patiëntkarakteristieken of timing van aneurysma behandeling (geadjusteerde median odds ratio 1.21, 95% betrouwbaarheidsinterval 1.11-1.44). Er waren geen statistisch significante verschillen tussen landen. Omdat de individuele schattingen op ziekenhuis niveau erg onzeker waren, was het lastig om ziekenhuizen te identificeren die beter of slechter presteerden dan anderen. De data waren relatief gedateerd, en de causaliteit tussen de geobserveerde uitkomst verschillen en variatie in behandeling van individuele patiënten (anders dan aneurysma behandeling) en kwaliteit van zorg kon niet worden geëvalueerd. Daarom was het niet mogelijk om aanbevelingen te doen voor de huidige klinische praktijk.

In **Hoofdstuk 8** werd de statistische efficiëntie van de utility-gewogen mRS (UW-mRS), een recent voorgestelde patiëntgerichte uitkomstmaat na een herseninfarct, onderzocht middels een simulatiestudie. De simulaties waren gebaseerd op de data van 500 patiënten vanuit een multicenter gerandomiseerde studie gericht op de effectiviteit van trombectomie na een herseninfarct. Lineaire analyse van de UW-mRS was minder efficiënt in het detecteren van behandeleffecten dan ordinale analyse van de mRS (power 85% versus 87%). Bovendien houdt de UW-mRS geen rekening met de individuele variatie in kwaliteit van leven binnen iedere mRS categorie, en kan deze uitkomstmaat de interpretatie van behandeleffecten bemoeilijken. Deze bevindingen benadrukken het belang van het bestuderen van de statistische efficiëntie en interpretatie van nieuwe patiëntgerichte uitkomstmaten, zoals ook beschreven in **Hoofdstuk 8.1**.

Deel IV Discussie

Het doel van dit proefschrift was het identificeren van patiënten met acute neurologische ziekten met een hoog risico op ongunstige uitkomst en het vergroten van onze kennis ten aanzien van variatie in uitkomsten en statistische efficiëntie van nieuwe uitkomstmaten. Leeftijd en neurologische status bij opname zijn de belangrijkste factoren geassocieerd met ongunstige klinische uitkomst na een herseninfarct, aSAB en matig ernstig en ernstig traumatisch hersenletsel. Prognostische modellen worden steeds vaker extern gevalideerd, wat van groot belang is voordat ze toegepast worden in de klinische praktijk. Het genereren van betrouwbare voorspellingen voor individuele patiënten met acute neurologische ziekten blijft echter lastig. Verder is er aanzienlijke variatie tussen ziekenhuizen in klinische uitkomsten na aSAB, welke niet verklaard kon worden door random variatie, patiëntkarakteristieken en timing van aneurysma behandeling. Tot slot heeft een simulatiestudie voor het bestuderen van de statistische efficiëntie van de UW-mRS aangetoond dat patiëntgerichte uitkomstmaten de power van een gerandomiseerde studie in het detecteren van behandeleffecten kunnen reduceren, en interpretatie van behandeleffecten kunnen bemoeilijken.

De grote hoeveelheid literatuur op het gebied van prognostische modellen in acute neurologische ziekten impliceert dat predictie van uitkomsten relevant wordt geacht voor de klinische praktijk. Prognostische modellen voor het herseninfarct, aSAB en traumatisch hersenletsel worden echter nauwelijks geïmplementeerd. Er zijn verschillende barrières voor het gebruik van prognostische modellen in de klinische praktijk, waaronder het gebrek aan kennis over beschikbaarheid en gebruik van prognostische modellen onder clinici, en het beperkte aantal evidence-based behandelingen voor acute neurologische ziekten. Besliskundige evaluaties en impactstudies moeten worden uitgevoerd om een indruk te krijgen van de klinische toepasbaarheid van prognostische modellen.

Verschillen in klinische uitkomsten tussen ziekenhuizen dienen, als volgende stap, gerelateerd te worden aan variatie in behandeling van individuele patiënten en kwaliteit van zorg om aanbevelingen te kunnen doen voor de klinische praktijk. Dit is mogelijk met vergelijkend effectiviteitsonderzoek, bij voorkeur op basis van een grote observationele cohortstudie. De heterogeniteit in definities en meetmethoden van klinische kenmerken en uitkomsten in acute neurologische ziekten kan gereduceerd worden door het standaardiseren van dataverzameling met behulp van common data elements.

Functionele uitkomstmaten zijn breed geïmplementeerd en voorzien in objectieve evaluatie van behandeleffecten. PROMs representeren zowel fysiek als mentaal welzijn, en houden rekening met individuele variatie in klinische uitkomsten. De ontwikkeling van PROMs is echter ingewikkeld, en het blijft de vraag of implementatie van deze uitkomstmaten in onderzoek of klinische praktijk haalbaar is gezien de problemen met betrekking tot validiteit. Daarom kan voor het evalueren van behandeleffecten het beste gebruik worden gemaakt van functionele uitkomstmaten, en dient individuele variatie in kwaliteit van leven apart beoordeeld te worden met gevalideerde uitkomstmaten.

Dankwoord

Dankwoord

Begin 2017 ben ik de uitdaging aangegaan om mijn proefschrift in een verkort traject te voltooien, en ik ben ontzettend blij en trots dat dit gelukt is! Gedurende de afgelopen 2.5 jaar heb ik veel (levenslessen) geleerd, met briljante mensen mogen samenwerken, en de gelegenheid gekregen om mijn onderzoek te presenteren op een groot aantal leuke congressen.

Dit proefschrift was er nooit gekomen zonder de begeleiding die ik heb gehad en de steun van familie en vrienden. Graag wil ik iedereen bedanken die heeft bijgedragen aan de totstandkoming van dit proefschrift, en in het bijzonder de volgende personen.

Ewout, ondanks je overstap naar Leiden ben je altijd betrokken geweest bij mijn traject en reageerde je snel en met goede input op mijn manuscripten. Bij projecten omtrent 'prognose' werd direct een link gelegd met mij. Dank ook voor de vrijheid en ruimte die je me gegeven hebt om mijn eigen draai te geven aan dit proefschrift.

Diederik, dank voor je input en betrokkenheid op de momenten dat ik daar om vroeg. Ik heb geleerd van je ontnuchterende (soms handgeschreven en lastig te ontcijferen) feedback. Tevens heb je me op de juiste momenten teruggefloten als ik bepaalde dingen net iets te sterk verwoordde in een manuscript of letter ("die Italianen zijn net zo blij dat ze in Stroke gepubliceerd hebben, dus die moet je niet zo neersabelen").

Hester, bedankt voor de leuke en leerzame wekelijkse besprekingen. Met name het brainstormen over nieuwe stukken en analyses gaf motivatie. Je hebt me de mogelijkheid gegeven om betrokken te zijn bij vele leuke projecten op het gebied van herseninfarct, subarachnoïdale bloeding en traumatisch hersenletsel. Daarnaast waardeer ik het heel erg dat er altijd ruimte was om mee te denken over mijn persoonlijke en wetenschappelijke ontwikkeling.

Mathieu, mijn enthousiasme voor wetenschappelijk onderzoek is gegroeid toen ik in 2014 bij jou mijn masteronderzoek mocht doen. Je gaf me de mogelijkheid tot het schrijven van twee manuscripten op het gebied van subarachnoïdale bloedingen als eerste auteur, nu onderdeel van dit proefschrift. Tijdens mijn coschappen heb je me voorgedragen bij CENTER-TBI, wat heeft geleid tot dit promotietraject. Dank daarvoor. Als copromotor heb je me vervolgens geloof ik wel 100 keer gevraagd "Heb je al een datum (voor je promotie)?" want "ik ging zo snel". Nu kan ik (eindelijk) zeggen: JA!

Aan al mijn supervisors: dank voor het vertrouwen de afgelopen jaren. De input vanuit verschillende invalshoeken (medische besliskunde, neurologie en intensive care) heeft ervoor gezorgd dat dit proefschrift een mooi coherent geheel is geworden.

Verder wil ik prof. dr. Peter Koudstaal, prof. dr. Saskia le Cessie en prof. dr. Geert Meyfroidt bedanken voor het beoordelen van dit proefschrift. De overige commissieleden wil ik hartelijk bedanken voor het lezen van mijn proefschrift en het deelnemen aan de oppositie. En natuurlijk iedereen die als coauteur een bijdrage heeft geleverd aan de manuscripten in mijn proefschrift en met wie ik heb mogen samenwerken: dank!

Voor de manuscripten die onderdeel zijn van dit proefschrift heb ik gebruik mogen maken van data van verschillende gerandomiseerde en observationele studies. Dank aan de PAIS, PRACTISE, PASS en MR CLEAN onderzoekers voor het beschikbaar stellen van de door jullie verzamelde data en de samenwerking. Further, I would like to thank prof. dr. Loch Macdonald and all members of the SAHIT collaboration for their patience and trust, and for giving me the opportunity to work with the SAHIT repository. Finally, two manuscripts in this thesis are part of the CENTER-TBI project. Many thanks to the CENTER-TBI participants and investigators for the collaboration and the possibility to work with this unique dataset.

Gelukkig was er naast (en tijdens) het schrijven van dit proefschrift ook voldoende tijd voor ontspanning, waarbij de volgende personen een belangrijke rol hebben gespeeld.

Heel veel dank gaat uit naar Kelly, Maaike en Vicky. Ik ben ontzettend blij dat ik toevallig op (toen nog) kamer 2424 terecht kwam en de gelegenheid heb gekregen om jullie te leren kennen. Jullie zijn enorm lieve, betrokken en ambitieuze meiden en ik heb jullie echt in mijn hart gesloten. We hebben samen heel veel gelachen, mooie maar ook verdrietige momenten gedeeld en leuk samengewerkt. Kelly, je hebt me wegwijs gemaakt op de afdeling en binnen CENTER-TBI, we hebben altijd goed kunnen sparren over onze gezamenlijke projecten, en onszelf in hilarische situaties gebracht. Maaike, jouw humor, de vele goede gesprekken (met natuurlijk de benodigde hoeveelheid cafeïne en RUMAG quotes), pogingen om R te verslaan, en altijd spot-on Netflix-tips waren (en zijn) zeer waardevol. En natuurlijk bedankt dat je mijn paranimf wilt zijn! Vicky, met jouw zorgzaamheid en doorzettingsvermogen heb je me enorm gesteund en geïnspireerd, en ik vond het altijd heel gezellig om (soms hele werkdagen) met je te kletsen. Mijn promotietraject is begonnen en geëindigd met een overload aan sushi in jullie gezelschap. We gaan nu allemaal een andere weg, maar ik hoop dat we nog lang bevriend zullen blijven! En om af te sluiten met een onmisbare (en iets aangepaste) RUMAG quote: JULLIE.ZIJN.FUCKING. GEWELDIG.

Eliza, Gwen, Jara en Mijna, ik weet eigenlijk niet eens waar ik moet beginnen. We kennen elkaar al vanaf de basisschool of middelbare school, en hebben heel veel met elkaar meegemaakt: van kinderfeestjes tot aan een examenreis. Hoewel we sinds onze studententijd door Nederland verspreid zijn vind ik het onwijs leuk en waardevol dat we elkaar nog steeds af en toe zien en belangrijke mijlpalen samen kunnen vieren. Ik hoop dat we in de toekomst nog meer mooie herinneringen kunnen maken!

Anna en Lisette, we zijn samen afgestudeerd, en de gezelligheid tijdens de studie en coschappen hebben we daarna voortgezet met leuke etentjes en weekendjes weg. Ik heb van jullie al wat mogen meekrijgen van het werken in de kliniek en begrijp dat me nog wel het een en ander te wachten staat. Maar heb het idee dat dat wel goedkomt, zeker als ik af en toe een avond kan ontspannen, lachen en genieten met jullie.

En dan mijn lieve familie, juist omdat we maar met een kleine groep zijn heb ik het extra gewaardeerd dat jullie altijd veel interesse tonen in (en soms ook rekening houden met) mijn studie en werk. Selma, Ria en Theo: ik heb vele goede herinneringen aan leuke verjaardagen, en het logeren in het uiterste noorden van het land. “Ria Curaçao”, zoals we je vaak noemen om verwarring te voorkomen:

je woont natuurlijk veel te ver weg, maar daarom is het extra speciaal wanneer we elkaar weer zien. En natuurlijk niet te vergeten lieve Yvonne, lieve “oma”, je hebt een enorm waardevolle bijdrage geleverd aan mijn opvoeding en ik geniet altijd weer van je verhalen over mijn jeugd. Dus daarom deze keer voor jou “een dikke knuffel en een zoen op iedere wang”.

Mijn lieve zus Maren, we zijn zo verschillend maar lijken toch zo veel op elkaar. Alles wat met geneeskunde te maken heeft vind jij maar niks, en ik snap niet zoveel van de consultancy wereld. Maar we hebben allebei enorm veel doorzettingsvermogen en we weten precies wat we willen. Ik heb altijd gezien dat je veel bewondering hebt voor mij, maar dat heb ik misschien nog wel meer voor jou. Mede daarom ben ik erg blij dat je mijn paranimf wilt zijn. Ik wens jou en Sander alle geluk van de wereld. Love you!

En tot slot mijn lieve ouders, dankzij jullie heb ik de mogelijkheid gekregen om de studie geneeskunde en ook dit promotietraject te doorlopen. Dat was zeker niet gelukt zonder de goede basis en mooie herinneringen die jullie me van jongs af aan hebben meegegeven. Mama, ik kan altijd bij je terecht, kan alles met je delen en je herinnert me er regelmatig aan dat ik goed voor mezelf moet blijven zorgen. Daarnaast waardeer ik het enorm dat je van elke mijlpaal, hoe klein ook, een feestje hebt gemaakt. Papa, op het moment dat jij je uitschreef uit het BIG-register mocht ik me daar registreren. Hoewel je het niet vaak zegt, zie en weet ik dat je trots op me bent. Jullie hebben altijd achter me gestaan en me in iedere keuze gesteund. Mam en pap, ik houd van jullie.

List of publications

List of publications

1. **Dijkland SA**, Jaja BNR, van der Jagt M, Roozenbeek B, Vergouwen MDI, Suarez JI, et al. Between-center and between-country differences in outcome after aneurysmal subarachnoid hemorrhage in the Subarachnoid Hemorrhage International Trialists (SAHIT) repository. *J Neurosurg.* 2019; Epub ahead of print.

2. **Dijkland SA**, van der Jagt M, Lingsma HF. Letter by Dijkland et al regarding article "Prediction of outcome after aneurysmal subarachnoid hemorrhage: Development and validation of the SAFIRE grading scale". *Stroke.* 2019; 50(7): e224.

3. **Dijkland SA**, Foks KA, Polinder S, Dippel DWJ, Maas AIR, Lingsma HF, et al. Prognosis in moderate and severe traumatic brain injury: A systematic review of contemporary models and validation studies. *J Neurotrauma.* 2019; Epub ahead of print.

4. Foks KA, **Dijkland SA**, Lingsma HF, Polinder S, van den Brand CL, Jellema K, et al. Risk of intracranial complications in minor head injury: the role of loss of consciousness and posttraumatic amnesia in a multicenter observational study. *J Neurotrauma.* 2019; Epub ahead of print.

5. **Dijkland SA**, Lingsma HF, Dippel DWJ. Response by Dijkland et al to letter regarding article, "Utility-weighted modified Rankin Scale as primary outcome in stroke trials: A simulation study". *Stroke.* 2018; 49(12): p. e338

6. **Dijkland SA**, Retel Helmrich IR, Steyerberg EW. Validation of prognostic models: challenges and opportunities. *J Emerg Crit Care Med.* 2018; 2: 91.

7. Foks KA, **Dijkland SA**, Steyerberg EW. Response to Walker et al.: Predicting long-term global outcome after traumatic brain injury. *J Neurotrauma.* 2019; 36(8): 1382-1383.

8. **Dijkland SA***, de Ridder IR*, Scheele M, den Hertog HM, Dirks M, Westendorp WF, et al. Development and validation of the Dutch Stroke Score for predicting disability and functional outcome after ischemic stroke: A tool to support efficient discharge planning. *Eur Stroke J.* 2018; 3(2), 165–173.

9. **Dijkland SA**, Voormolen DC, Venema E, Roozenbeek B, Polinder S, Haagsma JA, et al. Utility-weighted modified Rankin Scale as primary outcome in stroke trials: A simulation study. *Stroke.* 2018; 49(4): 965–971.

10. **Dijkland SA**, Dippel DWJ, Lingsma HF. Letter by Dijkland et al regarding article, "Development and validation of a predictive model for functional outcome after stroke rehabilitation: The Maugeri model". *Stroke.* 2018; 49(3), p. e133.

11. **Dijkland SA**, Roozenbeek B, Brouwer PA, Lingsma HF, Dippel DWJ, Vergouw LJ, et al. Prediction of 60-day case fatality after aneurysmal subarachnoid hemorrhage: External validation of a prediction model. *Crit Care Med.* 2016; 44(8): 1523–1529.

12. **Dijkland SA***, van Donkelaar CE*, van den Bergh WM, Bakker J, Dippel DWJ, Nijsten MWM, et al. Early circulating lactate and glucose levels after aneurysmal subarachnoid hemorrhage correlate with poor outcome and delayed cerebral ischemia: A two-center cohort study. *Crit Care Med.* 2016; 44(5): 966–972.

PhD portfolio

PhD portfolio

Name PhD student: Simone Dijkland

Erasmus MC Department: Public Health

Research School: COEUR

PhD period: February 2017 - August 2019

Promotors: Prof. dr. E.W. Steyerberg and Prof. dr. D.W.J. Dippel

Supervisors: Dr. H.F. Lingsma and Dr. M. van der Jagt

	Year	Workload (ECTS)
1. PhD training		
General academic courses		
Research Integrity Course, Erasmus MC	2017	0.3
Course Patient Oriented Research: design, conduct and analysis	2017	0.3
Basic course for clinical investigators (BROK®)	2019	1.5
NIHES / COEUR courses		
Biostatistical Methods I: Basic Principles Part A (NIHES)	2017	2.0
Advanced Analysis of Prognosis Studies (NIHES)	2017	0.9
Endovascular thrombectomy in acute ischemic stroke (COEUR)	2017	0.3
Regression Analysis (NIHES)	2017	1.4
Advanced Topics in Decision-making in Medicine (NIHES)	2018	2.4
Seminars and workshops		
PhD day, Erasmus MC Rotterdam	2017	0.3
Data curation workshop CENTER-TBI, University of Antwerp	2017	1.0
Weekly seminars at the Department of Public Health, Erasmus MC	2017-2019	3.0
General Assembly CENTER-TBI, University of Antwerp	2019	1.0
Presentations at national and international conferences		
35th International Symposium on Intensive Care and Emergency Medicine, Brussels, Belgium (two poster presentations)	2015	2.0
39th Annual North American Meeting of the Society for Medical Decision Making, Pittsburgh, USA (oral presentation)	2017	1.0
Wetenschapsdagen Nederlandse Vereniging voor Neurologie, Nunspeet, the Netherlands (poster presentation)	2017	1.0
4th European Stroke Organisation Conference, Gothenborg, Sweden (oral presentation)	2018	1.0
17th Biennial European conference of the Society for Medical Decision Making, Leiden, the Netherlands (oral and poster presentation)	2018	2.0
40th Annual North American Meeting of the Society for Medical Decision Making, Montréal, Canada (oral and poster presentation)	2018	2.0
Wetenschapsdagen Nederlandse Vereniging voor Neurologie, Nunspeet, the Netherlands (oral presentation)	2018	1.0
15th International Conference on SubArachnoid Hemorrhage, Amsterdam, the Netherlands (two oral presentations)	2019	2.0
Presentations for Department of Public Health / section Medical Decision Making, Erasmus MC Rotterdam, the Netherlands (at least four oral presentations)	2017-2019	4.0

2. Teaching activities		
Checking examinations (bachelor essays) from second year medical students	2017	1.0
Supervising community project for third year medical students	2018	1.5
Lecturer workshop Data Analytics CENTER-TBI	2018	1.0
Supervising PhD student Isabel Retel Helmrich (CENTER-TBI)	2018-2019	3.0
Peer review for several journals (Circulation, Neurology, Stroke)	2017-2019	1.5
Statistical reviewer for Intensive Care Medicine	2018-2019	1.0
Total ECTS		**39.4**

About the author

About the author

Simone Anna Dijkland was born in Utrecht, the Netherlands, on March 14th, 1992. During the last two years of secondary school at Revius Lyceum Doorn, she was selected to participate in Junior Med School at the Erasmus University Medical Center (Erasmus MC) in Rotterdam. This unique program is aimed at introducing talented 5th and 6th year VWO students to a career in medicine and science. The program consists of education on both medical and scientific topics, and allows students to perform their own research project. Through successful completion of Junior Med School, Simone was granted access to the Bachelor of Medicine at the Erasmus MC, which she started after finishing secondary school in 2010. In 2013, she obtained her Bachelor of Medicine (*Cum Laude*). In 2014, she conducted her master research project at the Department of Intensive Care (Erasmus MC) under supervision of dr. Mathieu van der Jagt. Her master thesis resulted in two publications in Critical Care Medicine (now part of this PhD thesis). After completing clinical rotations and receiving her medical degree in 2017, she got the opportunity to start as a PhD candicate at the Department of Public Health of the Erasmus MC under supervision of dr. Hester Lingsma, prof. dr. Ewout Steyerberg, dr. Mathieu van der Jagt (Department of Intensive Care) and prof. dr. Diederik Dippel (Department of Neurology). She worked with data from large international projects focused on improving care for patients with acute neurological diseases, such as the Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) study and the Subarachnoid Hemorrhage International Trialists (SAHIT) repository. She gained experience with a variety of statistical methods and presented her research at numerous (inter)national conferences. She aspires a career as clinician and researcher, and will start clinical work as a resident in Neurology (not in training, ANIOS) at the Albert Schweitzer Hospital in Dordrecht.