Prognostic research: Methodological Aspects and Applications in Acute Care

Isabel R.A. Retel Helmrich

Prognostic Research: Methodological aspects and applications in acute care

Isabel Rosalie Arianne Retel Helmrich

Layout and printing: ProefschriftMaken || www.proefschriftmaken.nl

This thesis was financially supported by the Erasmus University, the Erasmus university medical center and the Department of Public Health.

ISBN 978-94-6469-394-2

Prognose onderzoek: methodologische aspecten en toepassingen in de acute zorg

Prognostic Research: Methodological aspects and applications in acute care

Proefschrift

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

Prof.dr. A.L. Bredenoord

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

woensdag 4 oktober 2023 om 13:00 uur

door

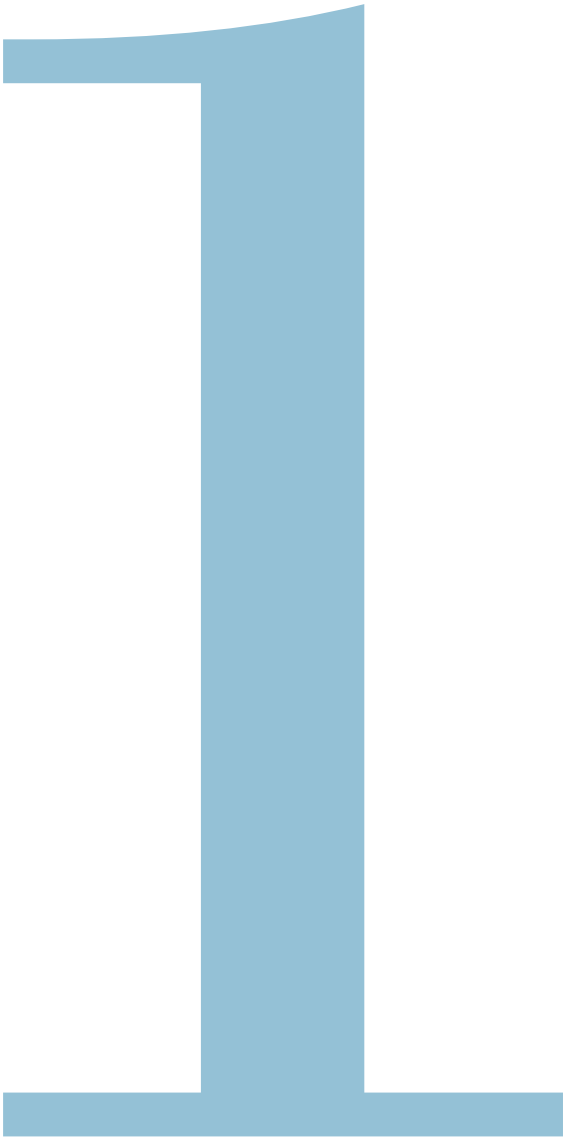
Isabel Rosalie Arianne Retel Helmrich
geboren te Rotterdam, Nederland

Erasmus University Rotterdam

Promotiecommissie:

Promotoren:	Prof.dr. E.W. Steyerberg Prof.dr. H.F. Lingsma
Overige leden:	Prof.dr.ir. H. Boersma Prof.dr. J. van der Naalt Prof.dr. B. van Calster
Copromotor:	Dr.ir. D. van Klaveren

Contents



Chapter 1

General Introduction

Prognosis

A prognosis can be defined as a prophecy, forecast or prediction of the risk of future conditions, such as the weather.[1] In medicine, prognosis is often expressed as the probability of an individual patient developing a particular outcome over a specific time, based on their clinical and non-clinical characteristics. The outcome of interest is usually a specific event, such as death. Predictions can also be made for other quantities, such as quality of life.

In clinical practice, health care providers frequently aim to predict a future outcome of an individual patient. The ability to accurately predict a patient's outcome is important and has several purposes. Outcome prognostication may support health care providers in presenting reliable information to patients and relatives, guide treatment decisions, and give insight in the quality of care by comparing observed and expected outcomes.[2] In research, predictions can be used for risk stratification of patients.

Prognostic research

In prognostic research we can distinguish between two types of studies: studies with a focus on the prognostic role of a specific characteristic in relation to the outcome ('prognostic factor research'), and studies with a focus on the combined effect of various prognostic factors in predicting the outcome ('prediction model research').[1] An example of 'prognostic factor research' is examining the role of a single characteristic such as age in relation to mortality, whereas the development of a model to predict mortality for an individual patient at 6 months based on multiple characteristics at presentation can be defined as 'prediction model research'. Prognostic models predict the outcome of an individual patient at a specific time based on multiple characteristics at presentation. In this thesis we will mainly focus on prediction model research.

In prediction model research we can identify distinct phases: model development, external validation, and impact analysis (Figure 1).[1, 3] Model development requires attention to methodological aspects. A prediction model needs to meet certain quality criteria to increase the chance of being useful for health care providers in clinical practice. When a model is developed, it should be externally validated.[4, 5] External validation provides information on the models' generalizability and (geographic or temporal) transportability, that is, how the model performs in new patients and settings. Validation, preferably across a range of settings, is required before the model should be considered for use in practice. Before considering developing yet a new model, the updating, adjusting or recalibrating of an established model should be

attempted.[2, 6] If the model performs adequate in external validation and is therefore deemed appropriate for implementation, the clinical impact of the model should be examined.[2] The clinical impact and implementation of prediction models has generally received little attention.

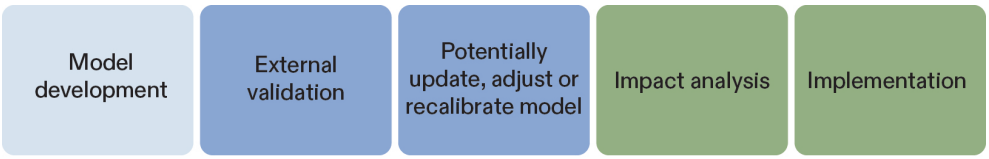


Figure 1: Process from model development to implementation, showing the three distinct phases in prognostic research: Model development, external validation and impact analysis.

Although there is an increasing number of prediction models being published for different prognostic purposes, the methodological quality is often suboptimal.[7-9] Despite recent guidelines for model development and reporting of prediction model studies,[3, 10, 11] several reviews showed shortcomings in model development and a general lack of external validation.[9, 12, 13]

Acute care

In the hospital, acute care services are provided to a patient with a severe illness or condition.[14] Patients are, for instance, treated briefly for a severe illness or condition that resulted from a disease or trauma at the emergency department or in the intensive care unit. A considerable proportion of this thesis came into being during the COVID-19 pandemic, in which research on COVID-19 emerged rapidly and took priority. In this thesis we will predominantly focus on traumatic brain injury, while also including results from a study on COVID-19 care.

Traumatic Brain Injury

Traumatic brain injury (TBI) is a major health concern with over 50,000,000 new cases reported globally every year.[15-17] TBI is defined as an injury to the brain induced by an external force.[18] In recent years, the epidemiology of TBI has changed substantially, especially regarding the mechanism of injury and age distribution.[15, 19, 20] The main cause of TBI has shifted from road traffic incidents to falls. Furthermore, TBI is increasingly reported in older patients and in women.[21-23]

Clinical severity of TBI is typically classified using the Glasgow Coma Scale (GCS).[24-26] This scale is used to assess impaired consciousness, based on eye, motor and verbal response, ranging from an unresponsive patient (GCS = 3) to a fully awake and oriented patient (GCS = 15). Based on the GCS, patients can be placed into three categories of injury severity: mild (13-15), moderate (9-12) or severe (3-8). Approximately 70–90% of patients with TBI can be categorized as 'mild'.[21, 27] A decline in mortality due to TBI has been observed, however, even following a 'mild' injury, long-term disability or residual complaints are common.[28-31] Because of the variation between patients in long-term outcome following TBI, personalized treatment and rehabilitation is required.

The Glasgow Outcome Scale Extended (GOSE) is the most widely used measure of global functional outcome following TBI (Text box 1).[32] In research, the GOSE is often dichotomized into clinically relevant endpoints: mortality (GOSE=1), unfavorable outcome (GOSE≤4) and incomplete recovery (GOSE<8). In TBI patients, prognostic models typically aim to predict global functional outcome, using the GOSE.[13, 33] As TBI patients often experience a combination of physical, emotional, and cognitive consequences, the need for prediction of more granular outcomes, including Health-related Quality of Life (HRQoL), has been emphasized.[15]

Two types of instruments are available to assess HRQoL; generic and condition-specific instruments.[34] Generic instruments, such as the Short Form-36 (SF-36), allow comparison with healthy individuals and various health states or conditions, as the items are not based on a particular disease or condition. A condition-specific instrument does consider key issues for patients following a certain disease or condition, such as the Quality of Life after Traumatic Brain Injury (QOLIBRI).[35, 36]

The role of prognostic models in TBI

There has been considerable interest in prognosis following TBI. TBI is said to be one of the most heterogeneous neurological conditions, with substantial variation in trauma mechanisms, pathophysiology and clinical presentation, which makes the prediction of outcome challenging.[15] It is important to identify patients who are at high risk of mortality or long-term consequences. Accurate and reliable prognostic models for outcome prediction after TBI have the potential to support health care providers and patients in making clinical decisions.

A recent systematic review reported that in the last two decades, over 42 different models have been developed to predict functional outcome following TBI.[13] Some

Text box 1: Outcomes and instruments

Instrument and description	Subscales	Domains
GOSE: A global measure of functional outcome and disability. Can be assessed as a structured interview or a questionnaire completed by the patient or a carer.		1. Dead 2. Vegetative state 3. Lower severe disability 4. Upper severe disability 5. Lower moderate disability 6. Upper moderate disability 7. Lower good recovery 8. Upper good recovery
SF-36/SF-12: A 36 or 12-item patient-reported HRQoL outcome which assesses multiple aspects of health-related functioning and well-being.	1. Physical functioning 2. Role limitations due to physical health 3. Bodily pain 4. General health perceptions	Physical component summary (PCS) score
	1. Vitality 2. Social functioning 3. Role limitations due to emotional health 4. General mental health.	Mental component summary (MCS) score
QOLIBRI-OS: A 6-item patient-reported HRQoL outcome specifically developed for patients following TBI.	Satisfaction with: 1. Cognition 2. Self 3. Daily life and autonomy 4. Social relationships 5. Current situation 6. Future prospects	

of these models have been externally validated frequently, such as the IMPACT and CRASH prognostic models.[13, 37, 38] The ability of the IMPACT and CRASH models to discriminate between high and low risk patients has been confirmed across a range of settings.[13] However, the agreement between observed and predicted outcomes varied, which reflects heterogeneity in the calibration of predictions. The IMPACT and CRASH models for moderate and severe TBI only explain 35% of the variance in outcome. [37, 38] Furthermore, prognostic models for mild TBI are less prevailing.[39, 40] A recent study externally validated five published models and showed that none of the prognostic models for early prediction of functional outcome and persistent post-concussive symptoms performed satisfactory in patients with mild TBI.[33] Improving prognostication has been considered critical by clinicians, researchers, and patients and caregivers alike.[15]

Prognostic models in TBI can potentially be improved by including novel predictors, such as biomarkers. [15] Furthermore, there is a need for prediction models for TBI patients of all severities, particularly for patients with mild TBI, and the prediction of more granular outcomes, such as HRQoL.

The CENTER-TBI study

The availability of large databases has given way to new opportunities for prognostic research. In this thesis we mainly use data from the Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) study. The CENTER-TBI study is a prospective observational cohort study in patients with mild, moderate and severe TBI.[41, 42] Inclusion criteria were a clinical diagnosis of TBI, presentation within 24 h after injury, and an indication for a CT scan. Participants were recruited between December 2014 and December 2017 in 18 countries across Europe and Israel. In the CENTER-TBI study, patients were differentiated by care pathway and assigned to the emergency room (ER) stratum (patients who were discharged from an emergency room), admission stratum (patients who were admitted to a hospital ward), or intensive care unit (ICU) stratum (patients who were admitted to the ICU).

Detailed and longitudinal information was measured for sociodemographic and injury-related characteristics, care path, blood-based biomarkers, and outcomes, including the GOSE and HRQoL which were assessed using generic and condition-specific instruments. The large sample size and richness of CENTER-TBI data allows for the development, update and external validation of prediction models for TBI patients of all severities. Consistent with the need for the improvement of characterization of TBI, CENTER-TBI data provides unique opportunities to increase our knowledge of prediction of outcomes in TBI patients.

COVID-19

The COVID-19 pandemic has been characterized by high levels of uncertainty in outcomes for those contracting the virus, including the severity of symptoms, disease trajectories, and risk of long-term consequences and mortality.[43, 44] Additionally, there are differences in governmental responses and restrictions over time and between countries.[45] Consequently, outcomes have varied temporally by 'wave' and by geographic region. This has further exacerbated uncertainty, making it difficult to predict outcomes among people with COVID-19 who are admitted to the hospital.

The role of prognostic models for COVID-19

In response to the COVID pandemic and related uncertainty,[44] prognostic models were being published rapidly. Since the start of the pandemic, hundreds of prognostic models have been developed, however the reporting and methodological quality of models was typically poor.[7] Almost all published models were identified as having high risk of bias, indicating that their reported predictive performance is likely to be overly optimistic. Some of the models have been externally validated, showing highly variable model performance in new patients and settings. Poorly calibrated models may lead to harm, as they provide misinformation on which clinical decision-making might be based.[4]

In the US (the New York City (NYC) area) and in the Netherlands, prognostic models were developed for predicting outcomes in patients hospitalized with COVID-19: The Northwell COVID-19 Survival (NOCOS) model and the COVID Outcome Prediction in the Emergency Department (COPE) model.[46] Both models were developed on large datasets, including over 12,000 hospitalized patients from the NYC region and over 5,000 hospitalized Dutch patients. Furthermore, unlike most prior models developed to predict COVID-19 outcomes, the models were developed consistent with methodological recommendations.

Before accurate prediction models can be considered for implementation in clinical practice, we must understand end-user perceptions, which include health care providers, patients, and surrogate decision-makers. Perceptions of stakeholders about the use of clinical prediction models and the models' impact should be considered before models are used in clinical practice.

Aims and outline of this thesis

The overall aim of this thesis is to increase our knowledge of prediction of outcome in acute care by exploring methodological aspects and applications of prognostic research.

Specific research questions are:
1. What methodological aspects are of key importance in prognostic research?
2. To what extent can we predict functional outcome and Health-Related Quality of Life after traumatic brain injury in contemporary patients?
3. Can blood-based biomarkers further improve prediction of functional outcome following traumatic brain injury?

4. How do health care providers, patients and surrogate decision makers perceive the use of prediction models to support clinical decision-making?

Part I – Methodological aspects

Part I focuses on the methodological aspects of prognostic research and answers research question 1. **Chapter 2** describes prediction of outcome following TBI, including prognostic factors and established prognostic models, in detail. **Chapter 3** gives a concise overview of the steps and considerations in prognostic research and introduces the reader to the concept of overfitting. In **Chapter 4**, methodological aspects of prognostic research are discussed in more detail. **Chapter 5** aims to examine the relation between methodological quality of model development studies and their performance at external validation.

Part II – Applications

Part II investigates several applications of prognostic research and answers research questions 2, 3 and 4. **Chapter 6** describes the external validation and potential application of two established prognostic models for outcome prediction after moderate and severe TBI in a contemporary cohort of patients across Europe. **Chapter 7** aims to examine the incremental prognostic value of serum biomarkers over demographic, clinical and radiological characteristics and over established prognostic models for the prediction of functional outcome after TBI. The relationship between disability and wellbeing following TBI is examined in **Chapter 8**. In **Chapter 9**, we aimed to identify predictors of, and develop prognostic models for the prediction of Health-Related Quality of Life after TBI. In **Chapter 10,** qualitative analyses are used to explore considerations of health care providers, patients, and surrogate decision makers (e.g. relatives and caregivers) about the use of prediction models to support clinical decision-making in COVID-19 care.

References

1. Moons K G M, Royston P, Vergouwe Y, Grobbee D E, Altman D G. Prognosis and prognostic research: what, why, and how? *Bmj*. 2009;**338**.
2. Moons K G M, Altman D G, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *Bmj*. 2009;**338**.
3. Steyerberg E W. Clinical prediction models: a practical approach to development, validation, and updating. New York: Springer Science & Business Media; 2009.
4. Steyerberg E W, Harrell F E. Prediction models need appropriate internal, internal–external, and external validation. *J Clin Epidemiol*. 2016;**69**:245-7.
5. Altman D G, Vergouwe Y, Royston P, Moons K G M. Prognosis and prognostic research: validating a prognostic model. *Bmj*. 2009;**338**.
6. Moons K G M, Kengne A P, Grobbee D E, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart*. 2012;**98**(9):691-8.
7. Wynants L, Van Calster B, Collins G S, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *bmj*. 2020;**369**.
8. Venema E, Wessler B S, Paulus J K, et al. Large-scale validation of the Prediction model Risk Of Bias ASsessment Tool (PROBAST) using a short form: high risk of bias models show poorer discrimination. *medRxiv*. 2021.
9. Bouwmeester W, Zuithoff N P A, Mallett S, et al. Reporting and methods in clinical prediction research: a systematic review. *PLoS Med*. 2012;**9**(5):e1001221.
10. Collins G S, Reitsma J B, Altman D G, Moons K G M. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMC med*. 2015;**13**(1):1.
11. Steyerberg E W, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *European heart journal*. 2014;**35**(29):1925-31.
12. Perel P, Edwards P, Wentz R, Roberts I. Systematic review of prognostic models in traumatic brain injury. *BMC medical informatics and decision making*. 2006;**6**(1):38.
13. Dijkland S A, Foks K A, Polinder S, et al. Prognosis in moderate and severe traumatic brain injury: a systematic review of contemporary models and validation studies. *J Neurotrauma*. 2019.
14. Hirshon J M, Risko N, Calvello E J B, et al. Health systems and services: the role of acute care. *Bulletin of the World Health Organization*. 2013;**91**:386-8.
15. Maas A I R, Menon D K, Adelson P D, et al. Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. *Lancet Neurol*. 2017.
16. Langlois J A, Rutland-Brown W, Wald M M. The epidemiology and impact of traumatic brain injury: a brief overview. *The Journal of head trauma rehabilitation*. 2006;**21**(5):375-8.
17. Dewan M C, Rattani A, Gupta S, et al. Estimating the global incidence of traumatic brain injury. *Journal of neurosurgery*. 2018;**130**(4):1080-97.
18. Menon D K, Schwab K, Wright D W, Maas A I. Position statement: definition of traumatic brain injury. *Archives of physical medicine and rehabilitation*. 2010;**91**(11):1637-40.
19. Roozenbeek B, Maas A I R, Menon D K. Changing patterns in the epidemiology of traumatic brain injury. *Nature Reviews Neurology*. 2013;**9**(4):231-6.
20. Peeters W, van den Brande R, Polinder S, et al. Epidemiology of traumatic brain injury in Europe. *Acta neurochirurgica*. 2015;**157**(10):1683-96.
21. Steyerberg E W, Wiegers E, Sewalt C, et al. Case-mix, care pathways, and outcomes in patients with traumatic brain injury in CENTER-TBI: a European prospective, multicentre, longitudinal, cohort study. *Lancet Neurol*. 2019;**18**(10):923-34.

22. Mollayeva T, Mollayeva S, Colantonio A. Traumatic brain injury: sex, gender and intersecting vulnerabilities. *Nature Reviews Neurology*. 2018;**14**(12):711-22.
23. Gardner R C, Dams-O'Connor K, Morrissey M R, Manley G T. Geriatric traumatic brain injury: epidemiology, outcomes, knowledge gaps, and future directions. *Journal of neurotrauma*. 2018;**35**(7):889-906.
24. Teasdale G, Maas A, Lecky F, Manley G, Stocchetti N, Murray G. The Glasgow Coma Scale at 40 years: standing the test of time. *Lancet Neurol*. 2014;**13**(8):844-54.
25. Jones C. Glasgow coma scale. LWW; 1979.
26. Shukla D, Devi B I, Agrawal A. Outcome measures for traumatic brain injury. *Clin Neurol Neurosurg*. 2011;**113**(6):435-41.
27. Cassidy J D, Carroll L, Peloso P, et al. Incidence, risk factors and prevention of mild traumatic brain injury: results of the WHO Collaborating Centre Task Force on Mild Traumatic Brain Injury. *Journal of rehabilitation medicine*. 2004;**36**(0):28-60.
28. Voormolen D C, Zeldovich M, Haagsma J A, et al. Outcomes after complicated and uncomplicated mild traumatic brain injury at three-and six-months post-injury: results from the CENTER-TBI study. *Journal of clinical medicine*. 2020;**9**(5):1525.
29. Carroll L, Cassidy J D, Peloso P, et al. Prognosis for mild traumatic brain injury: results of the WHO Collaborating Centre Task Force on Mild Traumatic Brain Injury. *Journal of rehabilitation medicine*. 2004;**36**(0):84-105.
30. McMahon P J, Hricik A, Yue J K, et al. Symptomatology and functional outcome in mild traumatic brain injury: results from the prospective TRACK-TBI study. *Journal of neurotrauma*. 2014;**31**(1):26-33.
31. Nelson L D, Temkin N R, Dikmen S, et al. Recovery after mild traumatic brain injury in patients presenting to US Level I trauma centers: A Transforming Research and Clinical Knowledge in Traumatic Brain Injury (TRACK-TBI) study. *JAMA neurology*. 2019;**76**(9):1049-59.
32. Wilde E A, Whiteneck G G, Bogner J, et al. Recommendations for the use of common outcome measures in traumatic brain injury research. *Archives of physical medicine and rehabilitation*. 2010;**91**(11):1650-60. e17.
33. Mikolić A, Polinder S, Steyerberg E W, et al. Prediction of Global Functional Outcome and Post-Concussive Symptoms after Mild Traumatic Brain Injury: External Validation of Prognostic Models in the Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) Study. *Journal of neurotrauma*. 2021;**38**(2):196-209.
34. Polinder S, Haagsma J A, van Klaveren D, Steyerberg E W, Van Beeck E F. Health-related quality of life after TBI: a systematic review of study design, instruments, measurement properties, and outcome. *Popul Health Metr*. 2015;**13**(1):4.
35. Von Steinbüchel N, Wilson L, Gibbons H, et al. Quality of Life after Brain Injury (QOLIBRI): scale development and metric properties. *Journal of neurotrauma*. 2010;**27**(7):1167-85.
36. Truelle J-L, Koskinen S, Hawthorne G, et al. Quality of life after traumatic brain injury: the clinical use of the QOLIBRI, a novel disease-specific instrument. *Brain Injury*. 2010;**24**(11):1272-91.
37. Steyerberg E W, Mushkudiani N, Perel P, et al. Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. *PLoS Med*. 2008;**5**(8):e165.
38. Collaborators M R C C T. Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients. *Bmj*. 2008;**336**(7641):425-9.
39. Silverberg N D, Gardner A J, Brubacher J R, Panenka W J, Li J J, Iverson G L. Systematic review of multivariable prognostic models for mild traumatic brain injury. *Journal of neurotrauma*. 2015;**32**(8):517-26.

40. Lingsma H F, Yue J K, Maas A I R, et al. Outcome prediction after mild and complicated mild traumatic brain injury: external validation of existing models and identification of new predictors using the TRACK-TBI pilot study. *Journal of neurotrauma*. 2015;**32**(2):83-94.
41. Maas A I R, Menon D K, Steyerberg E W, et al. Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) A Prospective Longitudinal Observational Study. *J Neurosurg*. 2014;**76**(1):67-80.
42. Wheble J L C, Menon D K. TBI—the most complex disease in the most complex organ: the CENTER-TBI trial—a commentary. British Medical Journal Publishing Group; 2016. 87-9.
43. Yelin D, Wirtheim E, Vetter P, et al. Long-term consequences of COVID-19: research needs. *The Lancet Infectious Diseases*. 2020;**20**(10):1115-7.
44. Koffman J, Gross J, Etkind S N, Selman L E. Clinical uncertainty and Covid-19: Embrace the questions and find solutions. Sage Publications Sage UK: London, England; 2020. 829-31.
45. Imtyaz A, Haleem A, Javaid M. Analysing governmental response to the COVID-19 pandemic. *Journal of Oral Biology and Craniofacial Research*. 2020;**10**(4):504-13.
46. van Klaveren D, Rekkas A, Alsma J, et al. COVID Outcome Prediction in the Emergency Department (COPE). *medRxiv*. 2021:2020.12. 30.20249023.

Part I

Methodological aspects

Chapter 2

Sequelae and Outcome in Traumatic Brain Surgery: Prognosis after Traumatic Brain Injury

I.R.A. Retel Helmrich, E.W. Steyerberg, A.I.R. Maas, and H.F. Lingsma

Published in: Youmans & Winn Neurological Surgery

Abstract

For patients who have sustained a traumatic brain injury (TBI) and their families, prognosis is of major importance. Prognostic analyses aim to support physicians' intuitive estimates of patients functional, emotional or cognitive status over time. In practice, prognostic models can be used to inform patients and relatives on prognosis, to stratify patients for clinical trials, and to support medical decision-making. In this chapter we aim to summarize the literature on state-of-the-art approaches to prognostic analysis, our current knowledge on prognosis in TBI and discuss the development, validation, application, and limitations of prognostic models for patients following TBI.

Prediction modelling considers combinations of prognostic factors, with challenges for model specification, estimation, evaluation, validation, and presentation. For moderate and severe TBI, most prognostic models have been developed to predict mortality and unfavorable outcome using the Glasgow Outcome Scale (GOS) or the GOS extended (GOSE). The strongest predictors are Glasgow Coma Scale (motor) score, age, and pupillary reactivity. For mild TBI, prognostic models are less well established, and patient characteristics might be of more relevance to prognosis than injury-related characteristics.

Advancements in prognostic research are likely to be made when key methodological principles are adhered to, and when research is conducted by a consortium of research groups. Prognostic models for mild TBI should be developed, validated and further improved. Model performance may be enhanced with the inclusion of biomarkers and advanced imaging, and the development of dynamic prediction models. Furthermore, prediction models should be developed for more multidimensional outcomes.

Key words: Traumatic Brain Injury, Prognostic research, Outcome, Methodology

No head injury is too severe to be despaired of, nor too trivial to be ignored.
—Hippocrates

Prognosis is an essential element of medicine, and estimates of prognosis are a frequent component in clinical decision making. Therapeutic and diagnostic actions all aim to improve prognosis. In ancient Greece, the quality of care was judged not so much by the result of treatment, but rather if the result was as the doctor had predicted. Much interest has been focused on prognosis following traumatic brain injury (TBI), but due to the heterogeneity of the condition, it has been considered difficult to say what the likely course of events will be in an individual patient. A seminal advance in the field of prognostic analysis in TBI was given by the Glasgow group in the 1970s, following the classic article on the Glasgow Coma Scale (GCS)[1], allowing quantification of impairment of consciousness, and the Glasgow Outcome Scale (GOS)[2], standardizing the assessment of outcome following severe brain damage. Over 40 years later, the GCS is still an integral part of clinical practice and research worldwide.[3]

The science of clinical decision making and advances in statistical modeling have made it possible to be more confident about what is likely to happen following a TBI, and to consider prognosis in terms of probabilities rather than prophecies. The availability of large databases has opened new opportunities for an evidence-based approach to prognostic analysis. In the current era, the availability of "big data," such as high-resolution data from the intensive care unit, might provide new opportunities for prognostic research.

Information about prognosis and predictive statements can be useful in a number of ways. Concern about likely outcome is often foremost in the mind of relatives; therefore realistic counseling is important. The place of prognosis in making decisions about the future management of individual patients is more controversial. Many neurosurgeons acknowledge that prognostic estimates have an important role in decision making, whereas others profess to attribute only a minor or even nonexistent role to prognosis, reflecting a range of attitudes arising from cultural and ethical differences as much as clinical convictions. Yet it is a fact that some form of estimation of prognosis is consciously or subconsciously used by physicians when allocating resources and prioritizing treatment—unfortunately, also now an increasing necessity in the high-income countries of the Western world. Caution remains appropriate in such circumstances.

Prognosis concerning an individual is informing about the expected person's future course of health, but outcome is further determined by the treatments chosen. Moreover, predictive equations can never include all items relevant to a particular

individual. Consequently, the estimated prognosis can be probabilistic only, whereas the outcome will be either favorable or unfavorable. The inherent uncertainty in prognosis needs to be considered when using prognostics models in clinical decision making in individual patients.

Estimates derived from evidence-based analysis of large data sets are preferable to relying on the gut feeling of a physician whose experience, no matter how vast, can never match the information contained in the data of thousands of patients entered into a database. Physician estimates of prognosis are often unduly optimistic, unnecessarily pessimistic, or inappropriately ambiguous.[4-7] Perhaps, however, the greatest application of prognostic analysis is not at the level of the individual patient, but at the "group" level for quantifying and classifying the severity of brain injury, as a reference for evaluating quality of care and for stratification and covariate adjustment in clinical trials.[8,9]

In this chapter, we summarize state-of-the-art approaches to prognostic analysis, review our current knowledge on prognosis in TBI, and discuss the development, application, and limitations of prognostic models for patients following TBI.

Approaches to Prognostic Analyses

In prognostic research we can distinguish between two types of studies: studies with a focus on the prognostic role of specific patient-related or disease-related characteristics in relation to outcome ("prognostic factor research"), and studies with a focus on the combined effect of various prognostic factors in predicting the outcome ("prediction model research").[10,11] In prognostic factor studies, we may start with assessing whether the factor is independently associated with the outcome of interest. Here "independently" refers to the association of the prognostic factor with the outcome separate from other prognostic indicators, and usually requires some form of statistical adjustment in the analysis. We might, for instance, be interested in the effect of motor score at admission on 6-month mortality. Typically, we first study univariable relations of the prognostic factor with the outcome of interest in a cross-table or regression model.

It should be stressed that a univariable association does not account for the role of other factors that might be more important for the observed association. The observed association therefore does not represent causality, and it may be secondary to other more relevant prognostic factors. Therefore univariable analysis is typically followed by multivariable regression analyses adjusting for confounding variables. Multivariable analysis allows us to explore the unique predictive value of that factor over and above

that of other predictors. Questions that require multivariable analysis are, for example: "What are the most important prognostic factors in a certain condition?" and "Are some prognostic factors correlated with each other, such that their apparent predictive effects are explained by other factors?" To perform multivariable analysis, multiple prognostic factors are included in the regression model as independent variables. Whereas prognostic factor research may provide insight in the relationship between a prognostic factor and the outcome of interest, prediction models aim to address a more pragmatic research question: "How well can we predict outcome based on a combination of prognostic factors?" Typically, combinations of prognostic factors are analyzed with multivariable models, followed by analyses of predictive performance, including measures for discrimination (e.g., concordance statistic) and calibration (e.g., graphics and calibration statistics). The relevance of a predictor is a function of the association of the predictor with the outcome, and the distribution of the predictor. For example, a dichotomous predictor with an odds ratio (OR) of 2.0 and 50% prevalence is more relevant for a prediction model than a dichotomous predictor, with an OR of 2.5 with 1% prevalence. The relationship between the predictors and the outcome can be quantified in several ways (Tables 399.1 and 399.2).

We often see that the positive predictive value is used as a measure for expressing prognostic performance. The positive predictive value, however, has limited usefulness because it does not take the frequency with which a predictor occurs within the population into account. Currently, the most widely used measure for expressing the strength of association in prognostic analysis is the OR, which can be obtained directly from the output of a regression model. In multivariable analysis, the ORs provided by the regression model are adjusted for the other predictors in the model. The coefficient of determination (R2) is also provided by the output of the regression model. The difference in R2 between a model without and with a certain predictor is the percentage of the variance that is explained by that predictor above the predictors in the former regression model and better represents both the prognostic strength and the frequency with which a predictor occurs within the population.

Methodologic Challenges in Prognostic Studies

Study Design and Sample Size

Prognostic studies are inherently longitudinal in nature, most often performed in cohorts of patients who are followed over time for an outcome to occur. The cohort is defined by the presence of one or more particular characteristics, such as having a certain disease or condition, living in a certain geographic region, or having a certain

TABLE 399.1: Performance Measures of Predictors

Measure	Definition	Interpretation
Relative risk (RR) $\frac{a/(a+b)}{c/(c+d)}$	Risk of outcome in group with predictor/Risk of outcome without predictor	For example, "RR of 2" means that the group with the predictor has twice the risk of the group without the predictor. When the predictor is continuous, RR represents the increase per unit.
Odds ratio (OR) $\frac{a \times d}{b \times c}$	Ratio of the odds for better versus poorer outcome in the presence of the parameter (a / b) compared to the odds in the absence of the parameter (c / d)	If the prognostic factor is not associated with outcome, the odds ratio will be 1. In reporting the odds ratio, the 95% confidence interval (CI) is frequently included. Statistical significance of the relationship is present if the CI does not include the value 1.
Nagelkerke R^2 $\frac{1-\exp\left(-\frac{LR}{n}\right)}{1-\exp\left(-\frac{L^0}{n}\right)}$	Model sum of squares (=parameter of regression model)/Total sum of squares (=parameter of the regression model)	Percentage of variability in the outcome that is explained by the predictors
Sensitivity $\frac{a}{a+c}$	Number of true positives/ Total number with the outcome	Proportion of patients with the outcome who have the predictor (true positive)
Specificity $\frac{d}{b+d}$	Number of true negatives/ Total number without the outcome	Proportion of patients without the outcome who do not have the predictor (true negative)
Positive predictive value $\frac{a}{a+b}$	Number of true positives/ Number of positives	Proportion of patients with the predictor who do have the outcome
Negative predictive value (NPV) $\frac{d}{c+d}$	Number of true negatives/ Number of negatives	Proportion of patients without the predictor who do not have the outcome

Data from Vittinghoff E. *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models*. New York: Springer; 2005; and Altman DG. *Practical Statistics for Medical Research*. New York: Chapman and Hall; 1991.

TABLE 399.2: 2 × 2 Table for Explanation of Performance Measures

	Dead	Alive
Predictor present	a	b
Predictor absent	c	d

age. For example, we might follow a cohort of adult patients admitted to the hospital with severe TBI to predict the risk of mortality 6 months postinjury.

A sufficient sample size is important to address any scientific question with empirical data. The effective sample size is mainly determined by the number of events in the study and not by the total number of subjects in a study. For example, when we study a disease with a 1% chance of mortality, a study with 1000 patients will contain only 10 events, and this number determines the effective sample size. It is important to match ambitions in research questions with the effective sample size that is available. When the sample size is very small, we should ask relatively simple questions. Questions that are more complex can be addressed with larger sample sizes.

Predictors

The choice for a predictor is based on subject matter knowledge: Is a certain factor expected to have an effect on outcome? The strongest predictors of mortality and unfavorable outcome following moderate and severe TBI are GCS motor score, age, and pupillary reactivity.[12] How the predictor should preferably be analyzed depends on the type; predictors can be continuous (age), ordinal (GCS), categorical (pupil reactivity), or binary (present/absent). Ideally, predictors are well defined, not too costly to obtain, and reliably measurable by any observer. In practice, observer variability is a problem for many measurements. In addition, some measurements are prone to biologic variability, and a single measurement may be misleading, as in the case of blood pressure.

In many studies, continuous or categorical predictors are collapsed into a binary variable, using threshold values. For example, the association between age and outcome has frequently been analyzed at a threshold value of 50. This approach has major disadvantages.[13] First, it is unnatural. Would risks be much different for patients who had their 50th birthday yesterday compared to patients having their 50th birthday tomorrow? In addition, a 30-year-old patient will have a different risk than a 49-year-old patient, yet both are below a threshold of 50.

Second, from a methodologic perspective, collapsing an ordinal or continuous scale into a binary variable (dichotomization) leads to loss of information and is therefore statistically inefficient.[14] In general, it is preferred to exploit the full information available and analyze the ordinal or continuous predictors. However, if a nonlinear function is expected based on clinical knowledge, several approaches can be considered in regression models, including polynomials, fractional polynomials, and splines.[15,16] These approaches leave the predictor continuous but allow a nonlinear relationship with the outcome.

Missing Data

Missing data are a common but underappreciated problem in medical scientific research. Missing values lead to a more limited set of patients with complete data compared to the ideal situation of complete original data.

The best approach is, of course, to prevent the occurrence of missing data. In a case of missing data, a common statistical approach is to delete patients with missing values from the analysis. This is often referred to as a *complete case analysis.*[17,18] Complete case analysis discards data from patients who have information on some, but not all, predictors. It is hence statistically inefficient, especially when we consider multiple predictors. Moreover, complete case analysis may lead to bias because of systematic differences between patients with complete data and patients with missing data. Bias occurs when missingness of a predictor is associated with the outcome.[19]

A more effective and sophisticated statistical approach to deal with missing values is single or multiple imputation.[15,20-22] Imputation methods substitute the missing values with plausible values so that the completed data can then be analyzed with.

With single imputation, missing values are substituted with the mean or with the mode, or based on a regression model, and only one completed data set is created. With multiple imputation procedures, *m* completed data sets are created. Multiple imputation is typically recommended, because single imputation ignores potential correlation of predictors and leads to an underestimation of variability of predictor values among subjects. Imputation methods are widely available in all standard software packages, and relatively easy to perform. As in any statistical analysis, the sensible judgment of the analyst is important, based on subject knowledge and the research question(s).

Outcome Measures for Prognostic Studies in Traumatic Brain Injury

In prognostic research, the outcome measure chosen should be clinically relevant, and "hard" end points are generally preferred. Mortality is often used as an end point in prognostic research, but global outcome measures (e.g., GOS and Extended Glasgow Outcome Scale [GOSE]), nonfatal events (e.g., disease recurrence), patient-centered outcomes (e.g., scores on quality-of-life questionnaires), or wider indicators of burden of disease (e.g., absence from work) also may be used. Whatever the end points chosen, assessment at a fixed time point is essential. Statistical power can also direct the choice of outcome. When an outcome is infrequent, it is not suited as an end point for statistical analysis.

Most, if not all, prognostic studies in TBI have used the GOS(E) or mortality as end points. In most cases the GOS, which can be considered an ordinal scale with five categories, was collapsed into a dichotomous variable, differentiating unfavorable versus favorable outcome (Table 399.3). With the use of a dichotomous outcome measure, statistical power is greatest when there is a 50:50 distribution between outcome categories. However, from a statistical point of view it is preferred to quantify prognostic effects across the full range of the GOS than after dichotomization into a binary variable.[23,24] The proportional odds methodology is appropriate for this purpose.[25] The eight-point GOSE has been introduced to increase sensitivity of outcome assessment. It should be noted that the potentially increased sensitivity of the GOSE is lost when this is again dichotomized to a binary scale.

Despite the increased sensitivity of the GOSE, it remains a global scale with broad categories aiming to capture functional reintegration without discriminating between physical and mental disabilities. Although TBI affects multiple outcome domains, current prognostic models cannot predict this range of outcomes. Therefore more granular outcome assessments have been proposed, including measures of cognitive functioning, neuropsychological tests, and quality-of-life assessments.[26-30] Further research should provide insight into which combination of outcome assessments is optimal for TBI—capturing all domains of outcome that are affected by TBI with sufficient sensitivity yet limiting the burden for patients of spending multiple hours

TABLE 399.3: Glasgow Outcome Scales

Score	Description
GLASGOW OUTCOME SCALE	
1	Dead
2	Vegetative
3	Severe disability (conscious but dependent)
4	Moderate disability (independent but disabled)
5	Good recovery (can resume normal activities)
EIGHT-POINT EXTENDED GLASGOW OUTCOME SCALE	
1	Dead
2	Vegetative
3	Lower severe disability
4	Upper severe disability
5	Lower moderate disability
6	Upper moderate disability
7	Lower good recovery
8	Upper good recovery

on tests. Development of multidimensional approaches to outcome classification should be a priority for future research.[31] This holds even more for mild TBI in which the GOSE is too insensitive to discriminate between patients with no, few, or more remaining symptoms after injury.[32]

Building Blocks for Prognostic Analysis

A wealth of literature has focused on the associations between predictors and outcome in univariable analysis. Most studies have concentrated on patients with severe and moderate TBI. Fewer studies have included multivariable analysis. The largest amount of evidence on univariable associations between predictors and outcome in moderate and severe TBI is provided by the International Mission for Prognosis and Analysis of Clinical Trials in TBI (IMPACT) study group. They reported results of extensive prognostic analysis performed in a meta-analysis of individual patient data from eight randomized controlled trials and three observational series, including a total of more than 9000 patients.[33] A series of papers reported on the relationship between the GOS and demographic characteristics,[34] cause of injury,[35] GCS and pupil response,[36] secondary insults,[37] blood pressure,[38] computed tomography (CT) scan features,[39] and laboratory parameters.[40] The results of multivariable analysis describing also the added predictive value were reported in the same series by Murray and colleagues.[41] The strength of the association of these predictors with outcome in both univariable and multivariable analyses as reported by the IMPACT investigators is summarized in Table 399.4. More recently, evidence on new predictors such as magnetic resonance imaging (MRI) features and genetic constitution has come from the Transforming Research and Clinical Knowledge in TBI study.[27]

We should further recognize that prognostic variables have mainly been identified in studies that used mortality or GOS as end points. It is plausible that other features may be more relevant for other outcomes (e.g., cognitive functioning, health- related quality of life) or when a multidimensional classification score is used as an end point.

Conceptually, the main predictors of outcome in TBI can be grouped together into "building blocks," some of which are modifiable and some not (Table 399.5). The current knowledge regarding these building blocks and parameters is summarized in the following sections.

Genetic Constitution

In this era of discovery of the human genome, several genes and their polymorphisms are under investigation in patients with TBI. However, examination of large numbers of genes results in high chance of type 1 error, underscoring the need for repeated studies of larger samples and high statistical power.[42] Furthermore, to be of prognostic

TABLE 399.4: Strength of the Association between Predictors and Outcome in TBI

Predictor	Reference Category	Univariable OR (95% CI)	Multivariable OR (Adjusted for A/M/P)
DEMOGRAPHICS			
Age	25%-75% IQR	2.14 (2.00-2.28)	—
Sex	Male	1.01 (0.92-1.11)	0.94 (0.85-1.04)
Race			
Black	Caucasian	1.30 (1.09-1.56)	1.44 (1.08-1.93)
Asian		1.09 (0.78-1.51)	1.22 (0.84-1.78)
CLINICAL SEVERITY			
Motor Score			
Absent	Localizing/Obey commands	5.30 (3.49-8.04)	—
Abnormal extension		7.48 (5.6-9.98)	—
Abnormal flexion		3.58 (2.71-4.73)	—
Flexion		1.74 (1.44-2.41)	—
Pupillary Reactivity			
One reacting	Both reacting	2.70 (2.07-3.53)	—
Both nonreacting		4.77 (3.46-6.57)	—
EXTRACRANIAL INJURIES			
Secondary Insults			
Hypotension	Absent	2.67 (2.09-3.41)	2.06 (1.64-2.59)
Hypoxia	Absent	2.08 (1.69-2.56)	1.65 (1.37-2.00)
Hypothermia	Absent	2.21 (1.56-3.15)	1.63 (1.11-2.40)
Structural Abnormalities			
CT Classification			
CT class I	CT class II	0.450 (0.350-0.067)	0.47 (0.32-0.70)
CT class III/IV	Absent	2.62 (2.13-3.21)	2.23 (1.83-2.72)
Mass lesion	No epidural	2.18 (1.83-2.61)	1.48 (1.27-1.71)
tSAH	Absent	2.64 (2.42-2.89)	2.01 (1.83-2.21)
Epidural hematoma	Absent	0.64 (0.56-0.72)	0.63 (0.55-072)
LABORATORY PARAMETERS			
Glucose		1.68 (1.54-1.83)	1.45 (1.36-1.55)
pH		0.80 (0.74-0.88)	0.84 (0.67-0.92)
Prothrombin time	25%-75% IQR	1.41 (0.99-1.99)	1.63 (1.40-1.89)
Hemoglobin		0.69 (0.60-0.78)	0.76 (0.66-0.88)
Sodium <137mmol/L	≥137 mmol/L	1.40 (1.22-1.60)	1.14 (0.91-1.43)

A/M/P, age/Glascow Coma Scale motor score/pupillary reactivity; CI, confidence interval; CT, computed tomography; IQR, interquartile range; OR, odds ratio; tSAH, traumatic subarachnoid hemorrhage.

TABLE 399.5: Building Blocks for Prognostic Analysis

Building Blocks	Items	Modifiable?
Genetic constitution	apoE	No
Demographics	Age, sex, race	No
Clinical severity	Glasgow Coma Scale score, pupillary reactivity, extracranial injuries	No
Secondary insults	Hypotension (blood pressure), hypoxia, hypothermia	Yes
Structural abnormalities	CT classification, traumatic subarachnoid hemorrhage, type of intracranial lesion	Sometimes
Laboratory parameters	Glucose, sodium, pH, coagulation parameters, hemoglobin	Yes
Biomarkers	Items under development	Uncertain
Omics	Items under development	Uncertain

CT, computed tomography.

value the independent effect of a gene on outcome beyond known predictors needs to be established. Thus large sample sizes and collection of comprehensive data, which allow for consideration of premorbid factors and assessment of injury severity, are essential.[43,44] This stadium has not been reached yet for most genes under study in TBI. Many studies have small sample sizes and fail to adjust for confounders.

However, evidence exists for an association between the presence of the *APOE* ε4 allele with poorer functional recovery.[45-47] The association of a common genetic variant within *ANKK1* with 6-month cognitive performance after TBI was also shown.[27] Other genes have been suggested to be associated with outcome, including the *P53, COMT, DND2,* and *CACNA1A* genes.[43]

Recent collaborative efforts such as the International Initiative for Traumatic Brian Injury Research (InTBIR) consortium provide opportunities for large-scale studies to explore the prognostic value of genetic constitution and biomarkers.[48]

Demographic Factors

Age is the strongest and one of the most extensively studied predictors of outcome in TBI. Many publications on the prognostic effects of age exist, all stating that greater age is correlated with poorer outcome (Fig. 399.1). It is remarkable that most studies have analyzed the association between age and outcome with threshold values. Different thresholds have been used, varying from 30 to 60 years of age.[49-59] Studies using higher threshold levels reported poorer outcome in the upper age group, and a mortality rate of greater than 75% has been described in patients over age 60 with

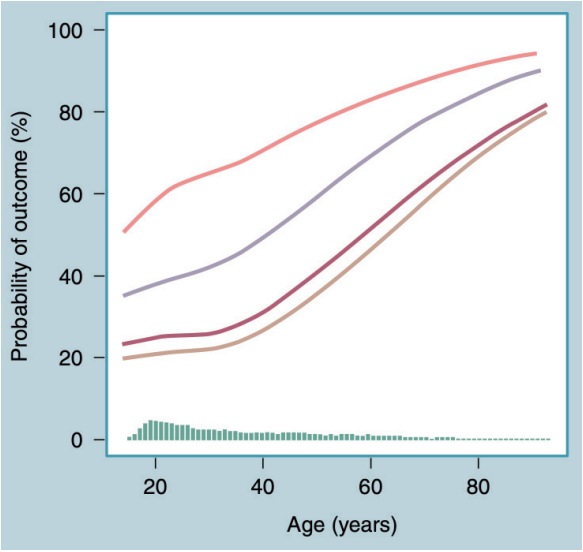


Figure 399.1: Continuous association between age and outcome. (From Mushkudiani NA, Engel DC, Steyerberg EW, et al. Prognostic value of demographic characteristics in traumatic brain injury: results from the IMPACT study. *J Neurotrauma.* 2007;24:259-269.)

severe TBI.[60-62] A continuous age dependency was described in only a few publications in the past.[63-65]

In the IMPACT study of more than 9000 patients, a continuous effect of age on outcome was described and could be approximated by a linear function. Threshold values could not be identified. However, the Corticosteroid Randomisation After Significant Head Injury (CRASH) study, which was of similar size, suggested an age effect starting at 40 years.[66]

Other demographic factors studied for their association with outcome after TBI include sex, race, and education. Men are more prone to suffer from TBIs because of a higher risk of road traffic accidents and assaults, and tend to acquire TBIs at a younger age. However, at an older age, a higher proportion of women suffer from fall-related TBIs.[67] Furthermore, women report worse 6-months outcomes than men.[68] The size of the differences in outcomes between men and women depends on TBI severity and age, and is more pronounced after mild TBI. The possible association between race and outcome after TBI has not been extensively studied. Two smaller studies showed poorer outcome in black patients,[69,70] but others did not find a clear association.[56,71,72] The IMPACT study group, however, studying data from 5320 patients, found a statistically significant association between race and outcome, with black

patients having a poorer outcome.[34] It was hypothesized that this might be due to differences in severity or cause of injury; however, this proved not to be the case and, following adjustment for cause of injury, age, motor score, and pupils, the prognostic effect was even stronger (Table 399.6). The response to injury may be different, because access to acute[73] and postacute[71]care may be more limited for black patients.

A weak association between level of education and outcome has been reported.[34,70] Mushkudiani and colleagues,[34] however, clearly showed that this weak association could be explained by other factors and disappeared upon adjustment.

TABLE 399.6: Strength of the Association between Race and Outcome in Traumatic Brain Injury

Type of Analysis	OR	95% CI
Univariable	1.30	1.09-1.56
Adjusted for cause of injury	1.31	1.08-1.57
Adjusted for age, motor score, and pupils	1.44	1.08-1.93
Adjusted for seven clinical predictors*	1.45	1.07-1.96

CI, confidence interval; OR, odds ratio.
*Seven predictors: age, Glasgow Coma Scale motor score, pupils, hypoxia, hypotension, computed tomography classification, traumatic subarachnoid hemorrhage.

Clinical Severity

Clinical severity is an important prognostic factor that, in theory, can be assessed in all patients. The severity relates to both extra- cranial and intracranial injuries. The overall severity of extracranial injuries is commonly assessed with the Abbreviated Injury Score[74] or the Injury Severity Score.[75] The prognostic value of extracranial injuries in TBI patients has been strongly debated in the literature for years. However, a meta-analysis of almost 40,000 patients showed that extracranial injury is an important prognostic factor for mortality in patients with TBI.[76] The effect varies by population, which explains the controversy in the literature. The strength of the effect is smaller in patients with more severe brain injury and depends on time of inclusion in a study.[76]

The clinical severity of intracranial injuries is reflected by the level of consciousness, assessed with the GCS.[1] Many studies have demonstrated an association between lower levels of the GCS and poorer outcome. In patients with more severe injuries, the motor component of the GCS has the greatest predictive value, because in these patients eye and verbal scores are commonly absent. It should be recognized that the GCS score can fluctuate early after injury, with some patients deteriorating and others improving. However, the prognostic effect of GCS and pupil reactivity appears to be robust for the time of assessment.[77] Reliable assessment with the GCS is further

increasingly obscured in the acute setting by confounders such as medical sedation, paralysis, or intoxication.[36,78,79]

Marmarou and colleagues[36] reported a stronger association with outcome for an abnormal extensor motor response compared to an absent motor response. The most likely explanation for this is that the category of patients scored as having an absent motor reaction will include "false-absent" scores, because of confounding effects of sedation and paralysis.[79]

Abnormalities in pupillary reactivity reflect brainstem compression, and they are strongly associated with poorer outcome.[36] Marmarou and colleagues[36] reported that pupillary reactivity, being less prone to influences of sedation and paralysis, was a more stable parameter in the early phase after injury than the GCS score.[36]

Secondary Insults

The injured brain is more vulnerable for systemic secondary insults, such as hypoxia and hypotension, than is a healthy brain. In experimental and clinical situations, the occurrence of secondary insults increases the degree of secondary damage after injury. The presence of secondary insults is associated with poorer outcome,[50,80,81] and the depth, duration, and number of hypotensive insults all contribute to poorer outcome.[37,82,83] Most studies have focused on early hypotensive and hypoxic events in which hypotension was defined as any episode with a systolic blood pressure less than 90 mm Hg. The association between the actual blood pressure on admission and outcome has been analyzed further in a continuous way by the IMPACT study group.[38] These studies, incorporating data from 6801 patients, show that the relation between blood pressure and outcome is continuous; low blood pressure and high blood pressure are both associated with poorer outcome. After adjusting for age, motor score, and pupillary reactivity, the effects of higher blood pressure largely disappeared, indicating that this association is most likely secondary to increasing severity of the injury. Various studies have shown that the combination of hypoxia and hypotension has a greater adverse effect on outcome than can be explained by either insult alone; however, the effects appear to be subadditive rather than synergistic.

Spreading depolarizations (SDs), which are pathologic waves of spreading mass neuronal depolarization arising in the injured gray matter, have been associated with unfavorable outcome. In 109 adults who needed neurosurgery for acute TBI, it was shown that patients with SDs had an increased risk of unfavorable outcome compared with patients without SD.[84] This finding was confirmed in a follow-up study in 138 patients.[85] However, SDs are not (yet) routinely measured, which limits their value for clinical decision making.

Structural Abnormalities

CT scanning is the investigation of choice in the acute phase after TBI to identify the presence and extent of structural damage. The relevance of CT scanning for the purpose of classification and prediction has increased with the growing difficulties in reliable assessment of clinical severity according to the GCS due to confounding effects of sedation and mechanical ventilation.[78,86,87] The prognostic value of individual CT characteristics in TBI is well documented, including status of basal cisterns, midline shift, the presence and type of intracranial lesions, and traumatic subarachnoid hemorrhage. In 1991, Marshall and colleagues[88] introduced a descriptive system of CT classification (Table 399.7) that focuses on the presence or absence of a mass lesion and differentiates diffuse injuries by signs of increased intracranial pressure (ICP; compression of basal cisterns, midline shift).

TABLE 399.7: Marshall CT Classification

Category	Definition
Diffuse injury I	No visible pathology
Diffuse injury II	Cisterns present, midline shift 0-5 mm and/or lesion densities present or no mass lesion >25 mL
Diffuse injury III (swelling)	Cisterns compressed or absent with midline shift 0-5 mm or no mass lesion >25 mL
Diffuse injury IV (shift)	Midline shift >5 mm, no mass lesion >25 mL
Evacuated mass lesion	Any lesion surgically evacuated
Nonevacuated mass lesion	High- or mixed-density lesion >25 mL, not surgically evacuated

From Marshall LF, Marshall SB, Klauber MR. A new classification of head injury based on computerized tomography. *J Neurosurg.* 1991;75:S14-S20.

This classification is also strongly related to outcome, with the poorest prognosis in patients with CT category IV (Signs of raised ICP + Shift) and the best outcome in patients without visible structural abnormalities. The Marshall CT classification has limitations, such as the broad differentiation between diffuse injuries and mass lesions, and the lack of specification of the type of mass lesion. Thus this classification can mask signs of raised ICP in addition to a mass lesion, and it does not fully use the prognostic information contained in the individual CT characteristics scored. Maas and colleagues[89] proposed a score chart for assessing the risk of poorer outcome based on individual CT characteristics, and they showed that this resulted in better discrimination between patients with better versus poorer outcome than the descriptive Marshall classification (Table 399.8). This advantage was confirmed in subsequent work from Flint and colleagues.[90] More recently, the Stockholm CT score

and Helsinki CT score have been developed. The Stockholm CT score was published in 2010 and was developed through retrospective analyses of 861 neuro–intensive care unit TBI patients at a single center between 1996 and 2001.[91] This score includes midline shift as a continuous parameter, CT-verified diffuse axonal injury in the basal ganglia or brainstem, presence of epidural hematoma, presence of dual-sided subdural hematoma, and a subscore of increasing amounts of subarachnoid hemorrhage and/or presence of intraventricular hemorrhage, as predictors of mortality and unfavorable outcome. The Helsinki CT score, published in 2014, was developed on a single-center retrospective analysis of 869 consecutive neuro–intensive care unit TBI patients and included bleeding type and size, intraventricular hemorrhage, and status of suprasellar cisterns as variables.[92]

TABLE 399.8: Rotterdam Prognostic CT Score

Predictor Value	Score
Basal Cisterns	
Normal	0
Compressed	1
Absent	2
Midline Shift	
No shift or shift ≤5 mm	0
Shift >5 mm	1
Epidural Mass Lesion	
Present	0
Absent	1
Intraventricular Blood or tSAH	
Absent	0
Present	1
Sum score*	Σ + 1

tSAH, traumatic subarachnoid hemorrhage.
*Sum score can be used to obtain the predicted probability of mortality from the formula below. We chose to add "plus 1" to make the grading numerically consistent with the grading of the motor score of the Glasgow Coma Scale and with the Marshall CT classification. The corresponding probabilities are calculated with the formula: Probability (mortality) = $1/[1+e^{-(-2.60 + 0.80 \times \text{sumscore})}]$

The Stockholm and Helsinki CT scores provide more detailed information on structural abnormalities following TBI than earlier classification systems. The Stockholm CT score predicted mortality and unfavorable outcome following TBI more accurately than its precursors, the Marshall and Rotterdam CT scores.[91] Discrimination of the Helsinki CT score was reported to be slightly better than that of the Rotterdam CT

score.[92] In a recent study comprising 1.115 neuro-intensive care TBI patients from Stockholm and Helsinki, both the Stockholm and Helsinki CT scores outperformed the Rotterdam CT score and Marshall CT classification systems.[93] The subarachnoid hemorrhage and diffuse injury markers were found to be stronger outcome predictors in comparison to focal injury markers such as mass lesions or basal cistern compression.

The prognostic relevance of traumatic subarachnoid hemorrhage was extensively described[94] following extensive analysis of the nimodipine studies. Later work confirmed the presence of traumatic subarachnoid hemorrhage as one of the strongest CT predictors of outcome in TBI. Most studies, however, have concentrated on the presence or absence of traumatic subarachnoid hemorrhage without differentiating as to the location (basal cisterns versus cortical) or extent. Cortical traumatic subarachnoid hemorrhage is frequently associated with underlying contusions and probably has a relevance different from traumatic subarachnoid hemorrhage in the basal cisterns, which can incur an increased risk for vasospasm. More detailed information about acute TBI lesions might improve outcome prediction. For instance, specific National Institute of Neurological Disorders and Stroke standardized imaging-based pathoanatomic descriptors were able to discriminate between patients with favorable and unfavorable outcomes 6 months after TBI, which indicates their potential added value in prediction models for outcome following TBI.[95]

In the past, technical possibilities in MRI have improved rapidly. This has also increased the interest in the prognostic value of MRI features. However, because MRI is still relatively costly, time consuming, and often a logistic challenge, large patient series are lacking. Recent work from the Transforming Research and Clinical Knowledge in TBI (TRACK-TBI) study showed that the addition of early CT and MRI markers to a prognostic model based on previously known demographic, clinical, and socioeconomic predictors resulted in a greater than twofold increase in the explained variance in 3-month GOSE.[96] However, the MRIs were made on average 14 days after injury and thus by default captured more prognostic information compared to baseline variables. Nevertheless, we consider it likely that in patients with mild TBI, biomarker levels in the blood may form indications for performing MRI.

Laboratory Parameters

Relatively few studies have investigated the relation between laboratory parameters on admission and outcome. This may seem surprising, because analysis of the prognostic value of laboratory parameters is particularly relevant as these are routinely measured and objective. Abnormal values may also be modifiable. The results from the IMPACT studies have shown that the addition of laboratory values to a prognostic

model increases discrimination. These studies showed the greatest discriminatory properties for coagulation abnormalities and glucose (see Table 399.5). Although laboratory values may be modifiable, the observed association between abnormal values and poorer outcome does not mean that correcting these abnormal values will indeed improve outcome. The observed abnormality may simply be an expression or surrogate marker of the severity of injury. Prospective studies, preferably randomized controlled trials, are required to prove such an effect. Currently, there is great interest in various biomarkers released from damaged or necrotic neurons and glial cells in the brain, such as S-100 calcium-binding protein B, ubiquitin C-terminal hydrolase L1, and glial fibrillary acidic protein. Various experimental and preliminary clinical studies have confirmed this potential.[28,42,97-101] A recent study has shown the incremental value of blood biomarkers for the prediction of CT abnormalities following TBI.[102] However, most clinical studies have focused on the diagnostic instead of prognostic value of these biomarkers. The value of biomarkers in clinical prediction models, and specifically their added value over known predictors, remains to be established in large studies.

Mild Traumatic Brain Injury

It should be noted that the knowledge regarding predictors of mortality and unfavorable outcome mainly applies to moderate and severe TBI. For mild TBI, the body of scientific evidence on prognostic factors is much smaller. Age appears to be a predictor of poor outcome,[32,66,103] as is more severe TBI. Preexisting psychiatric conditions are less often studied, but also have been found to predict poorer outcome.[32,104,105] Although speculative, it is possible that individuals with a preexisting mental health condition may have less reserve to overcome the additional strain of a mild TBI. Alternatively, symptoms that relate primarily to this comorbidity can falsely be attributed to the head injury.[106] Lower education has also been found to be predictive of worse outcomes after mild TBI. More highly educated patients may have more adaptive coping skills that allow them to return to their previous levels of functioning.[107]

Overall, important predictors in moderate and severe TBI, such as GCS, pupillary reactivity, and CT parameters, are less relevant in mild TBI. Here, indicators of social background, psychological factors, history of psychiatric conditions, and low education seem to be more predictive of poorer outcome.[105,108] It therefore has been suggested that in moderate and severe TBI, outcome is determined by what "the injury brings to the patient," whereas in mild TBI it is what "the patient brings to the injury."[32] The outcome measure for prognostic studies on mild TBI also deserves consideration. Previous studies in patients after mild TBI have mostly used the GOSE as a primary outcome measure. Although levels of disability on the GOSE are important, they remain a somewhat coarse measure of the consequences of injury in

patients with milder forms of TBI. Persisting symptoms, mental health problems, and specific functional limitations, however, may provide a more fine-grained way of identifying sequelae. A multidimensional construct, taking into consideration the patient's perspective of well-being and quality of life as well as the frequency and intensity of postconcussion symptoms, including fatigue and pain, would perhaps be more appropriate to study outcomes after mild TBI.

Development of Prognostic Models

Prediction models provide diagnostic and prognostic probabilities, and as such form an increasingly important tool in clinical medicine. From the perspective of prognosis in TBI, we want to estimate (or predict) the risk for death or unfavorable outcome of a certain patient admitted to the hospital with moderate or severe TBI. Single predictors often have insufficient predictive value to distinguish patients who will do well from those who will do poorly. Moreover, patients can have different characteristics that affect the prognosis in opposite directions. For example, for a 24-year-old patient with fixed pupils, we would predict a favorable outcome based on age, but an unfavorable outcome based on pupil reactivity. Thus estimation in prediction research is by definition a multivariable challenge in which multiple risk factors need to be considered jointly with multivariable analysis. To this purpose, relevant prognostic factors are combined in a prediction model and often presented as rules or nomograms. The most common approach to develop a prediction model is regression analysis. Machine learning approaches, such as support vector machine, random forest, and neural net algorithms, are becoming increasingly popular, but do not outperform regression analysis in prediction in TBI.[109,110]

When we consider the development of a new prediction model, we should first consider if it is appropriate to develop one. Key issues that should be addressed are: for whom is prediction needed? Are high-quality data available? Is no existing prediction model available to validate or update? Is sample size adequate? Are predictors known and commonly available in the setting of application?[15] If model development is pursued, a prediction model needs to meet stringent quality criteria to be useful for doctors in clinical practice. Valid development is important, with specific attention to prevention of overfitting. Overfitting means that the model described fits the study population well, but is unlikely to give reliable predictions for new patients. Internal and external validation should therefore be considered mandatory.[111]

Various prediction models have been proposed for use in TBI. Although guidelines for developing and reporting have been proposed,[11,112,113] several reviews showed many shortcomings in model development and lack of external validation.[12,114-116] The methodologic shortcomings in development of prognostic models could be considered

one of the major factors that have delayed more general acceptance of such models in clinical practice. Specific attention to the methodology of model development is therefore required. Seven logically distinct steps in the development of valid prediction models[15] with regression analysis are:

1. *Problem definition and data inspection:* What is the research question? What is the outcome of interest? What is already known about predictors? Are there missing values?

2. *Coding of predictors:* Do categorical and continuous variables require recoding? Dichotomization of a continuous predictor should be discouraged and a continuous approach to analysis preferred.

3. *Model specification:* What predictors should we include, considering what is known about predictors already, and what is observed in the data under study? The number of predictors that can be considered for inclusion in a prognostic model should be limited to prevent overfitting. This number can be approximated by dividing the number of events (outcome) by a factor 10 (also known as the events per variable rule). To reduce the set of predictors, stepwise selection methods are widely used, but they have many disadvantages, such as instability of the selection, bias in estimated regression coefficients ("testimation bias"), and underestimation of uncertainty in the selected model.

4. *Model estimation:* Estimation of model parameters is commonly done by regression analysis. Modern techniques have been developed that aim to reduce overfitting of a model to the available data. Examples of these techniques are the Least Absolute Shrinkage and Selection Operator (LASSO), ridge regression, and elastic net. LASSO penalizes for the absolute values of the regression coefficients. It shrinks coefficients of predictors with lesser contributions to zero, which means that predictors are dropped from the model.

5. *Model performance:* For a proposed model, we need to determine its quality with measures for model discrimination and calibration. *Discrimination* refers to the ability of a prediction model to separate subjects with and without the outcome; this can be quantified by the area under the receiver operating characteristic curve. This curve shows the relationship between sensitivity and specificity. *Calibration* refers to the reliability of predictions. If we predict 10%, on average 10% of the subjects with this prediction are expected to experience the outcome. Overall model performance measures include the R2 and the Brier score. It has been proposed

that the clinical usefulness of a model should also be quantified with, for example, decision curve analysis.[115,117]

6. *Model validity:* Because overfitting is a central problem in prediction modeling, we need to consider the validity of our model for new subjects rather than for those in the data set used for model development. Several statistical techniques are available to evaluate the internal validity of a model—that is, for the underlying population from which the data set was sampled. Internal validation can address statistical problems in the specification and estimation of a model (i.e., reproducibility). Split-sample validation, in which the derivation cohort is split into a development sample and a validation sample, is a common, but inefficient approach.[118] Recommended methods are cross-validation and bootstrap resampling procedures.[111]

7. *Model presentation:* A final step to consider is the presentation of a prediction model. Regression formulas can be used, but many alternatives are possible for easier applicability of a model, including score charts, nomograms, and web-based calculators.[119,120]

Validation of Prognostic Models

When a model is developed, the next requirement is external validation of the model. External validity relates to the generalizability of the prognostic model to another population. In other words, in external validation studies we aim to assess how the model performs in new patients and settings. We may, for instance, be interested in model performance in patients from the same center but over a different time period (temporal validation), in patients from different centers or countries (geographic validation), or in patients who differ from the derivation cohort on a particular characteristic, such as age, severity, or setting (domain validation). Validation, preferably across a range of settings, is required before application of a model can be considered. In external validation studies we can distinguish between three steps: (1) investigation of the extent of relatedness between the development and validation sample, (2) assessment of model performance, and (3) interpretation of model validation results.121 Interpretation of model validation results is important, because at external validation model performance is typically reduced, and the reasons for reduced performance need to be understood. Updating, adjusting, or recalibrating the model should be attempted before considering the development of yet another new model. Various methods are available to update models to a specific setting, including calibration-in-the large, adjustment of all regression coefficients, updating of individual predictor effects, and extending the model with new predictors.[15]

Application of Prognostic Models in Clinical Practice

The ability to accurately predict patient outcome after TBI is important in clinical practice and research. Outcome prognostication may, for instance, support clinicians in providing reliable information to patients and relatives, guide clinical management, and give insight into quality of care by comparing observed and expected outcomes, and can be used for risk stratification of patients and covariate adjustment in randomized controlled trials.

If a prognostic model is deemed appropriate for implementation, the clinical impact of the model should be studied. This means that we need to consider an impact analysis, wherein a prediction model is used as a decision rule and any improvement in physicians' decisions is determined.[122] Although validation of a prediction model can indicate the efficacy of the model across a range of settings, impact analysis is required to indicate the (cost-)effectiveness of the implementation of the model in practice.

Illustration of Prognostic Models after Moderate and Severe Traumatic Brain Injury

Over the years, many prediction models for moderate and severe TBI have been proposed. Two of these prediction models were developed using large patient series and have frequently been externally validated.[12] These models concern those presented by the MRC CRASH trial collaborators[123] and those proposed by the IMPACT study group[124] (Table 399.9). The CRASH models also include patient data from low- and middle-income countries. Importantly, both CRASH and IMPACT models were developed from data available upon admission, before providing specialist care. These models are therefore ideally suited for a baseline calculation of prognostic risk. Both models showed adequate performance in terms of discrimination and calibration. Both approaches confirmed that the largest amount of prognostic information was contained in a core set of three predictors: age, motor score, and pupillary reactivity. The IMPACT study group further evaluated the additional benefit of adding more building blocks, such as structural imaging (CT characteristics), secondary insults, and laboratory data. Slightly better performance was noted in models that included CT and laboratory data. The IMPACT and CRASH models are available online: http://www.tbi-impact.org and http://www.trialscoordinatingce ntre.lshtm.ac.uk/Risk%20calculator/index.html.

Table 399.9: Predictors of the CRASH and IMPACT Models

CRASH basic	CRASH CT	IMPACT core	IMPACT extended	IMPACT lab
Age, GCS total score, Pupillary reactivity, Major extracranial injury	Basic model predictors + Petechial hemorrhages, Obliteration of 3rd ventricle or basal cisterns, tSAH, Midline shift > 5 mm, Non-evacuated hematoma	Age, GCS motor score, Pupillary reactivity	Core model predictors + Hypoxia, Hypotension, Marshall CT classification, tSAH, EDH	Extended model predictors+ Glucose, Hemoglobin

Abbreviations: EDH, epidural hematoma; tSAH, traumatic subarchnoid hemorrhage.

The IMPACT model is also presented as a simple score chart for sequential application of the models (Fig. 399.2).

This score chart can be used to obtain an approximate prediction in individual patients. The predictive risk can then be derived by reading the predicted probability from nomograms124 (Fig. 399.3).

The CRASH trial collaborators and the IMPACT investigators reciprocally validated their prognostic models externally on the other data set, confirming good performance. In the last decade, the IMPACT and CRASH prognostic models have been externally validated by various research groups. Although results vary slightly across settings, the external validity of both the IMPACT and CRASH models is continuously confirmed.[125-136] The IMPACT and CRASH models have, for instance, been externally validated in contemporary TBI patients across Europe using the CENTER-TBI study.[136] The models showed adequate discrimination, but variable calibration. Adequate discrimination indicates that the models can identify patients at high risk for mortality or unfavorable outcome, which supports their use in research settings and for benchmarking in the context of quality- of-care assessment. Variable calibration means that there were discrepancies between observed and predicted rates of mortality and unfavorable outcome. Therefore adjustment of the models to local hospital and patient characteristics is strongly recommended before the models can be considered for practical application.

Direct application of a model in a new patient population should be done with caution and with consideration of the comparability between the development and application setting. It is up to the clinical field to adopt prognostic models for general clinical application, aiming at improving quality of care, challenging each physician to beat the prognostic estimate. However, the greatest application of prognostic models is not at the level of the individual patient, but at the “group” level to quantify and

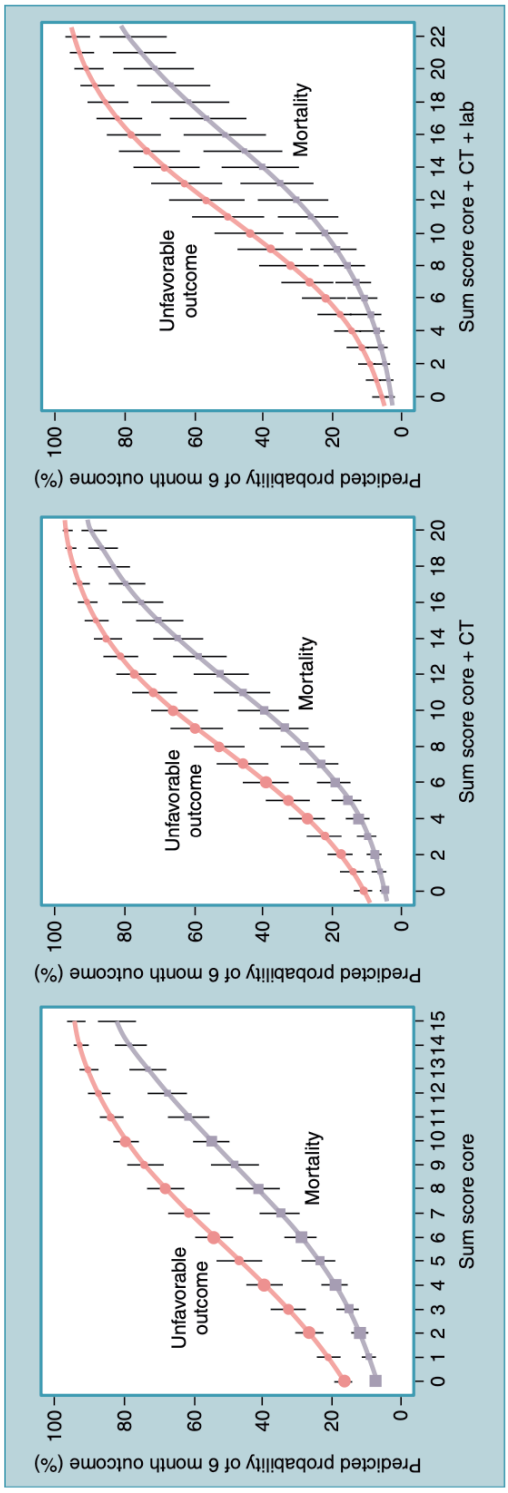


Figure 399.2: Impact model nomograms. CT, Computer tomography.

Characteristics	Value	Score	Sum
Age (years)	≤30	0	
	30-39	1	
	40-49	2	
	50-59	3	
	60-69	4	
	70+	5	
Motor score	None, extension	6	
	Abnormal flexion	4	
	Normal flexion	2	
	Localizes, obeys	0	
	Untestable, missing	3	
Pupillary reactivity	Both pupils reacted	0	
	One pupil reacted	2	
	No pupil reacted	4	
SUM SCORE CORE MODEL			
Hypoxia	Yes or suspected	1	
	No	0	
Hypotension	Yes or suspected	2	
	No	0	
CT classification	I	–2	
	II	0	
	III/IV	2	
	V/VI	2	
Traumatic subarachnoid hemorrhage	Yes	2	
	No	0	
Epidural hematoma	Yes	–2	
	No	0	
SUBSCORE COMPUTED TOMOGRAPHY			
SUM SCORE EXTENDED MODEL			
Glucose (mmol/L)	<6	0	
	6-8.9	1	
	9-11.9	2	
	12-14.9	3	
	15+	4	
Hemoglobin (g/dL)	<9	3	
	9-11.9	2	
	12-14.9	1	
	15+	0	
SUBSCORE LABORATORY			
SUM SCORE LABORATORY MODEL			

Figure 399.3: Score chart IMPACT model. CT, computed tomography.
Sum scores can be calculated for the core model (age, motor score, pupillary reactivity), the extended model (core + hypoxia + hypotension + CT characteristics), and a lab model (core + hypoxia + hypotension + CT + glucose + Hb). The probability of 6 mo outcome is defined as $1 / (1 + e^{-LP})$, where LP refers to the linear predictor in a logistic regression model. Six LPs were defined as follows:
$LP_{core,\ mortality} = -2.55 + 0.275 \times$ sum score core
$LP_{core,\ unfavorable\ outcome} = -1.62 + 0.299 \times$ sum score core
$LP_{extended,\ mortality} = -2.98 + 0.256 \times$ (sum score core + subscore CT)
$LP_{extended,\ unfavorable\ outcome} = -2.10 + 0.276 \times$ (sum score core + subscore CT)
$LP_{lab,\ mortality} = -3.42 + 0.216 \times$ (sum score core + subscore CT + subscore lab)
$LP_{lab,\ unfavorable\ outcome} = -2.82 + 0.257 \times$ (sum score core + subscore CT + subscore lab)

classify the severity of brain injury, as a reference for evaluating quality of care and for stratification and covariate adjustment in clinical trials.

Few prognostic models have been published for adult patients following mild TBI, and these performed unsatisfactory in external validation.[32,137,138] Development and validation of valid prediction models for patients following mild TBI requires further research efforts.

Future Directions

The knowledge about prognosis after TBI has expanded tremendously in the last decade, specifically for moderate and severe TBI. Prognostic models allow researchers and clinicians to better predict patient outcomes following TBI despite the heterogeneity of TBI populations. Still, current robust and well-validated prognostic models for moderate and severe TBI "only" explain up to 35% of the variance in outcome.[123,124] Therefore other key patient and injury characteristics could improve prognostication.

Various directions for prognostic research in TBI have been identified.[30] Improved predictions may come from new biomarkers; however, despite substantial research efforts in this direction, the added prognostic value of novel biomarkers has not been demonstrated convincingly. The same holds for genetic composition. Another direction is advanced imaging, including MRI, but MRI in the acute phase of severe TBI may be too logistically challenging to be implemented in routine clinical practice in the near future.

Prognostic research in TBI has focused on predictors available at baseline. Beyond doubt, additional prognostic information is captured in the clinical course in the first days after injury, but this is still largely unexplored. Such "dynamic prediction" can include repeated measures of predictors such as serum biomarkers. Because of large collaborative efforts, more big data are available in TBI, including high-resolution intensive care unit data. However, for some of the applications of prognostic models, such as covariate adjustment in randomized controlled trials and for evaluating quality of care, only predictors available at baseline can be considered.

Thus far, prognostic studies have mainly used the GOSE as outcome. Because TBI affects multiple outcome domains, the use of more granular outcome assessments has been proposed. Prognostic models are required, beyond the currently established models for GOS and GOSE, that predict cognitive, psychosocial, health-related quality-of-life, and other patient-reported outcomes.

For mild TBI the use of more granular outcomes will be even more important, because the GOSE is too insensitive in this population. Relevant outcomes following mild TBI include persistent postconcussive symptoms and return to normal life without TBI-related symptoms. Because prognostic models for mild TBI are less well established, research should focus on the development and validation of prognostic models starting with readily available predictors. Patient characteristics, such as level of education and preinjury mental health problems, may be more important in this population than injury-related characteristics and should therefore be considered in the development of prognostic models for mild TBI. TBI-related and psychological symptoms collected at 2 weeks could further improve prediction models.

The expanding knowledge of the effects of risk factors and predictors of TBI outcome awaits the ongoing initiatives of current multicenter studies, such as the TRACK-TBI study and the CENTER-TBI study,[134] which enrolled 3618 and 4509 patients, respectively, using an extensive standardized set of variables. These studies will allow us to converge and leverage research efforts to achieve the sample sizes needed to bring prognostic research in TBI again a step further.

References

1. Teasdale G, Jennett B. Assessment of coma and impaired conscious- ness. A practical scale. *Lancet*. 1974;2:81–84.
2. Jennett B, Bond M. Assessment of outcome after severe brain dam- age. *Lancet*. 1975;1:480–484.
3. Teasdale G, Maas A, Lecky F, et al. The Glasgow Coma Scale at 40 years: standing the test of time. *Lancet Neurol*. 2014;13:863.
4. Barlow P, Teasdale G. Prediction of outcome and the management of severe head injuries: the attitudes of neurosurgeons. *Neurosurgery*. 1986;19:989–991.
5. Chang RWS, Lee B, Jacobs S. Accuracy of decisions to withdraw therapy in critically ill patients: clinical judgement versus a computer model. *Crit Care Med*. 1989;17:1091–1097.
6. Dawes RM, Faust D, Meehl RE. Clinical versus actuarial judgement. *Science*. 1989;243:1668–1674.
7. Kaufmann MA, Buchmann B, Scheideggeer D, et al. Severe head injury: should expected outcome influence resuscitation and first day decisions? *Resuscitation*. 1992;23:199–206.
8. Maas AI, Murray GD, Roozenbeek B, et al. Advancing care for traumatic brain injury: findings from the IMPACT studies and perspectives on future research. *Lancet Neurol*. 2013;12:1200–1210.
9. Lingsma HF, Roozenbeek B, Steyerberg EW, et al. Early prognosis in traumatic brain injury: from prophecies to predictions. *Lancet Neurol*. 2010;9:543–554.
10. Riley RD, Hayden JA, Steyerberg EW, et al. Prognosis research strategy (PROGRESS) 2: prognostic factor research. *PLoS Medicine*. 2013;10(2):e1001380.
11. Steyerberg EW, Moons KGM, van der Windt DA, et al. Prognosis research strategy (PROGRESS) 3: prognostic model research. *PLoS Medicine*. 2013;10:e1001381.
12. Dijkland SA, Foks KA, Polinder S, et al. Prognosis in moderate and severe traumatic brain injury: a systematic review of contemporary models and validation studies. *J Neurotrauma*. 2020;37(1):1–13.
13. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med*. 2006;25:127–141.
14. Altman DG, Royston P. The cost of dichotomosing continuous variables. *BMJ*. 2006;332:1080.
15. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. 2nd ed. New York: Springer; 2008.
16. Collins GS, Ogundimu EO, Cook JA, Manach YL, Altman DG. Quantifying the impact of different approaches for handling continuous predictors on the performance of a prognostic model. *Stat Med*. 2016;35(23):4124–4135.
17. Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. 2nd ed. Hoboken, NJ: Wiley; 2002.
18. Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. *Am J Epidemiol*. 1995;142:1255–1264.
19. Vach W, Blettner M. Biased estimation of the odds ratio in case-control studies due to the use of ad hoc methods of correcting for missing values for confounding variables. *Am J Epidemiol*. 1991;134:895–907.
20. Marshall A, Altman DG, Royston P, Holder RL. Comparison of techniques for handling missing covariate data within prognostic modelling studies: a simulation study. *BMC Medical Research Methodology*. 2010;10(1):7.
21. Nielson JL, Cooper SR, Seabury SA, et al. TRACK-TBI investigators. Statistical guidelines for handling missing data in traumatic brain injury clinical research. *J Neurotrauma*. 2020.

22. Gravesteijn B, Sewalt C, Venema E, Nieboer D, Steyerberg EW. Missing data in prediction research: a five step approach for multiple imputation, illustrated in the CENTER-TBI study. *J Neurotrauma*. 2021;(ja).
23. McHugh GS, Butcher I, Steyerberg EW, 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.
24. Roozenbeek B, Lingsma HF, Perel P, et al. The added value of ordinal analysis in clinical trials: an example in traumatic brain injury. *Crit Care*. 2011;15:R127.
25. McCullaugh P. Regression-models for ordinal data. *J R Statist Soc Ser B Methodol*. 1980;42:109–142.
26. Wilde EA, Whiteneck GG, Bogner J, et al. Recommendations for the use of common outcome measures in traumatic brain injury re- search. *Arch Phys Med Rehabil*. 2010;91:1650–1660. e17.
27. Yue JK, Vassar MJ, Lingsma HF, 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.
28. McMahon P, Hricik A, Yue JK, et al. Symptomatology and functional outcome in mild traumatic brain injury: results from the prospective TRACK-TBI study. *J Neurotrauma*. 2014;31:26–33.
29. Von Steinbuechel N, Peteren C, Bullinger M, et al. Assessment of health-related quality of life in persons after traumatic brain injury—development of the Qolibri, a specific measure. *Acta Neurochir Suppl*. 2005;93:43–49.
30. Maas AIR, Menon DK, Adelson PD, et al. Traumatic brain injury: integrated approaches to improve prevention, clinical care, and re- search. *Lancet Neurol*. 2017;16(12):987–1048.
31. Manley GT, Maas AI. Traumatic brain injury: an international knowledge-based approach. *J Am Med Assoc*. 2013;310:473–474. 32. Lingsma HF, Yue JK, Maas AI, et al. TRACK-TBI Investigators.
32. Outcome prediction after mild and complicated mild traumatic brain injury: external validation of existing models and identification of new predictors using the TRACK-TBI pilot study. *J Neurotrauma*. 2015;32:83–94.
33. Marmarou A, Lu J, Butcher I, et al. IMPACT database of traumatic brain injury: design and description. *J Neurotrauma*. 2007;24:239–250.
34. Mushkudiani NA, Engel DC, Steyerberg EW, et al. Prognostic value of demographic characteristics in traumatic brain injury: results from the IMPACT study. *J Neurotrauma*. 2007;24:259–269.
35. Butcher I, McHugh GS, Lu J, et al. Prognostic value of cause of injury in traumatic brain injury: results from the IMPACT study. *J Neurotrauma*. 2007;24:281–286.
36. Marmarou A, Lu J, Butcher I, et al. Prognostic value of the Glasgow Coma Scale and pupil reactivity in traumatic brain injury assessed pre-hospital and on enrollment: an IMPACT analysis. *J Neurotrauma*. 2007;24:270–280.
37. McHugh GS, Engel DC, Butcher I, et al. Prognostic value of secondary insults in traumatic brain injury: results from the IMPACT study. *J Neurotrauma*. 2007;24:287–293.
38. Butcher I, Maas AI, Lu J, et al. Prognostic value of admission blood pressure in traumatic brain injury; results from the IMPACT study. *J Neurotrauma*. 2007;24:294–302.
39. 39. Maas AI, Steyerberg EW, Butcher I, et al. Prognostic value of computerized tomography scan characteristics in traumatic brain injury: results from the IMPACT study. *J Neurotrauma*. 2007;24:303–314.
40. Van Beek JG, Mushkudiani NA, Steyerberg EW, et al. Prognostic value of admission laboratory parameters in traumatic brain injury: results from the IMPACT study. *J Neurotrauma*. 2007;24:315–328.

41. Murray GD, Butcher I, McHugh GS, et al. Multivariate prognostic analysis in traumatic brain injury: results from the IMPACT study. *J Neurotrauma*. 2007;24:329–337.
42. Diaz-Arrastia R, Baxter VK. Genetic factors in outcome after traumatic brain injury: what the Human Genome Project can teach us about brain trauma. *J Head Trauma Rehabil*. 2006;21:361–374.
43. Jordan BD. Genetic influences on outcome following traumatic brain injury. *Neurochem Res*. 2007;32:905–915.
44. Dardiotis E, Fountas KN, Dardioti M, et al. Genetic association studies in patients with traumatic brain injury. *Neurosurg Focus*. 2010;28:E9.
45. Alexander S, Kerr ME, Kim Y, et al. Apolipoprotein E4 allele presence and functional outcome after severe traumatic brain injury. *J Neurotrauma*. 2007;24:790–797.
46. Zhou W, Xu D, Peng X, et al. Meta-analysis of APOE4 allele and outcome after traumatic brain injury. *J Neurotrauma*. 2008;25:279–290.
47. Zeng S, Jiang JX, Xu MH, et al. Prognostic value of apolipoprotein E epsilon4 allele in patients with traumatic brain injury: a meta-analysis and meta-regression. *Genet Test Mol Biomarkers*. 2014;18:202–210.
48. Tosetti P, Hicks RR, Theriault E, et al. Workshop Participants. Toward an international initiative for traumatic brain injury re- search. *J Neurotrauma*. 2013;30(14):1211–1222.
49. Vollmer DG, Torner JC, Eisenberg HM, et al. Age and outcome following traumatic coma: why do older patients fare worse? *J Neurosurg*. 1991;75:S37–S49.
50. Chesnut RM, Marshall LF, Klauber MR, et al. The role of secondary brain injury in determining outcome from severe head injury. *J Trauma*. 1993;34:216–222.
51. Gordon E, Von Holst H, Rudehill A. Outcome of head injury in 2298 patients treated in a single clinic during a 21-year period. *J Neurosurg Anesthesiol*. 1995;7:235–247.
52. Signorini DF, Andrews PJ, Jones PA, et al. Predicting survival using simple clinical variables: a case study in traumatic brain injury. *J Neurol Neurosurg Psychiatry*. 1999;66:20–25.
53. Ono J, Yamaura A, Kubota M, et al. Outcome prediction in severe head injury: analyses of clinical prognostic factors. *J Clin Neurosci*. 2001;8:120–123.
54. Andrews PJ, Sleeman DH, Statham PF, et al. Predicting recovery in patients suffering from traumatic brain injury by using admission variables and physiological data: a comparison between decision tree analysis and logistic regression. *J Neurosurg*. 2002;97:326–336.
55. Ratanalert S, Chompikul J, Hirunpat S, et al. Prognosis of severe head injury: an experience in Thailand. *Br J Neurosurg*. 2002;16:487–493.
56. Schreiber MA, Aoki N, Scott BG, et al. Determinants of mortality in patients with severe blunt head injury. *Arch Surg*. 2002;137:285–290.
57. Bahloul M, Chelly H, Ben Hmida M, et al. Prognosis of traumatic head injury in South Tunisia: a multivariate analysis of 437 cases. *J Trauma*. 2004;57:255–261.
58. Demetriades D, Kuncir E, Murray J, et al. Mortality prediction of head abbreviated injury score and Glasgow coma scale: analysis of 7,764 head injuries. *J Am Coll Surg*. 2004;199:216–222.
59. Kirkness CJ, Burr RL, Mitchell PH, et al. Is there a sex difference in the course following traumatic brain injury? *Biol Res Nurs*. 2004;5:299–310.
60. Narayan RK, Greenberg RP, Miller JD, et al. Improved confidence of outcome prediction in severe head injury. A comparative analysis of the clinical examination, multimodality evoked potentials, CT scanning, and intracranial pressure. *J Neurosurg*. 1981;54:751–762.
61. Lavati A, Farina ML, Vecchi G, et al. Prognosis of severe head injuries. *J Neurosurg*. 1982;57:779–783.

62. Berger MS, Pitts LH, Lovely M, et al. The outcome from severe head injury in children and adolescents. *J Neurosurg*. 1985;62:194–199.
63. Combes P, Fauvage B, Colonna M, et al. Severe head injuries: an outcome prediction and survival analysis. *Intensive Care Med*. 1996;22:1391–1395.
64. Gómez PA, Lobato RD, Boto GR, et al. Age and outcome after severe head injury. *Acta Neurochir (Wien)*. 2000;142:373–380; discussion 380–381.
65. Hukkelhoven CW, Steyerberg EW, Rampen AJ, et al. Patient age and outcome following severe traumatic brain injury: an analysis of 5600 patients. *J Neurosurg*. 2003;99:666–673.
66. MRC CRASH Trial Collaborators, Perel P, Arango M, et al. Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients. *BMJ*. 2008;336:425–429.
67. Peeters W, van den Brande R, Polinder S, Brazinova, et al. Epidemiology of traumatic brain injury in Europe. *Acta neurochirurgica*. 2015;157(10):1683–1696.
68. Mikolić A, van Klaveren D, Groeniger JO, et al. CENTER-TBI participants and investigators. Differences between men and women in treatment and outcome after traumatic brain injury. *J Neurotrauma*. 2021;38(2):235–251.
69. Jorge RE, Robinson RG, Starkstein SE, et al. Influence of major de- pression on 1-year outcome in patients with traumatic brain injury. *J Neurosurg*. 1994;81:726–733.
70. Wagner AK, Sasser HC, Hammond FM, et al. Intentional traumatic brain injury: epidemiology, risk factors, and associations with injury severity and mortality. *J Trauma*. 2000;49:404–410.
71. Burnett DM, Kolakowsky-Hayner SA, Slater D, et al. Ethnographic analysis of traumatic brain injury patients in the national Model Systems database. *Arch Phys Med Rehabil*. 2003;84:263–267.
72. Johnstone B, Mount D, Gaines T. Race differences in a sample of vocational rehabilitation clients with traumatic brain injury. *Brain Inj*. 2003;17:95–104.
73. Bazarian JJ, Pope C, McClung J, et al. Ethnic and racial disparities in emergency department care for mild traumatic brain injury. *Acad Emerg Med*. 2003;10:1209–1217.
74. The Abbreviated Injury Scale, 1990 revision. *Des Plaines*. IL: Association for the Advancement of Automotive Medicine; 1990:15–24.
75. Baker SP, O'Neill B, Haddon W, et al. The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. *J Trauma*. 1974;14:187–196.
76. van Leeuwen N, Lingsma HF, Perel P, 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.
77. Majdan M, Steyerberg EW, Nieboer D, et al. Glasgow coma scale motor score and pupillary reaction to predict six-month mortality in patients with traumatic brain injury: comparison of field and admission assessment. *J Neurotrauma*. 2015;32:101–108.
78. Balestreri M, Czosnyka M, Chatfield DA, et al. Predictive value of Glasgow Coma Scale after brain trauma: change in trend over the past ten years. *J Neurol Neurosurg Psychiatry*. 2004;75:161–162.
79. Stocchetti N, Pagan F, Calappi E, et al. Inaccurate early assessment of neurological severity in head injury. *J Neurotrauma*. 2004;21:1131–1140.
80. Miller JD, Sweet RC, Narayan R, et al. Early insults to the injured brain. *J Am Med Assoc*. 1978;240:439–442.
81. Walia S, Sutcliffe AJ. The relationship between blood glucose, mean arterial pressure and outcome after severe head injury: an observational study. *Injury*. 2002;33:339–344.
82. Manley G, Knudson MM, Morabito D, et al. Hypotension, hypoxia, and head injury: frequency, duration, and consequences. *Arch Surg*. 2001;136:1118–1123.

83. Chesnut RM. Secondary brain insults after head injury: clinical perspectives. *New Horiz*. 1995;3:366–375.
84. Hartings JA, Bullock MR, Okonkwo DO, et al. Spreading depolarisations and outcome after traumatic brain injury: a prospective observational study. *Lancet Neurol*. 2011;10:1058–1064.
85. Hartings JA, Andaluz N, Bullock MR, et al. Prognostic value of spreading depolarizations in patients with severe traumatic brain injury. *JAMA neurology*. 2020;77(4):489–499.
86. Buechler CM, Blostein PA, Koestner A, et al. Variation among trauma centers' calculation of Glasgow Coma Scale score: results of a national survey. *J Trauma*. 1998;45:429–432.
87. Moskopp D, Staehle C, Wassmann H. Problems of the Glasgow coma scale with early intubated patients. *Neurosurg Rev*. 1995;18:253–257.
88. Marshall LF, Marshall SB, Klauber MR. A new classification of head injury based on computerized tomography. *J Neurosurg*. 1991;75:S14–S20.
89. Maas AI, Hukkelhoven CW, Marshall LF, et al. 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.
90. Flint AC, Manley GT, Gean AD, et al. Post-operative expansion of hemorrhagic contusions after unilateral decompressive craniectomy in severe traumatic brain injury. *J Neurotrauma*. 2008;25:503–512.
91. Nelson DW, Nyström H, MacCallum RM, et al. Extended analysis of early computed tomography scans of traumatic brain injured patients and relations to outcome. *J Neurotrauma*. 2010;27(1):51–64.
92. Raj R, Sirronen J, Skrifvars MB, et al. Predicting outcome in traumatic brain injury: development of a novel computerized tomography classification system (Helsinki computerized tomography score). *Neurosurgery*. 2014;75:632–646; discussion 646–647.
93. Thelin EP, Nelson DW, Vehviläinen J, et al. Evaluation of novel computerized tomography scoring systems in human traumatic brain injury: an observational, multicenter study. *PLoS Medicine*. 2017;14(8):e1002368.
94. Kakarieka A, Braakman R, Schakel EH. Clinical significance of the finding of subarachnoid blood on CR scan after head injury. *Acta Neurochir (Wien)*. 1994;129:1–5.
95. Van de Vyvere T, De La Rosa E, Wilms G, et al. CENTER-TBI participants and investigators. Prognostic validation of the Ninds common data elements for the radiologic reporting of acute traumatic brain injuries: a CENTER-TBI study. *J Neurotrauma*. 2020;37(11):1269–1282.
96. Yuh EL, Mukherjee P, Lingsma HF, et al. Magnetic resonance imaging improves 3-month outcome prediction in mild traumatic brain injury. *Ann Neurol*. 2013;73:224–235.
97. Okonkwo DO, Yue JK, Puccio AM, et al. GFAP-BDP as an acute diagnostic marker in traumatic brain injury: results from the pro- spective transforming research and clinical knowledge in traumatic brain injury study. *J Neurotrauma*. 2013;30:1490–1497.
98. Papa L, Silvestri S, Brophy GM, et al. GFAP out-performs S100β in detecting traumatic intracranial lesions on computed tomography in trauma patients with mild traumatic brain injury and those with extracranial lesions. *J Neurotrauma*. 2014;31:1815–1822.
99. Wang KK, Yang Z, Zhu T, et al. An update on diagnostic and prognostic biomarkers for traumatic brain injury. Expert Rev Mol Diagn. 2018;18(2):165–180.
100. Okonkwo DO, Puffer RC, Puccio AM, et al. Transforming Research And Clinical Knowledge In Traumatic Brain Injury (TRACK-TBI) investigators. Point-of-Care platform blood biomarker testing of glial fibrillary acidic protein versus S100 calcium-binding protein B for prediction of

traumatic brain injuries: a transforming research and clinical knowledge in traumatic brain injury study. *J Neurotrauma*. 2020;37(23):2460–2467.

101. Xu LB, Yue JK, Korley F, Puccio, et al. TRACK-TBI investigators. High-sensitivity C-reactive protein is a prognostic biomarker of six-month disability after traumatic brain injury: results from the TRACK-TBI study. *J Neurotrauma*. 2020.
102. Czeiter E, Amrein K, Gravesteijn BY, CENTER-TBI Participants, Investigators, et al. Blood biomarkers on admission in acute traumatic brain injury: relations to severity, CT findings and care path in the CENTER-TBI study. *EBioMedicine*. 2020;56:102785.
103. Jacobs B, Beems T, Stulemeijer M, et al. Outcome prediction in mild traumatic brain injury: age and clinical variables are stronger predic- tors than CT abnormalities. *J Neurotrauma*. 2010;27:655–668.
104. Stulemeijer M, van der Werf S, Borm GF, et al. Early prediction of favourable recovery 6 months after mild traumatic brain injury. *J Neurol Neurosurg Psychiatry*. 2008;79:936–942.
105. Van der Naalt J, Timmerman ME, de Koning ME, et al. Early predictors of outcome after mild traumatic brain injury (UPFRONT): an observational cohort study. *Lancet Neurol*. 2017;16(7):532–540.
106. Mittenberg W, DiGiulio DV, Perrin S, et al. Symptoms follow- ing mild head injury: expectation as aetiology. *J Neurol Neurosurg Psychiatry*. 1992;55:200–204.
107. Stulemeijer M, van der Werf S, Bleijenberg G, et al. Recovery from mild traumatic brain injury: a focus on fatigue. *J Neurol*. 2006;253:1041–1047.
108. Silverberg ND, Gardner AJ, Brubacher JR, Panenka WJ, Li JJ, Iverson GL. Systematic review of multivariable prognos- tic models for mild traumatic brain injury. *J Neurotrauma*. 2015;32(8):517–526.
109. Van der Ploeg T, Nieboer D, Steyerberg EW. Modern modeling techniques had limited external validity in predicting mortality from traumatic brain injury. *J Clin Epidemiol*. 2016;78:83–89.
110. Gravesteijn B, Nieboer D, Ercole A, et al. *Machine Learning for Prediction of Mortality and Unfavorable Outcome in Traumatic Brain Injury: A CENTER-TBI Study*; 2019. Available at: SSRN 3441828.
111. Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol*. 2016:245–247.
112. Moons KGM, Altman DG, Reitsma JB, Collins GS. New guide- line for the reporting of studies developing, validating, or updating a multivariable clinical prediction model: the TRIPOD statement. *Adv Anat Pathol*. 2015;5:303–305.
113. Moons KG, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or di- agnosis (TRIPOD): explanation and elaboration. *Ann Intern Med*. 2015;162(1):W1–W73.
114. Mushkudiani NA, Hukkelhoven CWPM, Hernandez AV, 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.
115. Perel P, Edwards P, Wentz R, et al. Systematic review of prognos- tic models in traumatic brain injury. *BMC Med Inform Decis Mak*. 2006;6:38.
116. Bouwmeester W, Zuithoff NPA, Mallett S, et al. Reporting and methods in clinical prediction research: a systematic review. *PLoS Med*. 2012;5:e1001221.
117. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J*. 2014;35:1925–1931.
118. Steyerberg EW. Validation in prediction research: the waste by data-splitting. *J Clin Epidemiol*. 2018.

119. Bonnett LJ, Snell KI, Collins GS, Riley RD. Guide to presenting clin- ical prediction models for use in clinical settings. *Bmj*. 2019;365:l737. 120. Van Belle V, Van Calster B. Visualizing risk prediction models. *PloS One*. 2015;10(7):e0132614.
120. Debray TP, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg
121. EW, Moons KG. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol*. 2015;68(3):279–289.
122. Kappen TH, van Klei WA, van Wolfswinkel L, Kalkman CJ, Vergouwe Y, Moons KGM. Evaluating the impact of prediction models: lessons learned, challenges, and recommendations. *BMC Diagn Progn Res*. 2018;2:11.
123. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006;26:565–574.
124. Steyerberg EW, Mushkudiani N, Perel P, et al. Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. *PLoS Med*. 2008;5:e165.
125. Roozenbeek B, Chiu YL, Lingsma HF, 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.
126. Roozenbeek B, Lingsma HF, Lecky FE, et al. Prediction of outcome after moderate and severe traumatic brain injury: external validation of the IMPACT and CRASH prognostic models. *Crit Care Med*. 2012;40:1609–1617.
127. Lingsma H, Andriessen TM, Haitsema I, 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.
128. Panczykowski DM, Puccio AM, Scruggs BJ, et al. *J Neurotrauma*. 2012;29:47–52.
129. Yeoman P, Pattani H, Silcocks P, et al. Validation of the IMPACT outcome prediction score the nottingham head injury register data- set. *J Trauma*. 2011;71:387–392.
130. Raj R, Siironen J, Kivisaari R, et al. 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.
131. Wong GK, Teoh J, Yeung J, 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.
132. Majdan M, Lingsma HF, Nieboer D, et al. Performance of IMPACT, CRASH and Nijmegen models in predicting six months outcome of patients with severe or moderate TBI: an external vali- dation study. *Scand J Trauma Resusc Emerg Med*. 2014;22:68.
133. Harrison DA, Prabhu G, Grieve R, et al. Risk Adjustment in Neurocritical care (RAIN)–prospective validation of risk predic- tion 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;17:1– 350. vii-viii.
134. Han J, King NK, Neilson SJ, et al. External validation of the CRASH and IMPACT prognostic models in severe traumatic brain injury. *J Neurotrauma*. 2014;31:1146–1152.
135. Maas AI, Menon DK, Steyerberg EW, et al. Collaborative European NeuroTrauma Effectiveness Research In Traumatic Brain Injury (CENTER-TBI): a prospective longitudinal observational study. *Neurosurgery*. 2015;76:67–80.
136. Dijkland SA, Helmrich IRR, Nieboer D, et al. CENTER-TBI participants and investigators. Outcome prediction after moderate and severe traumatic brain injury: external validation of two estab- lished prognostic models in 1742 European patients [published on- line ahead of print, 2020 dec 14]. *J Neurotrauma*. 2020;10. http:// doi:10.1089/neu.2020.7300.

137. Cnossen MC, van der Naalt J, Spikman JM, et al. Prediction of per- sistent post-concussion symptoms after mild traumatic brain injury. *J Neurotrauma*. 2018;35(22):2691–2698.
138. Mikolić A, Polinder S, Steyerberg EW, Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) Study Participants and Investigators, et al. Prediction of global functional outcome and post-concussive symp- toms after mild traumatic brain injury: external validation of prognos- tic models in the collaborative European NeuroTrauma Effectiveness Research In Traumatic Brain Injury (CENTER-TBI) study. *J Neurotrauma.* 2020; 38(2):196–209.

Chapter 3

Prognostic Model Research: Overfitting, Validation and Application

I.R.A. Retel Helmrich, D. van Klaveren, and E.W. Steyerberg

Published in: Journal of Physiotherapy

Introduction

In physiotherapy, many prognostic models have been developed to predict future outcomes after musculoskeletal conditions, including neck pain.[1] Prognostic models combine several characteristics to predict the risk of an outcome for individual patients and may enable personalized prevention and care. In practice, they can be used to inform patients and relatives on prognosis, and to support clinical decision making. Moreover, models may be useful to stratify patients for clinical trials. Prediction models are increasingly published, including ninety-nine prognostic models for neck pain, predicting recovery (pain reduction, reduced disability, and perceived recovery).[2] Although developing and reporting guidelines have been proposed,[3, 4] many prognostic models in physiotherapy are prone to risks of bias,[2] according to a recently proposed assessment tool.[5]

Various limitations were noted regarding design, and analyses, which make models at risk of overfitting.[2] Overfitting relates to the notion of asking too much from the available data, which will result in overly optimistic estimates of model predictive performance; results that cannot be validated in underlying or related populations.[6] Consequently, the model may predict poorly, with serious limitations when the model is applied in clinical practice: it does not separate low from high risk patients (poor discrimination), and may give unreliable, or even misleading risk estimates (poor calibration).

We aim to describe a number of challenges related to the design and analysis in different stages of prognostic model research, and opportunities to reduce overfitting (summarized in Table 1). We emphasize validation before the application of prediction models is considered in medical practice. For illustration, we consider the Örebro Musculoskeletal Pain Screening Questionnaire (OMPQ) (Table 2).[7] The model has extensively been validated, and its use is recommended by clinical guidelines.[8] We also consider the Schellingerhout non-specific neck pain model predicting recovery after six months (Table 2),[9] which was indicated as one of the few externally validated models with a low risk of bias.[2]

Model development

The development of a prognostic model involves a number of steps. These include handling of missing data, selection and coding of predictor variables, choosing between alternative statistical models, and estimating model parameters.[10] Prognostic models are usually developed with multivariable regression techniques on data from (prospective) cohort studies, while machine learning techniques are gaining increased attention.

Missing data is common in prognostic research. A complete case analysis is often conducted, i.e. the exclusion of participants that have missing data on one or multiple predictor variables, resulting in smaller sample size. As a consequence, the number of events per variable (EPV) may drop below the number deemed necessary for reliable modelling (Table 1), increasing the risk of overfitting. Better approaches are imputation methods,[10] where missing values may be substituted with the mean or the mode with single imputation, and *m* completed data sets are created with multiple imputation procedures. Multiple imputation is recommended, because single imputation ignores potential correlation of predictors and leads to an underestimation of variability of predictor values among subjects.[11] This may lead to an overestimation of the precision of regression coefficients. Imputation methods are widely available through modern statistical software.

Table 1: Overview of challenges and opportunities categorized by the stage of prognostic model research in which they occur, and illustrated with two prediction models [7, 9]

Stage of prognostic model research	**Challenges**	**Opportunities**	Örebro Musculoskeletal Pain Screening Questionnaire	**Schellingerhout non-specific neck pain model**
Design	Insufficient sample size	Collaborative efforts to reach >10 events per variable (EPV), cross-validate across setting	No information on EPV	Restricted to 17 predictors based on EPV (10)
Development	Inappropriate handling of missing data; Complete case analysis	Multiple imputation methods	Complete case analysis	Multiple imputation with 5 repetitions
Development	Selection of predictors based on univariate analysis or stepwise selection procedures	Shrinkage and penalization in multivariable analysis	Univariate analysis	Backward stepwise selection
Internal validation	Apparent validation or inefficient internal validation procedures	Bootstrap resampling or cross-validation	Apparent validation	Apparent validation
External validation	Full model equation is not presented	Present full model equation	Yes	Yes
External validation	No external validation	Validation of models in cohort other than development cohort through collaborative research	Externally validated; AUC, but no calibration plot	Externally validated; AUC and calibration plot

Selecting the most promising predictors is difficult. Selection of candidate predictors based on literature and expert knowledge is often preferred over selection based on a relatively limited dataset.[10] Also, some related predictors can sometimes be combined in simple scores. For example, comorbid conditions are often combined in a comorbidity score,[12] and frailty in the elderly can be scored according to various characteristics.[13] After selection of candidate predictors we may consider to reduce the set of predictors. This can be done using univariate analysis and/or stepwise methods. However, both approaches do not truly reduce the problem of statistical overfitting, since the model specification is driven by findings in the data. Univariate analysis is common as a first step to select the most potent risk factors, which are then used in multivariable analysis. This approach was followed in the development of the OMPQ (Table 2). A common alternative is to use backward stepwise selection from a model that includes all candidate predictors, as was done by Schellingerhout to develop a model to predict non-specific neck pain (Table 2). Stepwise selection procedures are known to result in biased regression coefficient estimates ("*testimation bias*").[6] A modern approach to reduce such *testimation bias* and overfitting is by shrinkage of regression coefficients towards zero.[10] A key example of this approach is the Least Absolute Shrinkage and Selection Operator (LASSO). LASSO penalizes for the absolute values of the regression coefficients. It shrinks some coefficients to zero, which means that predictors are dropped from the model.

Validation: apparent, internal, and external performance

The aim of prognostic models is to provide accurate risk predictions for new patients. Therefore, validation of prognostic models is crucial. We can distinguish between three types of validation: apparent, internal and external validation.

Apparent validation entails the assessment of model performance directly in the derivation cohort. Because the regression coefficients are optimized for the derivation cohort this provides optimistic estimates of the model's performance (overfitting). To correct for overfitting several internal validation procedures are available. Bootstrap resampling and cross-validation provide stable estimates with low bias and are therefore recommended.[10]

Before a prognostic model can be applied in practice it is crucial to explore how the model performs outside the setting in which it was developed, preferably across a range of settings. External validity relates to the generalizability of the prognostic model to another population.[10] A cross-validation across different non-random parts of the development data gives an indication of external validity.[14] Heterogeneity in predictor effects across settings indicates that the model should be calibrated to each specific setting, to achieve robust model performance across settings. To enable

Table 2: Overview of prognostic model characteristics of the Orebro Musculoskeletal Pain Screening Questionnaire and the Schellingerhout non-specific neck pain model

	Örebro Musculoskeletal Pain Screening Questionnaire	**Schellingerhout non-specific neck pain model**
Development		
Patient population of development cohort	N=137; adult patients; acute/subacute back pain; Sweden [7]	N=468; adult patients (18-70 years); non-specific neck pain; primary care; The Netherlands [9]
Outcome	Accumulated sick leave; Six months follow-up	Global perceived recovery; dichotomized into "recovered or much improved" versus "persistent complaints"; Six months follow-up
Predictors	21 predictors; Physical functioning, fear-avoidance beliefs, the experience of pain, work, and reactions to the pain	9 predictors; Age, pain intensity, previous neck complaints, radiation of pain, accompanying low back pain, accompanying headache, employment status, health status, and cause of complaints
External validation		
External validation	N=106; Adult patients; Acute/subacute low back pain; workers' compensation and medical practitioner referral; Observational study; Australia [17]	N = 346; Adult patients (18-70); Non-specific neck pain; primary care; Randomized Controlled Trial; PANTHER trail; United Kingdom [9]
Model performance	AUC 0.80 (CI 95%, 0.66–0.93); No calibration plot	AUC 0.65 (CI 95%, 0.59-0.71); Calibration plot
Application		
Practical application	Recommended in clinical guidelines as screening instrument [8], and used to select trial participants [19]	Score chart [9]

external validation of the model the full model equation should be presented in the paper (Table 1). The OMPQ has extensively been validated in international cohorts,[15] while such external validation is rare for other prognostic models for musculoskeletal conditions.[2, 16]

Performance measures

Model performance at internal and external validation is commonly expressed with discrimination and calibration. Discrimination indicates the ability of the model to differentiate between high and low risk patients. It can be measured by the concordance

statistic (C-statistic, or area under the receiver operating characteristic curve: AUC). The AUC ranges between 0.50 (no discrimination) and 1.0 (perfect discrimination). For instance, the OMPQ was validated in an observational study of patients with acute back pain in Australia.[17] At external validation of the OMPQ the AUC was 0.80 (CI 95%, 0.66–0.93) for absenteeism at six months (Table 2).[17] The discriminative ability of the Schellingerhout non-specific neck pain model was lower: AUC 0.66 (CI 95%, 0.61–0.71) at development, and validation cohort AUC 0.65 (CI 95%, 0.59–0.71).[9]

Calibration refers to the agreement between predicted and observed probabilities. This agreement can be illustrated with a calibration graph. Ideally, the plot shows a 45 degree line with calibration slope 1 and intercept 0. Calibration is more informative at external than internal validation, because a model is expected to provide correct predictions for the derivation cohort it is fitted on. At external validation, the Schellingerhout non-specific neck pain score chart showed reasonable calibration (Figure 1); it slightly overestimated the risk of persistent complaints in adult patients presenting with non-specific neck pain.[9] More severe miscalibration is common for prediction models.[18]

Application of prognostic models in practice

A prognostic model is more likely to be applicable for implementation in practice if the model was developed with high quality data from an appropriate study design, and with careful statistical analysis.[10] Even better is when the model is externally validated in the setting where it is to be used.[14] For instance, the OMPQ is recommended in clinical guidelines to be applied in screening to predict delayed recovery,[8] and was used to select trial participants,[19] likely motivated by the extensive and positive external validation studies across multiple settings. When a prognostic model is deemed appropriate for implementation, the impact (clinical effectiveness and costs) of the use of the model in clinical practice should be studied.[4] Although recommended, these clinical impact studies are scarce, and some prediction models have been recommended to be used in clinical practice without adequate evaluation of their (cost) effectiveness.

The presentation of clinical prediction models is important to facilitate the implementation of prognostic models in practice. The Schellingerhout model was presented as a score chart that can readily be used by physicians. Although the score chart may be easy to use, predictions of risks are only approximate, because continuous predictors are categorized, and regression coefficients are rounded. The score chart is ideally externally validated across various settings before it can be considered for use in broader

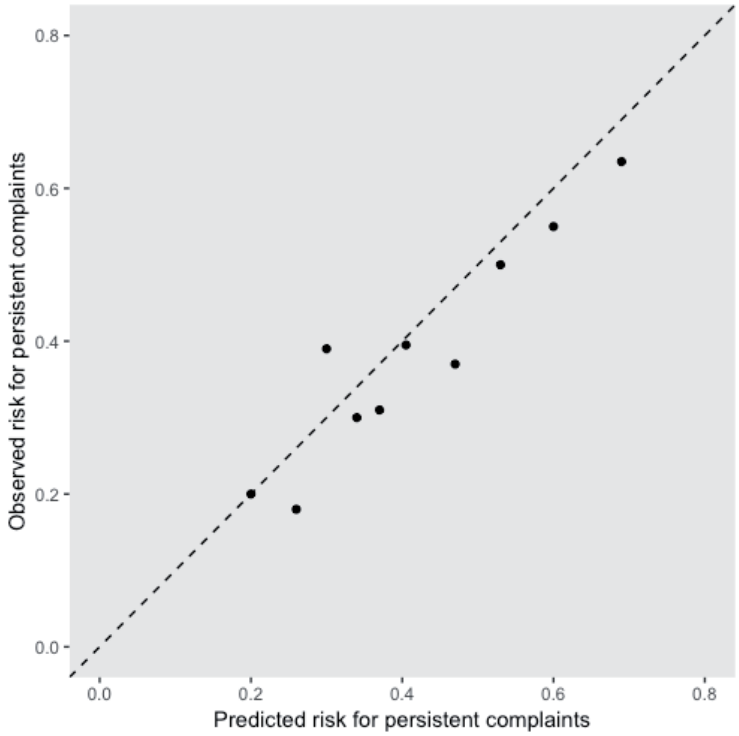


Figure 1: Calibration of the Schellingerhout non-specific neck pain score chart in external validation cohort. Deciles of risk (●). Perfect calibration (----). Adapted from Schellingerhout et al.[9]

practice. Other common formats include web-based calculators and apps for mobile devices.[10, 20]

Summary

The aim of prognostic models for predicting future outcomes after musculoskeletal conditions is to provide accurate and patient-specific estimates of the risk of relevant clinical outcomes, such as delayed recovery. These models may be applied in primary care to identify patients likely to have poor outcomes. Most models in physiotherapy have been judged to be at moderate to high risk of bias.[2] Approaches to reduce overfitting should be better utilized. These include appropriate handling of missing data, careful selection of predictors with domain knowledge, as well as internal and external validation (Table 1). Assessment of performance across a range of settings may show suboptimal results, specifically with respect to calibration of predictions. Such suboptimal performance may motivate updating of a model before it can be considered for application in a specific setting.[10] Furthermore, clinical impact studies are recommended to assess the (cost)effectiveness of a prognostic model in clinical

practice. The presentation format of a prognostic model is also important, as this can facilitate implementation of prognostic models in clinical practice to the improvement of decision making and outcome by personalized medicine.

References

1. Kelly, J, Ritchie, C, Sterling, M. Clinical prediction rules for prognosis and treatment prescription in neck pain: A systematic review. *Musculoskelet Sci Pract.* 2017: 155-164.
2. Wingbermühle, RW, van Trijffel, E, Nelissen, PM, Koes, B, Verhagen, AP. Few promising multivariable prognostic models exist for recovery of people with non-specific neck pain in musculoskeletal primary care: a systematic review. *J Physiother.* 2018;1: 16-23.
3. Collins, GS, Reitsma, JB, Altman, DG, Moons, KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMC med.* 2015;1: 1.
4. 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;2: e1001381.
5. Wolff, RF, Moons, KGM, Riley, RD, Whiting, PF, Westwood, M, Collins, GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med.* 2019;1: 51-58.
6. Babyak, MA. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosom Med.* 2004;3: 411-421.
7. Linton, SJ, Halldén, K. Can we screen for problematic back pain? A screening questionnaire for predicting outcome in acute and subacute back pain. *Clin J Pain* 1998;3: 209-215.
8. Corporation, AC. *New Zealand acute low back pain guide, incorporating the guide to assessing psychosocial yellow flags in acute low back pain*. 2004; Available from: https://www.acc.co.nz/assets/provider/ff758d0d69/acc1038-lower-back-pain-guide.pdf [Accessed June 14, 2019]
9. Schellingerhout, JM, Heymans, MW, Verhagen, AP, Lewis, M, de Vet, HCW, Koes, BW. Prognosis of patients with nonspecific neck pain: development and external validation of a prediction rule for persistence of complaints. *Spine J.* 2010;17: E827-E835.
10. Steyerberg, EW. Clinical prediction models: a practical approach to development, validation, and updating;2008. Springer Science & Business Media: New York.
11. Ambler, G, Omar, RZ, Royston, P. A comparison of imputation techniques for handling missing predictor values in a risk model with a binary outcome. *Stat Methods Med Res.* 2007;3: 277-298.
12. Charlson, ME, Pompei, P, Ales, KL, MacKenzie, CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis Manag.* 1987;5: 373-383.
13. Searle, SD, Mitnitski, A, Gahbauer, EA, Gill, TM, Rockwood, K. A standard procedure for creating a frailty index. *BMC Geriatr.* 2008;1: 24.
14. Steyerberg, EW, Harrell, FE. Prediction models need appropriate internal, internal–external, and external validation. *J Clin Epidemiol.* 2016: 245-247.
15. Hockings, RL, McAuley, JH, Maher, CG. A systematic review of the predictive ability of the Orebro Musculoskeletal Pain Questionnaire. *Spine J.* 2008;15: E494-E500.
16. van Oort, L, van den Berg, T, Koes, BW, de Vet, RHCW, Anema, HJR, Heymans, MW, et al. Preliminary state of development of prediction models for primary care physical therapy: a systematic review. *J Clin Epidemiol.* 2012;12: 1257-1266.
17. Linton, SJ, Boersma, K. Early identification of patients at risk of developing a persistent back problem: the predictive validity of the Örebro Musculoskeletal Pain Questionnaire. *Clin J Pain* 2003;2: 80-86.

18. Riley, RD, Ensor, J, Snell, KIE, Debray, TPA, Altman, DG, Moons, KGM, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *bmj.* 2016: i3140.
19. Schmidt, CO, Chenot, J-F, Pfingsten, M, Fahland, RA, Lindena, G, Marnitz, U, et al. Assessing a risk tailored intervention to prevent disabling low back pain-protocol of a cluster randomized controlled trial. *BMC Musculoskelet Disord.* 2010;1: 5.
20. Bonnett, LJ, Snell, KIE, Collins, GS, Riley, RD. Guide to presenting clinical prediction models for use in clinical settings. *bmj.* 2019: l737.

Chapter 4

Prognostic Research in Traumatic Brain Injury: Markers, Modeling and Methodological Principles

I.R.A. Retel Helmrich, H.F. Lingsma, A.F. Turgeon, J.M. Yamal, and E.W. Steyerberg

Published in: Journal of Neurotrauma

doi.org/10.1089/neu.2019.6708

Abstract

Prognostic assessment in traumatic brain injury (TBI) is imbedded deeply in clinical care. Considering the limitations of current prognostic indicators, there is increasing interest in understanding the role of new biomarkers, and in finding other prognostic indicators of long-term outcomes following TBI. New prognostic indicators may result in the development of more accurate prediction models that could be useful for both risk stratification and clinical decision-making. We aimed to review methodological issues and provide tentative guidelines for prognostic research in TBI.

Prognostic factor research focuses on the role of a specific patient or disease-related characteristic in relation to outcome. Typically, univariable relations of the prognostic factor are studied, followed by analyses adjusting for other variables related to the outcome. Following existing guidelines, we emphasize the importance of transparent reporting of patient and specimen characteristics, study design, clinical endpoints, and statistical analysis. Prognostic model research considers combinations of predictors, with challenges for model specification, estimation, evaluation, validation, and presentation. We highlight modern approaches and opportunities, related to missing values, exploration of non-linear effects, and assessing between-study heterogeneity.

Prognostic research in TBI can be improved if key methodological principles are adhered to and when research is performed in collaboration among multiple centers to ensure generalizability.

Key words: Prognostic research, Traumatic Brain Injury, Outcome, Markers, REMARK, TRIPOD

Introduction

Establishing a reliable prognosis early after traumatic brain injury (TBI) is notoriously difficult due to the heterogeneity of the condition. Clinicians involved in the care of patients with severe TBI are not always in agreement when predicting long-term functional outcomes.[1] More so, mortality after severe TBI has been observed to be variable across centers in this population, while most TBI patients die following the decision to withdraw life-sustaining therapies.[2] The lack of appropriate prognostic information was one of the factors shown to influence decisions regarding the level of care in patients with severe TBI.[3]

To predict outcomes after moderate and severe TBI various prediction models have been developed.[4] Prediction models combine clinical characteristics and data to predict the risk of an outcome for individual patients. Prediction models may support early clinical decision-making. They may also facilitate reliable comparison of outcomes between different patient cohorts and variations in results over time. Furthermore, prediction models have been used for risk stratification of patients and covariate adjustment in randomized controlled trials (RCTs).[5, 6]

Over the years, several prediction models for moderate and severe TBI were proposed.[4] Among those, the CRASH (Corticosteroid Randomisation After Significant Head injury) and IMPACT (International Mission for Prognosis and Analysis of Clinical Trials in Traumatic Brain Injury) models formed a contrast to previously developed models.[7, 8] Previous models were commonly developed on relatively small samples, often originated from a single center or region, and lacked external validation.[4] Simple and more extensive versions of the CRASH and IMPACT prediction models were proposed, with increasing discriminative ability (Supplements, Table 1). Blood biomarkers, imaging biomarkers, and dynamic predictors have been suggested as promising indicators in TBI research, which have the potential to further improve these models.

Although guidelines have been proposed for model development and reporting,[5, 9-11] prognostic research studies in TBI often have methodological limitations.[4, 12, 13] We aimed to review methodological issues and provide tentative guidelines for prognostic research in TBI. We first consider prognostic factor research.[11] Such research focuses on the prognostic role of a single or multiple markers in combination with clinical characteristics and other prognostic indicators. Next, we consider prediction model research.[5] Since prognostic research is increasingly done in collaborative initiatives, such as the International Initiative for Traumatic Brain Injury Research (InTBIR),[14]

we explicitly consider challenges of multicenter data and analyses for model development and validation.

Methods for 'prognostic factor' versus 'prediction model' research

Prognostic research studies can be separated in two main categories: studies with a focus on the prognostic role of specific patient or disease-related characteristics in relation to outcome ('prognostic factor research'), and studies with a focus on the combined effect of various prognostic factors in predicting the outcome ('prediction model research') (Table 1).[5, 11, 15] In prognostic factor studies, we may start with assessing whether the factor is independently associated with the outcome of interest. 'Independently' here refers to the association of the prognostic factor with the outcome separate from other prognostic indicators, and typically requires some form of statistical adjustment in the analysis. We may also analyze whether the risk of the outcome uniformly increases or decreases, or has a more complex relationship when considering a continuous predictor. Moreover, we may be interested in a quantification of the incremental predictive value.

In prognostic factor research, it is typical to first study univariable relations of the prognostic factor in a cross-table or regression model, followed by regression analyses adjusting for other variables related to the outcome. The effect measure commonly is

Table 1: Prognostic research: characteristics of prognostic factor and prediction model research

Characteristic	Prognostic factor	Prediction model
Research question	a) Is this factor independently associated with the outcome? b) What is the shape of the association? c) What is the incremental predictive value?	How well can we predict outcome based on a combination of prognostic factors?
Effect measure	a) Relative risk (e.g. OR, HR) with 95% confidence interval; b) Graphical assessment of continuous prognostic factors; c) Improvement in performance measures (e.g. c statistic, Nagelkerke R^2)	Predictive performance, including discrimination (e.g. c statistic) and calibration (e.g. graphical assessment)
Analysis	Univariate analysis and adjusted analysis including confounders	Multivariable modeling and validation

relative, for example, an odds ratio (OR, in a logistic regression model) when considering a binary outcome, or hazard ratio (HR, in a Cox regression model) when considering a survival outcome. A p-value may support claims of statistical significance, which can also be inferred if the 95% confidence interval does not include the value 1. Graphical assessments are helpful to study the shape of an association, while the incremental value can be noted from the improvement in performance measures such as the concordance (c) statistic (equivalent to the area under the ROC curve for binary outcomes), or Nagelkerke's R^2 .

While prognostic factor research may give rise to speculation on causal effect and mechanism of action, prediction models commonly address a more pragmatic research question: How well can we predict the outcome based on a combination of prognostic factors? In prediction model research it is typical to analyze combinations of prognostic factors in multivariable models, followed by analyses of predictive performance, including measures for discrimination (e.g. c statistic) and calibration (e.g. graphics and calibration statistics).[6] The assessment of performance needs validation in independent data. We can distinguish between internal and external validation. With internal validation procedures, such as bootstrap resampling and cross-validation, we aim to correct the performance estimates. In external validation studies, we study the generalizability of the model in different but related settings, for instance by assessing the performance of a proposed model in another cohort.[16, 17]

Methodological guidance for prognostic research is dispersed throughout the literature. We take two previously proposed reporting guidelines as a basis: REMARK, which was originally intended for reporting of marker research in oncology, and TRIPOD, which was proposed for reporting of prediction model development and validation.[9, 18]

Guidance for prognostic factor research

The original REMARK guideline consists of 20 items that need to be reported in studies that focus on one or more prognostic factors (Supplement, Table 2). More specifically, the guideline was developed for prognostic model studies in oncology that include tumor markers. The REMARK guideline is applicable to prognostic factor research in fields other than oncology and is especially relevant when tissue biomarkers are included as candidate predictors. The guideline was endorsed by multiple journals.[10, 18, 19] An 'Exploration and Elaboration' document provides more detail on the choice of the items, their relevance, and examples of good practice.[20] We focus

on the items that are most relevant to prognostic factors in TBI research and provide examples for illustration (Table 2).

Table 2: Specific elements for prognostic factor research in TBI, building on the REMARK guideline (MCshane, 2005)

Topic	Description
Selection of patients	Inclusion and exclusion criteria may vary between studies in relevant aspects, such as age, severity, setting, region, and treatment policies. Inclusion of patients is ideally consecutive.
Prognostic factors considered	Typical prognostic factors may include clinical indicators (vital signs, intracranial pressure, cerebroperfusion pressure); radiological imaging (CT-scan, MRI); electrophysiological tests (EEG, SSPEP); tissue biomarkers (in blood, cerebrospinal fluid). Timing of assessment after trauma, method of acquisition (technology), as well as methods of handling and storage are important to report and address in statistical analyses.
Study design	All candidate variables need to be reported if examined or considered for inclusion in statistical models. Outcome measures must be chosen based on the severity of the TBI and may vary from mild to more severe TBI. Rationale for sample size should be provided.
Statistical analysis methods	A core set of predictors needs to be considered for adjustment of prognostic factor associations, depending on the severity of disease and availability of data. Core predictors for moderate and severe TBI include age, motor score (or full GCS), and pupillary reactivity. For mild TBI, the core set is less well defined and may depend on the outcome considered. Missing values in the variables in the core set may be imputed to gain efficiency. The relation of the marker to the core set variables needs to be studied, e.g. with correlation analyses and graphical inspections. Marker values are commonly continuous in nature. The shape of the association with the outcome needs to be examined with sufficiently flexible functions, such as splines, with graphical inspection.
Descriptive results	Consider the flow of patients through each stage of the analysis, with the number of events, and reasons for dropout. Describe characteristics in sufficient detail, including demographics, standard prognostic variables, and the prognostic factors considered, including the numbers of missing values.
Statistical results	Univariate and adjusted analyses show the relation between the marker and outcome, with the estimated association (for example, odds ratio plus confidence interval). Adjustment should be for the core set of variables, irrespective of statistical significance. Incremental predictive value can be indicated by measures for discrimination, such as the increase in c statistic, and overall fit, such as explained variability (R^2).
Interpretation	Results should be interpreted in the context of the pre-specified hypotheses and other relevant studies. Ideally, replication is done in similar studies. Limitations of the study need to be considered, and implications for future research.

Patient selection

When designing prognostic studies, the study population must represent the targeted population in which these models will be used. However, the selection of patients may vary substantially between studies. Inclusion and exclusion criteria may differ between studies in relevant aspects, such as age (pediatric, adult, geriatric), severity (e.g. based on the Glasgow Coma Scale score (GCS), setting (Emergency department, ward, Intensive Care Unit), region (low/middle/high income country), and treatment policies. For instance, the CRASH study included 7526 patients from low and middle-income countries, and 2482 from high-income countries, where mortality at 14 days was 21 vs 16% ($p<0.001$).[7] The IMPACT study included 11 cohorts (3 RCTs, 8 observational studies),[8] and substantial differences were found in outcome between 265 centers in this study.[21] In the InTBIR consortium, inclusion criteria also vary substantially between studies considering the different objectives and targeted populations of TBI patients (See Box 1: A selection of InTBIR studies and their characteristics).

Prognostic factors

Prognostic factors may range from clinical indicators (e.g. disease severity, vital signs, intracranial pressure, cerebral perfusion pressure) to radiological imaging (e.g. CT, MRI), electrophysiological tests (EEG, SSEP), and tissue biomarkers (in blood, or cerebrospinal fluid) (Supplements, Table 4). The timing of measurements may vary from the acute phase to several weeks after trauma.

Biomarkers have received increased attention in the last decade. Specifics of the data acquisition need to be considered carefully, and this may be challenging when conducting multicenter studies. Timing and method of acquisition, as well as methods of preservation and storage, are important. Apart from these aspects being reported, they may also need to be addressed in statistical analyses, for instance, by adjusting for the time between data acquisition and trauma. If control samples are used, their characteristics also need to be described carefully, including their selection.

The assay methods used should be provided, preferably with a detailed protocol, including specific reagents or kits used, quality control procedures, reproducibility assessments, quantitation methods, and scoring and reporting protocols. Furthermore, it is important to perform assays blinded to the study endpoint for an unbiased assessment.

As an example, biomarkers were sampled at admission up to 24 months post-injury in the Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) study. Samples of whole blood, serum, and plasma for genetic, biomarker and hemostasis analyses were stored in a specific biobank (Pecs, Hungary).

Box 1: InTBIR consortium

InTBIR is a collaborative effort of the European Commission (EC), the National Institutes of Health (NIH), the Canadian Institutes of Health Research (CIHR), the US Department of Defense (DoD), the Ontario Brain institute (OBI) and OneMind that aims to coordinate and leverage international clinical research activities on traumatic brain injury (TBI) research. InTBIR's goal is to improve health care and lessen the global burden of TBI through the discovery of causal relationships between treatments and clinically meaningful outcomes. InTBIR therefore focuses on collecting, standardizing, and sharing clinical TBI data for comparative effectiveness research.

Table I: A selection of InTBIR studies and their characteristics

Study	Geographical region(s)	Centers	Inclusion criteria
CREACTIVE	Europe, Israel	72	Moderate to severe TBI
CENTER-TBI	Europe, Israel	59	Clinical diagnosis of TBI; <24 hours after injury; Clinical indication for CT scan
ADAPT	Australia, Europe, India, New Zealand, South Africa, United States	49	<18 years; Severe TBI; Require ICP monitoring
TRACK-TBI	United States	20	Clinical diagnosis of TBI; Clinical indication for CT scan
TBI-Prognosis	Canada	17	Severe TBI; Admitted to the intensive care unit
GNRG	Latin America	14	≥13 years; Severe TBI; Non-penetrating TBI; Absence of ICP monitoring
TBIcare	United States	2	≥18 years; Mild to severe TBI; Clinically significant pain over the last 6 months

Information obtained from https://intbir.nih.gov/projects

A detailed description of the data acquisition, including the timing, method, preservation, and storage has been described elsewhere.[22]

Study design

Cohort studies are the preferred design for prognostic research. Ideally, we measure a prognostic factor in a prospective cohort of consecutive patients and evaluate the relationship with the outcome while minimizing potential confounding. Confounding may occur if these prognostic factors (from clinical data or test results) are evaluated according to clinical indications. In TBI research, we might consider a number of different outcomes, depending on the research question and population under study (See Box 2: A selection of outcomes in prognostic research in TBI). For efficiency, case-control or nested case-control designs can also be used for prognostic factor studies, especially if measurements are relatively expensive. For reporting, all clinical

Box 2: A selection of outcomes to be considered in prognosis research in traumatic brain injury

Outcome	Instrument
Mortality	-
Functional status	GOS(-E)
Generic HRQoL	EQ5-D, SF36, SF12
TBI-specific HRQoL	QoLIBRI, QoLIBRI-OS
Post-concussion symptoms	Rivermead post-concussion questionnaire
Posttraumatic stress disorder	PCL-5
Depression	HADS, PHQ-9
Anxiety	HADS, GAD-7
Neuropsychological testing	GOAT, RAVLT, TMT, CANTAB, 10 m walk and timed up and go
Return to work	-

Adapted from Maas et al., (2014)
CANTAB, Cantab neuropsychological assessment tests; GOAT, Galveston Orientation and Amnesia Test; GOSE, Glasgow Outcome Scale Extended; HADS,Hospital Anxiety and Depression Scale; HRQoL, Health-Related Quality of Life; PCL-5, PTSD Check List; PHQ-9, Patient Health Questionnaire; QOLIBRI, Quality of Life after Brain Injury; QOLIBRI-OS, QOLIBRI-Overall Scale; RAVLT, Rey Auditory Verbal Learning Test; SF12v2, Short-Form 12 version 2; SF36v2, Short-Form 36 version 2; TMT, Trail Making Test.

endpoints, and all candidate variables need to be mentioned if examined or considered for inclusion in any form of statistical analysis. Such a transparent report is essential for proper interpretation of a specific prognostic factor – outcome relationship from a large set of potential relationships as examined in the study. For subjective outcomes measure, blinding of the assessor is important. This means that the assessor should be unaware of the values of the results of the prognostic factors studied.

Sample size

A methodological and ethical rationale for the sample size should be provided (See Box 3: Example of a rational for sample size). A formal approach may consider a pre-specified effect size for the prognostic factor, combined with the anticipated distribution of the factor and the endpoint. A pragmatic rationale can also be provided.

Statistical analysis methods

Some key predictors need to be considered for adjustment of prognostic factor associations ('adjustment model'), with the aim to disentangle the 'independent' association of the prognostic factor. The choice of the set will depend on the severity of TBI and the availability of data. Core predictors for moderate and severe TBI may include age, motor score (or full GCS), and pupillary reactivity (based on literature, and the CRASH and IMPACT Core models) (Supplements, Table 1). Other important

Box 3: Example of a rationale for sample size

Sample size calculation for the CENTER-TBI study

The sample size estimate (N = 5400) for the CENTER-TBI study was motivated by:

- Practical logistic considerations; higher numbers would imply too large a burden on local, national and international infrastructure.
- Power calculations for the different strata, targeting comparative effectiveness analyses, assuming a between-center and between-country heterogeneity as identified in previous research (expressed by variance parameter from a random effects model, tau of 0.43).
- Postulated odds ratios for intervention effects of approximately 5% improvement in outcome, to be evaluated in comparative effectiveness research.

Overall, a sample size of 5400 subjects would provide statistical power to detect odds ratios of 1.2 associated with differences in process characteristics of specific interventions with a power of 80%.

prognostic factors may include CT characteristics, secondary insults, and biomarker measurements.[23] For mild TBI, a core set is less well defined and may depend on the outcome considered, such as post-concussive symptoms, neurocognitive functioning, and health-related quality of life. In defining confounders, we should follow epidemiological principles, and not adjust for intermediate factors, which are positioned in-between the prognostic factor and the outcome.

The analysis with the adjustment model should ideally be described in detail, not only with respect to the selection of potentially confounding factors, but also on their coding, and the approach to missing values (see Box 4: Missing values). Furthermore, the relations of the prognostic factors to the variables in the adjustment model need to be studied, including correlation analyses and graphical inspections. A final issue is how we deal with continuous variables (see Box 5: Continuous variables).

Descriptive results

It is important to show the flow of patients through the study, including the number of patients included in each stage of the analysis and number of dropouts. The number of patients and the number of events need to be clear for each of the analyses performed. Baseline characteristics of the patients must be described in sufficient detail, including distributions of basic demographic characteristics (age, sex, and GCS), standard (disease-specific) prognostic variables, and the prognostic factors considered, including the numbers of missing values.

Box 4: Missing values

For missing values, multiple imputation has evolved as a standard statistical tool. For many research questions it is suboptimal to simply drop records with a missing value (complete case analysis). It may be reasonable to drop a variable with high numbers of missing observations.

Multiple imputation

- Multiple imputation may often be reasonable for missing values in the variables in the adjustment model, to maximize the available sample size for the adjusted analysis. The assumption is that missingness may be related to other variables, but not to unmeasured confounders. To make this 'Missing At Random' assumption plausible, it is advised to let the imputation model have a rich set of variables: prognostic factors, context factors (e.g. place (site) and time (year) of inclusion), and the outcome.
- Multiple imputation may also be used for the prognostic factor under study. Such imputation may be more controversial since we may not want to project findings from the prognostic factor – outcome relation in the complete data on the incomplete data. On the other hand, imputation is efficient, especially if the prognostic factor is correlated to other factors. Repeating the imputations multiple times should appropriately capture uncertainty in the process.
- Even more controversial is the imputation of missing outcome data. Statistically, this approach is especially useful if correlates of outcome are available, such that the correlation structure can be exploited. For example, missing 6 month GOS might be imputed for a patient if 3 and 12 month GOS are available in the data. Again, uncertainty should be captured appropriately by repeating the imputation procedure several times. (Richter et al., 2019)
- Each imputed data set is analyzed as a complete set, with combination of results according to Rubin's rules. (Van Buuren, 2018)

Statistical results

Results of univariable and adjusted analyses should be shown to document the relationship between the prognostic factor and the outcome, with the estimated effect size (for example, odds ratio and confidence interval). If imputation of missing observations must be done, modeling results should be compared to the results from a complete case analysis. In complete case analysis, participants that have missing data on one or multiple predictor variables are excluded. A comparison of the results between imputation and complete case analysis is especially important if participants have missing data on the prognostic factor under study. To evaluate the impact of missing values, it may also be insightful to present patterns of missingness.

Incremental predictive value is often of interest. This can be indicated by measures for discrimination, such as the increase in the concordance statistic (c statistic, or area under the receiver operating characteristic (ROC) curve (AUC)), and overall fit, such as pseudo R^2.[24] The c statistic or AUC ranges between 0.50 (no discrimination) and 1.0 (perfect discrimination).

Box 5: Continuous variables

A common approach is to dichotomize prognostic factors as normal / abnormal, or normal / elevated. Such dichotomization implies a loss of information if the original variable was continuous. (Royston et al., 2006) This loss can be quantified by comparing model fit with a continuous version of the prognostic factor and a dichotomized version, expressed e.g. as explained variability (R^2 statistics). There are many arguments why dichotomization should be avoided in medical research. (Royston, 2006)

Instead, the shape of the association of a continuous predictor with the outcome needs to be examined carefully. A linear association may be considered as a starting point. Log transformations are common to consider for biomarkers. Various other types of non-linear functions can be used, which provide greater flexibility, such as square terms, splines or fractional polynomials. Graphical inspection is also useful to visualize relations that are difficult to grasp from formulas. Differences in fit can be examined e.g. by R^2 statistics.

Continuous variables in the IMPACT study

Linear relations with outcome were good approximations after assessment of nonlinearity using restricted cubic splines for the continuous predictors age and glucose. A positive linear relation was observed for age and glucose, with higher values being associated with poorer prognosis (Figure 1).

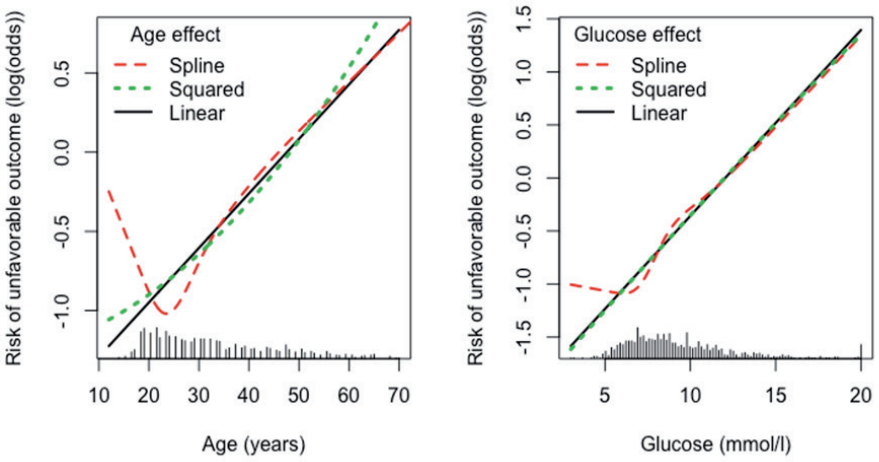


Figure 1 Nonlinearity assessment of continuous variables age (left) and glucose (right) in parts of the International Mission for Prognosis and Analysis of Clinical Trials in Traumatic Brain Injury (IMPACT) study.

For instance, the discriminatory power of the IMPACT models was calculated with a cross-validation procedure within the IMPACT data.[8] The discriminative ability, indicated with the AUC, increased with increasing model complexity; the AUC was 0.74 for the core, 0.77 for the CT, and 0.79 for the lab model for mortality in the Tirilazad US trial (Figure 2). These results showed the incremental predictive value of the predictors in the CT and lab IMPACT models.

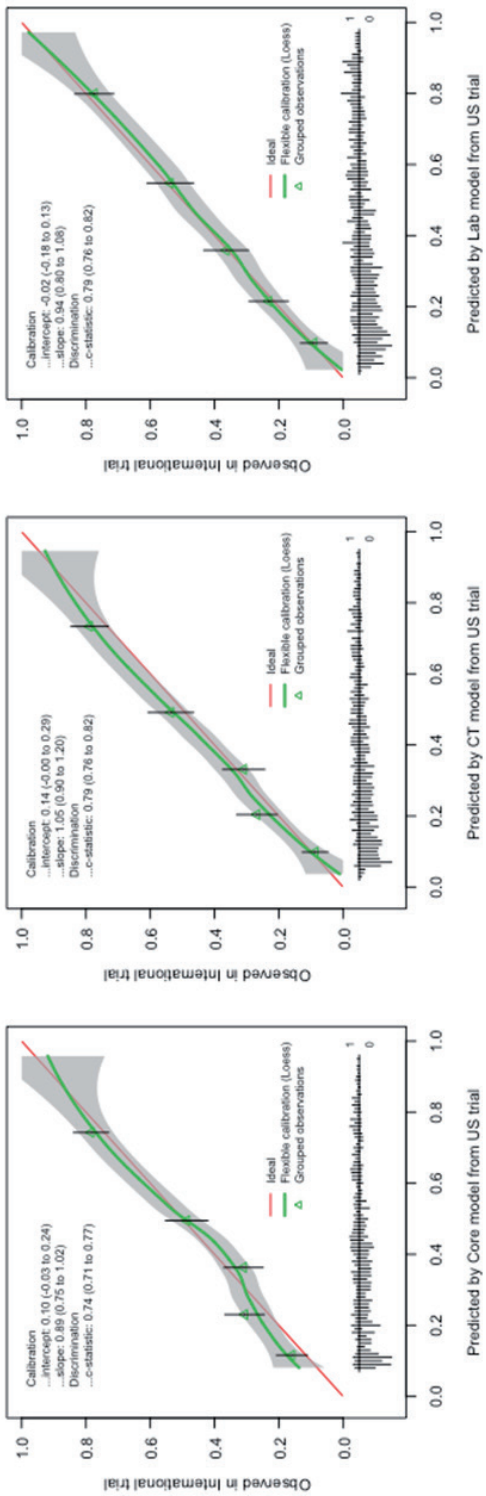


Figure 2: Model performance of the International Mission for Prognosis and Analysis of Clinical Trials in Traumatic Brain Injury (IMPACT) core (left), computed tomography (CT) extended (middle) and laboratory (right) models at external validation in the tirilazad United States trial.

Interpretation

Although studies on prognostic factors or biomarkers often have quite positively framed conclusions, only a few prognostic factors or markers have been validated and proven clinically useful.[25, 26] Results should hence be interpreted with caution, and in the context of the pre-specified hypotheses and other relevant studies. Ideally, replication is done in similar studies. Limitations of the study need to be discussed, and implications for future research, including the need for validation studies. A prognostic factor may hint at a biological mechanism, opening new avenues for further research, including potential new interventions. A prognostic factor may also simply indicate an association without a deeper explanation. Such a factor may still be useful by allowing for more accurate risk estimation, specifically as a building block with other prognostic factors in prediction models.

Guidance for prediction model research

The TRIPOD guideline consists of 22 items that need to be reported in a study that focuses on the development and/or validation of a prediction model (Supplements, Table 4).[9] Similar to REMARK, the guideline was endorsed by multiple journals,[27, 28] and an extensive 'Explanation and Elaboration' document is available.[29] We focus on the items that are most relevant to prediction models in TBI research (Table 3), and provide examples for illustration.

Modeling context

Prediction model development or validation studies share many aspects with prognostic factor research. In both, the selection of patients, the measurements aspects of prognostic factors, and the study design need attention. Prediction models aim to estimate absolute risk and are best developed or validated in cohort studies. The lowest risk of bias is expected in prospective, and cohorts of consecutive patients. Prospective cohort studies are increasingly being conducted including in the context of the InTBIR initiative.[14] Data from a RCT has the advantage of careful, standardized, high-quality data collection. However, a RCT typically has specific eligibility criteria, which results in a relatively narrow patient selection. Furthermore, not all consecutive patients may consent to participate. A cohort of patients that is not representative of all consecutive patients is not optimal for prediction modeling. An optimal cohort for prediction models consists of unselected consecutive patients which meet the eligibility criteria for the target population and was collected through prospective, careful, standardized and uniformed high-quality data.

Table 3: Specific considerations for prognostic models in TBI, building on the TRIPOD guideline (Collins, 2015)

Topic	Description
Candidate predictors and modeling	All prognostic factors considered in developing or validating the multivariable prediction model need to be mentioned, including how and when they were measured. Discuss approaches to dealing with missing values.
Model development	The modeling approach needs careful description, including dealing with continuous predictors and statistical interactions. Estimation of associations may include shrinkage and penalization techniques to prevent that predictions for new patients are too extreme, a problem that occurs with traditional estimation methods. Finally, clustering of data needs to be considered, e.g. in a multicenter context, or multi-cohort context, with fixed or random effect modeling.
Model performance	Important measures include discrimination (c statistic), calibration (graphical assessment, with intercept and slope as key parameters), overall performance (e.g. R^2), and indications of decision quality (decision curve analysis, if clinical decisions are to be supported by the model).
Model validation	Methods for internal validation, such as bootstrap resampling, are important at model development to correct performance estimates for statistical optimism. External validation is a stronger test, and starts with a clear description of how the predictions were calculated and a comparison of the validation and development samples. "Internal-external" validation should be considered for multicenter or multi-cohort studies.
Interpretation	Multivariable prediction models have a long tradition in TBI research, but the clinical role of the developed or validated model needs careful discussion, e.g. merely informing, benchmarking, or supporting decision making.

The outcome considered for a prediction model should be patient centered. Mortality and the Glasgow Outcome Scale(-extended) (GOS and GOSe) six months post-injury have often been used, in line with the primary endpoints in RCTs (see Box 2: A selection of outcomes in prognostic research in TBI). In prognostic model research, the outcome measure depends several factors, including the clinical severity of patients, clinical endpoint, and the purpose of the model. The GOSe is a valuable outcome scale in moderate and severe TBI. It is, however, a relatively simplistic scale for assessment of global outcome after TBI, which lacks sensitivity for mild TBI patients. We may, therefore, consider to include other outcomes, such as health-related quality of life, post-concussive symptoms, and neurocognitive functioning, especially for patients following mild TBI.

As for prognostic factor research, the sample size needs to be justified. In addition to reasoning based on the anticipated effect size of a prognostic factor, sample size can

also be motivated by rules of thumb, such as having at least ten events per candidate predictor.[6]

Candidate predictors

The CRASH and IMPACT models considered only three to four predictors for their Basic and Core models respectively (Supplements, Table 1). This narrow selection of predictors was motivated by a literature review to identify the most common and strong key predictors in moderate to severe TBI.[13] A wider selection may be of interest to improve the predictive ability. It may then make sense to consider extensions upon the small set of key predictors rather than start a 'de novo model' building exercise. In mild TBI, prediction models and key predictors are less well established, which may motivate a more exploratory approach.

If the number of candidate predictors is limited (say less than 20), this is considered as modeling of low dimensional data. A different and challenging field is that of high dimensional data, where numbers of candidate predictors may exceed the numbers of patients ("p>n situation", see Box 6: Modeling techniques). In any case, all candidate predictors that are used in developing or validating the multivariable prediction model should be listed, with a description of how and when they were measured.

As discussed for prognostic factors, missing values are a key issue. A candidate predictor may be excluded from further modeling if it has many missing values (e.g. more than 50% missing values). Traditional multivariable statistical analysis may suffer severely from missing values: a single missing value in one of the candidate predictors causes the full patient record to be discarded. Multiple imputation is a well-known strategy to fill in missing values. It allows for statistically appropriate use of records with one or more missing values and should hence be preferred over a complete case analysis.

Model development

Prognostic models are typically developed using regression analysis. Recently, machine learning approaches, such as support vector machines, random forests and neural nets, have gained increased attention. It is yet unclear whether and in which scenarios such methods outperform regression analysis in prediction of outcome following TBI.[30]

Modeling a combination of predictors poses challenges that include potential non-linear associations, interaction effects of predictors, optimal estimation, and dealing with clustering. When we consider a continuous predictor as a linear term in a prediction model, we assume that the effect is the same at each part of the range of the predictor. For instance, we may assume that the effect of being 10 years older is the same at the age of 40 (50 versus 40) and 70 (80 versus 70) years for patients following

Box 6: Modeling techniques

Prognostic models are usually developed with multivariable regression techniques on data from (prospective) cohort studies. Prognostic models for TBI most commonly predict binary outcomes (e.g. mortality, unfavourable outcome). For such binary outcomes the logistic regression model is the most widely used statistical technique. Recently, modern modeling techniques, such as penalized estimation and machine learning, are gaining increased attention. Machine learning techniques aim to learn more directly from the data, without assuming some type of underlying statistical model. These techniques are more flexible than traditional modeling techniques (e.g. can adjust the complexity of the method according to the data, and can capture complex relationships between the factors and the outcome). However, more data is required to obtain accurate estimates of model parameters, and reduce the risk of overfitting.

Table Traditional and modern modeling techniques

Modeling techniques	Methods
Regression	Logistic regression Generalized Additive Model
Penalized estimation	Ridge regression Lasso Elastic net
Machine learning techniques	Neural network Random forest Support Vector Machine

TBI. If a non-linear relation is expected, we can use flexible functions such as splines or fractional polynomials in regression models. Modern modeling and algorithmic approaches, such as generalized additive models (gam) or Support Vector Machines (SVMs), may include even more flexible high dimensional smoothness in the relations of continuous predictors to the outcome.

Combinations of prognostic factors may have differential effects, meaning that statistical interactions may be present. Multiplicity is a threat in attempts of modeling such interactions. For example, five predictors imply that ten potential two-way interactions could be studied, while ignoring higher-order interactions. The number of potential interactions rises quickly with a large number of predictors (e.g. 45 for 10 predictors, and 190 for 20 predictors). Currently, available prediction models often ignore such potential interactions and merely rely on the main effects of predictors. If assessed, it may be useful to perform overall tests of significance (e.g. considering all potential interactions with age and sex).[31] If such a single overall test does not show significant results, we may decide to ignore the interactions. Alternatively, tree-based methods like CART and Random forests indirectly consider interactions between factors. However, these methods are at risk for overfitting.[30]

Another challenge lies in the optimal estimation of prognostic associations. Statistical overfitting of available data is a key problem in prediction models: patterns in the data are described that do not generalize outside the specific data set considered.[32] As the complexity of the model (e.g. the number of coefficients or parameters estimated) increases, there is a greater risk for overfitting. Such overfitting may be reduced by reducing the number of examined prognostic associations in a model. Regression coefficients can be shrunk towards zero for less extreme and more stable predictions.[33, 34] Similarly, penalization procedures can be followed, such as ridge regression, and penalized estimation. A particularly promising approach is the LASSO (Least Absolute Shrinkage and Selection Operator), which shrinks some coefficients to zero, hence effectively reducing the set of candidate predictors. Such selection through penalization is a statistical improvement over classical stepwise selection methods (e.g. backward stepwise selection based on p-values). Stepwise selection methods have many disadvantages; for instance, these methods lead to too extreme estimates of the effect of selected predictors.[6]

Finally, the statistical analysis may need to consider the clustering of data, for instance, in the context of a multi-center or multi-cohort study. Stratification by cluster can be achieved by a fixed effect approach (by conditioning on the cluster, e.g. with dummy variables for the studies), or by random effect modeling (also known as hierarchical modeling, or mixed effect modeling). Random effect modeling is advised in case of a larger number of clusters, for instance over five clusters, which allows for a quantification of the between cluster heterogeneity. Heterogeneity is commonly found in the baseline risk, while heterogeneity in prognostic effects may be less relevant, specifically, if clusters are similar in basic attributes (setting, inclusion criteria).[35] Updating of a model to a specific setting may be motivated by substantial heterogeneity across settings.[36, 37]

Model performance

Performance of prediction models is most commonly assessed with respect to discriminative ability: 'How well can we separate low risk (favourable outcome) from high-risk patients (unfavourable outcome)?' Discriminative ability is then measured with a concordance (c) statistic (equivalent to the area under the ROC curve for binary outcomes). The c statistic is a rank order statistic that ranges between 0.5 (no discrimination) and 1 (perfect discrimination). Limits for 'satisfactory' or 'good' discrimination are inherently subjective, and context-dependent. Values around 0.8 have been achieved for rather simple prediction models (Basic CRASH and Core IMPACT models, Supplements, Table 1). We note that the discriminative ability depends not only on model characteristics (strong prognostic associations), but also on the heterogeneity of the sample of patients (between patient differences and case-

mix). Hence, discrimination was substantially higher in unselected cohort studies compared to RCTs.[38]

Performance can further be quantified by calibration: 'Do close to *x* of 100 patients with a risk prediction of *x*% have the outcome?' A graphical assessment of calibration is attractive, with predictions on the x-axis and the outcome on the y-axis. Perfect predictions should be on the 45-degree line. For binary outcomes, the plot contains 0 and 1 values for the y-axis. Smoothing techniques can be used to estimate the observed probabilities of the outcome in relation to the predicted probabilities (e.g., using the loess algorithm or polynomials).[39, 40]

The calibration plot can be characterized by an intercept *a*, which indicates whether predictions are systematically too low or too high ('calibration-in-the-large'), and a calibration slope *b*, which should be 1.[41] At model development, a=0 and b=1 for classical regression models. At validation, calibration-in-the-large problems are common, as well as *b* smaller than 1, reflecting overfitting of a model.

Performance of the CRASH model was assessed in terms of calibration (calibration graph) and discrimination (c statistic). In the derivation sample, the CRASH models showed excellent discrimination, with c statistics above 0.80. Moreover, the models showed good calibration graphically.

Overall performance measures relate to goodness of fit. For binary outcomes, a model's goodness of fit can be assessed through measures similar to overall performance measures for linear models that indicate the explained variability, here labeled 'pseudo R^2'. Pseudo R^2 is based on the improvement in model likelihood over a null model. [42] These measures are especially useful to assess incremental value of predictors. Several variants are available, including Nagelkerke's and McFadden's R^2 , which both provide for a natural scaling between 0 and 100%.[43]

Recent developments include decision-analytic approaches that explicitly consider the relative weight of false-positive and true-positive classifications. Such weighting is essential if prediction models are used to help guide clinical decisions. However, prediction models have not been perceived clinically useful to help guide decisions with regard to level of care. However, they could be eventually used as part of decision support tools or aid for level of care decisions. 'Decision curve analysis' has been proposed as a technique to quantify clinical utility, if clinical decisions are to be supported by the model. Here, the clinical utility of a prediction model, or the extension of a model with a prognostic factor, is examined over a range of plausible decision thresholds.[44]

Model validation

Model performance can be estimated from the sample where the model was developed ('apparent performance'). Such assessments are usually optimistic since the model was optimized for the same data where performance is evaluated. Methods for internal validation, such as bootstrap resampling or cross-validation, are important during model development to correct the performance estimates.[45] Bootstrap procedures can estimate statistical optimism if all modeling steps are replayed per bootstrap resample, such as selection from a set of candidate predictors.[46] Both bootstrap resampling and cross-validation use all available data for model development and are therefore preferred methods for internal validation. A split sample approach uses only part of the data for model development and is not recommended because of inefficiency.[45]

External validation is a stronger test for a prediction model. Such an analysis starts with a clear description of how the predictions were calculated and a comparison of the validation and development samples. Specifically, the heterogeneity in case-mix is relevant: more heterogeneous validation samples will increase the expected discrimination.[47] Temporal validation includes validation in a more recent data set, while geographic validation includes validation in another place. Large scale multicenter studies provide good opportunities for what has been labeled 'internal-external validation'. With internal-external validation, every center or cohort is left out once when developing the models. Models are evaluated in the centers or cohorts that were not used for model development. This process is repeated until all participants have been used for model validation, so all centers are left out once, and model performance is estimated over all validations. This procedure is a variant of cross-validation, a technique that can also be used for internal validation. At internal validation, parts of the data are left out at random, while internal-external validation leaves non-random parts of the data out. A sufficient number of events is required for reliable assessment of performance.[48]

As an example, the IMPACT models were cross-validated across IMPACT cohorts and externally validated in the CRASH trial. [6] Across the IMPACT cohorts, the best performance was seen for the three observational studies, with AUCs over 0.80, whereas the RCTs showed lower discriminative abilities. At external validation in the CRASH trial, the discriminatory ability of the models increased with increasing complexity; the AUCs ranged between 0.77 and 0.80.

Interpretation

The clinical role of the model needs careful consideration. In the design of RCTs, prediction models may assist in patient selection or stratification.[49] In the analysis of RCTs, prognostic factors can be used for covariate adjustment. In that case, we advise

that the model is refitted in the sample under study, preferably with the prognostic factors as individual variables.[50] One of the aims of this adjustment is to correct for any imbalance that may have arisen (by chance) between the treatment arms. Moreover, the statistical power of an adjusted analysis of the treatment effect may be larger than that of an unadjusted analysis.[51]

In observational studies, a prediction model may be used for confounder adjustment in comparing outcomes between centers or at a patient level. Prediction models may also serve as a reference in the evaluation of the incremental value of a new prognostic factor. A prognostic factor should be added to the reference model for a fair evaluation.[52] Such an evaluation should also consider the time of assessment. For example, when prognostic factors are measured during the first 3 days of hospital admission, a reference model can be based on all available information until 3 days rather than a model that focuses on the admission phase.

Validity of estimated associations and baseline risk is required if predictions are used to provide prognostic estimates to patients or relatives, benchmarking, or decision support. Similarly, validity of predictions is essential if models are used to support the decision-making process at the bedside. For example, models are used to decide whether or not a patient should have a CT performed following mild TBI when consulting in the emergency room.[53]

Model validations in various medical fields have shown that baseline risk often varies between settings.[54] For TBI, substantial between-center differences in baseline risk have been shown.[55] Using a prediction model in a specific setting hence requires consideration of the plausibility of applying absolute risk predictions from the development setting. Ideally, a validation study is performed, to verify whether the average observed and predicted risks are similar (calibration-in-the-large, reflecting correctness of baseline risk). Often, we find that some statistical updating of the baseline risk is needed.[37]

Summary

Establishing a reliable prognosis early after TBI is challenging due to several factors including the heterogeneity of the condition. Furthermore, prognostic research in TBI often has methodological limitations, which has resulted in a lack of reliable prognostic information for patients with moderate and severe TBI. We aimed to review methodological issues and provide tentative guidelines for prognostic research in TBI. For this purpose, we have considered two existing reporting guidelines: the REMARK

and TRIPOD guidelines, from which advice on appropriate methods can be inferred. For prognostic factor research, we emphasize the importance of transparent reporting of patient and specimen characteristics, study design, clinical endpoints, and statistical analysis. Prediction model research especially brings challenges for model specification, estimation, evaluation, validation, and presentation. The TRIPOD guidelines underscore the importance of transparent reporting of these aspects of model development and validation. Furthermore, we have highlighted modern approaches and opportunities, related to missing values, exploration of non-linear effects, and assessing between-study heterogeneity by leave-one-study-out cross-validation.

Discussion

Our review presents various methodological aspects of prognostic research and may provide a solid foundation for future studies of prognostic factors and prediction models in TBI. For prognostic factor research, we took the REMARK guidelines as a foundation. These guidelines have been developed by methodological experts and have been received positively by other scientists and editors of journals.[10, 18, 19] Similarly, we used TRIPOD as a foundation for methodological guidance for prediction model research.[9, 27, 28] These guidelines are primarily intended for transparent reporting. Additionally, advice on appropriate methods can be inferred from the items listed, and from the motivation for the inclusion of the items in these guidelines. Also, more in-depth material is available in the 'Explanation and Elaboration' documents of these guidelines.[20, 29]

In prognostic research it is important to carefully describe the selection of patients, the prognostic factors considered, and the study design. Previous reviews showed poor methodological quality of many model development studies in TBI,[12, 13] and improvements can be made with respect to dealing with missing values (multiple imputation), assessment of non-linear relations (using splines or other flexible functions), and estimation of prognostic associations (e.g. using LASSO). For prediction models, modern machine learning algorithms may prove to be useful for modeling of high dimensional data. No benefit is expected from such methods in low dimensional data.[30] Classical regression models may then be adequate, especially if the selection from candidate predictors is done carefully, and modern shrinkage or penalization techniques are used to prevent too extreme and optimistic predictions.

Validation of prognostic claims is essential in prognostic factor research and prediction model research. Prediction models need external validation to assess discriminative ability and reliability (calibration) of predictions in new settings. Heterogeneous

model performance can be interpreted as a warning signal that simple universal applicability of a 'global' prediction model should be reconsidered, and this indicates that the prediction model should be updated per setting.[55] Various methods are available to update models to a specific setting, including calibration-in-the large, adjustment of all regression coefficients, updating of individual predictor effects, and extending the model with new predictors.[37] Recent collaborative efforts, such as the InTBIR consortium, provide opportunities for validation. For prognostic factors, we suggest providing forest plots with estimates per study, as is standard for genomic analyses. For prediction models, we suggest internal-external validation procedures, where performance is assessed in a study that is left out of the model development process.

Prognostic research in TBI can be improved if the described key methodological principles are adhered to, and when research is performed in collaboration among multiple centers. Recent collaborative initiatives provide new opportunities for large-scale studies with cross-validation of promising findings.

R script

Script available from the authors.

Acknowledgements

We would like to thank the international initiative for TBI Research (InTBIR) for facilitating collaboration between the authors. Furthermore, we would like to thank the InTBIR collaborators, specifically Goundappa Balasubramani, Greta Carrara, Chris Lindsell, Andrew Maas, and Giovanni Nattino for their support for the InTBIR data analytics working group.

Author Disclosure Statement

IRARH, HFL, and EWS were supported by the European Union 7th Framework program (Center-TBI, EC grant 602150). AFT is the chairholder of the Canada Research Chair in Critical Care Neurology and Trauma and is supported by the Canadian Institutes of Health Research (CIHR) (TBI-Prognosis study, CIHR grant #275553 and CIHR Foundation Scheme Grant # 354039; HEMOTION Trial, CHIR grant #367141).

Supplementary material

Supplementary material is available at: https://www.liebertpub.com/doi/10.1089/neu.2019.6708

References

1. Turgeon, A.F., Lauzier, F., Burns, K.E.A., Meade, M.O., Scales, D.C., Zarychanski, R., Moore, L., Zygun, D.A., McIntyre, L.A., and Kanji, S. (2013). Determination of neurologic prognosis and clinical decision making in adult patients with severe traumatic brain injury: a survey of Canadian intensivists, neurosurgeons, and neurologists. Crit Care Med 4, 1086-1093.
2. Turgeon, A.F., Lauzier, F., Simard, J.-F., Scales, D.C., Burns, K.E.A., Moore, L., Zygun, D.A., Bernard, F., Meade, M.O., and Dung, T.C. (2011). Mortality associated with withdrawal of life-sustaining therapy for patients with severe traumatic brain injury: a Canadian multicentre cohort study. Cmaj 14, 1581-1588.
3. Turgeon, A.F., Dorrance, K., Archambault, P., Lauzier, F., Lamontagne, F., Zarychanski, R., Fowler, R., Moore, L., Lacroix, J., and English, S. (2019). Factors influencing decisions by critical care physicians to withdraw life-sustaining treatments in critically ill adult patients with severe traumatic brain injury. Cmaj 24, E652-E663.
4. Dijkland, S.A., Foks, K.A., Polinder, S., Dippel, D.W.J., Maas, A., Lingsma, H., and Steyerberg, E.W. (2019). Prognosis in moderate and severe traumatic brain injury: a systematic review of contemporary models and validation studies. J Neurotrauma.
5. Steyerberg, E.W., Moons, K.G.M., van der Windt, D.A., Hayden, J.A., Perel, P., Schroter, S., Riley, R.D., Hemingway, H., Altman, D.G., and Group, P. (2013). Prognosis Research Strategy (PROGRESS) 3: prognostic model research. PLoS Med 2, e1001381.
6. Steyerberg, E.W. (2008). Clinical prediction models: a practical approach to development, validation, and updating, Springer Science & Business Media: New York.
7. Collaborators, M.R.C.C.T. (2008). Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients. Bmj 7641, 425-429.
8. Steyerberg, E.W., Mushkudiani, N., Perel, P., Butcher, I., Lu, J., McHugh, G.S., Murray, G.D., Marmarou, A., Roberts, I., and Habbema, J.D.F. (2008). Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. PLoS Med 8, e165.
9. Collins, G.S., Reitsma, J.B., Altman, D.G., and Moons, K.G.M. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMC med 1, 1.
10. McShane, L.M., Altman, D.G., Sauerbrei, W., Taube, S.E., Gion, M., and Clark, G.M. (2005). Reporting recommendations for tumor marker prognostic studies (REMARK). J Natl Cancer Inst 16, 1180-1184.
11. Riley, R.D., Hayden, J.A., Steyerberg, E.W., Moons, K.G.M., Abrams, K., Kyzas, P.A., Malats, N., Briggs, A., Schroter, S., and Altman, D.G. (2013). Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. PLoS Med 2, e1001380.
12. Perel, P., Edwards, P., Wentz, R., and Roberts, I. (2006). Systematic review of prognostic models in traumatic brain injury. BMC medical informatics and decision making 1, 38.
13. Mushkudiani, N.A., Hukkelhoven, C.W.P.M., Hernández, A.V., Murray, G.D., Choi, S.C., Maas, A.I.R., and Steyerberg, E.W. (2008). A systematic review finds methodological improvements necessary for prognostic models in determining traumatic brain injury outcomes. J Clin Epidemiol 4, 331-343.
14. Tosetti, P., Hicks, R.R., Theriault, E., Phillips, A., Koroshetz, W., Draghia, A., and the Workshop Participants, R. (2013). Toward an international initiative for traumatic brain injury research. J Neurotrauma 14, 1211-1222.

15. Bouwmeester, W., Zuithoff, N.P.A., Mallett, S., Geerlings, M.I., Vergouwe, Y., Steyerberg, E.W., Altman, D.G., and Moons, K.G.M. (2012). Reporting and methods in clinical prediction research: a systematic review. PLoS Med 5, e1001221.
16. Justice, A.C., Covinsky, K.E., and Berlin, J.A. (1999). Assessing the generalizability of prognostic information. Ann. Intern. Med. 6, 515-524.
17. Steyerberg, E.W. and Harrell, F.E. (2016). Prediction models need appropriate internal, internal–external, and external validation. J Clin Epidemiol, 245-247.
18. Hayes, D.F., Ethier, S., and Lippman, M.E. (2006). New guidelines for reporting of tumor marker studies in breast cancer research and treatment: REMARK. Breast Cancer Res Treat 2, 237-238.
19. McShane, L.M., Altman, D.G., Sauerbrei, W., Taube, S.E., Gion, M., and Clark, G.M. (2006). REporting recommendations for tumor MARKer prognostic studies (REMARK). Breast Cancer Res Treat 2, 229-235.
20. Altman, D.G., McShane, L.M., Sauerbrei, W., and Taube, S.E. (2012). Reporting recommendations for tumor marker prognostic studies (REMARK): explanation and elaboration. BMC med 1, 51.
21. Lingsma, H.F., Roozenbeek, B., Li, B., Lu, J., Weir, J., Butcher, I., Marmarou, A., Murray, G.D., Maas, A.I.R., and Steyerberg, E.W. (2011). 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. J Neurosurg 3, 601-608.
22. Maas, A.I.R., Menon, D.K., Steyerberg, E.W., Citerio, G., Lecky, F., Manley, G.T., Hill, S., Legrand, V., and Sorgner, A. (2014). Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) A Prospective Longitudinal Observational Study. J Neurosurg 1, 67-80.
23. Winn, H.R. (2011). Youmans Neurological Surgery E-Book, Elsevier Health Sciences.
24. Steyerberg, E.W., Pencina, M.J., Lingsma, H.F., Kattan, M.W., Vickers, A.J., and Van Calster, B. (2012). Assessing the incremental value of diagnostic and prognostic markers: a review and illustration. EJCI 2, 216-228.
25. Ioannidis, J.P.A. (2013). Biomarker failures. Clin Chem 1, 202-204.
26. Ioannidis, J.P.A. and Bossuyt, P.M.M. (2017). Waste, leaks, and failures in the biomarker pipeline. Clin Chem 5, 963-972.
27. Localio, A.R. and Stack, C.B. (2015). TRIPOD: a new reporting baseline for developing and interpreting prediction models. Ann. Intern. Med. 1, 73-74.
28. Moons, K.G.M., Altman, D.G., Reitsma, J.B., and Collins, G.S. (2015). New guideline for the reporting of studies developing, validating, or updating a multivariable clinical prediction model: the TRIPOD statement. Adv Anat Pathol 5, 303-305.
29. Moons, K.G.M., Altman, D.G., Reitsma, J.B., Ioannidis, J.P.A., Macaskill, P., Steyerberg, E.W., Vickers, A.J., Ransohoff, D.F., and Collins, G.S. (2015). Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann. Intern. Med. 1, W1-W73.
30. van der Ploeg, T., Nieboer, D., and Steyerberg, E.W. (2016). Modern modeling techniques had limited external validity in predicting mortality from traumatic brain injury. J Clin Epidemiol, 83-89.
31. Harrell Jr, F.E. (2015). Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis, Springer.
32. Babyak, M.A. (2004). What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. Psychosom Med 3, 411-421.

33. Copas, J.B. (1985). Prediction equations, statistical analysis, and shrinkage. Prediction in criminology, 232-255.
34. Van Houwelingen, J.C. (2001). Shrinkage and penalized likelihood as methods to improve predictive accuracy. Stat Neerl 1, 17-34.
35. Austin, P.C., van Klaveren, D., Vergouwe, Y., Nieboer, D., Lee, D.S., and Steyerberg, E.W. (2017). Validation of prediction models: examining temporal and geographic stability of baseline risk and estimated covariate effects. Diagn Progn Res 1, 12.
36. Steyerberg, E.W., Borsboom, G.J.J.M., van Houwelingen, H.C., Eijkemans, M.J.C., and Habbema, J.D.F. (2004). Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. Stat Med 16, 2567-2586.
37. Vergouwe, Y., Nieboer, D., Oostenbrink, R., Debray, T.P.A., Murray, G.D., Kattan, M.W., Koffijberg, H., Moons, K.G.M., and Steyerberg, E.W. (2017). A closed testing procedure to select an appropriate method for updating prediction models. Stat Med 28, 4529-4539.
38. Roozenbeek, B., Lingsma, H.F., Lecky, F.E., Lu, J., Weir, J., Butcher, I., McHugh, G.S., Murray, G.D., Perel, P., and Maas, A.I.R. (2012). Prediction of outcome after moderate and severe traumatic brain injury: external validation of the IMPACT and CRASH prognostic models. Crit Care Med 5, 1609.
39. Austin, P.C. and Steyerberg, E.W. (2014). Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. Stat Med 3, 517-535.
40. Nattino, G., Finazzi, S., and Bertolini, G. (2014). A new calibration test and a reappraisal of the calibration belt for the assessment of prediction models based on dichotomous outcomes. Statistics in medicine 14, 2390-2407.
41. Cox, D.R. (1958). Two further applications of a model for binary regression. Biometrika 3/4, 562-565.
42. Hemmert, G.A.J., Schons, L.M., Wieseke, J., and Schimmelpfennig, H. (2018). Log-likelihood-based pseudo-R 2 in logistic regression: Deriving sample-sensitive benchmarks. Sociol Methods Res 3, 507-531.
43. Smith, T.J. and McKenna, C.M. (2013). A comparison of logistic regression pseudo R2 indices. Multiple Linear Regression Viewpoints 2, 17-26.
44. Vickers, A.J. and Elkin, E.B. (2006). Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making 6, 565-574.
45. Steyerberg, E.W., Harrell Jr, F.E., Borsboom, G.J.J.M., Eijkemans, M.J.C., Vergouwe, Y., and Habbema, J.D.F. (2001). Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol 8, 774-781.
46. Steyerberg, E.W., Bleeker, S.E., Moll, H.A., Grobbee, D.E., and Moons, K.G.M. (2003). Internal and external validation of predictive models: a simulation study of bias and precision in small samples. J Clin Epidemiol 5, 441-447.
47. Debray, T.P.A., Vergouwe, Y., Koffijberg, H., Nieboer, D., Steyerberg, E.W., and Moons, K.G.M. (2015). A new framework to enhance the interpretation of external validation studies of clinical prediction models. J Clin Epidemiol 3, 279-289.
48. Vergouwe, Y., Steyerberg, E.W., Eijkemans, M.J.C., and Habbema, J.D.F. (2005). Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. J Clin Epidemiol 5, 475-483.
49. Roozenbeek, B., Maas, A.I.R., Lingsma, H.F., Butcher, I., Lu, J., Marmarou, A., McHugh, G.S., Weir, J., Murray, G.D., and Steyerberg, E.W. (2009). Baseline characteristics and statistical power in randomized controlled trials: selection, prognostic targeting, or covariate adjustment? Crit Care Med 10, 2683-2690.

50. Turner, E.L., Perel, P., Clayton, T., Edwards, P., Hernández, A.V., Roberts, I., Shakur, H., Steyerberg, E.W., and Collaborators, C.T. (2012). Covariate adjustment increased power in randomized controlled trials: an example in traumatic brain injury. J Clin Epidemiol 5, 474-481.
51. Hernández, A.V., Steyerberg, E.W., and Habbema, J.D.F. (2004). Covariate adjustment in randomized controlled trials with dichotomous outcomes increases statistical power and reduces sample size requirements. J Clin Epidemiol 5, 454-460.
52. Xanthakis, V., Sullivan, L.M., Vasan, R.S., Benjamin, E.J., Massaro, J.M., D'Agostino Sr, R.B., and Pencina, M.J. (2014). Assessing the incremental predictive performance of novel biomarkers over standard predictors. Stat Med 15, 2577-2584.
53. Stiell, I.G., Wells, G.A., Vandemheen, K., Clement, C., Lesiuk, H., Laupacis, A., McKnight, R.D., Verbeek, R., Brison, R., and Cass, D. (2001). The Canadian CT Head Rule for patients with minor head injury. Lancet 9266, 1391-1396.
54. Riley, R.D., Ensor, J., Snell, K.I.E., Debray, T.P.A., Altman, D.G., Moons, K.G.M., and Collins, G.S. (2016). External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. bmj, i3140.
55. Steyerberg, E.W., Nieboer, D., Debray, T.P.A., and van Houwelingen, H.C. (2019). Assessment of heterogeneity in an individual participant data meta-analysis of prediction models: An overview and illustration. Stat Med 22, 4290-4309.

Chapter 5

Does Poor Methodological Quality Lead to Poor Model Performance? An illustration in Traumatic Brain Injury

I.R.A. Retel Helmrich, A. Mikolic, D.M. Kent, H.F. Lingsma, L. Wynants, E.W. Steyerberg, and D. van Klaveren

Published in: Diagnostic and Prognostic Research

Abstract

Background: Prediction modeling studies often have methodological limitations, which may compromise model performance in new patients and settings. We aimed to examine the relation between methodological quality of model development studies and their performance at external validation.

Methods: We systematically searched for externally validated multivariable prediction models that predict functional outcome following moderate or severe traumatic brain injury. Risk of bias and applicability of development studies was assessed with the Prediction model Risk Of Bias Assessment Tool (PROBAST). Each model was rated for its presentation with sufficient detail to be used in practice. Model performance was described in terms of discrimination (AUC), and calibration. Delta AUC (dAUC) was calculated to quantify the percentage change in discrimination between development and validation for all models. Generalized estimation equations (GEE) were used to examine the relation between methodological quality and dAUC while controlling for clustering.

Results: We included 54 publications, presenting ten development studies of 18 prediction models, and 52 external validation studies, including 245 unique validations. Two development studies (four models) were found to have low risk of bias (RoB). The other eight publications (14 models) showed high or unclear RoB. The median dAUC was positive in low RoB models (dAUC 8%, [IQR -4% to 21%]) and negative in high RoB models (dAUC -18%, [IQR -43% to 2%]). The GEE showed a larger average negative change in discrimination for high RoB models (-32% (95% CI: -48 to -15) and unclear RoB models (-13% (95% CI: -16 to -10)) compared to that seen in low RoB models.

Conclusion: Lower methodological quality at model development associates with poorer model performance at external validation. Our findings emphasize the importance of adherence to methodological principles and reporting guidelines in prediction modeling studies.

Key words: Prediction modeling studies; Prognosis; PROBAST; Traumatic Brain Injury

Introduction

Prediction models estimate an individual's risk of a certain outcome based on a combination of (clinical) characteristics. Despite numerous efforts to provide guidelines and recommendations for the reporting and analyses of prediction modeling studies (1, 2), these studies often suffer from methodological limitations. Prior reviews have judged the methodological quality of prediction modeling studies generally as poor (3-5), due to the small sample size of the derivation cohort, and a lack of internal and external validation. Furthermore, prediction modeling studies often suffer from incomplete reporting, which could indicate that specific methodological aspects were not considered.

Prognostic models that predict functional outcome after moderate and severe TBI are abundant in the literature; 67 prognostic models for moderate and severe TBI have been developed, of which 31 were externally validated over the past decades (6). The ability to accurately predict patient outcome after traumatic brain injury (TBI) has an important role in clinical practice and research. Outcome prognostication may support clinicians in providing reliable information to patients and relatives, and guide clinical management and study design.

Satisfactory methodological quality of prediction modeling studies is considered a prerequisite before implementation of the model in clinical practice should be advocated. Usability of a prediction model, which could be determined by whether sufficient information is provided about the model to enable use in practice, is expected to stimulate its implementation. The reporting of the full model equation enables validation, whereas the development of an online calculator might facilitate use in clinical practice. Assessing the quality of included studies and model usability are therefore important steps in systematic reviews of prediction models.

Recently, the PROBAST tool has become available to assess the risk of bias and concerns regarding applicability of studies that develop and/or validate a multivariable prediction model in systematic reviews (7). Risk of bias indicates that shortcomings in the study design, conduct, or analysis may lead to systematically distorted estimates of model predictive performance. Methodological quality of prediction modeling studies might therefore be related to model performance, with lower methodological quality resulting in poor performance, especially in new patients and settings.

The aim of our study was to empirically examine the relation between the methodological quality of a model development study and model performance at external validation.

Methods

Systematic Search

We used data from a recent systematic review of multivariable prediction models based on admission characteristics (first 24 hours after injury), for patients after moderate and severe TBI (Glasgow Coma Scale ≤ 12) that were published between 2006-2018 (6) (Supplementary Table 1 and 2). The protocol of this systematic review has been registered on PROSPERO (registration number 2016: CRD42016052100). Studies were eligible for inclusion if they reported on the development, validation or extension of multivariable prognostic models for functional outcome in patients aged ≥ 14 years with moderate and severe TBI. There were no limitations concerning outcome measurement, provided that functional outcome was measured between 14 days and 24 months after injury.

We updated the systematic search for 2019-2021 (December 2018-June 2021). One investigator (IRH) independently screened records for possibly relevant studies based on title and abstract. Subsequently, full texts of potentially relevant articles were assessed for eligibility. In case of doubt, a second investigator (AM) was consulted.

Study Selection

We selected externally validated prediction models for moderate and severe TBI (Supplementary Table 1) as previously identified by Dijkland et al., (2019) or identified through the updated search. To be included, the model development study had to report model performance in terms of discriminative ability. The external validation could be described in the same publication that described model development, or in a separate publication.

Data Extraction

One investigator (IRH) extracted data from the included studies. A check for all included studies was performed by a second investigator (AM). For the development studies, the data extraction form was based on the Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modeling Studies (CHARMS) checklist (8), and included the source of data, participants, outcome, sample size, predictors, missing data, model development, performance measures, and presentation. For the validation studies, data was extracted on the study design, setting, inclusion criteria, sample size, and model performance. To ensure consistency of the data extraction, the form was tested on two studies by both investigators.

If one publication reported on multiple prediction models, data extraction was performed separately for each model. Prediction models were classified as separate if they

included a different set of predictors (e.g. IMPACT core, and IMPACT extended (9)). Models with identical set of predictors, but for different outcomes (e.g. mortality and unfavorable outcome) were not classified as separate models.

Risk of Bias and Applicability

Risk of bias and applicability of included development studies were assessed with the Prediction model Risk Of Bias Assessment Tool (PROBAST) (7). Judgements on high, low, or unclear risk of bias for the model development studies were made for five key domains (participant selection, predictors, outcome, sample size and participant flow, and analysis) using 20 signaling questions (Supplementary Table 3). We also used a short form based on the PROBAST including 8/20 signaling questions, which was recently proposed and validated, and showed high sensitivity (98%) and perfect specificity to identify high RoB (10).

To determine if there was a reasonable number of outcome events in a logistic regression (PROBAST item 4.1), The lowest number of events in the smallest group of two outcome frequencies (patients with the outcome versus without the outcome) was divided by the total degrees of freedom used during the whole modeling process. The total degrees of freedom was based on the number of variables (continuous variables) or categories (categorical variables) in the model; henceforth referred to as Events Per Parameter (EPP). All candidate predictors were considered as part of the modeling process, including those not selected for the multivariable model based on univariable regression analysis or selection procedures. We assumed a reasonable number of outcome events when EPP ≥ 10.

Concerns regarding the applicability of an included study to the review question can arise when the population, predictors, or outcomes of the included study differ from those specified in the review question (7). Applicability was judged based on three key domains (participant selection, predictors, and outcome).

Two reviewers (IRH and AM) independently completed the PROBAST checklist (Supplementary Table 3). A third independent reviewer (LW) scored two of the model development studies (17%). Discrepancies between reviewers were resolved through discussion or by consultation with a senior member (DvK) of the review team. The RoB, applicability and usability were reported per study, in which we presented one assessment for models described in the same publication, but with a different set of predictors (e.g. IMPACT core, and IMPACT extended) and models with identical set of predictors, but for different outcomes (e.g. mortality and unfavorable outcome). An overall judgement about risk of bias and applicability of the prediction model

study was reached based on a summative rating across all domains according to the PROBAST criteria (low, high, or unclear).

Usability

A model's usability in research and clinical practice was rated for its presentation with sufficient detail to be used in the intended context and target population. The model was deemed usable in research if the full model equation or sufficient information to extract the baseline risk (intercept) and individual predictor effects was reported, and usable in clinical practice if an alternative presentation of the model was included (e.g. a nomogram, score chart or web calculator).

Relatedness

For validation studies, we assessed the similarity between the derivation population and the validation population for each study, which we refer to as "relatedness". To judge relatedness we created a rubric, aiming to capture various levels or relatedness by dividing the validation studies into three categories: 'related', 'moderately related', and 'distantly related' (6) (Supplementary Table 4). The rubric contained three domains: I) setting (Intensive Care Unit, Emergency Department, Ward; Country; Not specified), II) inclusion criteria and III) outcome assessment and timing. Studies that did not meet the domain about setting were judged 'moderately related', whereas studies that did not meet the domains about inclusion criteria and/or outcome assessment and timing were judged 'distantly related'.

Model Performance

Model performance was summarized in terms of discrimination and calibration. In prior studies, discrimination was assessed in terms of the c statistic or AUC, which ranges between 0.50 (no discrimination) and 1.0 (perfect discrimination). In prior studies, calibration was typically assessed with the calibration intercept *a*, which indicates whether predictions are systematically too low or too high, and should ideally be 0. Prior studies also reported the calibration slope *b* which indicates whether the overall prognostic effect of the linear predictor of the developed model is over- or underestimated, and should ideally be 1.

Relation between Methodological Quality and Model Performance

To quantify the relation between methodological quality at development and model performance at external validation, we first calculated the change in discriminative performance between the derivation cohort and the validation cohort. The percent change in discrimination was calculated as follows:

$$\%\ change\ in\ discrimination = \frac{(validation\ AUC - 0.5) - (derivation\ AUC - 0.5)}{(derivation\ AUC - 0.5)}\ x\ 100$$

For instance, when the AUC decreases from 0.70 in derivation to 0.60 in validation, this drop of 0.10 points represents a 50% loss in discriminative power (since 0.50 represents the lowest possible value). We calculated the median and interquartile range (IQR) of the change in discrimination for low, high and unclear RoB models.

We used generalized estimated equations (GEE) to estimate the effect of the RoB classification (Low; High; Unclear RoB based on the original PROBAST) on the observed change in discrimination, taking into account the correlation between validations of the same model and similarity in study design between the development and validation study (Similar; Cohort to trial; Trial to cohort).

Evidence synthesis

A synthesis was provided for the included development and external validation studies. Extracted data, RoB, applicability and usability were presented in summary tables and where appropriate in graphical representations. Figures were constructed with R software version 3.6 (R Foundation for Statistical Computing, Vienna, Austria).

Results

Study Selection

We included 54 publications comprising 18 multivariable regression models (Figure 1). The publications include ten (10/54) model development papers, describing 18 models, and 52 (52/54) validation papers, describing 245 external validations. These 18 models were previously described by Dijkland et al., (2020), and no additional models were included based on the updated search strategy.

Study Characteristics

The 18 multivariable prognostic models predict mortality or unfavorable outcome at discharge or up to twelve months after hospital admission and were published between 1985 and 2021 (Supplementary Table 5). Four models (4/18; 22%) were developed in adult patients (aged > 14 years) who were admitted to the ICU (11-13), and fourteen models (14/18; 78%) were specifically developed in patients with TBI (9, 14-19). Data for model development were collected through single or multi-center observational cohort studies, randomized controlled trials (RCTs) or pooled data

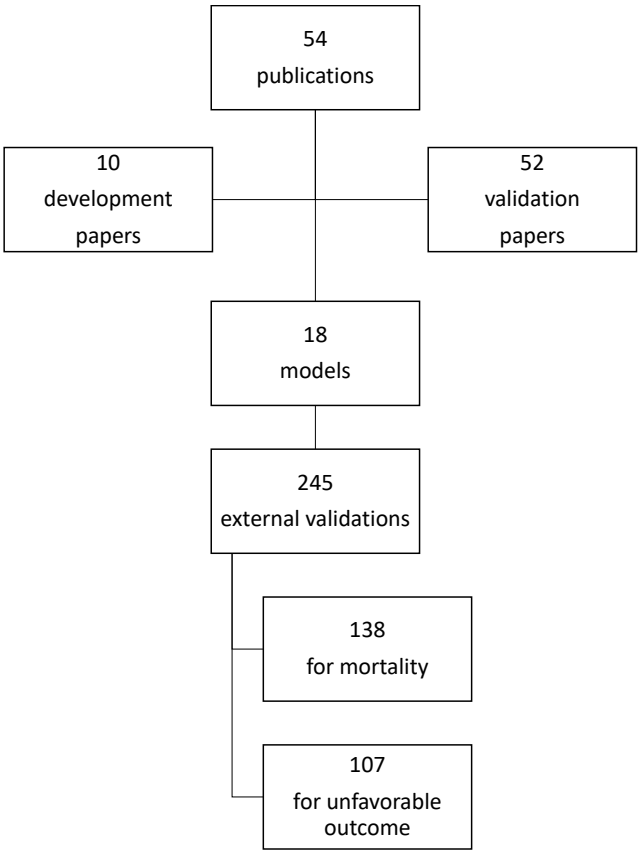


Figure 1: Flowdiagram of included studies based on the systematic search

derived from both cohort studies and RCTs. All studies, except for Yuan et al., (19) used prospective data.

Candidate predictors of outcome following TBI were collected at admission and typically included a combination of demographic, clinical and radiology characteristics. The number of missing predictor and outcome data was not reported in three studies (3/10; 30%) (Supplementary Table 5 continued). Three studies (3/10; 30%) applied imputation methods for handling missing data. Seven studies (7/10; 70%) used a selection procedure, for instance stepwise selection, to reduce the number of predictors that were included in the final model.

Five studies (5/10; 50%) used an internal validation procedure (e.g. bootstrap validation procedure or cross-validation), whereas in the other five studies (5/10; 50%) the internal validation procedure was lacking or inefficient (split-sample procedure).

The AUCs at development ranged between 0.71 to 0.90 for the prediction of mortality, and between 0.65 to 0.90 for the prediction of unfavorable outcome. Of the nine development studies that described model performance in terms of calibration, three studies (3/9; 33%) exclusively reported the Hosmer-Lemeshow goodness-of-fit test and one (1/9; 11%) exclusively showed calibration graphically using a calibration plot, whereas five studies (5/9; 55%) reported both the Hosmer-Lemeshow goodness-of-fit test and a calibration plot.

Methodological quality of model development studies

Methodological quality of model development studies was assessed in terms of applicability and Risk Of Bias (RoB) with the PROBAST checklist (Table 1). Of the ten model development studies, eight (8/10; 80%) were judged high RoB (Table 2). In each case (8/8), the statistical analysis (analysis domain) resulted in a high RoB, due to insufficient sample size, suboptimal handling of missing data and lack of or insufficient internal validation procedures (e.g. split-sample procedure). Four model development studies (4/10; 40%) were deemed high RoB in terms of applicability as these models were developed for patients admitted to the ICU and not strictly for patients following moderate and severe TBI. For most studies (9/10), the overall judgment on a short form based on the PROBAST, including 8/20 signaling questions, was consistent with the original PROBAST (Supplementary Table 6). Based on the short form, one study was identified as low RoB, but unclear RoB (CRASH models) on the original PROBAST, due to key information that was not reported.

Usability

Just over half of the model development studies (6/10; 60%) did not provide the full model equations or sufficient information to extract the baseline risk (intercept) and individual predictor effects (regression coefficients). Almost half of the studies (4/10; 40%) included sufficient information to externally validate the models (Table 2). Most (9/10; 90%) studies included a presentation of the final prediction models, such as a nomogram or score chart, which makes implementation of the model in clinical practice more feasible (Table 2).

External validation

The 18 prognostic models were externally validated 245 times (Supplementary Table 7). The IMPACT prognostic models were externally validated most extensively (127 times), followed by the CRASH models (56 times). Most (164/245, 67%) of the validation studies were judged 'distantly related' (Table 2), indicating that the validation cohort substantially differed from the model development study in terms of inclusion criteria and/or outcome assessment. Furthermore, 45/245 (18%) of the validation

Table 1: Methodological quality of model development studies for outcome following moderate and severe traumatic brain injury in terms of Applicability and Risk of Bias assessed with the original PROBAST and models' usability in research and clinical practice.

Study	Models	Applicability				Risk of bias					Usability	
		Participant selection	Predictors	Outcome	Overall applicability	Participant selection	Predictors	Outcome	Analysis	Overall RoB	Research	Practice
Knaus	APACHE II	H	L	L	H	L	L	L	H	H	y	y
Le Gall	SAPS II	H	L	L	H	L	L	L	H	H	n	y
Lemeshow	MPM II models	H	L	L	H	L	L	L	H	H	n	n
Signorini	Signorini	L	L	L	L	L	L	L	H	H	n	y
Hukkelhoven	Hukkelhoven model	L	L	L	L	L	L	L	L	L	y	y
Maas	Rotterdam CT score	L	L	L	L	L	L	L	U	U	n	y
Perel	CRASH models	L	L	L	L	L	L	L	U	U	n	y
Steyerberg	IMPACT models	L	L	L	L	L	L	L	L	L	y	y
Jacobs	Nijmegen models	L	L	L	L	L	L	L	H	H	y	y
Yuan	Yuan models	L	L	L	L	L	U	L	H	H	n	y

Risk of Bias: Low = L; High = H; Unclear = U
Usability: No = n; Yes = y
All models within the same publication were judged the same on applicability, risk of bias and usability and therefore results are reported per publication.

Table 2: Overview of risk of bias, applicability, usability, and similarity in study design of development and validation studies

Model development studies (N=10 development studies)		
Overall Risk of Bias of development studies		
High	6	60%
Low	2	20%
Unclear	2	20%
Applicability of development studies		
High	3	30%
Low	7	70%
Unclear	0	0%
Usability of models		
Research		
Yes	4	40%
No	6	60%
Clinical practice		
Yes	9	90%
No	1	10%
External validation studies (N=245)		
Similarity in study design between development and validation cohorts		
Similar	147	60%
Cohort to trial	26	11%
Trial to cohort	71	29%
NA	1	
Relatedness		
Related	35	14%
Moderately related	45	18%
Distantly related	164	67%
NA	1	

Risk of bias: Risk of bias was assessed with the original PROBAST (Supplementary Table 3).
Usability: The model was deemed usable in research if the full model equation or sufficient information to extract the baseline risk (intercept) and individual predictor effects was reported, and usable in clinical practice if an alternative presentation of the model was included (e.g. a nomogram, score chart or web calculator).
Relatedness: To judge relatedness we created a relatedness rubric, aiming to capture various levels or relatedness by dividing the validation studies into three categories: 'related', 'moderately related', and 'distantly related' (Supplementary Table 3).

studies were judged 'moderately related', as the models were validated in a different setting (e.g. country) than the model was originally developed in.

The discriminative ability of the models showed substantial variation (Supplementary Table 8; Figure 2). Overall, the AUCs at external validation ranged between 0.47 to 0.94 for the prediction of mortality, and between 0.61 to 1.00 for the prediction of unfavorable outcome.

There was substantial variation in the agreement between observed and predicted probabilities. The reported calibration intercept ranged between -1.27 to 0.93 for mortality, and between –0.51 to 2.39 for the prediction of unfavorable outcome. The reported calibration slopes ranged between 0.72 to 2.3 for mortality and between 0.71 to 2.5 for unfavorable outcome.

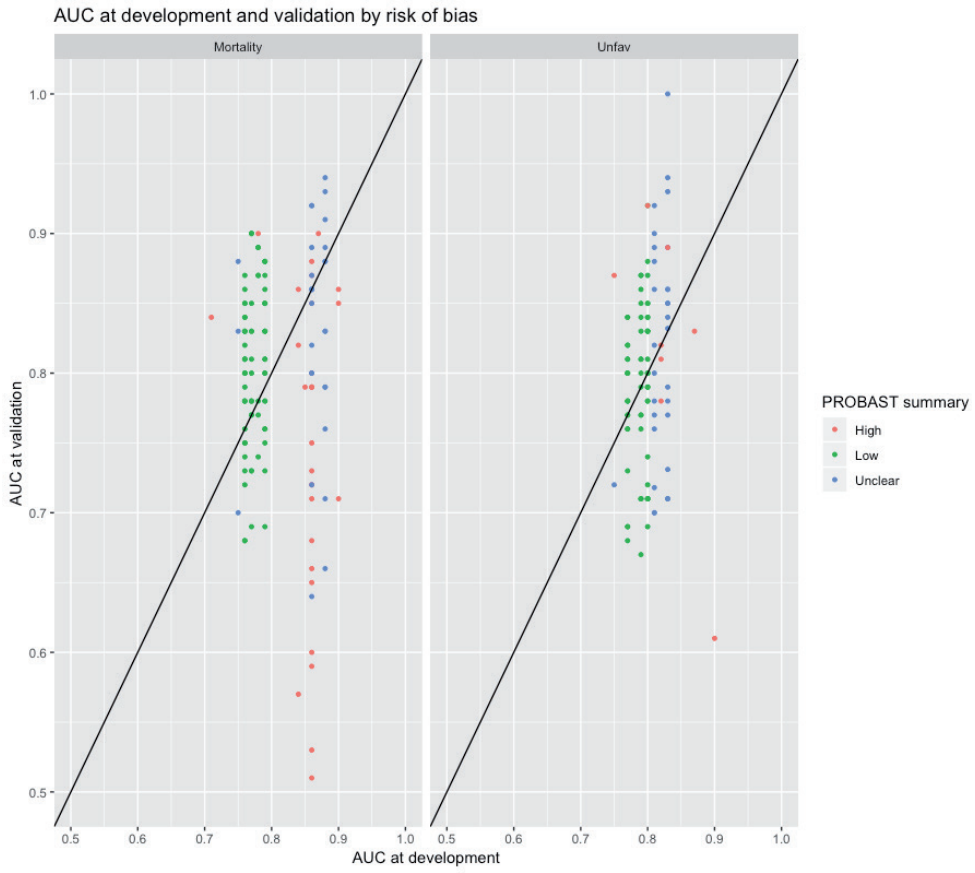


Figure 2: AUC of 18 models at development and in 242 validation studies by Risk of Bias assessed with the PROBAST.

Relation between methodological quality and model performance

The difference between the AUC at development and validation was highly variable (Figure 2). The median change in the discriminative ability in low RoB models was positive (N=139 validation studies, dAUC 8%, [IQR -4% to 21%]) compared to a negative median dAUC in high ROB models (N=45 validation studies, dAUC -18%, [IQR -43% to -2%]) (Table 3).

Table 3: The median AUC at development and external validation and the absolute and percentage change between development AUC and validation AUC stratified by Risk of Bias (RoB) of model development studies based on the original PROBAST.

	N	Median AUC at development (N=12) [IQR]	Median AUC at external validation (N=245) [IQR]	Median delta AUC [IQR]	Median AUC change in percentage [IQR]
Low RoB	139	0.78 [0.77, 0.79]	0.80 [0.76, 0.84]	0.02 [-0.01, 0.06]	8% [-4, 21]
High RoB	45	0.86 [0.84, 0.86]	0.79 [0.69, 0.84]	-0.06 [-0.16, -0.01]	-18% [-43, -2]
Unclear RoB	61	0.83 [0.81, 0.86]	0.83 [0.77, 0.88]	0.00 [-0.06, 0.04]	0.0% [-19, 10]

Using the GEE, we found a larger average negative change in discrimination for high ROB models (-32% (95% CI: -48 to -15) and unclear RoB models (-13% (95% CI: -16 to -10)) compared to that seen in low RoB models (Table 4), while taking into account the correlation between validations of the same model and similarity in study design between the development and validation study. Models that were developed in a cohort and validated in a trial had an estimated change in discrimination of -18%

Table 4: Results of generalized estimated equations (GEE) for the percentage change in AUC between 18 development and 245 validation studies.

	Percentage change in AUC (95% CI)
Intercept	9.5% (5.5, 13.4)
Risk of bias (Low)	
High	-31.7% (-48.2, -15.2)
Unclear	-13.4% (-16.4, -10.3)
Study design (Similar)	
Cohort to trial	-18.5% (-26.2, -10.8)
Trial to cohort	0.19% (-3.7, 4.1)

The generalized estimated equations (GEE) model includes a random intercept on model level (N=21), Risk of Bias assessment (Low, High, Unclear based on the original PROBAST) and similarity in study design between the development and validation study (Similar, Cohort to trial, Trial to cohort) to estimate the percentage change in AUC between the development and validation studies. The intercept indicates the percentage change in AUC for low risk of bias models with a similar study design between the development and validation study.

(95% CI: -26 to -10), whereas models that were developed in a trial and validated in a cohort had an estimated change in AUC of 0.4% (95% CI: -3 to 4), compared to models that were developed and validated in data derived from a similar study design.

Discussion

We examined the relation between methodological quality of prediction model development studies and performance at external validation for prognostic models predicting outcome of patients after moderate or severe traumatic brain injury (TBI). Of the ten included model development studies, two studies (four models) were found to have low risk of bias (RoB) and were applicable for patients after moderate and severe TBI. The other eight publications (fourteen models) showed 'high' or 'unclear' RoB and had limited usability or applicability for patients after moderate and severe TBI. At external validation model performance is typically reduced (20). However, our findings showed that, on average, the change in discriminative ability was positive in validations of 'low' RoB models meaning that the models performed better at external validation. Conversely, the change in discriminative ability was negative for 'high' RoB models, which means that the models performed worse at external validation. Methodological quality of model development studies was associated with discriminative ability at external validation, implying that poor methodological quality results in poorer model performance in new patients and settings. A recent large-scale validation study of a short form based on the PROBAST in the field of cardiovascular disease showed that high RoB was associated with poorer discrimination (10). Our study confirms these findings for prognostic models in the field of TBI.

We critically appraised and assessed methodological quality of model development studies using the PROBAST (7). Since its publication the PROBAST has, for instance, been applied in the field of rehabilitation (21), cardiology (10) and infectious diseases (COVID-19) (22). Consistent with prior studies, the overall judgement on the 20 PROBAST questions was often 'unclear' or 'high' (21-24), due to key details that were not reported (5). These findings emphasize the importance of adherence to reporting guidelines, such as the TRIPOD reporting guideline (25). Additionally, the PROBAST checklist, which includes 20 items on participant selection, study design, predictors, outcome and statistical analysis, can inform investigators on what should be reported in prognostic model studies. A short form based on the PROBAST, consisting of 8/20 items, was recently validated and could distinguish well between high and low RoB (10). In our study, the overall judgment on the short form was consistent with the original PROBAST for almost all studies.

A prior study reported that the majority of prediction studies in high impact journals did not follow methodological recommendations based on reporting statements, checklists and quality assessment tools (26). Similarly, in most model development studies included in our study the statistical analyses were suboptimal due to insufficient sample size, suboptimal handling of missing data, stepwise selection procedures, and lack of or insufficient internal validation procedures, resulting in a high RoB. Consistent with prior studies that have critically appraised model development studies in TBI, internal validation studies of models developed before 2005 were often lacking or inefficient (3, 4). In contrast, models that were developed more recently, between 2005 and 2021, did more often include an internal validation procedure. In recent years, the importance of internal validation has been stressed (27, 28) and internal validation procedures are accessible through free statistical software such as R (29). These developments may have resulted in a higher uptake of these practices.

External validation aims to examine how the model performs in new patients from different settings (30). This may relate to model performance in patients from different regions or countries (geographical validation), or in patients that differ from the derivation cohort on a characteristic (domain validation) (2). External validation, preferably across a range of settings, is required before clinical application of a model can be recommended. Varying levels of relatedness between the development and validation study are expected. We used a relatedness rubric to define the consistency between development and validation studies, using three categories: 'related', 'moderately related' and 'distantly related' (13). Most of the validation studies differed substantially from the model development study in terms of inclusion criteria and/or outcome assessment, and were judged 'distantly related'.

Differences in case-mix (distribution of patient characteristics) might arise from various levels of relatedness between the development and validation study and differences in study design between the development and validation study. Case-mix differences typically affect the observed change in discrimination (31). Differences in case-mix are expected between observational cohort studies and RCTs, with cohort studies being more heterogeneous. We found that similarity in study design between the development and validation study was associated with the observed change in discriminative ability. For instance, models that were developed in a cohort and validated in a trial had worse discriminative ability at external validation, whereas models that were developed in a trial and validated in a cohort had better discriminative ability at external validation, compared to models that were developed and validated in data derived from a similar study design. These findings reflect larger case-mix heterogeneity in cohorts versus trials. Differences in case-mix can be measured through the model based concordance (c) statistic (mbc) (32), which provides insight into the influence

of case-mix heterogeneity on the discriminative ability. In our study, the mbc was reported in only two of the validation studies published after its introduction in 2016 (33, 34).

Prior systematic reviews found that calibration, the agreement between observed and predicted outcomes, is described less often than discrimination (5, 26, 35). Similarly, a number of the external validation studies did not assess model performance in terms of calibration. When reported, calibration was assessed with the Hosmer-Lemeshow goodness-of-fit test (36) or shown graphically with a calibration plot. The Hosmer-Lemeshow statistic has poor power to detect various violations of model assumptions (37). Although broadly used as a measure of calibration in validation studies, this statistic is not recommended for this purpose (38). To be able to compare model performance between validation studies, reporting the calibration intercept and slope is preferred. Dijkland et al. (6) concluded that the calibration of models for moderate and severe TBI was highly variable, reflecting heterogeneity in reliability of predictions, which motivates continuous validation and updating if clinical implementation is pursued.

Strengths and Limitations

The key strength of this study is that a risk of bias assessment (PROBAST) was related to model performance in external validation studies. Although the 'Explanation and Elaboration' form provides extensive instructions for the scoring of PROBAST, many items are open for interpretation and the overall judgement is dependent on decisions that are made throughout the reviewing process. For instance, to determine if there was a reasonable number of outcome events relative to the number of predictors, we used EPP ≥ 10, which is widely adopted in prediction modeling studies as the minimal guideline criterion for binary logistic regression analysis. However, more recently, authors have suggested higher EPP's of at least 20 and criteria that consider the outcome prevalence, overall model performance, and predictor distributions to determine the sample size required (39). In our study, two of the twelve model development papers were assessed by a third independent reviewer (LW) (Cohen's kappa = 0.64). In each case the disagreement between the reviewers were 'no information' versus '(probably) yes', and they did not influence the overall RoB score.

We included 18 prognostic models for functional outcome following moderate and severe TBI that were externally validated at least once. Although the assessment of model performance in new patients and settings is crucial, external validation is often lacking (20). Therefore, we could include only a limited number of models. In our study, we decided to examine the association between methodological quality and performance in terms of discrimination and not calibration for several reasons. First,

calibration is less often described than discrimination. The calibration at external validation using the calibration intercept and slope was reported for only 8 of 18 models. Second, different measures (e.g. Hosmer-Lemeshow goodness-of-fit test, calibration plot, calibration intercept (calibration-in-the-large) and slope) are used to assess calibration, which makes it more difficult to compare calibration between validation studies. These different calibration measures, such as the calibration intercept and slope, are likely to be affected differently by methodological quality of the development study. Third, apart from methodological quality of the development study, calibration is likely influenced by relatedness between the development and validation study. Thus, calibration can be highly variable between external validation studies because of differences in setting and patient characteristics. For instance, it can be strongly influenced by differences in outcome rates between development and validation, beyond what is predicted by the model. Furthermore, consistent with prior studies, there was low variability in the PROBAST overall judgements as well as the relatedness assessment. Because of the limited sample size and low variability additional variables that might have an effect on the observed change in discrimination (e.g. relatedness) were not included in the GEE. Other variables (e.g. usability and applicability) were not included in the GEE as they were not expected to have an effect on the observed change in discrimination. The models with low RoB, the Hukkelhoven model and IMPACT models (9, 15), were externally validated more frequently than the models classified as high RoB. This implies that the number of external validations might be related to methodological quality of the model development study. Apart from low RoB, these models were also presented with sufficient information to be externally validated. Our results are limited in terms of number of models, but confirm findings from a larger study, which showed that most published prediction models are at high RoB and that high RoB is associated with poorer discrimination. A previous study by Venema et al., (2021) included 556 prediction models for cardiovascular disease, with 1147 validations from the Tufts Predictive Analytics and Comparative Effectiveness (PACE) CPM Registry (10). Venema et al., also corrected for other factors that could be related to the difference in model performance between development and external validation, including overlap in authors between development and validation study, sample size at validation, and years between the development and validation study. In our study, we did not assess methodological quality of the validation studies, which could also influence the difference in model performance between the development and validation study. Future research should further explore the association between methodological quality of external validation studies and model performance.

Conclusion

Higher methodological quality of model development studies is associated with better model performance at external validation in the field of TBI. Our findings support

the importance of adherence to methodological principles at model development and following guidelines for reporting of prediction modeling studies.

Ethical Approval and Consent to participate

Not applicable

Consent for publication

Not applicable

Availability of supporting data

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that there is no conflict of interest.

Funding

The work by IRH, AM, HL and ES was supported by the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 602150 (CENTER-TBI). LW is a post-doctoral fellow of the Research Foundation Flanders (FWO).

Authors' contributions

IRH and AM extracted the data, and conducted the RoB assessment with help of LW. IRH took the lead in writing the manuscript. DvK supervised the project. All authors contributed to writing the manuscript, and they all read and approved the final manuscript.

Acknowledgments

The authors wish to thank Dr. Wichor, M. Bramer and colleagues from the Erasmus MC Medical Library for developing and updating the search strategies.

Supplementary material

Supplementary material is available at:
https://diagnprognres.biomedcentral.com/articles/10.1186/s41512-022-00122-0#Sec22

References

1. 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(1):W1-W73.
2. Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating. New York: Springer Science & Business Media; 2008.
3. Perel P, Edwards P, Wentz R, Roberts I. Systematic review of prognostic models in traumatic brain injury. BMC medical informatics and decision making. 2006;6(1):38.
4. Mushkudiani NA, Hukkelhoven CWPM, Hernández AV, Murray GD, Choi SC, Maas AIR, et al. A systematic review finds methodological improvements necessary for prognostic models in determining traumatic brain injury outcomes. J Clin Epidemiol. 2008;61(4):331-43.
5. 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 medical research methodology. 2014;14(1):40.
6. Dijkland SA, Foks KA, Polinder S, Dippel DWJ, Maas A, Lingsma H, et al. Prognosis in moderate and severe traumatic brain injury: a systematic review of contemporary models and validation studies. J Neurotrauma. 2019.
7. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med. 2019;170(1):51-8.
8. Moons KGM, de Groot JAH, 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 medicine. 2014;11(10):e1001744.
9. 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(8):e165.
10. Venema E, Wessler BS, Paulus JK, Salah R, Raman G, Leung LY, et al. Large-scale validation of the Prediction model Risk Of Bias ASsessment Tool (PROBAST) using a short form: high risk of bias models show poorer discrimination. medRxiv. 2021.
11. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Critical care medicine. 1985;13(10):818-29.
12. Le Gall J-R, Lemeshow S, Saulnier F. A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study. Jama. 1993;270(24):2957-63.
13. 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(20):2478-86.
14. Signorini DF, Andrews PJD, Jones PA, Wardlaw JM, Miller JD. Predicting survival using simple clinical variables: a case study in traumatic brain injury. Journal of Neurology, Neurosurgery & Psychiatry. 1999;66(1):20-5.
15. Hukkelhoven CWPM, Steyerberg EW, Habbema JDF, 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. Journal of neurotrauma. 2005;22(10):1025-39.
16. Maas AIR, Hukkelhoven CWPM, 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(6):1173-82.

17. Collaborators MRCCT. Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients. Bmj. 2008;336(7641):425-9.
18. 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. Neurocritical care. 2013;19(1):79-89.
19. Yuan F, Ding J, Chen H, Guo Y, Wang G, Gao W-W, et al. Predicting outcomes after traumatic brain injury: the development and validation of prognostic models based on admission characteristics. Journal of Trauma and Acute Care Surgery. 2012;73(1):137-45.
20. Siontis GCM, Tzoulaki I, Castaldi PJ, Ioannidis JPA. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. Journal of clinical epidemiology. 2015;68(1):25-34.
21. Wingbermühle RW, van Trijffel E, Nelissen PM, Koes B, Verhagen AP. Few promising multivariable prognostic models exist for recovery of people with non-specific neck pain in musculoskeletal primary care: a systematic review. J Physiother. 2018;64(1):16-23.
22. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. bmj. 2020;369.
23. Cooray SD, Wijeyaratne LA, Soldatos G, Allotey J, Boyle JA, Teede HJ. The unrealised potential for predicting pregnancy complications in women with gestational diabetes: a systematic review and critical appraisal. International journal of environmental research and public health. 2020;17(9):3048.
24. Bellou V, Belbasis L, Konstantinidis AK, Tzoulaki I, Evangelou E. Prognostic models for outcome prediction in patients with chronic obstructive pulmonary disease: systematic review and critical appraisal. bmj. 2019;367.
25. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMC med. 2015;13(1):1.
26. Bouwmeester W, Zuithoff NPA, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, et al. Reporting and methods in clinical prediction research: a systematic review. PLoS Med. 2012;9(5):e1001221.
27. Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal–external, and external validation. J Clin Epidemiol. 2016;69:245-7.
28. Steyerberg EW, Harrell Jr FE, Borsboom GJJM, Eijkemans MJC, Vergouwe Y, Habbema JDF. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol. 2001;54(8):774-81.
29. Harrell FE J. rms: Regression Modeling Strategies. R package version 3.4-0. 2012.
30. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. Ann Intern Med. 1999;130(6):515-24.
31. Vergouwe Y, Moons KGM, Steyerberg EW. External validity of risk models: use of benchmark values to disentangle a case-mix effect from incorrect coefficients. American journal of epidemiology. 2010;172(8):971-80.
32. van Klaveren D, Gönen M, Steyerberg EW, Vergouwe Y. A new concordance measure for risk prediction models in external validation settings. Statistics in medicine. 2016;35(23):4136-52.
33. Dijkland SA, Helmrich IRAR, Nieboer D, van der Jagt M, Dippel DWJ, Menon DK, et al. Outcome prediction after moderate and severe traumatic brain injury: external validation of two established prognostic models in 1742 European patients. Journal of Neurotrauma. 2020.
34. Mikolić A, Polinder S, Steyerberg EW, Retel Helmrich IRA, Giacino JT, Maas AIR, et al. Prediction of Global Functional Outcome and Post-Concussive Symptoms after Mild Traumatic Brain

Injury: External Validation of Prognostic Models in the Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) Study. Journal of neurotrauma. 2021;38(2):196-209.
35. Wessler BS, Paulus J, Lundquist CM, Ajlan M, Natto Z, Janes WA, et al. Tufts PACE clinical predictive model registry: update 1990 through 2015. Diagnostic and prognostic research. 2017;1(1):1-8.
36. Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied logistic regression: John Wiley & Sons; 2013.
37. Hosmer DW, Hosmer T, Le Cessie S, Lemeshow S. A comparison of goodness-of-fit tests for the logistic regression model. Statistics in medicine. 1997;16(9):965-80.
38. Vergouwe Y, Steyerberg EW, Eijkemans MJC, Habbema JDF. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. J Clin Epidemiol. 2005;58(5):475-83.
39. van Smeden M, Moons KGM, de Groot JAH, Collins GS, Altman DG, Eijkemans MJC, et al. Sample size for binary logistic prediction models: beyond events per variable criteria. Statistical methods in medical research. 2019;28(8):2455-74.

Part II

Applications

6

Chapter 6

External validation of the IMPACT and CRASH prognostic models in the CENTER-TBI study

S.A. Dijkland, I.R.A. Retel Helmrich, D. Nieboer, M. van der Jagt, D.W. Dippel, D.K. Menon, N. Stocchetti, A.I. Maas, H.F. Lingsma, E.W. Steyerberg, and the CENTER-TBI Participants and Investigators

Published in: Journal of Neurotrauma

doi.org/10.1089/neu.2020.7300

Abstract

Objective: The International Mission on Prognosis and Analysis of Clinical trials in Traumatic brain injury (IMPACT) and Corticoid Randomisation After Significant Head injury (CRASH) prognostic models predict functional outcome after moderate and severe traumatic brain injury (TBI). We aimed to assess their performance in a contemporary cohort of patients across Europe.

Design: External validation study.

Setting: The Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) core study is a prospective observational cohort study in patients presenting with TBI and an indication for brain computed tomography. The CENTER-TBI core cohort consists of 4509 TBI patients available for analyses from 59 centers in 18 countries across Europe and Israel.

Patients: The IMPACT validation cohort included 1173 patients with GCS≤12, age≥14 and 6-month Glasgow Outcome Scale Extended (GOSE) available. The CRASH validation cohort contained 1742 patients with GCS≤14, age≥16 and 14-day mortality or 6-month GOSE available.

Interventions: None.

Measurements and Main Results: Performance of the three IMPACT and two CRASH model variants was assessed with discrimination (area under the receiver operating characteristic curve; AUC) and calibration (comparison of observed versus predicted outcome rates). For IMPACT, model discrimination was good, with AUCs ranging between 0.77-0.85 in 1173 patients and between 0.80-0.88 in the broader CRASH selection (n=1742). For CRASH, AUCs ranged between 0.82-0.88 in 1742 patients and between 0.66-0.80 in the stricter IMPACT selection (n=1173). Calibration of the IMPACT and CRASH models was generally moderate, with calibration-in-the-large and calibration slopes ranging between -2.02-0.61 and between 0.48-1.39, respectively.

Conclusions: The IMPACT and CRASH models adequately identify patients at high risk for mortality or unfavorable outcome, which supports their use in research settings and for benchmarking in the context of quality of care assessment.

Introduction

Traumatic brain injury (TBI) is a heterogeneous disease with substantial variation in trauma mechanisms, pathophysiology and clinical presentation (1). Early outcome prediction is important in research settings, e.g. for selecting patients for clinical trials (2). Informed predictions could also facilitate risk communication with patients or relatives and case-mix adjustment for benchmarking quality of care (3). Many prognostic models for functional outcome after moderate and severe TBI have been developed and validated (4-6). Of these, the International Mission on Prognosis and Analysis of Clinical trials in Traumatic brain injury (IMPACT) models and the Corticoid Randomisation After Significant Head injury (CRASH) models are the most widely known (7, 8). These models were developed a decade ago on large multicenter cohorts using state-of-the-art statistical methodology. The models combine clinical, radiological and laboratory admission characteristics to predict risk of mortality and unfavorable outcome. The IMPACT and CRASH models have shown highly variable model performance across different settings (6). Moreover, previous validation studies were mostly performed in small observational cohorts or randomized clinical trials (RCTs) that may not represent the current TBI population. We aimed to gain insight in the performance of the IMPACT and CRASH prognostic models in contemporary patients across Europe.

Materials and Methods

IMPACT and CRASH models

Details of the development of the IMPACT and CRASH prognostic models have been reported (7, 8). In short, the IMPACT models were developed on 8,509 patients with moderate or severe TBI (Glasgow Coma Scale [GCS] ≤12) from eight RCTs and three observational studies (8). The IMPACT models comprise three variants (core, extended and lab) with increasing complexity (Table 1). The models predict mortality and functional outcome at 6 months post-injury.

The two versions of the CRASH prognostic model (basic and computed tomography [CT]) (Table 1) were developed on 10,008 TBI patients with GCS ≤14 from one RCT (7). The models predict mortality at 14 days and functional outcome at 6 months post-injury.

Study design and population

We used data from the Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) core study, a prospective observational

Table 1: Variables included in the International Mission on Prognosis and Analysis of Clinical Trials (IMPACT) and Corticoid Randomisation After Significant Head injury (CRASH) prognostic models.

IMPACT core	IMPACT extended	IMPACT lab	CRASH basic	CRASH CT
Age GCS motor score Pupillary reactivity	Core model predictors + Hypoxia Hypotension Marshall CT classification tSAH EDH	Extended model predictors + Glucose Hemoglobin	Age GCS total score Pupillary reactivity Major extracranial injury	Basic model predictors + Petechial hemorrhages Obliteration of 3rd ventricle or basal cisterns tSAH Midline shift > 5 mm Non-evacuated hematoma

GCS, Glasgow Coma Scale; CT, computed tomography; tSAH, traumatic subarachnoid hemorrhage; EDH, epidural hematoma

cohort study in patients with TBI presenting within 24 hours of injury and with an indication for brain CT. Participants were recruited from December 2014 through December 2017 from 59 centers in 18 countries across Europe and Israel. The study protocol of CENTER-TBI has been described (9). Informed consent by patients and/or legal representative/next of kin was obtained, according to local legislations, for all patients recruited in the CENTER-TBI core dataset and documented in the electronic case report form (e-CRF). Ethical approval was obtained for each recruiting site. The sites, Ethical Committees, approval numbers and approval dates are listed on the website: https://www.center-tbi.eu/project/ethical-approval.

Because the IMPACT and CRASH models were developed on different selections of TBI patients, the models were validated on separate cohorts with inclusion criteria corresponding to the development cohorts. For the IMPACT core model, we included patients aged≥14 years with admission GCS≤12 and available functional outcome. The validation cohort for the CRASH basic model included patients aged≥16 years with admission GCS≤14 and available functional outcome. For validation of the IMPACT and CRASH models that included admission CT and laboratory characteristics, patients without CT scan or blood samples in the first 24 hours after injury were excluded. To directly compare performance of the IMPACT and CRASH models, we additionally validated the IMPACT models in the CRASH validation cohort and vice versa.

In CENTER-TBI, functional outcome at 6 months post-injury was assessed with the Glasgow Outcome Scale Extended (GOSE). In line with the original IMPACT and CRASH models, we dichotomized the 6-month GOSE into mortality (GOSE 1)

versus survival (GOSE 2-8), and unfavorable (GOSE 1-4) versus favorable (GOSE 5-8) outcome. For the CRASH models, mortality was assessed at 14 days post-injury.

Predictor effects

Definitions and coding of the predictors in the validation cohorts were similar to those in the IMPACT and CRASH development cohorts (Supplemental Digital Content 1, 2, 3) (7, 8). Major extracranial injury was defined as a score of ≥3 on at least one of the extracranial domains of the Abbreviated Injury Scale (10).

The IMPACT and CRASH logistic regression models were refitted in the validation data to enable comparison of predictor effects between development and validation cohorts. Associations between predictors and outcomes were expressed as odds ratios (ORs) with 95% confidence intervals (CIs).

Validation

The IMPACT and CRASH models were validated by applying the coefficients of the original models to the validation data (Supplemental Digital Content 2, 3). Because participating centers in CENTER-TBI were mainly situated in western countries, we used the CRASH models for high-income countries (7). Model performance was assessed with discrimination and calibration. Discrimination was expressed with the area under the receiver operating characteristic curve (AUC). The AUC ranges from 0.5 for a non-discriminative model to 1.0 for a perfect model (11). Calibration indicates the agreement between predicted and observed outcome probabilities. It was assessed graphically by plotting observed frequencies of mortality and unfavorable outcome versus predicted risk. Additionally, we calculated the calibration slope and calibration-in-the-large. The calibration slope is ideally equal to 1 and represents the overall predictor effects in the validation cohort versus the development cohort. Calibration-in-the-large indicates whether predictions are systematically too high or too low, and should ideally be zero (12).

Model discrimination at external validation may be affected by the distribution of patient characteristics (case-mix) in the validation cohort (13, 14). Distinguishing patients with good versus poor outcome is more difficult in a homogeneous cohort than in a heterogeneous population leading to higher AUCs in heterogeneous cohorts. We therefore calculated the case-mix-corrected AUC, which reflects model discrimination under the assumption that the regression coefficients are correct for the validation population (13).

Statistical analyses were performed with R software, version 3.4.3 (R foundation for statistical computing, Vienna, Austria). Calibration plots were created with an updated

version of the *val.prob* function (*rms* library in R) (15). Missing 6-month GOSE as a consequence of loss-to-follow-up (in patients with at least one GOSE observation at another time point) were imputed with a Bayesian mixed effect model (Supplemental Digital Content 3). Patients without any GOSE observation were excluded from the analyses. Derived variables for GCS (motor) score and pupillary reactivity were generated based on methodology as used in the IMPACT database (Supplemental Digital Content 3) (16). The remaining missing predictor values were statistically imputed with multiple imputation based on the predictors and outcomes included in the IMPACT and CRASH models (*mice* package in R). CENTER-TBI data was collected through the Quesgen e-CRF (Quesgen Systems Inc, USA), hosted on the International Neuroinformatics Coordinating Facility (INCF) platform and extracted via the INCF Neurobot tool (INCF, Sweden). Version Core 1.1 of the CENTER-TBI dataset was used in this study.

Results

Study population

In total, 4509 patients included in the CENTER-TBI core study could be analyzed. Of those, 1173 and 1742 patients met the inclusion criteria for the IMPACT and CRASH validation cohort, respectively (Supplemental Digital Content 4). Missing predictor values for the IMPACT (5%) and CRASH (4%) cohorts were imputed (Supplemental Digital Content 5).

The IMPACT validation cohort consisted mainly of severe TBI patients (72%). At 6 months, 347 patients had died (30%), and 644 patients (55%) had unfavorable outcomes (Table 2). In the CRASH validation cohort, one-third of the patients had an admission GCS of 13-14. At 14 days, 266 patients had died (15%), and at 6 months 751 patients (43%) had unfavorable outcomes (Table 2).

Compared to the IMPACT and CRASH development cohorts, patients in the CENTER-TBI validation cohorts were on average 20 years older and had more severe TBI (Table 2). More patients had major extracranial injury in the CRASH validation cohort (49%) than the development cohort (22%). Traumatic subarachnoid hemorrhage occurred almost twice as often in the CENTER-TBI validation cohorts versus the IMPACT and CRASH development cohorts. Overall, functional outcomes at 6 months were poorer in CENTER-TBI, with a higher proportion of unfavorable outcomes in both validation cohorts compared to the development cohorts (Table 2).

Table 2: Characteristics of patients in the International Mission on Prognosis and Analysis of Clinical Trials (IMPACT) and Corticoid Randomisation After Significant Head injury (CRASH) development cohorts and the IMPACT and CRASH validation cohorts in the Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) core study.

Admission characteristics	Measure or category	IMPACT development cohort (n=8509)	CENTER-TBI IMPACT validation cohort (n=1173)	CRASH development cohort (n=10008)	CENTER-TBI CRASH validation cohort (n=1742)
Age, years	Median (IQR)	30 (21-45)	49 (29-66)	33 (23-47)	51 (32-67)
GCS motor score	None (1)	1395 (16%)	527 (45%)	-	-
	Extension (2)	1042 (12%)	66 (6%)	-	-
	Abnormal flexion (3)	1085 (13%)	67 (6%)	-	-
	Normal flexion (4)	1940 (23%)	118 (10%)	-	-
	Localizes/obeys (5/6)	2591 (30%)	395 (34%)	-	-
	Untestable/ missing (9)	456 (5%)	0 (0%)	-	-
GCS total score (3-14)	Mild (13-14)	-	-	3019 (30%)	582 (33%)
	Moderate (9-12)	-	324 (28%)	3041 (30%)	316 (18%)
	Severe (3-8)	-	849 (72%)	3948 (40%)	844 (48%)
Pupillary reactivity	Both pupils reacted	4486 (53%)	817 (71%)	8057 (81%)	1338 (77)
	One pupil reacted	886 (10%)	99 (8%)	588 (6%)	111 (6%)
	No pupil reacted	1754 (21%)	216 (18%)	825 (8%)	228 (13%)
Major extracranial injury	Yes	-	-	2216 (22%)	845 (49%)
Hypoxia	Yes or suspected	1116 (13%)	198 (17%)	-	-
Hypotension	Yes or suspected	1171 (14%)	187 (16%)	-	-
Marshall CT classification	I	360 (4%)	66 (6%)	-	-
	II	1838 (22%)	413 (35%)	-	-
	III/IV	1050 (12%)	124 (11%)	-	-
	V/VI	1944 (23%)	377 (32%)	-	-
Traumatic subarachnoid hemorrhage	Yes	3313 (39%)	764 (65%)	2458 (25%)	1009 (58%)

Table 2: Continued

Admission characteristics	Measure or category	IMPACT development cohort (n=8509)	CENTER-TBI IMPACT validation cohort (n=1173)	CRASH development cohort (n=10008)	CENTER-TBI CRASH validation cohort (n=1742)
Epidural hematoma	Yes	999 (12%)	170 (14%)	-	-
≥1 petechial hemorrhages	Yes	-	-	2238 (22%)	215 (12%)
Obliteration of 3rd ventricle or basal cisterns	Yes	-	-	1827 (18%)	474 (27%)
Midline shift > 5mm	Yes	-	-	1136 (11%)	347 (20%)
Non-evacuated hematoma	Yes	-	-	2111 (21%)	480 (28%)
Glucose (mmol/l)	Median (IQR)	8.2 (6.7-10.4)	7.8 (6.5-9.6)	-	-
Hemoglobin (g/dL)	Median (IQR)	12.7 (10.8-14.3)	13.0 (11.3-14.2)	-	-
Mortality at 14 days	Yes	-	-	1948 (19%)	266 (15%)
Outcome at 6 months	Dead	2396 (28%)	347 (30%)	2323 (23%)	394 (23%)
	Vegetative[a]	351 (4%)	0 (0%)	272 (3%)	0 (0%)
	Lower severe disability	-	243 (21%)	-	291 (17%)
	Upper severe disability	1335 (16%)	54 (5%)	962 (10%)	66 (4%)
	Lower moderate disability	-	91 (8%)	-	138 (8%)
	Upper moderate disability	1666 (20%)	148 (13%)	1664 (17%)	212 (12%)
	Lower good recovery	-	147 (13%)	-	267 (15%)
	Upper good recovery	2761 (32%)	143 (12%)	4333 (43%)	374 (22%)
	Death or severe disability	4082 (48%)	644 (55%)	3557 (36%)	751 (43%)

IMPACT, International Mission on Prognosis and Analysis of Clinical Trials; CRASH, Corticoid Randomisation After Significant Head injury; CENTER-TBI, Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury; IQR, interquartile range; GCS, Glasgow Coma Scale; CT, computed tomography.

[a]Vegetative state and lower severe disability combined (GOSE categories 2 and 3)

IMPACT models

In CENTER-TBI, associations of the predictors in the IMPACT models with 6-month outcome were similar to those reported for the IMPACT development cohort (Supplemental Digital Content 6). However, presence of hypoxia and traumatic subarachnoid hemorrhage did not significantly increase risk of poor outcome in the CENTER-TBI cohort.

The IMPACT models distinguished well between patients who died and patients who were alive, indicated by AUCs>0.80 (Table 3). Addition of CT variables to the core model for mortality increased discriminative ability (core model: 0.81, 95% CI 0.79-0.84; extended model: AUC 0.85, 95% CI 0.82-0.87). The IMPACT lab model for mortality also had an AUC of 0.85 (95% CI 0.82-0.87, Table 3). The IMPACT models had slightly lower discriminative ability for unfavorable outcome (core 0.77 (95% CI 0.74 to 0.80); extended 0.80 (95% CI 0.78-0.83); lab 0.81 (95% CI 0.78-0.84), Table 3).

Calibration showed that observed mortality risk was lower than predicted (calibration-in-the-large: core -0.74 (95% CI -0.88 to -0.60); extended -0.73 (95% CI -0.88 to -0.57); lab -0.37 (95% CI -0. 52 to -0.21)) and the IMPACT models slightly overestimated (core and extended) or underestimated (lab) risks for unfavorable outcome (calibration-in-the-large: core -0.10 (95% CI -0.23 to 0.04); extended -0.03 (95% CI -0.17 to 0.11); lab 0.12 (95% CI -0.03 to 0.27)) (Table 3, Figure 1). The calibration slopes ranged between 1.20-1.32 for the models for mortality and 0.97-1.02 for the models for unfavorable outcome (Table 3, Figure 1), reflecting stronger (mortality) or similar (unfavorable outcome) predictor effects in CENTER-TBI versus the IMPACT development cohort.

We observed higher AUCs for the IMPACT models for mortality in the validation cohort compared to the development cohort (e.g. for the lab model: AUC 0.85 versus 0.79, respectively, Table 3). When calculating the case-mix-corrected AUC, these differences in discriminative ability disappeared (Table 3). For the models for unfavorable outcome, the AUC at external validation and the case-mix-corrected AUC were similar, indicating comparable case-mix.

CRASH models

Associations between some predictors and outcomes varied between the CENTER-TBI validation cohort versus the CRASH development cohort. For instance, presence of major extracranial injury did not significantly increase mortality risk in CENTER-TBI, and the effect of midline shift was non-significant (Supplemental Digital Content 7).

Table 3: Performance of the International Mission on Prognosis and Analysis of Clinical Trials (IMPACT) and Corticoid Randomisation After Significant Head injury (CRASH) models in the Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) core study.

Mortality					
Performance measure	**IMPACT core (n=1173)**	**IMPACT extended (n=1030)**	**IMPACT lab (n=1006)**	**CRASH basic (n=1742)**	**CRASH CT (n=1542)**
AUC – development (internal validation)	0.77	0.81	0.79	0.86	0.88
AUC – external validation	0.81 (0.79 to 0.84)	0.85 (0.82 to 0.87)	0.85 (0.82 to 0.87)	0.86 (0.83 to 0.88)	0.88 (0.86 to 0.90)
Calibration slope	1.20 (1.04 to 1.36)	1.23 (1.06 to 1.39)	1.32 (1.14 to 1.50)	0.95 (0.84 to 1.06)	0.75 (0.66 to 0.84)
Calibration-in-the-large	-0.74 (-0.88 to -0.60)	-0.73 (-0.88 to -0.57)	-0.37 (-0. 52 to -0.21)	-0.01 (-0.16 to 0.15)	-2.02 (-2.21 to -1.83)
Observed vs. predicted	30% vs. 43%	29% vs. 41%	29% vs. 35%	15% vs. 15%	15% vs. 33%
AUC – case-mix-corrected	0.77 (0.75 to 0.80)	0.80 (0.76 to 0.82)	0.79 (0.77 to 0.83)	0.86 (0.84 to 0.88)	0.91 (0.87 to 0.91)
Unfavorable outcome					
Performance measure	**IMPACT core (n=1173)**	**IMPACT extended (n=1030)**	**IMPACT lab (n=1006)**	**CRASH basic (n=1742)**	**CRASH CT (n=1542)**
AUC – development (internal validation)	0.78	0.81	0.81	0.81	0.83
AUC – external validation	0.77 (0.74 to 0.80)	0.80 (0.78 to 0.83)	0.81 (0.78 to 0.84)	0.82 (0.80 to 0.84)	0.84 (0.82 to 0.86)
Calibration slope	0.97 (0.84 to 1.10)	1.01 (0.87 to 1.15)	1.02 (0.89 to 1.16)	0.97 (0.88 to 1.07)	0.85 (0.76 to 0.93)
Calibration-in-the-large	-0.10 (-0.23 to 0.04)	-0.03 (-0.17 to 0.11)	0.12 (-0.03 to 0.27)	-0.02 (-0.14 to 0.09)	-0.93 (-1.06 to -0.79)
Observed vs. predicted	55% vs. 57%	54% vs. 55%	54% vs. 52%	43% vs. 43%	43% vs. 56%
AUC – case-mix-corrected	0.78 (0.74 to 0.79)	0.80 (0.79 to 0.84)	0.81 (0.78 to 0.84)	0.83 (0.81 to 0.85)	0.86 (0.84 to 0.88)

IMPACT, International Mission on Prognosis and Analysis of Clinical Trials; CRASH, Corticoid Randomisation After Significant Head injury; CT, computed tomography; AUC, area under the receiver operating characteristic curve.
All performance values for external validation are reported with a 95% confidence interval.

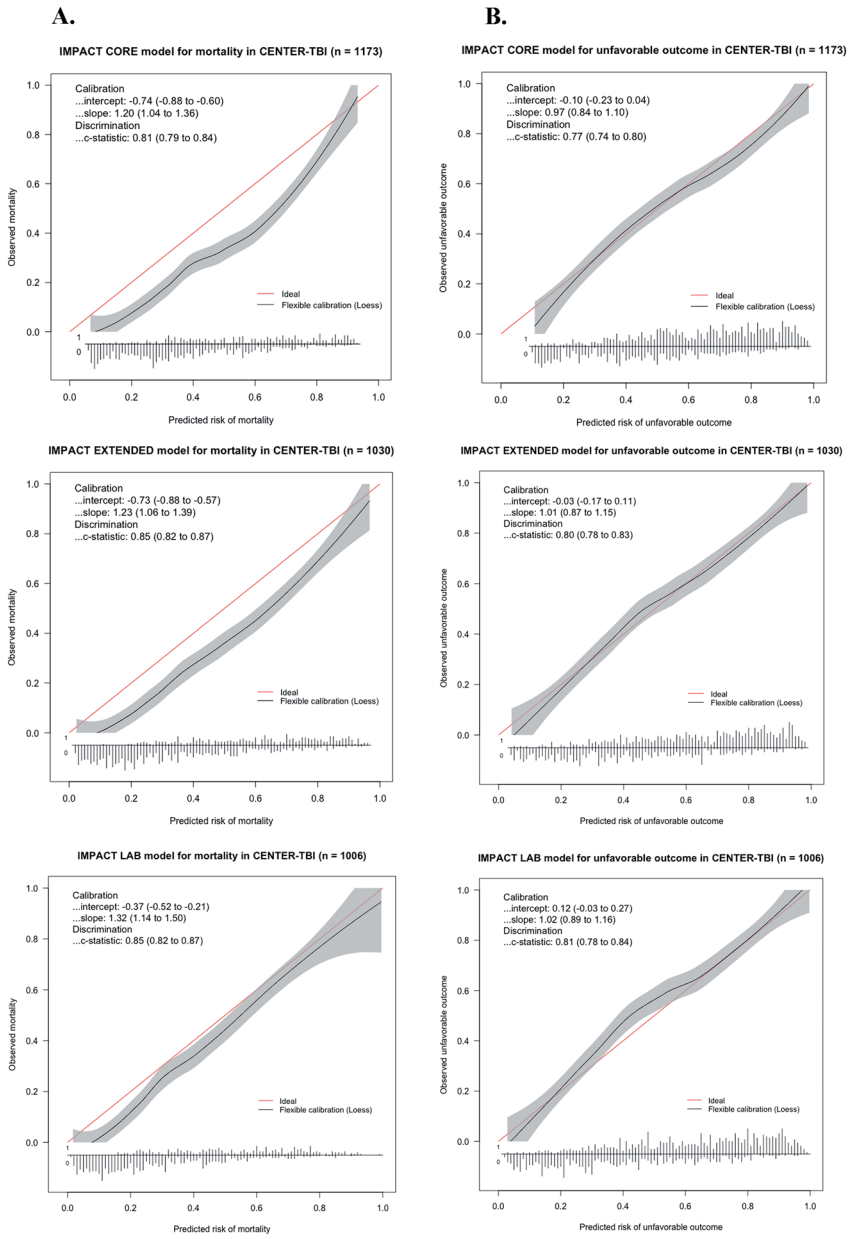


Figure 1: Calibration plots of the International Mission on Prognosis and Analysis of Clinical Trials (IMPACT) models for (**A**) mortality and (**B**) unfavorable outcome at 6 months. Predicted probabilities are on the x-axis and observed outcomes on the y-axis. The distribution of the predicted probabilities is shown at the bottom of the graphs, separate for those with (= 1) and without (= 0) the outcome of interest. The 45-degree line with intercept 0 and slope 1 represents perfect agreement between predicted and observed outcome rates. Deviation above or below this line indicates that the model underestimates or overestimates mortality or unfavorable outcome rates, respectively. For instance, the calibration plots in panel A show that all three IMPACT models tend to overestimate mortality rates in the CENTER-TBI validation cohort. CENTER-TBI, Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury.

Discriminative ability of the CRASH models was good for both mortality and unfavorable outcome (Table 3). We observed comparable AUCs for the CT model (mortality: 0.88, 95% CI 0.86-0.90; unfavorable outcome: 0.84, 95% CI 0.82-0.86) versus the basic model (mortality: 0.86, 95% CI 0.83-0.88; unfavorable outcome: 0.82, 95% CI 0.80-0.84).

Assessment of model calibration revealed differences between observed and predicted risk of mortality and unfavorable outcome for the CRASH CT model (Supplemental Digital Content 7, Figure 2). The CRASH basic model adequately predicted mortality and unfavorable outcome (calibration-in-the-large -0.01, 95% CI -0.16-0.15 and -0.02, 95% CI -0.14-0.09, respectively), whereas the CT model strongly overestimated risk of mortality and unfavorable outcome (calibration-in-the-large, mortality: -2.02, 95% CI -2.21 to -1.83; unfavorable outcome: -0.93, 95% CI -1.06 to -0.79). The moderate calibration slopes for the CRASH CT model (mortality: 0.75, 95% CI 0.66-0.84; unfavorable outcome: 0.85, 95% CI 0.76-0.93) reflect the smaller predictor effects in CENTER-TBI compared to the CRASH development cohort (Table 3, Figure 2).

Discriminative ability was similar in the validation versus development cohort, although the validation cohort had a somewhat more homogeneous case-mix (Table 3).

Comparison IMPACT and CRASH

When validating the IMPACT models in the broader CRASH selection in CENTER-TBI (n=1742), performance of the IMPACT and CRASH models for mortality and unfavorable outcome was similar (Supplemental Digital Content 8, 9).

Validation of the CRASH models in the stricter IMPACT selection within CENTER-TBI (n=1173) yielded lower AUCs and larger discrepancies between observed and predicted rates of mortality and unfavorable outcome for the CRASH models compared to the IMPACT models (Supplemental Digital Content 8, 10).

Discussion

We performed detailed evaluations of the external validity of the IMPACT and CRASH prognostic models in a large contemporary European cohort of TBI patients. Both sets of models showed good discriminative ability, which modestly improved with addition of CT variables to the IMPACT core and CRASH basic models. There were substantial differences between observed and predicted outcome risk, specifically for the CRASH CT model.

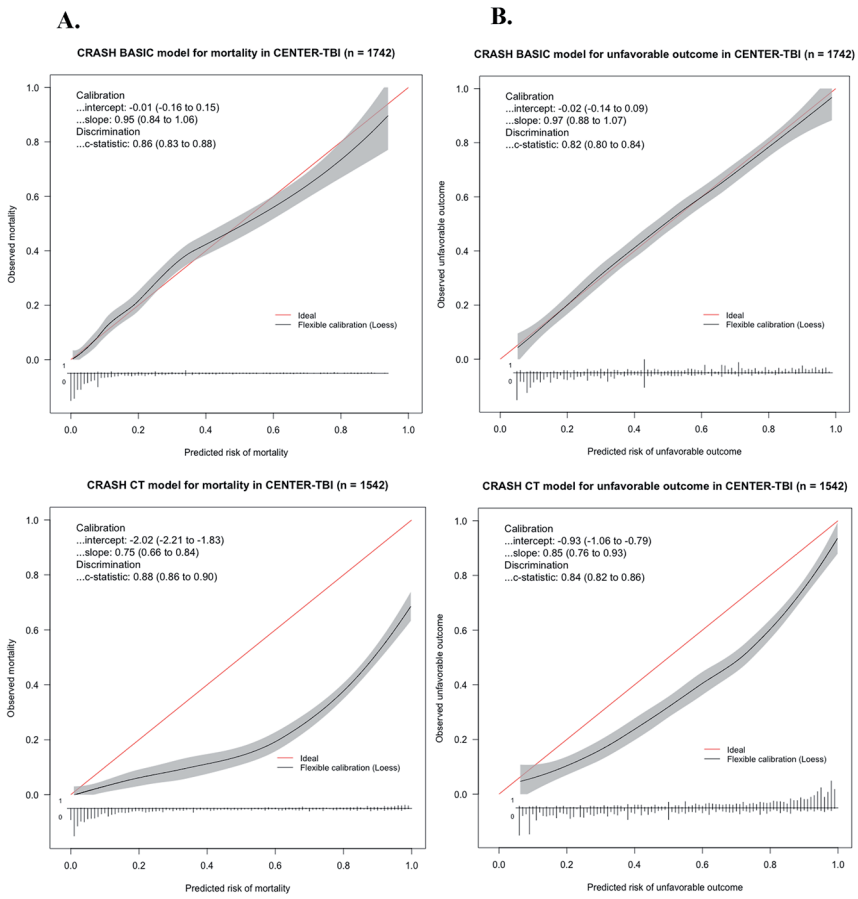


Figure 2: Calibration plots of the Corticoid Randomisation After Significant Head injury (CRASH) models for (**A**) mortality at 14 days and (**B**) unfavorable outcome at 6 months. Predicted probabilities are on the x-axis and observed outcomes on the y-axis. The distribution of the predicted probabilities is shown at the bottom of the graphs, separate for those with (= 1) and without (= 0) the outcome of interest. The 45-degree line with intercept 0 and slope 1 represents perfect agreement between predicted and observed outcome rates. Deviation above or below this line indicates that the model underestimates or overestimates mortality or unfavorable outcome rates, respectively. For instance, the CRASH CT model overestimates mortality and unfavorable outcome rates in the CENTER-TBI validation cohort.

Over the past decade, the IMPACT and CRASH models have been externally validated in many different, but mostly small, selected or single-country cohorts. A recent systematic review on prognostic models in moderate and severe TBI showed that discriminative ability of the IMPACT and CRASH models at external validation was moderate to good across different settings (mean AUCs weighted for sample size 0.77-0.82 over 91 validations) (6). Calibration was however highly variable and substantial miscalibration was observed in subgroups of TBI patients (e.g. patients who

underwent decompressive craniectomy). Compared to previous external validation studies, the IMPACT and CRASH models performed generally well in the CENTER-TBI validation cohort, indicating that the models stood the test of time (6). Overall, observed mortality was lower than predicted and observed unfavorable outcome was similar as predicted, which may indicate that survival has improved over time but more patients survive with (severe) disabilities.

Our validation cohort was part of a large and unique multicenter observational study with data from contemporary TBI patients throughout Europe (17). We could validate the original IMPACT and CRASH models due to availability of all included predictors and outcomes. However, discrepancies might still exist in the assessment method and definitions of predictors and outcomes. For example, imaging techniques may have improved or changed over time (14). Another limitation of our study is that the CRASH models for low-middle income countries could not be validated because mainly high-income countries participated in CENTER-TBI.

Model performance at external validation is sensitive to several study characteristics (14). Differences in case-mix in the validation cohorts compared to the development cohorts influenced the discriminative ability of the IMPACT and CRASH models. The CENTER-TBI validation cohort generally consisted of older and more severely affected TBI patients, and was more heterogeneous than the IMPACT database which predominantly included RCTs (8, 16). The CENTER-TBI cohort was somewhat more homogeneous than the CRASH trial, which fits with the relatively broad inclusion criteria in that trial (18). We observed substantial miscalibration for the IMPACT and CRASH models in CENTER-TBI. This could be explained by differences in prevalence and effects of predictors between the derivation and validation cohorts. Major extracranial injury, traumatic subarachnoid hemorrhage and midline shift were more prevalent in CENTER-TBI than in the CRASH development cohort, while mortality at 14 days was similar (Table 2). Presence of midline shift was not associated with mortality and unfavorable outcome in CENTER-TBI (Supplemental Digital Content 2, 7). This may explain the substantial overestimation of mortality and unfavorable outcome by the CRASH CT model (13).

Overall, discriminative ability of the IMPACT and CRASH models only marginally improved with increasing model complexity. This observation confirms that the core clinical predictors (age, GCS (motor) score and pupillary reactivity) are essential for adequate identification of TBI patients at high risk of mortality or unfavorable outcome and that additional predictors add relatively little prognostic information. Calibration of the IMPACT core models was similar or inferior compared to the more complex models (Table 3). This underscores the need for model updating (e.g. refitting

the model intercept or refitting the coefficients) to adjust models to specific clinical settings (11, 19). Extension of the IMPACT and CRASH models with new predictors has been attempted previously but did not yield substantial improvement in model performance (6). In CENTER-TBI, updating the IMPACT (and CRASH) models may be pursued (20, 21). For instance, performance of the IMPACT extended model may be improved by replacing the Marshall CT classification with a more recent CT score (e.g. Rotterdam or Helsinki) or a combination of individual CT characteristics (22, 23). Also, the models could be enriched with promising biomarkers or dynamic characteristics obtained during the clinical course (24).

Continuous external validation of prognostic models for moderate and severe TBI in recent cohorts has been recommended (6, 24, 25). The IMPACT and CRASH models were developed on relatively historic data, while the epidemiology of TBI has changed substantially over the last years, e.g. regarding age distribution (1). This study adds to the existing evidence by showing that the IMPACT and CRASH models are valid for outcome prediction in contemporary TBI patients across Europe. Nevertheless, discrepancies between observed and predicted rates of mortality and unfavorable outcome exist for both sets of models. Adjustment of the models to local hospital and patient characteristics is therefore strongly recommended.

Performance of the IMPACT and CRASH models in the broadest selection of TBI patients was comparable. The additional effect of major extracranial injury in CRASH seems limited, probably because patients in CENTER-TBI were selected based on TBI and not any trauma (10). The decision on which model to use should mainly be guided by the characteristics of a specific setting or population (e.g. TBI severity, country economic status). Use of either the IMPACT or CRASH model and degree of complexity of the model also depends on availability of predictors. Given the substantial uncertainty on likely outcomes in individual patients, the IMPACT and CRASH models are not recommended for clinical decision making. Treatment options for TBI patients are scarce and documenting prognosis in the intensive care setting does not seem to substantially affect treatment decisions (26-28). On the other hand, there is an increasing recognition that estimates of prognosis by clinicians are often unduly pessimistic for TBI patients (29), and regular comparison of outcome predicted by these models with clinical expectations may help individual clinicians calibrate their prognostication and practice. Based on the good discriminative ability of the IMPACT and CRASH models, potential applications in research settings are risk stratification in trials and covariate adjustment in statistical analyses to increase statistical power. The models may also provide a point of reference for quality of care by comparing observed versus expected outcomes (3).

Conclusions

The IMPACT and CRASH models adequately identify patients at high risk for mortality or unfavorable outcome, which supports their use in research settings and for benchmarking in the context of quality of care assessment.

Acknowledgments

We are grateful to all patients that participated in the CENTER-TBI study to help us in our efforts to improve care and outcome for TBI.

Supplementary material

Supplementary material is available at: https://www.liebertpub.com/doi/abs/10.1089/neu.2020.7300

References

1. Maas AIR, Menon DK, Adelson PD, et al. Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. *Lancet Neurol* 2017;16(12):987-1048.
2. Thompson DD, Lingsma HF, Whiteley WN, et al. Covariate adjustment had similar benefits in small and large randomized controlled trials. *J Clin Epidemiol* 2015;68(9):1068-1075.
3. Huijben JA, Wiegers EJA, de Keizer NF, 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(1):95.
4. Mushkudiani NA, Hukkelhoven CW, Hernandez AV, et al. A systematic review finds methodological improvements necessary for prognostic models in determining traumatic brain injury outcomes. *J Clin Epidemiol* 2008;61(4):331-343.
5. Perel P, Edwards P, Wentz R, et al. Systematic review of prognostic models in traumatic brain injury. *BMC Med Inform Decis Mak* 2006;6:38.
6. Dijkland SA, Foks KA, Polinder S, 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.
7. Perel PA, Olldashi F, Muzha I, et al. Predicting outcome after traumatic brain injury: Practical prognostic models based on large cohort of international patients. BMJ 2008;336(7641):425-429.
8. Steyerberg EW, Mushkudiani N, Perel P, et al. Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. *PLoS Med* 2008;5(8):e165; discussion e165.
9. Maas AIR, Menon DK, Steyerberg EW, et al. Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) A Prospective longitudinal observational study. *Neurosurgery* 2014;76(1):67-80.
10. van Leeuwen N, Lingsma HF, Perel P, 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(4):811-818; discussion 818.
11. Steyerberg EW. *In: Validation of Prediction Models In: Clinical Prediction Models: A Practical Approach to Development, Validation and Updating* New York: Springer, 2009 p. 299-310.
12. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;21(1):128-138.
13. 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(8):971-980.
14. Damen J, Debray TPA, Pajouheshnia R, et al. Empirical evidence of the impact of study characteristics on the performance of prediction models: a meta-epidemiological study. *BMJ Open* 2019;9(4):e026160.
15. Van Calster B, Nieboer D, Vergouwe Y, et al. A calibration hierarchy for risk models was defined: from utopia to empirical data. *J Clin Epidemiol* 2016;74:167-176.
16. Marmarou A, Lu J, Butcher I, et al. IMPACT database of traumatic brain injury: design and description. *J Neurotrauma* 2007;24(2):239-250.
17. Maas AI, Menon DK, Steyerberg EW, et al. Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI): a prospective longitudinal observational study. *Neurosurgery* 2015;76(1):67-80.
18. Roberts I, Yates D, Sandercock P, et al. Effect of intravenous corticosteroids on death within 14 days in 10008 adults with clinically significant head injury (MRC CRASH trial): randomised placebo-controlled trial. *Lancet* 2004;364(9442):1321-1328.

19. Vergouwe Y, Nieboer D, Oostenbrink R, et al. A closed testing procedure to select an appropriate method for updating prediction models. *Stat Med* 2017;36(28):4529-4539.
20. Siregar S, Nieboer D, Versteegh MIM, et al. Methods for updating a risk prediction model for cardiac surgery: a statistical primer. *Interact Cardiovasc Thorac Surg* 2019.
21. Nieboer D, Vergouwe Y, Ankerst DP, et al. Improving prediction models with new markers: a comparison of updating strategies. *BMC Med Res Methodol* 2016;16(1):128.
22. Maas AI, Hukkelhoven CW, Marshall LF, et al. 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(6):1173-1182; discussion 1173-1182.
23. Thelin EP, Nelson DW, Vehvilainen J, et al. Evaluation of novel computerized tomography scoring systems in human traumatic brain injury: An observational, multicenter study. *PLoS Med* 2017;14(8):e1002368.
24. Lingsma HF, Roozenbeek B, Steyerberg EW, et al. Early prognosis in traumatic brain injury: from prophecies to predictions. *Lancet Neurol* 2010;9(5):543-554.
25. Roozenbeek B, Lingsma HF, Lecky FE, 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(5):1609-1617.
26. Letsinger J, Rommel C, Hirschi R, et al. The aggressiveness of neurotrauma practitioners and the influence of the IMPACT prognostic calculator. *PLoS One* 2017;12(8):e0183552.
27. Turnbull AE, Hayes MM, Brower RG, et al. Effect of Documenting Prognosis on the Information Provided to ICU Proxies: A Randomized Trial. *Crit Care Med* 2019.
28. Carter EL, Hutchinson PJ, Kolias AG, et al. Predicting the outcome for individual patients with traumatic brain injury: a case-based review. *Br J Neurosurg* 2016;30(2):227-232.
29. Harvey D, Butler J, Groves J, et al. Management of perceived devastating brain injury after hospital admission: a consensus statement from stakeholder professional organizations. *Br J Anaesth* 2018;120(1):138-145.

Chapter 7

Incremental Predictive Ability of Acute Serum Biomarkers for Functional Outcome Following Traumatic Brain Injury (CENTER-TBI): an Observational Cohort Study

I.R.A. Retel Helmrich, E. Czeiter, K. Amrein, A. Büki, H.F. Lingsma, D.K. Menon, S. Mondello, Ewout W. Steyerberg, N. von Steinbüchel, K.K.W. Wang, Lindsay Wilson, H. Xu, Z. Yang, D. van Klaveren, A.I.R. Maas, and the CENTER-TBI investigators and participants

Published in: The Lancet Neurology

doi.org/10.1016/S1474-4422(22)00218-6

Abstract

Background: Several studies have reported an association between serum biomarker values and functional outcome following traumatic brain injury (TBI). We aimed to examine the incremental (added) prognostic value of serum biomarkers over demographic, clinical and radiological characteristics and over established prognostic models, such as IMPACT and CRASH, for prediction of functional outcome.

Methods: We used data from the Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) Core study. Incremental prognostic value of six serum biomarkers (S100B, NSE, GFAP, UCH-L1, NFL and T-Tau), collected <24h of injury, was determined separately, and in combination. The primary outcome was the Glasgow Outcome Scale Extended (GOSE) six-months post-injury. Incremental prognostic value, using proportional odds and a dichotomized analysis, was assessed by delta concordance (C) statistic and delta R^2 between models with and without serum biomarkers, corrected for optimism with a bootstrapping procedure.

Findings: Serum biomarker values and 6-month GOSE were available in 2283/4509 patients. Higher biomarker levels were associated with worse outcome. Adding biomarkers improved the C-statistic and R^2 compared to demographic, clinical and radiological characteristics by 0·014 (95% CI 0·009-0·020) and R^2 by 4·9% (95% CI 3·6%-6·5%) for predicting GOSE. UCH-L1 had the greatest incremental prognostic value. Adding biomarkers to established prognostic models resulted in a relative increase in R^2 of 48%-65% for IMPACT and 30%-34% for CRASH prognostic models, respectively.

Interpretation: Serum biomarkers have incremental prognostic value for functional outcome following TBI. Our findings support integration of biomarkers, in particular UCH-L1, in established prognostic models.

Funding: The CENTER-TBI study was supported by the European Union 7th Framework program (EC grant 602150).

Key words: Prognosis; Prognostic model; Serum biomarkers; Traumatic Brain Injury; Glasgow Outcome Scale Extended.

Introduction

Traumatic brain injury (TBI) poses a major and increasing health burden with global socio-economic implications,[1] and represents a leading cause of death. In those who survive, long-term disability or residual complaints are common, even if they experienced 'mild' TBI as indicated by a Glasgow Coma Score of 13-15.[2]

Functional outcome following TBI depends on many different aspects, including patient and injury characteristics, mechanisms of trauma, patient response and the quality of care provided.[1] Establishing a reliable prognosis early after injury is challenging, but can be facilitated by the use of a prognostic model. Prognostic models combine information from multiple predictors to support clinicians in providing reliable information to patients and their relatives, help guide clinical decision making, inform benchmarking quality of care, and guide the design and analysis of clinical trials. Validated models are available to predict functional outcome following moderate and severe TBI,[3] including the IMPACT and CRASH models.[4, 5] However, these models only explain 35% of variance in outcome. Prognostic models for mild TBI (mTBI) are less well established.[6] Improving prognostication has been recognized as a high priority by clinicians and researchers.[7]

Prognostic value may increase by adding biomarkers. Over the past decade, blood-based protein biomarkers, and in particular S100 calcium-binding protein B (S100B), glial fibrillary acidic protein (GFAP), and ubiquitin C-terminal hydrolase L1 (UCH-L1) have received much attention for their role in diagnosing mTBI and triaging patients for computed tomography (CT) scanning of the head.[8] S100B has been implemented in the Scandinavian TBI Guidelines, and the combination of GFAP and UCH-L1 was approved by the FDA as a diagnostic test in patients suspected of mTBI based on the results of the ALERT-TBI study.[9]

In addition to the diagnostic role of biomarkers in TBI, an increasing body of evidence indicates the potential for a prognostic role. A substantial number of studies have shown an association between serum biomarkers and functional outcome following TBI.[10-16] However, most prior studies have mainly focused on the unadjusted prognostic effect of biomarkers rather than estimating their value over and above established prognostic factors, which is considered essential.[17] As a consequence, the independent prognostic value of biomarkers remains uncertain and their incremental value unknown.

The aim of our study was to determine the incremental prognostic value of six serum biomarkers (S100B, GFAP, UCH-L1, NSE, NFL, T-Tau) over patient's demographic,

clinical and radiological characteristics for the prediction of six-month functional outcome after TBI. Furthermore, we aimed to examine the incremental prognostic value of biomarkers when added to the IMPACT core and CRASH basic models for predicting mortality and unfavorable outcome after TBI.

Methods

Study population and design

Participants were drawn from the Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) Core study (version 3.0). CENTER-TBI was a prospective, multicenter, longitudinal, observational cohort study,[18, 19] that recruited patients between December 2014 and December 2017 from 18 countries across Europe and Israel. Inclusion criteria for the Core study were 1) a clinical diagnosis of TBI; 2) a clinical indication for computed tomography (CT) scanning; and 3) presentation within 24h of injury. Patients with a severe pre-existing neurological disorder were excluded. For the current analysis, selection of patients was limited to those with 1) blood sampling within 24h of injury, 2) availability of results from CT scan, 3) and for whom outcome assessment according to the Glasgow Outcome Scale-Extended (GOSE) was available at six-months.

Patients were stratified at enrollment by care path into the Emergency Department (ER) (assessed in the ER and discharged out of hospital), Admission (admitted to hospital ward), and Intensive Care Unit (ICU) strata (primary admission to the ICU). Informed consent was obtained from all participants or their legal representative according to local and national requirements. The use of biological samples was in accordance with the terms of the informed consent. The study was registered with ClinicalTrials.gov (NCT02210221), and is reported in accordance with the STROBE recommendations (see Supplementary material).

Clinical data were collected using a web-based electronic case report form (eCRF), with variables coded in accordance with the Common Data Elements (CDE) scheme (https://commondataelements.ninds.nih.gov/). Data were entered on the Quesgen e-CRF (Quesgen Systems Inc, USA), hosted on the International Neuroinformatics Coordinating Facility (INCF) platform and extracted via the INCF Neurobot tool (INCF, Sweden). We extracted data on demographic, clinical and radiological predictors of outcome, results of biomarker assays and outcome. The selection of predictors was based on established prognostic models for functional outcome after mild and moderate to severe TBI (Suppl Table 1).[3, 6] Radiological parameters were obtained from central readings of the first CT scan. Missing predictor values were

imputed with five iterations with multiple imputation using the mice package.[20] All demographic, clinical, and radiological characteristics, serum biomarkers, stratum, injury severity score (ISS) and six-month GOSE were included in the imputation model. Most observations showed low missingness (2 - 5%); the only exception being level of education, where missingness was higher (18%).

The primary outcome was the Glasgow Outcome Scale Extended (GOSE), the most widely used measure of global functional outcome following TBI, six-months post-injury. The GOSE was assessed by structured interview, conducted either by face to face or telephone interview, or by postal questionnaire (Suppl Table 2). Data collection for the GOSE interview was standardized using a manual for CENTER-TBI.[21] GOSE interviews and questionnaires were scored centrally using an algorithm to derive the GOSE rating. In subjects for whom both interview and questionnaire assessments were available, we used the interview-based rating. Categories 2 (vegetative state) and 3 (lower severe disability) were combined. Using a multi-state model, missing GOSE values for six months were imputed based on GOSE measurements obtained at other time points up to 18 months post-injury.[22] Biomarker values were not available at the time of outcome assessment, so all ratings were blinded to biomarker values.

We analysed the association of biomarkers with six-month GOSE adjusted for demographic, clinical and radiological parameters, and determined their incremental prognostic value. GOSE was analysed across all severities, and dichotomized into clinically relevant endpoints, namely mortality (GOSE=1), unfavorable outcome (GOSE≤4) and incomplete recovery (GOSE<8). Subgroup analyses were performed by stratum, and by injury severity. Finally, we determined the incremental prognostic value of biomarkers when added to the IMPACT core and CRASH basic models.

Ethical approval

The CENTER-TBI study (EC grant 602150) has been conducted in accordance with all relevant laws of the EU if directly applicable or of direct effect and all relevant laws of the country where the Recruiting sites were located, including but not limited to, the relevant privacy and data protection laws and regulations (the "Privacy Law"), the relevant laws and regulations on the use of human materials, and all relevant guidance relating to clinical studies from time to time in force including, but not limited to, the ICH Harmonised Tripartite Guideline for Good Clinical Practice (CPMP/ICH/135/95) ("ICH GCP") and the World Medical Association Declaration of Helsinki entitled "Ethical Principles for Medical Research Involving Human Subjects". Informed Consent by the patients and/or the legal representative/next of kin was obtained, accordingly to the local legislations, for all patients recruited in the Core Dataset of CENTER-TBI and documented in the e-CRF. Ethical approval

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

Sample Collection and Biomarker Measurements

Blood samples were collected using gel-separator tubes for serum and centrifuged within 60 (45±15) minutes. The serum was processed, aliquoted (8x0·5ml), and stored at –80°C locally until shipment on dry ice to the CENTER-TBI serum biobank (Pécs, Hungary).

We assayed S100B, NSE, GFAP, UCH-L1, NFL, and T-Tau. Details of the analyses procedures have been previously described.[23] In brief: S100B and NSE were measured with a clinical-use automated system, using an electrochemiluminecesence immunoassay kit (Elecsys S100 and NSE assays on the Cobas 8000 modular analyzer, Roche Diagnostics, Mannheim, Germany). GFAP, UCH-L1, t-tau, and NFL were analysed using Single Molecule Arrays (SiMoA) based assay on the SR-X benchtop assay platform (Quanterix Corp., Lexington, MA). Unique aliquots were used for analyses on two platforms to avoid repeated freeze-thaw cycles and analyzed in one round of experiments using the same batch of reagents by qualified laboratory technicians blinded to clinical information. All biosamples have reportable values above the LLOD value for the respective markers, with the exception of n=19 samples that have UCH-L1 values below its LLOD (1·34 pg/mL). For those samples, we assigned their UCH-L1 levels as 1·34 pg/mL. A technical summary of the biomarker sample collection and measurements can be found in the Supplementary materials (Suppl Table 3).

Statistical analysis

Descriptive statistics are presented as means, medians or frequencies. Differences in biomarker values by stratum (ER, Admission, ICU) and injury severity (mild and moderate/severe TBI) were compared using independent sample t tests.

We used proportional odds analysis to quantify the relationship between serum biomarkers and six-month GOSE across all severities, adjusted for demographic, clinical and radiological parameters, and binary logistic regression for the GOSE dichotomized for mortality (GOSE=1), unfavorable outcome (GOSE≤4) and incomplete recovery (GOSE<8). For the serum biomarkers we assessed nonlinearity with spline functions. The six biomarkers were considered separately, and in combination, with a particular focus on the combination of GFAP and UCH-L1, as this combination has been approved by the FDA as a diagnostic test for patients after mTBI in the US.[9]

Model performance was expressed in terms of discrimination (C-statistic), which indicates how well the model can differentiate between patients with a low and high risk of the outcome, and the R^2 (quantified as a percentage from 0-100 (%)), which indicates the goodness of fit of a logistic regression model.[24]

The incremental value of biomarkers in prognosticating outcome was assessed by calculating the difference in C-statistic (delta C; ΔC) and R^2 (delta R^2; ΔR^2) between the models with and without the serum biomarkers ('reference model'). A bootstrapping procedure was used to reduce optimistic model performance estimates.[24] Bootstrapping entails drawing random samples (n = 200) with replacement from the derivation cohort, with sample size equal to that of the derivation cohort. We also used bootstrapping to obtain confidence intervals for C, ΔC, R^2, and ΔR^2. Finally, we assessed the incremental prognostic value of biomarkers relative to the IMPACT core (Age, GCS motor, pupillary reactivity),[4] and CRASH basic (Age, GCS, pupillary reactivity, major extracranial injury (MEI)) models.[5]

Sensitivity analyses

We accounted for differences in predictor effects following mild, and moderate/severe TBI by fitting the models with interaction terms for GCS and the demographic, clinical and radiological parameters.

Subgroup analyses

The following subgroup analyses were performed:

- by care path as defined by stratum (ER; Admission; ICU)
- by injury severity, differentiated as moderate to severe (GCS 3-12) and mild (GCS 13-15)
- uncomplicated very mTBI (GCS=15, no traumatic abnormalities on first CT))
- mTBI with and without traumatic abnormalities on first CT

Statistical analysis was performed using R statistical software (http://www.r-project.org, version 3.6.0) in RStudio (http://www.rstudio.com, version 1.1.456). We used the 'rms' package to fit the logistic regression models.[25]

Role of funding source

The funders had no role in the collection, analysis and interpretation of data, nor in the writing of the report or in publication decisions. The authors had full access to study data and the senior authors had final responsibility for the decision to publish.

Results

Study Population

We included 2283/4509 (51%) adult patients (≥14 years) with available serum biomarker values within 24h after injury and six-month GOSE (Suppl Fig 1). Patients had a median age of 51 years (IQR = 32-67), 68% were male, and most (67%) were diagnosed with mild TBI (mTBI; GCS 13-15) (Table 1). More than a third (37%) experienced major extracranial injury. Baseline characteristics were largely similar to those previously described in the overall cohort (Suppl Table 4).[23] Characteristics of patients not included (n=2226) were similar to those analyzed (n=2283), although the percentage of patients with severe TBI was lower (20% versus 24%), and serum biomarker values were generally lower in patients not included.

The time from admission to sampling was shortest in the ER stratum (Median 5·0, IQR= [3·5-9·5]), compared to the admission (15·5 [(9·9-19·9]) and ICU strata (14·3 [7·7-19·6]) (Suppl Table 5; Suppl Fig 2).

At six months, 270 (12%) patients had died, 593 (26%) had unfavorable outcome, and 1443 (63%) patients had an incomplete recovery (Table 1).

Serum biomarkers and functional outcome following TBI

Higher biomarker levels were associated with poorer outcome overall, and when differentiated by stratum and injury severity (Fig 1; Suppl Table 5; Suppl Fig 3). Associations were stronger for UCH-L1, NFL, S100B, T-tau, and GFAP compared to NSE. Biomarker levels scaled with the intensity of care (as defined by stratum), and with TBI severity (higher after moderate-severe TBI compared to those with mTBI). All serum biomarkers were negatively correlated with six-month GOSE (Spearman rank correlations: S100B -0·43; NSE -0·28; GFAP -0·50; UCHL1 -0·54; T-tau -0·52; NFL -0·56; Suppl Fig 4).

Incremental prognostic value of serum biomarkers for prediction of GOSE

In proportional odds logistic regression analysis, biomarkers improved the prognostic value in addition to demographic, clinical and radiological characteristics for the prediction of six-month GOSE (Fig 2; Suppl Table 6). The C-statistic for the reference model was 0·781 (95% CI 0·768, 0·794), and increased with the addition of biomarkers. Improvements in C-statistic ranged from 0·002 (95% CI 0·000, 0·004) for NSE to 0·010 (95% CI 0·006, 0·015) for UCH-L1 (Suppl Table 6). Similarly, the addition of the biomarkers increased the R^2 of the reference model (44·8% (95% CI 41·4%, 47·8%), with improvements ranging from 0·8% R^2 (95% CI 0·3, 1·4) for NSE, to 3·8% R^2 (95% CI 2·8%, 5·1%) for UCH-L1 (Fig 2; Suppl Table 6). All six biomarkers

Table 1: Patients' demographic, clinical and radiological characteristics at admission, serum biomarker values within 24h and six-months functional outcome for all patients and by stratum (ER, Admission and ICU).

Characteristics	Overall[a] (n =2283)	ER (n =505, 22%)	Admission (n =624, 27%)	ICU (n =1154, 51%)
Age (14-95) (Median [IQR])	51 [32-67]	50·00 [32-66]	54 [35-69]	49 [31,66]
% Male sex	68% (1559)	57% (287)	67% (420)	74% (852)
Level of education	(N=1881)	(N=479)	(N=538)	(N=864)
College/Uni degree	467 (25)	156 (33)	141 (26)	170 (20)
Currently in school/With diploma or degree-oriented program	395 (21)	84 (18)	129 (24)	182 (21)
None/primary school	347 (18)	94 (20)	100 (19)	153 (18)
Secondary/High school	672 (36)	145 (30)	168 (31)	359 (42)
Pre-injury mental health problems	272 (12)	60 (12)	68 (11)	144 (13)
GCS baseline	*(N=2209)*	(N=503)	(N=605)	(N=1101)
Mild (13-15)	1472 (67)	499 (99)	578 (96)	395 (36)
Moderate (9-12)	186 (8)	2 (0·4)	21 (4)	163 (15)
Severe (3-8)	551 (25)	2 (0·4)	6 (1)	543 (49)
GCS motor score	(N=2241)	(N=503)	(N=606)	(N=1132)
None	361 (16)	2 (0·4)	2 (0·3)	357 (32)
Extension	35 (2)	0 (0·0)	1 (0·2)	34 (3)
Abnormal flexion	40 (2)	0 (0·0)	1 (0·2)	39 (3)
Normal flexion	89 (4)	0 (0·0)	4 (1)	85 (8)
Localizes	235 (11)	4 (1)	14 (2)	217 (19)
Obeys	1481 (66)	497 (99)	584 (96)	400 (35)
Reaction of Pupils	(N=2178)	(N=483)	(N=591)	(N=1104)
Both	1944 (89)	474 (98)	577 (98)	893 (81)
One	90 (4)	1 (0·2)	8 (1)	81 (7)
None	144 (7)	8 (2)	6 (1)	130 (12)
Marshall CT	(N=2182)	(N=497)	(N=597)	(N=1088)
I	836 (38)	428 (86)	292 (49)	116 (11)
II	834 (38)	67 (14)	252 (42)	515 (47)
III	90 (4)	0 (0)	6 (1)	84 (8)
IV	19 (1)	0 (0)	0 (0)	19 (2)
V	6 (0·3)	0 (0)	1 (0·2)	5 (0·5)
VI	397 (18)	2 (0·4)	46 (8)	349 (32)
Traumatic Subarachnoid Hemorrhage	1015 (47)	44 (9)	195 (32)	776 (73)

Table 1: Continued

Characteristics	Overall[a] (n =2283)	ER (n =505, 22%)	Admission (n =624, 27%)	ICU (n =1154, 51%)
Epidural Hematoma	233 (11)	1 (0·2)	42 (7)	190 (18)
Hypotension	172 (9)	3 (1)	9 (2)	160 (15)
Hypoxia	163 (8)	1 (0·2)	9 (2)	153 (14)
Glucose (Median [IQR])	7·1 [6·0-8·6]	6·0 [5·3-7·0]	6·6 [5·8-7·8]	7·7 [6·4-9·3]
Hemoglobin (Median [IQR])	13·5 [12·0-14·6]	14·1 [12·8-14·9]	13·9 [12·8-14·9]	13·2 [11·6-14·5]
ISS (0-75) (Median, [IQR])	16 [9-29]	4 [2-6]	10 [9-16]	29 [25-41]
MEI[b]	848 (37)	17 (3)	172 (28)	659 (57)
Serum biomarkers within 24 hours				
S100B mg/L	0·12 [0·07-0·26]	0·09 [0·05-0·15]	0·08 [0·06-0·2]	0·19 [0·10-0·43]
NSE ng/ml	15·5 [11·7-23·4]	13·7 [10·9-17·5]	13·6 [11·1-18·4]	19·3 [13·4-29·5]
GFAP ng/ml	3·0 [0·48-15·7]	0·30 [0·11-0·91]	1·3 [0·32-4·8]	12·3 [3·4-38·0]
UCH-L1 pg/ml	88·5 [35·1-281·3]	35·8 [15·8-62·6]	49·1 [22·2-108·2]	232·6 [93·4-563·1]
T-Tau pg/ml	2·6 [1·2-7·0]	1·1 [0·63-1·7]	1·7 [0·99-3·2]	5·9 [2·7-13·8]
NFL pg/ml	23·7 [9·4-74·6]	8·7 [5·3-15·1]	13·7 [7·3-25·9]	58·8 [27·7-139·9]
Sampling time (h) ((Median [IQR])	12·6 [6·0-18·9]	5·0 [3·5-9·5]	15·5 [9·9-19·9]	14·3 [7·7-19·6]
Functional outcome six months post-injury				
Death	270 (12)	3 (1)	31 (5)	236 (21)
Vegetative state/Lower Severe disability	221 (10)	9 (2)	18 (3)	194 (17)
Upper Severe disability	102 (5)	7 (1)	14 (2)	81 (7)
Lower Moderate disability	225 (9)	15 (3)	42 (7)	168 (15)
Upper Moderate disability	209 (9)	26 (5)	50 (8)	133 (12)
Lower good recovery	416 (18)	94 (19)	160 (26)	162 (14)
Upper good recovery	840 (37)	351 (70)	309 (50)	180 (16)

[a] Patients <14 years of age (N=43) were excluded.

[b] Patients with an Abbreviated Injury Scale ≥ 3 regarding the following body regions; face, cervical spine, thorax/chest, abdomen/pelvic contents, extremities and pelvic girdle, or external (skin), thus excluding head and neck.

Abbreviations: GCS, Glasgow Coma Scale; ISS, Injury Severity Score; N, Number; MEI, Major Extracranial Injury; SD, Standard Deviation; IQR, Interquartile range

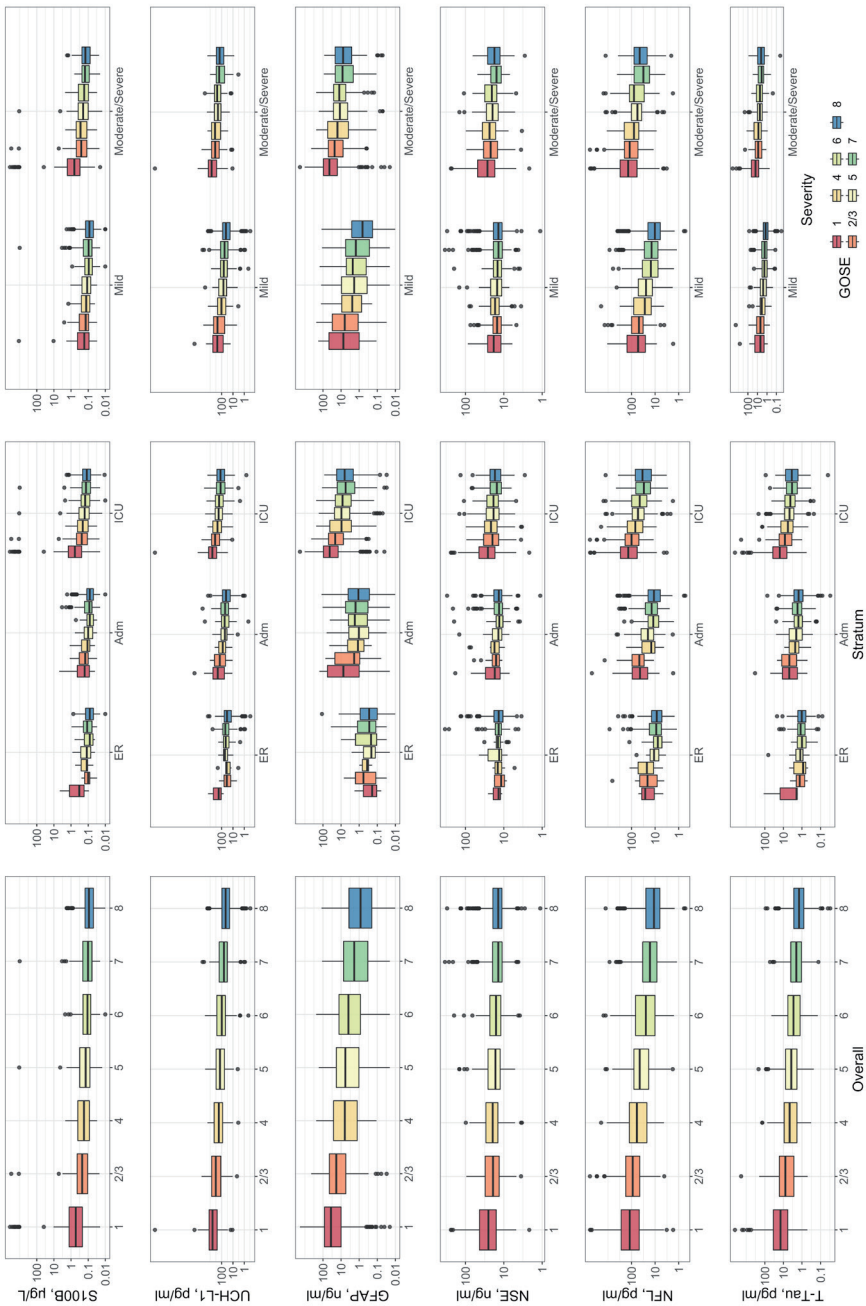


Figure 1: Biomarker values by stratum and by clinical severity, differentiated for the Glasgow Outcome Scale Extended (GOSE). The serum biomarkers are shown on the log scale. The boxplots show the median serum biomarker value (thick black line). The first quartile is indicated by the line above and the third quartile by the line below the median. The whiskers are the minimum and maximum. The dots are outliers.

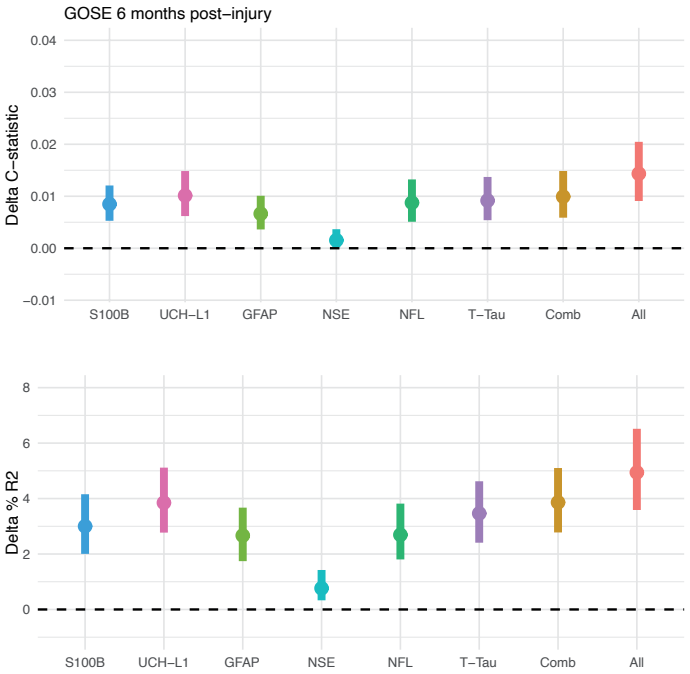


Figure 2: The difference (delta) in C-statistic and % R^2 between the reference model and models including serum biomarkers of ordinal regression models adjusted for demographic, clinical and radiology parameters (see Supplementary Table 1 for model parameters) for the Glasgow Outcome Scale Extended 6 months post-injury. Six biomarkers are considered separately, in combination (Comb; GFAP + UCHL1) and taken together ("all"). The absolute values are presented in Supplementary Table 4. The points illustrate the delta C-statistic (above) and R^2 (below) and the vertical lines above and below the points illustrate the 95% CI around the estimate.

taken together had substantial incremental value over single biomarkers (ΔC-statistic 0·014 (95% CI 0·009, 0·020; ΔR^2 4·9% (95% CI 3·6%, 6·5%)). Combinations of UCH-L1 with NFL, NFL with T-tau, and NFL with T-tau and S100B showed similar performance as all biomarkers together (Suppl Table 7). The combination of GFAP with UCH-L1 did not improve discrimination compared to UCH-L1 alone.

In binary logistic regression analysis, the reference model discriminated very well: 0·922 (95% CI 0·906, 0·936) for mortality, 0·883 (95% CI 0·866, 0·898) for unfavorable outcome, and 0·802 (95% CI 0·783, 0·819) for incomplete recovery (Table 2). Nevertheless, biomarkers showed incremental prognostic value (Table 2, Suppl Fig 5). Incremental value was highest for UCH-L1 and T-Tau in predicting mortality (ΔC-statistic for both biomarkers: 0·011 (95% CI 0·005, 0·017); ΔR2: 3·8% (95% CI 2·1%, 5·9%) for UCH-L1 and 3·8% (95% CI 2·0%, 6·2%) for T-Tau), and for

Table 2: Discriminative ability (C-statistic) and R^2 of serum biomarkers adjusted for demographic. clinical and radiology parameters to predict functional outcome six-months following traumatic brain injury for three clinically relevant thresholds; mortality, unfavorable outcome (GOSE ≤4) and incomplete recovery (GOSE <8).

	Mortality (N=270)				**Unfavorable outcome (N=593)**				**Incomplete recovery (N=1443)**			
	C-statistic (95% CI)		**R^2 (%) (95% CI)**		**C-statistic (95% CI)**		**R^2 (%) (95% CI)**		**C-statistic (95% CI)**		**R^2 (%) (95% CI)**	
Reference model	0·922	(0·906, 0·936)	51·0%	(45·7%, 56·1%)	0·883	(0·866, 0·898)	49·7%	(45·3%, 53·8%)	0·802	(0·783, 0·819)	33·8%	(29·8%, 37·7%)
Serum biomarkers	**Delta C-statistic (95% CI)**		**Delta R^2 (%) (95% CI)**		**Delta C-statistic (95% CI)**		**Delta R^2 (%) (95% CI)**		**Delta C-statistic (95% CI)**		**Delta R^2 (%) (95% CI)**	
S100B	0·009	(0·004, 0·015)	3·3%	(1·4%, 5·7%)	0·008	(0·004, 0·012)	2·4%	(1·3%, 3·8%)	0·007	(0·003, 0·012)	1·6%	(0·8%, 2·5%)
UCH-L1	0·011	(0·005, 0·017)	3·8%	(2·1%, 5·9%)	0·014	(0·009, 0·021)	4·6%	(3·0%, 6·5%)	0·007	(0·003, 0·012)	1·9%	(1·1%, 3·0%)
GFAP	0·006	(0·002, 0·012)	2·3%	(0·9%, 4·1%)	0·010	(0·005, 0·015)	3·0%	(1·8%, 4·5%)	0·003	(0·000, 0·006)	1·0%	(0·4%, 1·8%)
NSE	0·004	(0·001, 0·009)	1·6%	(0·4%, 3·3%)	0·003	(0·001, 0·007)	1·1%	(0·3%, 2·1%)	0·001	(0·000, 0·002)	0·2%	(0·1%, 0·5%)
NFL	0·005	(0·001, 0·009)	1·5%	(0·3%, 3·0%)	0·015	(0·009, 0·022)	4·2%	(2·9%, 5·9%)	0·007	(0·003, 0·013)	1·8%	(0·9%, 3·0%)
T-tau	0·011	(0·005, 0·017)	3·9%	(2·0%, 6·2%)	0·013	(0·008, 0·019)	4·2%	(2·6%, 5·9%)	0·006	(0·002, 0·011)	1·7%	(0·9%, 2·7%)
GFAP + UCH-L1	0·011	(0·006, 0·018)	4·1%	(2·2%, 6·2%)	0·014	(0·008, 0·020)	4·5%	(2·9%, 6·3%)	0·006	(0·002, 0·012)	1·8%	(0·9%, 2·9%)
All	0·012	(0·005, 0·019)	4·3%	(1·9%, 6·9%)	0·019	(0·012, 0·027)	5·6%	(3·7%, 7·7%)	0·010	(0·004, 0·018)	2·5%	(1·2%, 4·1%)

NFL in predicting unfavorable outcome (ΔC-statistic: 0·015 (95% CI 0·009, 0·022) ; ΔR2: 4·2% (95% CI 2·9%, 5·9%)). Single biomarkers had lower incremental value for the prediction of incomplete recovery, and was highest for S100B, UCH-L1 and NFL, and lowest for NSE (Table 2). Results were similar for the prediction of incomplete recovery in patients with mTBI and uncomplicated very mTBI; the incremental prognostic value of biomarkers was highest for S100B, UCH-L1 and NFL (Suppl Table 8).

Results were consistent across strata and injury severity (Suppl Tables 9 and 10). Serum biomarkers had incremental prognostic value for the prediction of six-month functional outcome for patients in the ER, admission, and ICU strata (Suppl Table 9), and in patients following mild and moderate/severe TBI (Suppl Table 10). The incremental prognostic value of biomarkers was similar in mTBI and moderate/severe TBI (Suppl Table 10). In patients following mTBI with and without traumatic abnormalities on CT the incremental value remains, but the added value of biomarkers is more pronounced in mTBI patients with CT abnormalities (Suppl Table 11). The addition of interaction terms for GCS led to a decrease in incremental value of serum biomarkers for prediction of six-month GOSE (Suppl Table 12).

Incremental prognostic value of serum biomarkers relative to established prognostic models

The incremental value of biomarkers when added to the IMPACT core and CRASH basic models, for prediction of mortality and unfavorable outcome in patients with moderate to severe TBI, was substantial (Table 3). For mortality, improvements in C-statistic ranged from 0·016 (95% CI 0·000, 0·036) for NFL to 0·053 (95% CI 0·029, 0·080) for UCHL-L1 for the IMPACT model, and from 0·013 (95% CI 0·003, 0·026) for NFL to 0·035 (95% CI 0·019, 0·052) for UCH-L1 for the CRASH model. For unfavorable outcome, improvements in C-statistic ranged from 0·030 (95% CI 0·015, 0·048) for NSE to 0·066 (95% CI 0·041, 0·093) for UCH-L1 for IMPACT, and from 0·018 (95% CI 0·009, 0·029) for NSE to 0·041 (95% CI 0·026, 0·058) for UCH-L1 for CRASH. The R^2 for the IMPACT and CRASH models was 30·7% (23·5%, 37·7%) and 22·6% (95% CI 15·6%, 29·1%) for mortality, and 35·2% (28·8%, 41·8%) and 33·8% (28·4%, 39·7%) for unfavorable outcome. For mortality, adding all biomarkers increased the prognostic value with 14·6% R^2 (95% CI 8·6%, 20·6%) for IMPACT and 10·7% R^2 (6·4%, 15·2%) for CRASH, corresponding to a relative increase of 48% (14·6/30·7) for IMPACT and of 30% (10·7/35·2) for CRASH. For unfavorable outcome, adding all biomarkers increased model performance with 14·6% R^2 (95% CI 9·5%, 20·2%) for IMPACT and 11·6% R^2 (95% CI 7·8%, 15·8%) for CRASH, corresponding to a relative increase of 65% (14·6/22·6) for IMPACT and 34% (11·6/33·8) for CRASH. Of single biomarkers,

UCH-L1 had the greatest incremental value in R^2: 12·5% (95% CI 7·3%, 17·8%) when added to IMPACT, and 12·8% (95% CI 8·3%, 18·0%) when added to CRASH for predicting mortality.

Table 3: Change in discriminative ability (C-statistic) and R^2 of serum biomarkers compared to IMPACT core and CRASH basic models to predict mortality and unfavorable outcome six-months following traumatic brain injury.

	IMPACT core (Age, GCS motor, GCS pupils)							
	GCS ≤ 12 (N=737)							
	Mortality				**Unfavorable outcome**			
	C-statistic (95% CI)		**R^2 (%) (95% CI)**		**C-statistic (95% CI)**		**R^2 (95% CI)**	
Reference model	0·803	(0·766, 0·836)	30·7%	(23·5%, 37·7%)	0·736	(0·696, 0·772)	22·6%	(15·6%, 29·1%)
Serum biomarkers	**Delta C-statistic (95% CI)**		**Delta R^2 (%) (95% CI)**		**Delta C-statistic (95% CI)**		**Delta R^2 (95% CI)**	
S100B	0·059	(0·037, 0·084)	12·6%	(7·2%, 17·9%)	0·046	(0·025, 0·070)	8·7%	(4·7%, 13·2%)
UCH-L1	0·053	(0·029, 0·080)	12·5%	(7·3%, 17·8%)	0·066	(0·041, 0·093)	12·8%	(8·3%, 18·0%)
GFAP	0·037	(0·018, 0·061)	8·8%	(4·4%, 13·4%)	0·048	(0·026, 0·071)	8·7%	(4·6%, 13·1%)
NSE	0·033	(0·015, 0·054)	8·3%	(4·4%, 12·9%)	0·030	(0·015, 0·048)	6·1%	(3·2%, 9·3%)
NFL	0·016	(0·000, 0·036)	4·1%	(0·9%, 8·3%)	0·046	(0·023, 0·070)	8·7%	(4·6%, 13·1%)
T-tau	0·050	(0·026, 0·076)	12·5%	(7·5%, 17·9%)	0·058	(0·034, 0·083)	11·1%	(6·7%, 15·9%)
GFAP + UCH-L1	0·053	(0·030, 0·080)	12·8%	(7·7%, 18·2%)	0·065	(0·040, 0·093)	12·6%	(8·0%, 17·9%)
All	0·061	(0·033, 0·089)	14·6%	(8·6%, 20·6%)	0·075	(0·047, 0·105)	14·6%	(9·5%, 20·2%)

Table 3: Continued

	CRASH basic (Age, GCS, GCS pupils, MEI)							
	GCS < 15 (N= 1083)							
	Mortality				Unfavorable outcome			
	C-statistic (95% CI)		R^2 (%) (95% CI)		C-statistic (95% CI)		R^2 (%) (95% CI)	
Reference model	0·835	(0·805, 0·863)	35·2%	(28·8%, 41·8%)	0·798	(0·772, 0·825)	33·8%	(28·4%, 39·7%)
Serum biomarkers	**Delta C-statistic (95% CI)**		**Delta R^2 (%) (95% CI)**		**Delta C-statistic (95% CI)**		**Delta R^2 (%) (95% CI)**	
S100B	0·034	(0·019, 0·050)	8·5%	(4·7%, 12·3%)	0·025	(0·012, 0·039)	6·0%	(3·4%, 9·3%)
UCH-L1	0·035	(0·019, 0·052)	9·2%	(5·6%, 13·2%)	0·041	(0·026, 0·058)	10·0%	(6·7%, 13·8%)
GFAP	0·023	(0·010, 0·040)	6·6%	(3·5%, 10·4%)	0·028	(0·015, 0·042)	6·8%	(3·8%, 9·8%)
NSE	0·021	(0·009, 0·036)	5·9%	(3·2%, 9·5%)	0·018	(0·009, 0·029)	4·4%	(2·3%, 7·0%)
NFL	0·013	(0·003, 0·026)	3·2%	(0·8%, 6·1%)	0·031	(0·018, 0·047)	7·5%	(4·5%, 10·8%)
T-tau	0·033	(0·017, 0·051)	9·0%	(5·2%, 13·0%)	0·038	(0·023, 0·054)	9·3%	(6·0%, 13·1%)
GFAP + UCH-L1	0·035	(0·019, 0·052)	9·4%	(5·7%, 13·5%)	0·040	(0·025, 0·057)	9·8%	(6·4%, 13·7%)
All	0·040	(0·022, 0·059)	10·7%	(6·4%, 15·2%)	0·049	(0·031, 0·068)	11·6%	(7·8%, 15·8%)

Discussion

We examined the incremental prognostic value of serum biomarkers, independent of patient's demographic, clinical and radiological characteristics, for prediction of six-month GOSE following TBI. All examined serum biomarkers – UCH-L1, S100B, GFAP, NFL, t-tau, and NSE - obtained within 24h after injury, improved the prognostic value for functional outcome. We found that UCH-L1 had the greatest incremental prognostic value. Combining all six biomarkers resulted in small further increments in C-statistic and R^2, compared to the best performing individual biomarkers separately. Adding biomarkers to the IMPACT and CRASH models resulted in an R^2 up to 45% and 46% for mortality and 37% and 45% for unfavorable outcome, respectively.

Previous studies have reported associations between serum biomarker levels and functional outcome following TBI.[11, 15, 26] These studies typically focused on the unadjusted effect of biomarkers rather than estimating their value over and above known predictors of outcome following TBI. We showed that the addition of biomarkers can improve prognostication over and above demographic, clinical, and radiological characteristics. We also provide greater detail on the context-specific performance and potential clinical application of our findings.

We showed that the prognostic performance of individual biomarkers may vary with injury severity. NFL provided the greatest incremental prognostic value in patients after mTBI for predicting incomplete recovery, followed by S100B, UCH-L1 and T-tau. However, in moderate to severe TBI, the greatest incremental value was provided by UCH-L1 for predicting unfavorable outcome,[27] closely followed by T-Tau, NFL, and S100B. Future studies should further examine differences in prognostic value of serum biomarkers between patients following mild versus moderate/severe TBI. As S100B can also be present outside the central nervous system,[28] questions have been raised about the specificity of S100B as a biomarker in TBI, particularly in patients with extracranial injuries. However, our results suggest that S100B has added value for the prediction of functional outcome after TBI, relative to known predictors, including major extracranial injury.

The prognostic performance of individual biomarkers may not be concordant with their diagnostic utility. In a prior CENTER-TBI study of the incremental value of these six serum biomarkers for the prediction of CT abnormalities, GFAP outperformed the other markers.[23] This is consistent with other studies of the diagnostic performance of GFAP.[9, 29, 30] The association between biomarkers and imaging phenotypes was described in greater detail in a prior CENTER-TBI publication.[31] Lesion volume showed stronger associations with biomarkers than pathoanatomical type of injury. Overall, GFAP showed the highest value in all pathology groups. In contrast, in the current study, GFAP showed relatively little added value for the prediction of functional outcome following TBI. Our findings indicate that GFAP is more relevant for diagnostic purposes, and less so for predicting functional outcome following TBI. Different pathobiological roles, marker-specific features (e.g., kinetics, abundance, localization), and their link with distinct injury types and pathophysiological mechanisms could underlie these differences in performance. Accordingly, previous studies have demonstrated different GFAP and UCH-L1 release patterns as a result of different patterns of structural damage, which in turn imply different clinical relevance and ensuing outcomes.[15, 31, 32] Previously, UCHL-1, assessed over the first 5 days after injury, displayed the best discrimination for predicting outcome in univariate analysis, outperforming other known predictors.[33] On multivariable analysis, however, GFAP

and NFL added most independent information to predict unfavorable outcome. The differential diagnostic and prognostic effects of biomarkers may have various explanations. First, UCHL-1 and NFL are neuronal markers, whilst GFAP is an astroglial marker. Conceptually a marker that reflects neuronal damage could be expected to be better correlated with outcome than an astroglial marker. Second, temporal trajectories may be relevant. Future research should focus on the validation of our findings and explaining differences in the diagnostic and prognostic value of different biomarkers. It has been suggested that a panel of biomarkers, based on a multi-marker approach, might improve prognostic accuracy.[17] We found that a multi-marker approach of all six biomarkers together indeed has most incremental prognostic value. However, the combination of GFAP with UCH-L1, which has been proposed as a useful combination for diagnosis of TBI,[9] did not improve discrimination of outcome when compared to UCH-L1 alone. Based on our findings, combinations of UCH-L1 with S100B, NFL and T-tau may provide better opportunities in future research for the prediction of functional outcome following TBI. However, when compared to the best performing individual biomarkers, the incremental discrimination provided by combining the entire biomarker panel was relatively small. Consequently, the use of a single marker or a combination of two markers, might be preferred in clinical practice, especially in low- and middle-income countries and austere environments.

The improvement in prognostic value by combining biomarker data with conventional predictors of outcome may translate into clinical application. First, integrating biomarker data with established prognostic models for the prediction of death or unfavorable outcome, resulted in a relative increase in R^2 of 48%-65% for IMPACT and 30%-34% for CRASH, respectively. These models are widely used to stratify patients in clinical trials, and for benchmarking quality-of-care assessments. These improvements in prognostic value were for IMPACT core and CRASH basic models, when only age, initial injury severity (based on GCS, motor score and pupillary reactivity) and MEI were considered. Second, even when all demographic, clinical and radiological parameters were used, biomarkers were still able to provide incremental value not just for the GOSE overall, but also for mortality, unfavorable outcome, and incomplete recovery, which are relevant to clinicians and patients; Biomarkers resulted in R^2 up to 55% for mortality, 55% for unfavorable outcome, and 36% for incomplete recovery. These results make a strong case for integrating serum biomarker data when developing or updating prognostic models for functional outcome following TBI.

Strengths and limitations

Strengths of our study include the use of a longitudinal prospective international cohort study (the CENTER-TBI study), resulting in an unprecedented large number

of patients following TBI with available serum biomarkers obtained within 24h. Our sample included 2283/4509 (51%) patients from the overall CENTER-TBI cohort. Baseline characteristics were largely similar to those previously described in the overall CENTER-TBI cohort.[23] The analyses were performed across all severities of TBI, including mostly patients following mTBI, which reflects contemporary clinical practice. Furthermore, the CENTER-TBI study includes a relatively high percentage of patients with traumatic abnormalities on CT, reflecting the type of patients seen in large trauma referral centres. To study generalizability, our findings should be further validated in new patients and settings. In contrast to prior studies of serum biomarkers in TBI, we adjusted for known predictors of outcome following TBI. Furthermore, we examined the incremental value of six serum biomarkers that have been studied most extensively in recent studies, both in isolation and in combination (including the specific combination of GFAP and UCH-L1, thought to have specific diagnostic utility). Prior CENTER-TBI studies have examined and explained differences between men and women in outcome after TBI.[34, 35] Future studies should explore the relationship between serum biomarkers and differences in outcomes between men and women following TBI.

Several limitations of our study must be considered. Most patients were categorized as mTBI based on the GCS. However, predictors of outcome following mTBI are less well established than those for moderate and severe TBI. Most demographic, clinical and radiological characteristics included in our study are relevant to predict outcome in patients following moderate and severe TBI, but less so in patients following mTBI. Therefore, we also included MEI, level of education and pre-injury mental health problems, which are known predictors outcome in patients following mTBI. Differences were noted in predictor effects for patients following mild versus moderate and severe TBI. Second, in the CENTER-TBI study the time of biomarker sampling is widely varying and typically late (Mean 12.6 hours after injury). Serial sampling of serum biomarkers, including S100B and NSE, has revealed different temporal trajectories.[36] Future research should consider mixed model approaches for the prediction of functional outcome following TBI including repeated measures of serum biomarkers. Third, the Quanterix platform on which we measured four of the six biomarkers is a research-use only device, and this platform currently cannot be used in clinical practice. Robust clinical assay platforms are required before biomarkers can be broadly implemented into clinical practice for either diagnostic or prognostic purposes. The high coefficients of variation (CVs) reported for the assays performed on the Quanterix platform are of some concern. However, we consider that these high CVs would be more likely to dilute prognostic effects than to inflate these. S100B and NSE tests are available as clinical lab tests, and have been cleared in the US and Europe as cancer marker tests. Procedures for regulatory approval of assays for other

biomarkers are ongoing. Recently, a point-of-care assay for UCH-L1/GFAP obtained FDA clearance in the US and CE mark by EMN/European Medicines Agency as in vitro diagnostic test for mTBI patients with suspected brain lesions. Fourth, we recognize that levels of some biomarkers (e.g. NSE) could be artificially elevated in haemolytic samples, and that this may have contributed to the relatively low prognostic strength of NSE. As the current study aims to assess the incremental value of biomarkers in clincal practice, and haemolysis may sometimes occur despite strict procedures for sampling, pre-processing and processing of samples, we opted not to exclude haemolytic samples.

Conclusion

Serum biomarkers obtained within 24h after injury have incremental prognostic value relative to demographic, clinical and radiological characteristics in predicting functional outcome following mild, moderate and severe TBI. Our findings support the integration of biomarkers in established models for predicting outcome after TBI.

Authors' contributions

All authors certify that they have participated in the concept, design, analysis, writing, or revision of the manuscript. AB and EC conceived the original idea. DvK and AIRM supervised the project. IRARH and DvK analysed the data. All authors participated in the interpretation of results relevant to their domain of interest. IRARH, AIRM, and DM prepared the draft manuscript and coordinated its finalisation. All authors approved the final manuscript. IRARH and DvK verified the underlying data, all main authors had full access to study data and IRARH, DvK and AIRM had final responsibility for the decision to publish

Conflicts of interest

The authors declare that there is no conflict of interest.

Data sharing

Individual participant data will be available immediately following publication, conditional to approved study proposal, with no end date. Data will be available to researchers who provide a methodologically sound study proposal that is approved by the management committee to achieve the aims in the approved proposal. Proposals can be submitted online at https://www.center-tbi.eu/data. A data access agreement is required and all access must comply with regulatory restrictions imposed on the original study.

Funding and role of the funding source

The research leading to these results was supported by the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 602150 (CENTER-TBI). Additional funding was obtained from the Hannelore Kohl Stiftung (Germany), from OneMind (USA), from Integra LifeSciences Corporation (USA) and from NeuroTrauma Sciences (US). The funders had no role in the collection, analysis and interpretation of data, nor in the writing of the report or in publication decisions. IRARH and DvK verified the underlying data, all main authors had full access to study data and IRARH, DvK and AIRM had final responsibility for the decision to publish

Acknowledgements

We are grateful to all patients and investigators that participated in the CENTER-TBI study.

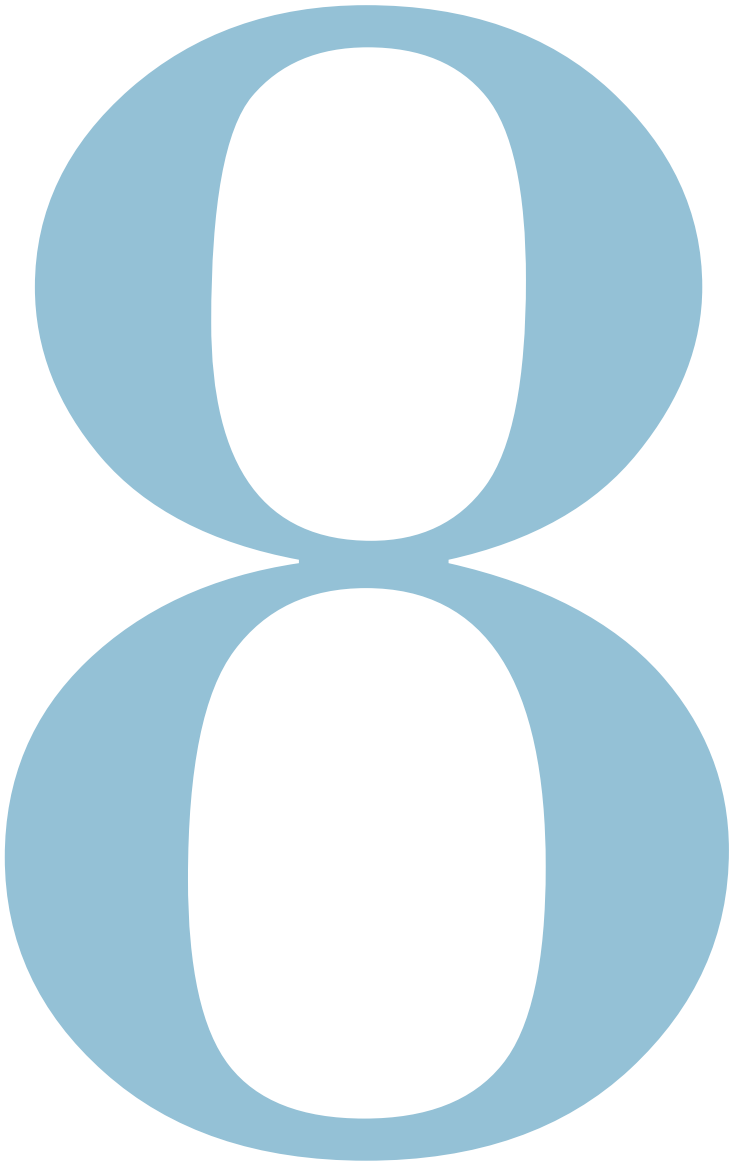
Supplementary material

Supplementary material is available at: https://www.thelancet.com/journals/laneur/article/PIIS1474-4422(22)00218-6/fulltext

References

1. Maas A I R, Menon D K, Adelson P D, et al. Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. *Lancet Neurol*. 2017.
2. Buki A, Kovacs N, Czeiter E, et al. Minor and repetitive head injury. *Adv Tech Stand Neurosurg*. 2015;**42**:147-92.
3. Dijkland S A, Foks K A, Polinder S, et al. Prognosis in moderate and severe traumatic brain injury: a systematic review of contemporary models and validation studies. *J Neurotrauma*. 2019.
4. Steyerberg E, Mushkudiani N, Perel P, et al. Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. *PLoS Med*. 2008;**5**(8):e165; discussion e.
5. Perel P, Arango M, Clayton T, et al. Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients. *BMJ*. 2008;**336**(7641):425-9.
6. Mikolić A, Polinder S, Steyerberg E W, et al. Prediction of Global Functional Outcome and Post-Concussive Symptoms after Mild Traumatic Brain Injury: External Validation of Prognostic Models in the Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) Study. *J Neurotrauma*. 2020.
7. Kim D Y, Lissauer M, Martin N, Brasel K. Defining the surgical critical care research agenda: Results of a gaps analysis from the Critical Care Committee of the American Association for the Surgery of Trauma. *J Trauma Acute Care Surg*. 2020;**88**(2):320-9.
8. Mondello S, Sorinola A, Czeiter E, et al. Blood-based protein biomarkers for the management of traumatic brain injuries in adults presenting to emergency departments with mild brain injury: a living systematic review and meta-analysis. *J Neurotrauma*. 2021;**38**(8):1086-106.
9. Bazarian J J, Biberthaler P, Welch R D, et al. Serum GFAP and UCH-L1 for prediction of absence of intracranial injuries on head CT (ALERT-TBI): a multicentre observational study. *Lancet Neurol*. 2018;**17**(9):782-9.
10. Vos P E, Jacobs B, Andriessen T, et al. GFAP and S100B are biomarkers of traumatic brain injury: an observational cohort study. *Neurology*. 2010;**75**(20):1786-93.
11. Zhang Z, Zhang L, Dong X, et al. Comparison of the performances of copeptin and multiple biomarkers in long-term prognosis of severe traumatic brain injury. *Peptides*. 2014;**60**:13-7.
12. Di Battista A P, Buonora J E, Rhind S G, et al. Blood biomarkers in moderate-to-severe traumatic brain injury: potential utility of a multi-marker approach in characterizing outcome. *Front Neurol*. 2015;**6**:110.
13. Gradisek P, Carrara G, Antiga L, et al. Prognostic value of a combination of circulating biomarkers in critically ill patients with traumatic brain injury: results from the European CREACTIVE study. *J Neurotrauma*. 2021;**38**(19):2667-76.
14. Graham N S N, Zimmerman K A, Moro F, et al. Axonal marker neurofilament light predicts long-term outcomes and progressive neurodegeneration after traumatic brain injury. *Sci Transl Med*. 2021;**13**(613):eabg9922.
15. Mondello S, Papa L, Buki A, et al. Neuronal and glial markers are differently associated with computed tomography findings and outcome in patients with severe traumatic brain injury: a case control study. *Crit Care*. 2011;**15**(3):1-10.
16. Takala R S K, Posti J P, Runtti H, et al. Glial fibrillary acidic protein and ubiquitin C-terminal hydrolase-L1 as outcome predictors in traumatic brain injury. *World Neurosurg*. 2016;**87**:8-20.
17. Huie J R, Mondello S, Lindsell C J, et al. Biomarkers for traumatic brain injury: data standards and statistical considerations. *J Neurotrauma*. 2020.

18. Steyerberg E W, Wiegers E, Sewalt C, et al. Case-mix, care pathways, and outcomes in patients with traumatic brain injury in CENTER-TBI: a European prospective, multicentre, longitudinal, cohort study. *Lancet Neurol*. 2019;**18**(10):923-34.
19. Maas A I R, Menon D K, Steyerberg E W, et al. Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) A Prospective Longitudinal Observational Study. *J Neurosurg*. 2014;**76**(1):67-80.
20. van Buuren S. Flexible Imputation of Missing Data: CRC Press, Taylor & Francis Group; 2018.
21. Wilson L, Boase K, Nelson L D, et al. A manual for the glasgow outcome scale-extended interview. *Journal of neurotrauma*. 2021;**38**(17):2435-46.
22. Kunzmann K, Wernisch L, Richardson S, et al. Imputation of ordinal outcomes: a comparison of approaches in traumatic brain injury. *J Neurotrauma*. 2021;**38**(4):455-63.
23. Czeiter E, Amrein K, Gravesteijn B Y, et al. Blood biomarkers on admission in acute traumatic brain injury: Relations to severity, CT findings and care path in the CENTER-TBI study. *EBioMedicine*. 2020;**56**:102785.
24. Steyerberg E W. Clinical prediction models: a practical approach to development, validation, and updating. New York: Springer Science & Business Media; 2008.
25. Harrell FE J. rms: Regression Modeling Strategies. R package version 3.4-0. 2012.
26. Takala R, Posti J, Runtti H, et al. Glial Fibrillary Acidic Protein and Ubiquitin C-Terminal Hydrolase-L1 as Outcome Predictors in Traumatic Brain Injury. *World Neurosurgery*. 2016;**87**:8-20.
27. Papa L, Akinyi L, Liu M C, et al. Ubiquitin C-terminal hydrolase is a novel biomarker in humans for severe traumatic brain injury. *Crit Care Med*. 2010;**38**(1):138.
28. Haimoto H, Hosoda S, Kato K. Differential distribution of immunoreactive S100-alpha and S100-beta proteins in normal nonnervous human tissues. *Lab Invest*. 1987;**57**(5):489-98.
29. Okonkwo D O, Puffer R C, Puccio A M, et al. Point-of-Care Platform Blood Biomarker Testing of Glial Fibrillary Acidic Protein versus S100 Calcium-Binding Protein B for Prediction of Traumatic Brain Injuries: A Transforming Research and Clinical Knowledge in Traumatic Brain Injury Study. *J Neurotrauma*. 2020;**37**(23):2460-7.
30. Yue J K, Yuh E L, Korley F K, et al. Association between plasma GFAP concentrations and MRI abnormalities in patients with CT-negative traumatic brain injury in the TRACK-TBI cohort: a prospective multicentre study. *Lancet Neurol*. 2019;**18**(10):953-61.
31. Whitehouse D P, Monteiro M, Czeiter E, et al. Relationship of admission blood proteomic biomarkers levels to lesion type and lesion burden in traumatic brain injury: A CENTER-TBI study. *EBioMedicine*. 2022;**75**:103777.
32. Mondello S, Jeromin A, Buki A, et al. Glial neuronal ratio: a novel index for differentiating injury type in patients with severe traumatic brain injury. *J Neurotrauma*. 2012;**29**(6):1096-104.
33. Thelin E, Al Nimer F, Frostell A, et al. A serum protein biomarker panel improves outcome prediction in human traumatic brain injury. *J Neurotrauma*. 2019;**36**(20):2850-62.
34. Mikolić A, van Klaveren D, Groeniger J O, et al. Differences between Men and Women in Treatment and Outcome after Traumatic Brain Injury. *J Neurotrauma*. 2021;**38**(2):235-51.
35. Mikolić A, Groeniger J O, Zeldovich M, et al. Explaining outcome differences between men and women following mild traumatic brain injury. *J Neurotrauma*. 2021;**38**(23):3315-31.
36. Thelin E P, Zeiler F A, Ercole A, et al. Serial sampling of serum protein biomarkers for monitoring human traumatic brain injury dynamics: a systematic review. *Front Neurol*. 2017;**8**:300.



Chapter 8

Discrepancy between Disability and Reported Wellbeing After Traumatic Brain Injury

I.R.A. Retel Helmrich, D. van Klaveren, N. Andelic, H.F. Lingsma, A.I.R. Maas, D. Menon, S. Polinder, C. Røe, E.W. Steyerberg, E. van Veen, L. Wilson, and the CENTER-TBI investigators and participants

Published in: Journal of Neurology, Neurosurgery and Psychiatry

doi.org/10.1136/jnnp-2021-326615

Abstract

Background: Following traumatic brain injury (TBI), the clinical focus is often on disability. However, patients' perceptions of well-being can be discordant with their disability level, referred to as the 'disability paradox'. We aimed to examine the relationship between disability and health-related quality of life (HRQoL) following TBI, while taking variation in personal, injury-related, and environment factors into account.

Methods: We used data from the Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury study. Disability was assessed six months post-injury by the Glasgow Outcome Scale -Extended (GOSE). HRQoL was assessed by the SF-12v2 physical and mental component summary scores (PCS and MCS) and the Quality of Life after Traumatic Brain Injury overall scale (QOLIBRI-OS). We examined mean total and domain HRQoL scores by GOSE. We quantified variance in HRQoL explained by GOSE, personal, injury-related and environment factors with multivariable regression.

Results: Six-month outcome assessments were completed in 2075 patients, of whom 78% had mild TBI (Glasgow Coma Scale 13-15). Patients with severe disability had higher HRQoL than expected on the basis of GOSE alone, particularly after mild TBI. Up to 50% of patients with severe disability, reported HRQoL scores within the normative range. GOSE, personal, injury-related and environment factors explained a limited amount of variance in HRQoL (up to 29%).

Conclusion: Contrary to the idea that discrepancies are unusual, many patients with poor functional outcomes reported well-being that was at or above the boundary considered satisfactory for the normative sample. These findings challenge the idea that satisfactory HRQoL in patients with disability should be described as 'paradoxical' and question common views of what constitutes 'unfavourable' outcome.

Key words: Disability paradox; Traumatic brain injury; Glasgow Outcome Scale Extended; Health-Related Quality of Life

Introduction

Disability relates to a set of difficulties a person may experience when interacting with their social and physical environments (1, 2). Disability is common following moderate and severe traumatic brain injury (TBI), and increasingly recognized as a consequence of mild TBI (3). Following TBI, individuals often experience impairments in different aspects of their life, including physical, social and cognitive limitations, which may impact their well-being (4-8).

Clinical decisions about the management of TBI are often based on the likelihood of the person remaining dependent on others in daily life and therefore having impaired quality of life (9). However, healthy people can overestimate the emotional impact that chronic illness and disability will have on a persons' well-being (10). Furthermore, patients' perceptions of quality of life can be discordant with their objective health status (11). This phenomenon has been described as the 'disability paradox': a discrepancy between severe disability that is observable by others and good quality of life reported by the patient (11). However, critics argue that the 'paradox' depends on the assumption that disability determines well-being (12).

Previous reports consistent with the idea of a 'disability paradox' indicate that patients with severe disability several months following TBI can experience good or excellent well-being (13). A common explanation for this phenomenon is anosognosia: lack of awareness of disability, as a result of neurological impairment (14). In the classic descriptions of anosognosia the individual may, for example, deny having hemiparesis after stroke (15). Anasognosia following TBI might be related to behavioural disorders, frontal lobe syndromes and/or problems with social cognition. Other explanations for the 'disability paradox' include psychological processes such as coping (11), and personal and environment factors (16): for instance, how patients experience disabilities might be affected by employment, pre-injury mental health, and satisfaction with social support (12). This is in agreement with the way in which the relationship between health and disability is described by the World Health Organization: disability is a complex construct involving an interaction between the person and their environment (1).

To date, the discordance between disability level and well-being and the 'disability paradox' have mainly been described as a theoretical construct (11-13, 16), or observed in practice without receiving much attention in empirical studies.

We aimed to examine the relationship between functional outcome, and HRQoL in individuals six months following TBI, while taking variation in personal, injury-

related, and environment factors into account. We hypothesized that the relationship between disability and health-related quality of life (HRQoL) differs by injury severity. Predictors of functional outcome for mild injuries differ from those for more severe injuries (17), suggesting that these subgroups have distinctive characteristics. Further, we hypothesized that contextual factors, including personal, injury-related, and environment factors contribute to explaining variation in HRQoL.

Methods

Study Population

We analyzed data from the Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) study. This is a prospective, multicenter, longitudinal, observational study (18, 19). Data was collected for patients with a clinical diagnosis of TBI and an indication for computed tomography (CT), presenting within 24 hours of injury in one of the 59 participating centers.

Participants were recruited from December 2014 to December 2017 in 18 countries across Europe and Israel. In our study, patients were included if they were aged ≥ 16 years and had available GOSE, and SF-12v2 or QOLIBRI-OS scores at 6 months post-injury.

Data for the CENTER-TBI study was entered by participating sites on the Quesgen e-CRF (Quesgen Systems Inc, USA), hosted on the International Neuroinformatics Coordinating Facility (INCF) platform, and extracted via the INCF Neurobot tool (INCF, Sweden) (database Core 2.1). Informed consent was obtained from all participants according to local and national requirements.

In our study, we included 2075 patients aged 16 years or over who had completed the outcome assessments at 6 months postinjury (online supplemental Figure 1). Patients with missing questionnaires or with proxy responses on HRQoL assessments were excluded.

Outcome assessment

Disability

The Glasgow Outcome Scale -Extended (GOSE) is widely used as a global measure of functional outcome and disability. The scale has eight categories: 1) death, 2) vegetative state, 3) lower severe disability, 4) upper severe disability, 5) lower moderate disability, 6) upper moderate disability, 7) lower good recovery, and 8) upper good

recovery (20) (Supplementary Table 1). In CENTER-TBI, the GOSE was assessed as a structured interview or a questionnaire completed by the patient or a carer. At six months follow-up the format of the assessment was an interview in 79% cases and a questionnaire in 20% (Supplementary Table 2). The respondent for the GOSE was almost always the patient, either alone or with a relative or carer (98%). The GOSE was scored centrally combining the ratings of the interviews and the questionnaires. Missing GOSE values were imputed based on GOSE measurements at other time points if available (21).

Health-Related Quality of Life

We used the Short Form-12 version 2 (SF-12v2) and the QOLIBRI-OS to assess Health-Related Quality of Life (HRQoL). The SF-12v2 is a 12-item patient-reported HRQoL outcome which assesses multiple aspects of health-related functioning and well-being (22). The SF-12v2 comprises eight subscales and two summary scores: physical functioning, role limitations due to physical health, bodily pain, and general health perceptions, are included in the physical component summary (PCS) score, and vitality, social functioning, role limitations due to emotional health, and general mental health, are included in the mental component summary (MCS) score. The PCS emphasizes aspects of functional status, while the MCS incorporates well-being including mental health (23). The norm-based T-scores (standardized to mean 50 and standard deviation (SD) of 10) were calculated for the MCS and PCS. MCS and PCS scores range between 2 (poorest possible HRQoL) and 74 (best possible HRQoL). For the SF-12v2, scores of 45 and above are considered within the normative range for the general population, scores of 40-45 are borderline, and scores below 40 are considered impaired (22).

The QOLIBRI-OS is a 6-item patient-reported HRQoL outcome specifically developed for patients following TBI (24). The QOLIBRI-OS assesses satisfaction with aspects of life (cognition, self, daily life and autonomy, social relationships, current situation and future prospects), and ranges from 0 (poorest possible HRQoL) to 100 (best possible HRQoL). Scores of 61 and above are considered within the normative range, scores of 52-60 are considered borderline, and scores below 52 are considered low or impaired (25).

Contextual factors related to HRQoL following TBI

We studied the following personal and injury-related factors that are relevant to HRQoL: age (26), sex (26), marital status, level of education (27), type of employment pre-injury (27), pre-injury mental health problems (28), pre-injury substance abuse (29), preinjury health status (The American Society of Anesthesiologists - physical status classification system (ASA-PS)), cause of injury, injury severity (29, 30),

the presence of intracranial abnormality, and major extracranial injury (31). Initial injury severity was assessed with the Glasgow Coma Scale (GCS). TBI was considered mild in patients with GCS 13-15, moderate in patients with GCS 9-12, and severe in patients with GCS of 3-8 (19). The definition of 'mild' injury allows that patients may have an abnormality on CT (3). Pre-injury health status was assessed with the American Society of Anesthesiologists - physical status classification system (ASA-PS); patients are categorized as 'normal healthy patient', 'mild systemic disease', 'severe systemic disease', or 'severe systemic disease that is a constant threat to life'. The categories 'severe systemic disease' and 'severe systemic disease that is constant threat to life' were combined. Major extracranial injury was defined as an Abbreviated Injury Scale ≥ 3 regarding the following body regions; face, thoracic/ lumbar spine, thorax/ chest, abdomen/pelvic contents, extremities and pelvic girdle, or external (skin), thus excluding head and neck. Environment factors involve satisfaction with social support, satisfaction with support from the hospital and health services, and satisfaction with support from rehabilitation services six months post-injury (26, 27, 32).

Statistical analyses

Descriptive statistics are presented as medians (Interquartile range [IQR]) or frequencies (percentage).

We examined the relationships between disability and HRQoL in three ways: I) we calculated the percentage of patients by GOSE category that have scores in the normative range on the QOLBRI-OS and MCS; II) we examined differences between the PCS and the MCS as a measure of dissociation between physical and mental HRQoL; and III) we studied the association of the GOSE and HRQoL using linear regression analysis, including personal, injury-related and environment factors.

All analyses were performed separately for individuals with mild (Glasgow Coma Scale (GCS) 13-15), and moderate/severe (GCS 3-12) TBI. The decision to combine patients with moderate and severe TBI was motivated by the sample size (Moderate/ severe TBI N= 466), and the limited number of patients classified as moderate TBI (N=149). To account for differences in the relationship between GOSE and HRQoL following mild, moderate and severe TBI, we performed two-way ANOVA for SF-12 PCS, MCS and QOLIBRI-OS. The relationship between HRQoL following TBI and the GOSE, personal, injury-related and environment factors were analyzed with linear regression analyses. The contribution of predictors to the explained variance (R^2) for each outcome was shown graphically by the partial R^2. Furthermore, the associations between the GOSE and the MCS and QOLIBRI-OS total score, adjusted for personal, injury-related and environment factors were shown graphically.

Analyses are performed with R statistical software (R version 3.6.0). We used the *rms* package to fit the regression models (33).

Results

Study Sample

We included 2075 adult patients who completed the GOSE and SF-12v2 or the QOLIBRI-OS six months post-injury (Supplementary Figure 1). SF-12v2 and QOLIBRI-OS completion rates at follow-up differed by GOSE category (Supplementary Table 3): patients with GOSE 3 had the lowest completion rates (QOLIBRI-OS: 60%, SF-12v2: 65%), while completion rates for patients with higher levels of functioning were higher, and generally above 75%.

The median age was 51 years (IQR = 32-64) (Table 1). Most patients (78%) were classified as having a mild TBI. A third (35%) had major extracranial injury (MEI). 53% was employed, 23% was retired, and 18% unemployed. About 10% had pre-injury mental health problems. Moreover, 40% reported pre-injury comorbid health issues.

Patients following moderate/severe TBI were younger, more often male and more often involved in traffic accidents than patients after mild TBI (Table 1). Rehabilitation was less often received by patients after mild TBI (24%) compared to those after moderate/severe TBI (79%) (Table 1).

Six months after TBI, 186 patients experienced severe disability (9%) (GOSE 3-4), 528 patients experienced moderate disability (25%) (GOSE 5-6), and 1361 (66%) could be classified as having a good recovery (GOSE 7-8) (Table 2).

Health-Related Quality of Life Stratified by Injury Severity and Disability

Overall, SF-12 PCS, MCS and QOLIBRI-OS scores six months following TBI increased with the GOSE (Figure 1). In both severity groups the PCS showed an almost linear relationship with the GOSE. This contrasts with the relationship with the MCS particularly at lower levels of outcome. Specifically, following mild TBI, patients with a GOSE of 3-4, reported higher MCS scores than patients with a GOSE of 5 (mean 42 [95% CI 38-47] and 48 [41-47] for GOSE 3 and 4 versus 38 [36-40] for GOSE 5) (Supplementary Table 4). The results for the QOLIBRI-OS in the mild group mirror those of the MCS (QOLIBRI-OS mean 45 [95% CI 37-54] and 54 [48-60] for GOSE 3 and 4, versus 48 [44-52] for GOSE 5).

Table 1: Patients' demographic and injury characteristics.

Characteristics	All patients* 2075	Mild TBI (GCS 13-15)† 1609	Moderate & Severe TBI (GCS 3-12)† 466	
				p-value‡
Demographics				
Age median [IQR]	51 [32-64]	53 [35-66]	41 [26-55]	<.001
% Male sex	65	63	70	>.05
Marital status N (%)				>.05
Married	1069 (52)	856 (53)	213 (46)	
Missing	117 (6)	87 (5)	30 (6)	
Highest level of Education				<.001
College/Uni degree	548 (26)	453 (28)	95 (20)	
Currently in school/with diploma or degree-oriented program	440 (21)	340 (21)	100 (22)	
None/primary school	246 (12)	202 (13)	44 (9)	
Secondary/high school	620 (30)	463 (29)	157 (34)	
Missing	221 (11)	151 (9)	70 (15)	
Employment type N (%)				<.001
Working	1109 (53)	842 (52)	267 (57)	
Homemaker	29 (1)	25 (2)	4 (1)	
Retired	469 (23)	412 (26)	57 (12)	
Sick leave/Unable to work	49 (2)	36 (2)	13 (3)	
Student	199 (10)	142 (9)	587(12)	
Unemployed	91 (4)	66 (4)	25 (5)	
Missing	129 (6)	86 (5)	43 (9)	
Employment status N (%)				<.001
Yes	1109 (53)	842 (52)	267 (57)	
Retired	469 (23)	412 (26)	57 (12)	
No	368 (18)	269 (17)	99 (21)	
Missing	129 (6)	86 (5)	43 (9)	
ASA Pre-injury health status§ N (%)				
Healthy	1223 (59)	917 (57)	307 (66)	
Mild disease	663 (32)	538 (33)	125 (27)	
Severe disease	175 (8)	146 (9)	29 (6)	
Missing	14 (1)	8 (1)	6 (1)	
Pre-injury substance abuse¶				<.001
Yes	45 (2)	27 (2)	18 (4)	
Missing	19 (1)	8 (1)	11 (2)	

Table 1: Continued

Characteristics	All patients* 2075	Mild TBI (GCS 13-15)† 1609	Moderate & Severe TBI (GCS 3-12)† 466	
				p-value‡
Pre-injury mental health problems N (%)**				<.01
Yes	205 (10)	169 (11)	36 (8)	
Missing	23 (1)	8 (1)	11 (2)	
Injury characteristics				
Cause of injury N (%)				<.001
Road traffic incident	851 (41)	618 (38)	233 (50)	
Incidental fall	908 (44)	751 (47)	157 (34)	
Other non-intentional injury	174 (8)	136 (8)	38 (8)	
Violence/assaults	104 (5)	79 (5)	25 (5)	
Missing	38 (2)	25 (2)	13 (3)	
Major extracranial injury†† N (%)				<.001
Yes	744 (35)	450 (28)	269 (58)	
ISS	13 [8-25]	10 [5-18]	29 [25-41]	<.001
Any intracranial abnormality‡‡ N (%)				<.001
Present	863 (42)	711 (44)	385 (83)	
Missing	116 (6)	80 (5)	36 (8)	

Statistics are for the difference between mild and moderate/severe subgroups. *Patients<16 years of age (n=149), proxy responses (n=251), patients with missing GOSE (n=8) and those that did not complete the HRQoL questionnaires (n=476) were excluded.

†Initial injury severity was assessed with the GCS. TBI was considered mild in patients with GCS 13–15, moderate in patients with GCS 9–12, and severe in patients with GCS of 3–8.

‡P values from ANOVA for continuous and $\chi 2$ statistics for categorical variables. §Preinjury health status was assessed with the American Society of Anesthesiologists—physical status classification system (ASA-PS).

¶Patients with a history of substance abuse disorder prior to the injury.

**Patients with a history of anxiety, depression, sleep disorders, or schizophrenia prior to the injury.

††Patients with an Abbreviated Injury Scale≥3 regarding the all body regions excluding head and neck.

‡‡The presence of intracranial traumatic abnormalities was assessed through
the first CT scan after injury, and indicates whether any of the 12 following abnormalities was present: mass lesion, hematoma, epidural hematoma, acute or subacute subdural hematoma, subdural collection mixed density, contusion, TAI, traumatic subarachnoid haemorrhage, intraventricular haemorrhage, midline shift or cisternal compression.

AIS, Abbreviated Injury Scale; ASA-PS, The American Society of Anesthesiologists- physical status classification system; GCS, Glasgow Coma Scale; ISS, Injury Severity Score; MEI, major extracranial injury; N, number; TBI, traumatic brain injury.

Table 2: Patients' satisfaction with social support, use of rehabilitation services and outcomes 6 months post-injury.

Characteristics	All patients 2075	Mild TBI (GCS 13-15)† 1609	Moderate & Severe TBI (GCS 3-12)† 466	
				p-value‡
Social support 6 months post-injury*				
Satisfaction with social support N (%)				**>.05**
Low	265 (13)	219 (14)	46 (10)	
High	1755 (85)	1347 (84)	408 (88)	
Missing	55 (3)	43 (3)	12 (3)	
Satisfaction with social support from hospital and health services N (%)				**<.05**
Low	202 (10)	172 (11)	30 (6)	
High	1800 (87)	1386 (86)	414 (89)	
Missing	73 (4)	51 (3)	22 (5)	
Satisfaction with social support from rehabilitation services N (%)				**<0.001**
Low	404 (20)	322 (20)	82 (18)	
High	1473 (71)	1108 (69)	365 (78)	
Missing	198 (10)	179 (11)	19 (4)	
Type of rehabilitation services received N (%)				**<0.001**
No rehabilitation	1290 (64)	1194 (76)	96 (21)	
In-patient/residential	408 (20)	150 (10)	258 (57)	
Out-patient/community	234 (16)	221 (14)	98 (22)	
Six-month functional outcome				
Glasgow Outcome Scale -Extended 6 months post-injury				**<0.001**
Lower severe disability	77 (4)	35 (2)	42 (9)	
Upper severe disability	109 (5)	58 (4)	51 (11)	
Lower moderate disability	225 (11)	116 (7)	109 (23)	
Upper moderate disability	303 (15)	203 (13)	100 (22)	
Lower good recovery	491 (24)	417 (26)	74 (16)	
Upper good recovery	870 (42)	780 (49)	90 (19)	

‡Statistics are for the difference between mild and moderate/ severe subgroups. *Satisfaction with social support in general, from hospital and health services and from rehabilitation services were assessed 6-month post-injury. The response categories 'not at all', 'slightly' and 'moderately' were classified as 'low' satisfaction with social support, and the response categories 'quite', and 'very' were classified as 'high' satisfaction with social support.

†Initial injury severity was assessed with the GCS. TBI was considered mild in patients with GCS 13-15, moderate in patients with GCS 9-12, and severe in patients with GCS of 3-8. GCS, Glasgow Coma Scale; TBI, traumatic brain injury.

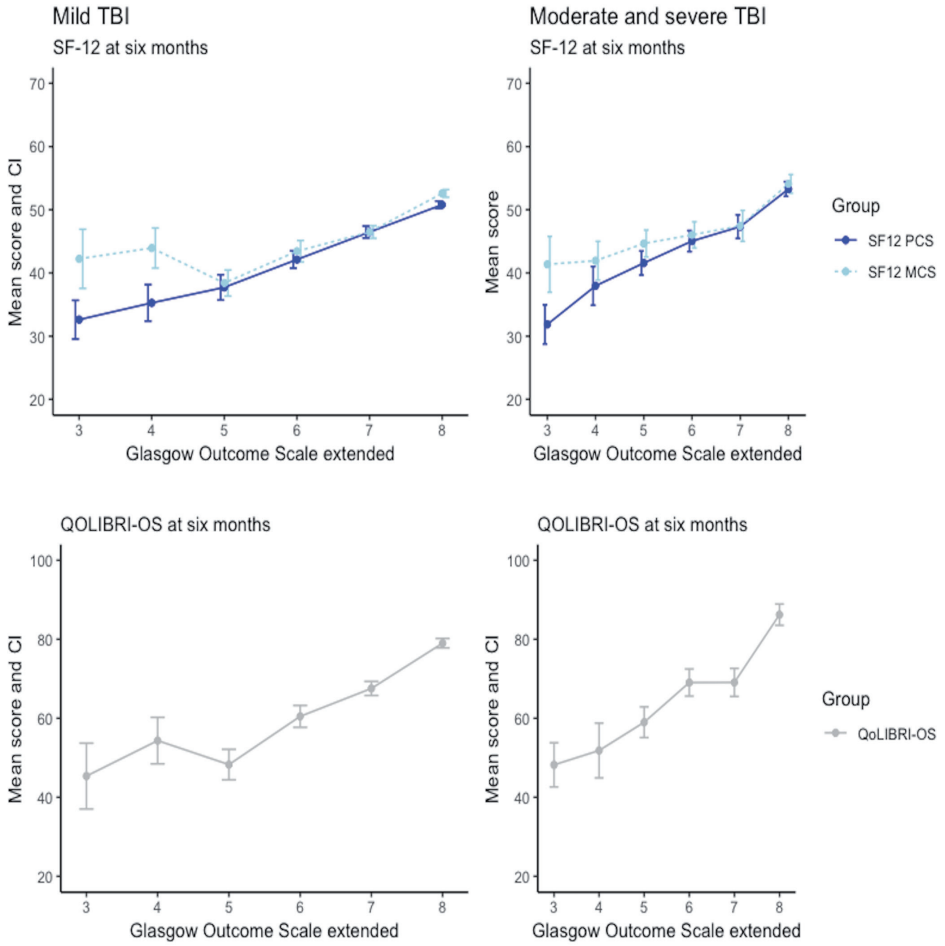


Figure 1: Plots of the SF-12v2 physical and mental health component summary scores (top) and the Quality of Life after Traumatic Brain Injury OS (bottom) by time point for mild (left) and moderate and severe TBI (right). The points are means and error bars are 95% confidence intervals.

Based on the ANOVA, there were significant differences on all HRQoL outcomes by GCS and GOSE. The interaction between GCS and GOSE was significant for MCS ($F=4.137$, df 1, $p<0.01$) but not for QOLIBRI-OS ($F=0.55$, df 1, $p=0.46$) and PCS ($F=0.098$, df 1, $p=0.755$).

For patients following mild TBI, the lowest mean score on the MCS was reported for those with lower moderate disability (GOSE 5) (Supplementary Table 4) (mean 38 [95% CIs 36,40] compared to >42 [95% CI 38,47]). Following moderate and severe TBI, patients with lower severe disabilities (GOSE 3) reported the lowest mean MCS scores (mean 41 [95% CIs 39,45] compared to >42 [39,45]). For four SF-12

subscales, namely 'bodily pain', 'general health', 'role emotional' and 'mental health', and the QOLIBRI-OS items 'how your brain is working', 'feelings and emotions', 'social life' and 'current situation and future prospects' individuals following mild TBI with lower moderate disability (GOSE 5) scored lower than patients with upper severe disability (GOSE 4) (Supplementary Table 4). The median score on the PCS increased with recovery level on the GOSE. Similarly, the MCS and QOLIBRI-OS scores generally increased with recovery level on the GOSE, but in patients following mild TBI HRQoL scores did not increase from GOSE 3 to 5.

Discordance between Disability and Health-Related Quality of Life: The 'Disability Paradox'

Similar to the trends depicted in Figure 1, a higher percentage of patients following mild TBI with upper severe disability (GOSE 4) reported HRQoL scores within the normative range than patients with lower moderate disability (GOSE 5) (MCS 50% versus 30%; QOLIBRI-OS 42% versus 35%) (Table 3).

Following mild TBI, up to half of the individuals with severe disability (N=93) had normative QOLIBRI-OS and MCS scores six months following TBI (QOLIBRI-OS 29% and 42%, MCS 40% and 50%) (Table 3). In contrast, a smaller proportion of individuals with severe disabilities had normative PCS scores (11% and 24%). Following moderate and severe TBI, more than a third of individuals with severe disability (N=88) had normative QOLIBRI-OS and MCS scores six months following TBI (QOLIBRI-OS 40% and 37%; MCS 26% and 13%) (Table 3).

Second, we calculated the difference between the PCS and the MCS by recovery level on the GOSE. Patients with severe disability had larger mean differences between the MCS and PCS compared to patients with moderate disability and good recovery (Table 3). The difference for patients with severe disability was nearly 10 points, which is equivalent to one SD at the population level. This implies that severely disabled individuals have a substantial discordance between the PCS and MCS.

The Relation between Disability, Contextual Factors and HRQoL

The GOSE had the largest contribution to explaining the variance of HRQoL compared to personal, injury-related and environment factors (Figure 2).
While adjusting for personal, injury-related and environment factors in patients with mild TBI, estimates of the MCS and QOLIBRI-OS for patients with GOSE 5 were lower than estimates for patients with GOSE 3-4 (Figure 3). Thus, personal and injury-related factors (including MEI) and satisfaction with social support did not explain the discrepancies between GOSE and HRQoL in patients following mild TBI.

Table 3: Number and percentage of patients with HRQoL scores within the normative range six months post-injury, and mean differences between the MCS and PCS.

Mild TBI (N=1609)				
GOSE	**QOLIBRI-OS >61**	**SF-12 MCS >45**	**SF-12 PCS >45**	**Mean MCS – PCS (SD)**
3 (N=35)	9 (29)	14 (40)	4 (11)	9.62 (15.58)
4 (N=58)	24 (42)	29 (50)	14 (24)	8.68 (17.34)
5 (N=116)	41 (35)	35 (30)	33 (29)	0.68 (17.20)
6 (N=203)	109 (54)	93 (46)	89 (44)	1.31 (16.21)
7 (N=417)	281 (68)	244 (59)	259 (62)	0.00 (14.44)
8 (N=780)	671 (88)	631 (82)	605 (79)	1.79 (11.42)
Moderate and severe TBI (N=466)				
GOSE	**QOLIBRI-OS >61**	**SF-12 MCS >45**	**SF-12 PCS >45**	**Mean MCS – PCS (SD)**
3 (N=42)	13 (32)	16 (38)	5 (12)	9.50 (20.78)
4 (N=51)	19 (38)	20 (41)	13 (27)	3.97 (15.82)
5 (N=109)	56 (52)	57 (53)	46 (43)	3.09 (14.97)
6 (N=100)	71 (72)	57 (58)	52 (53)	0.97 (12.78)
7 (N=74)	52 (72)	44 (59)	45 (61)	0.12 (14.51)
8 (N=90)	84 (95)	76 (85)	83 (93)	0.82 (9.20)

The data are shown by Glasgow Outcome Scale-Extended categories separately for mild and moderate/severe TBI.
GOSE, Glasgow Outcome Scale-Extended; HRQoL, health-related quality of life; MCS, mental component summary; PCS, physical component summary; QOLIBRI-OS, Quality of Life after Traumatic Brain Injury overall scale; TBI, traumatic brain injury.

Besides the GOSE, satisfaction with social support six months following TBI contributed to explaining the variance in HRQoL (Figure 2). Independent of initial injury severity based on GCS, patients with lower moderate disabilities (GOSE 5) were least satisfied with the support they received from rehabilitation (67% vs. ≥70% for mild and 75% vs. ≥79% for moderate and severe TBI) (Supplementary Table 5). As expected, patients with moderate disability (GOSE 5-6) were less likely than patients with severe disability (GOSE 3-4) to receive rehabilitation six months post-injury (54-62% vs. <51% for mild TBI respectively; 9-19% vs. <5% for moderate/severe TBI respectively) (Supplementary Table 6).

Up to 29% (mild) and 28% (moderate and severe) of the variance in QOLIBRI-OS and 21% (mild) and 11% (moderate and severe) of the variance in MCS was explained by the combination of GOSE, personal and injury related characteristics, and satisfaction with social support at 6 months post-injury.

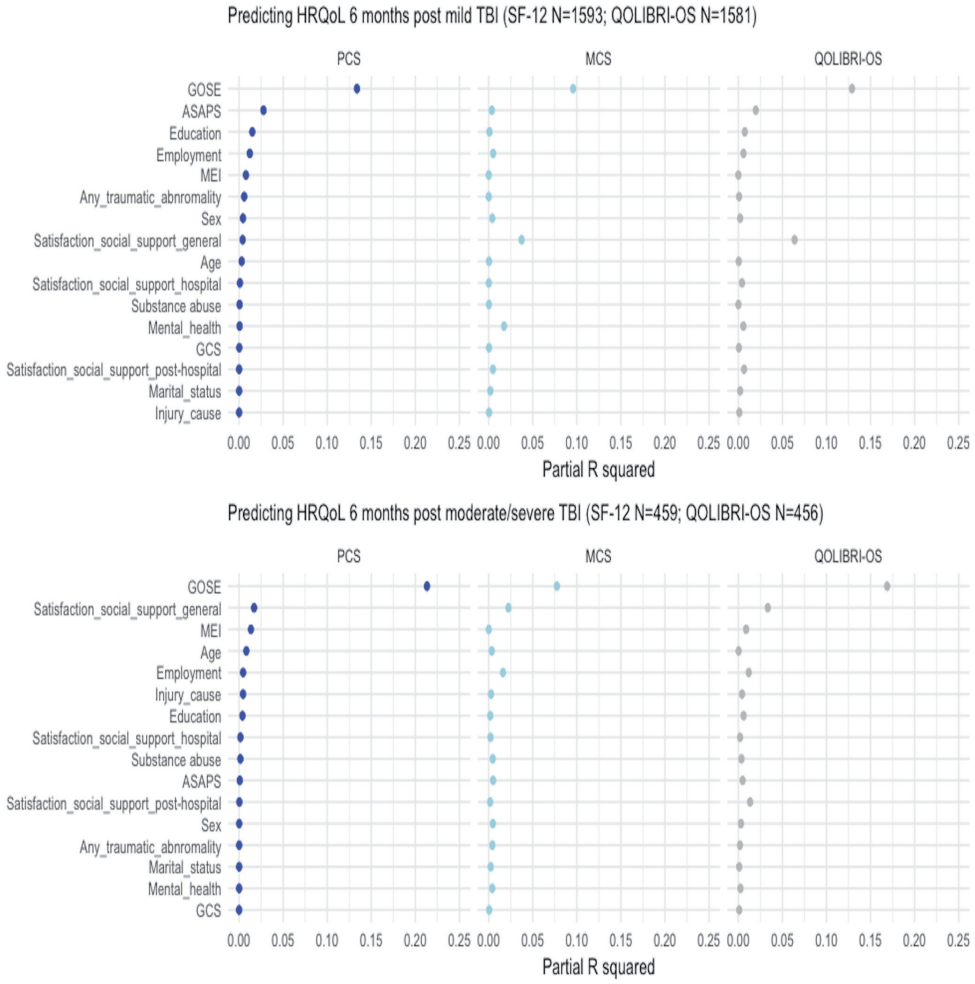


Figure 2: Contribution of predictors to explained variance (partial R^2) of the models for SF-12 PCS (left), SF-12 MCS (middle) and QOLIBRI-OS (right). The partial R^2 is calculated as follows: Total R2 of multivariable model – R^2 multivariable model without individual predictor: Total R2 of multivariable model without individual predictor = Partial R^2.

Discussion

We examined the relationship between disability assessed with the GOSE and HRQoL measured with the SF12v2 MCS and QOLIBRI-OS six months following TBI in the CENTER-TBI study. Following mild TBI, patients can have poor functional outcomes, which is consistent with growing awareness that patients classified as mild by GCS criteria can suffer a range of problems (3). In patients following mild TBI, HRQoL did not decrease linearly with greater disability. Specifically, patients with

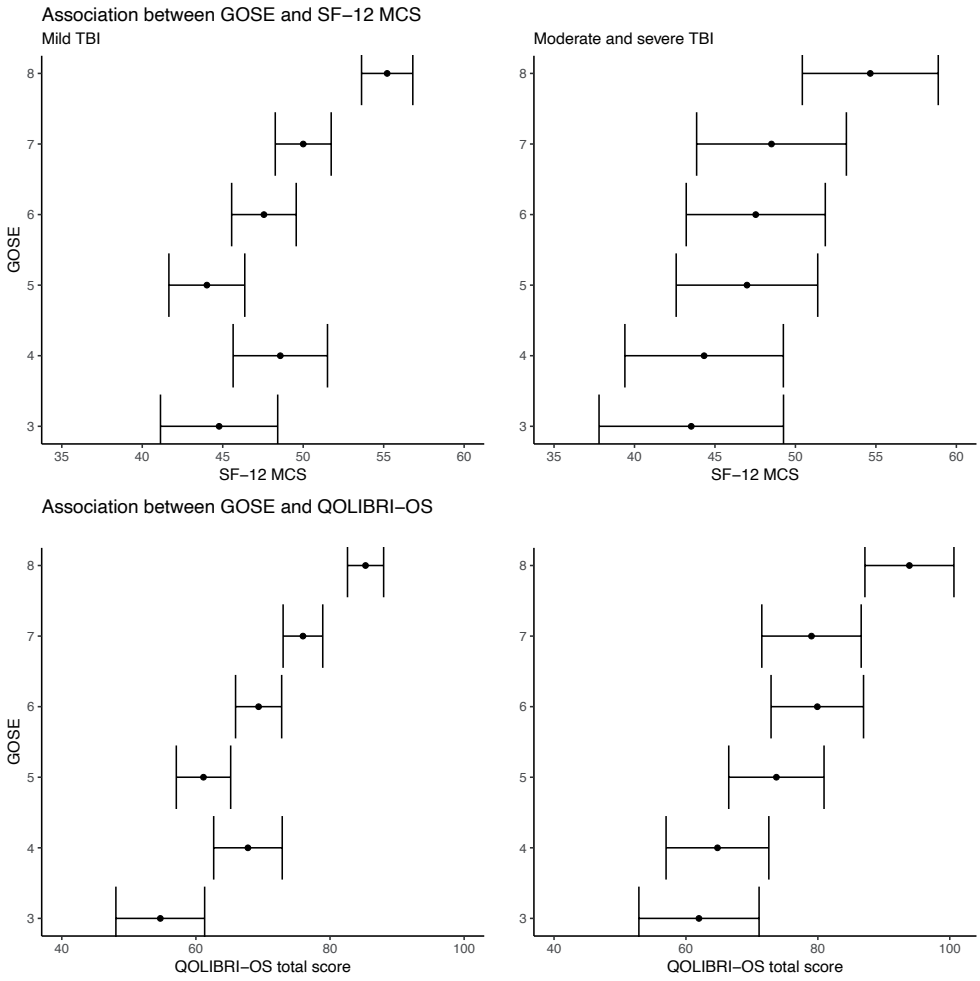


Figure 3: Adjusted association between the GOSE and the SF-12 MCS (upper) and QOLIBRI-OS (lower) for the 'average' patient (Sex=Male; Age=51; Marital status=Married; Highest level of education=Second/high school, Type of employment=Working, Pre-injury mental health problems=No, Pre-injury substance abuse=No; Pre-injury health status (ASAPS)=Healthy; Injury severity (GCS)=15; Cause of injury=Incidental fall; Major extracranial injury=No; Presence of intracranial traumatic abnormalities=Present; Satisfaction with social support=High; Satisfaction with support from the hospital and health services=High; Satisfaction with support from rehabilitation services=High.)

severe disability on the GOSE reported higher MCS and QOLIBRI-OS scores than patients with moderate disabilities. Furthermore, between a third and half of patients with severe disabilities reported HRQoL within the normative range. Our study therefore confirms that individuals' perceptions of aspects of well-being and mental health are often discordant with their objective functioning following TBI.

Our findings are consistent with prior studies describing good or excellent well-being and quality of life following TBI (13, 34). Furthermore, our findings imply that satisfactory HRQoL in patients with disabilities is not a 'paradox', since individuals frequently report HRQoL within the normative range following TBI. Discordance between disability and HRQoL should therefore be regarded as a characteristic of TBI outcomes. Characterizing HRQoL within the normative range despite severe disability as a 'paradox' has serious shortcomings, as it implies that patients with severe disability cannot normally experience satisfactory HRQoL (13). Discrepancies between disability and HRQoL have been observed in prior studies in TBI (35-37). To provide quantification of the discordance between physical and mental health, we therefore examined the difference between the SF-12v2 MCS and PCS. Similarly, patients with severe disabilities had the largest discordance between the MCS and PCS.

It is often suggested that patients with severe disability after TBI have lower self-awareness or anosognosia and a bias towards responding positively on outcome assessments (14, 38). This might explain, for example, positive ratings on the QOLIBRI-OS among more disabled individuals. Although impairments of self-awareness can be present after TBI, Sasse et al. (2013) (38) found that the influence on reported HRQoL was weak. Furthermore, in our study patients showed awareness of functional limitations on the PCS, and nonetheless gave positive ratings of HRQoL on the MCS. The dissociation observed for two summary components of the same self-reported outcome, appears to rule out an account in terms of global lack of awareness. That is, the discrepancy means that patients were not simply responding with positive ratings across all items, in a way that one might expect if the person had profound loss of awareness, and would imply that the responses were meaningless. Nonetheless, more selective limitations of awareness may play a role, for example, lack of awareness of cognitive impairment or mental health problems (39). Alterations in awareness may thus contribute to discrepancies, and this deserves further study.

Besides deficits in general functional outcome cognitive impairments are likely to play a role in perception of wellbeing after TBI. A prior CENTER-TBI study found that MCS scores generally decreased with increasing cognitive impairment, and apparently reached a plateau in the severely disabled group (37). Cognition may play a number of different roles, and it is possible that cognitive impairment has some protective role in the most severely disabled patients (38). Data on cognitive impairments from severely disabled patients (GOSE 3-4) was too limited to allow us to examine this issue, and it remains an important topic for future research. Furthermore, prevalence of cognitive impairment is likely to be a key difference between the two severity groups that we studied (40). Notably, discrepancies were observed in both groups, and were not more pronounced in more severely injured patients than the group with mild injuries.

Following TBI, disability is often assessed using functional outcome scales such as the GOSE. The SF-12v2 and QOLIBRI-OS also try to capture the patient's subjective experience of their well-being in daily life (7). Decisions about the management of TBI are sometimes founded on the likelihood of the person remaining dependent, under the assumption this will lead to impaired HRQoL, and therefore classified as an "unfavourable" outcome. In contrast, our findings showed that HRQoL does not simply follow functioning. Our results thus represent a strong caution against adopting a negative view of potential HRQoL and well-being in patients who are severely disabled based on the GOSE.

We found the lowest levels of HRQoL in patients with moderate disability. Similarly, in a study of patients after severe TBI, Mailhan and colleagues (35) found the lowest level of life satisfaction in patients with moderate disability, which they attribute to lower satisfaction in the domains social and family life. Our results also indicated that patients with moderate disability might be less satisfied with their social support and were less likely to receive rehabilitation. As expected, access to rehabilitation services is more likely among patients following moderate and severe TBI and patients with severe disability compared to their respectively less severely injured and disabled counterparts (41). A previous study showed that patients after less severe TBI report more unmet rehabilitation needs than those following severe TBI (42). Patients with moderate disability are independent, but are unable to return to work, and experience activity limitations (20, 43). Although these patients experience activity limitations, the injury and its consequences might be less visible to their environment compared to patients with severe disability, which could result in less (social) support. To be unable to work and be isolated in the community, may well be worse for well-being than being dependent in daily life but well-supported by others. Our results thus suggest that patients with lower moderate disability living in the community should be a particular target for additional support, rehabilitation and interventions. Furthermore, as perceptions of well-being are often discordant with disability level following TBI, recovery should be based on a multidimensional outcome measure including disability on multiple domains including physical, cognitive and social disabilities and HRQoL.

The disability 'paradox' has more than once been described as good well-being 'against all odds', implying that physical disabilities are the main driver of well-being (11). However, we found that personal, injury-related and environment factors explain a proportion of HRQoL outcomes beyond functional outcome. Nevertheless, only up to 29% of the variance in QOLIBRI-OS and 21% of the variance in MCS was explained by GOSE, personal and injury-related characteristics, and satisfaction with social support. Furthermore, personal, injury-related and environment factors did not

explain the discrepancies between the GOSE and HRQoL in patients following mild TBI. Injury-related factors included major extracranial injury, which is known to have a dominant effect on outcome after mild TBI (31). As the majority of variance remained unexplained, future research should consider the effect of coping, resilience, adaptation, and cognitive impairments on HRQoL following TBI. To further explain HRQoL in patients following TBI, it is crucial to involve patients and their relatives. The focus on mixed methods research, combining quantitative and qualitative methods, might help to elucidate patients' perceptions of satisfactory quality of life following TBI.

Strengths

The strengths of this study include the use of data from a large international, multicenter observational study. Consequently, we made use of a standardized collection of data, and a well described and contemporary cohort of patients. Furthermore, the CENTER-TBI study enrolled patients following mild, moderate and severe TBI, which enabled us to compare HRQoL outcomes by injury severity. Moreover, to describe HRQoL following TBI we used generic (SF-12v2) and disease-specific (QOLIBRI-OS) instruments. The combination of generic and disease-specific instruments has been recommended to more fully capture patients' HRQoL following TBI (7). Furthermore, we demonstrated the dissociation between physical and mental HRQoL using two scales from the same instrument, arguing against the idea that the discordance results from compromised self-awareness following TBI (12, 13).

Limitations

Several limitations of our study have to be considered. Patients with lower functional outcome on the GOSE and lower HRQoL were less likely to complete the questionnaires, potentially resulting in a response bias. Furthermore, the SF-12v2 is not suitable for patients with major cognitive impairment or language difficulties. Thus, the most severely disabled patients, who are likely to be among the most distressed, are not represented in the data. Taken together, the results of our study can only be generalized to patients who are able to respond to follow-up questionnaires, implying that our findings will not apply to a subgroup of patients with profound disability, severe neurological problems, or language difficulties.

Conclusion

Our study confirms that patients' perceptions of HRQoL are often discordant with level of disability following TBI. Contrary to the idea that discrepancies are unusual, many patients with poor functional outcomes report satisfactory wellbeing, particularly in patients after mild injury. These results indicate that the effects of 'mild' TBI can be extensive and warrant further investigation. Furthermore, the findings chal-

lenge the idea that good quality of life in patients with disability should be described as 'paradoxical', and question common views of what constitutes "unfavourable" outcome.

Funding

The research leading to these results was supported by the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 602150 (CENTER-TBI). Additional funding was obtained from the Hannelore Kohl Stiftung (Germany), from OneMind (USA) and from Integra LifeSciences Corporation (USA).

Ethical approval

The CENTER-TBI study (EC grant 602150) has been conducted in accordance with all relevant laws of the EU if directly applicable or of direct effect and all relevant laws of the country where the Recruiting sites were located, including but not limited to, the relevant privacy and data protection laws and regulations (the "Privacy Law"), the relevant laws and regulations on the use of human materials, and all relevant guidance relating to clinical studies from time to time in force including, but not limited to, the ICH Harmonised Tripartite Guideline for Good Clinical Practice (CPMP/ICH/135/95) ("ICH GCP") and the World Medical Association Declaration of Helsinki entitled "Ethical Principles for Medical Research Involving Human Subjects". Informed Consent by the patients and/or the legal representative/next of kin was obtained, accordingly to the local legislations, for all patients recruited in the Core Dataset of CENTER-TBI and documented in the e-CRF. Ethical approval was obtained for each recruiting site. The list of sites, Ethical Committees, approval numbers and approval dates can be found on the website: https://www.center-tbi.eu/project/ethical-approval

Acknowledgements

We are grateful to all patients and investigators that participated in the CENTER-TBI study.

Conflict of interest

The authors declare that there is no conflict of interest.

Supplementary material

Supplementary material is available at: https://jnnp.bmj.com/content/93/7/785

References

1. World Health O. International Classification of Functioning, Disability, and Health: ICF: World Health Organization; 2001.
2. Pierce CA, Hanks RA. Life satisfaction after traumatic brain injury and the World Health Organization model of disability. American Journal of Physical Medicine & Rehabilitation. 2006;85(11):889-98.
3. Nelson LD, Temkin NR, Dikmen S, Barber J, Giacino JT, Yuh E, et al. Recovery after mild traumatic brain injury in patients presenting to US Level I trauma centers: A Transforming Research and Clinical Knowledge in Traumatic Brain Injury (TRACK-TBI) study. JAMA neurology. 2019;76(9):1049-59.
4. Stocchetti N, Zanier ER. Chronic impact of traumatic brain injury on outcome and quality of life: a narrative review. Critical Care. 2016;20(1):1-10.
5. 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.
6. Scholten AC, Haagsma JA, Andriessen T, Vos PE, Steyerberg EW, Van Beeck EF, et al. Health-related quality of life after mild, moderate and severe traumatic brain injury: patterns and predictors of suboptimal functioning during the first year after injury. Injury. 2015;46(4):616-24.
7. Polinder S, Haagsma JA, van Klaveren D, Steyerberg EW, Van Beeck EF. Health-related quality of life after TBI: a systematic review of study design, instruments, measurement properties, and outcome. Popul Health Metr. 2015;13(1):4.
8. Andelic N, Sigurdardottir S, Schanke A-K, Sandvik L, Sveen U, Roe C. Disability, physical health and mental health 1 year after traumatic brain injury. Disability and rehabilitation. 2010;32(13):1122-31.
9. Gillett GR, Honeybul S, Ho KM, Lind CRP. Neurotrauma and the RUB: where tragedy meets ethics and science. Journal of medical ethics. 2010;36(12):727-30.
10. Ubel PA, Loewenstein G, Schwarz N, Smith D. Misimagining the unimaginable: the disability paradox and health care decision making. Health Psychology. 2005;24(4S):S57.
11. Albrecht GL, Devlieger PJ. The disability paradox: high quality of life against all odds. Social science & medicine. 1999;48(8):977-88.
12. Koch T. The illusion of paradox: commentary on Albrecht, GL and Devlieger, PJ (1998). The disability paradox: high quality of life against all odds. Social Science & Medicine 48, 977-988. Social science & medicine (1982). 2000;50(6):757-9.
13. Honeybul S, Gillett GR, Ho KM, Janzen C, Kruger K. Is life worth living? Decompressive craniectomy and the disability paradox. Journal of neurosurgery. 2016;125(3):775-8.
14. Prigatano GP. Anosognosia after traumatic brain injury. The study of anosognosia. 2010:229-54.
15. Vallar G, Ronchi R. Anosognosia for motor and sensory deficits after unilateral brain damage: A review. Restorative neurology and neuroscience. 2006;24(4-6):247-57.
16. Fellinghauer B, Reinhardt JD, Stucki G, Bickenbach J. Explaining the disability paradox: a cross-sectional analysis of the Swiss general population. BMC Public Health. 2012;12(1):655.
17. Lingsma HF, Yue JK, Maas AIR, Steyerberg EW, Manley GT, including T-TI, et al. Outcome prediction after mild and complicated mild traumatic brain injury: external validation of existing models and identification of new predictors using the TRACK-TBI pilot study. Journal of neurotrauma. 2015;32(2):83-94.
18. Maas AIR, 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. J Neurosurg. 2014;76(1):67-80.

19. Steyerberg EW, Wiegers E, Sewalt C, Buki A, Citerio G, De Keyser V, et al. Case-mix, care pathways, and outcomes in patients with traumatic brain injury in CENTER-TBI: a European prospective, multicentre, longitudinal, cohort study. The Lancet Neurology. 2019;18(10):923-34.
20. Wilson JTL, Pettigrew LEL, Teasdale GM. Structured interviews for the Glasgow Outcome Scale and the extended Glasgow Outcome Scale: guidelines for their use. Journal of neurotrauma. 1998;15(8):573-85.
21. Kunzmann K, Wernisch L, Richardson S, Steyerberg EW, Lingsma H, Ercole A, et al. Imputation of ordinal outcomes: a comparison of approaches in traumatic brain injury. Journal of neurotrauma. 2021;38(4):455-63.
22. Ware JE, Kosinski M, Bjorner JB, Turner-Bowker DM, Gandek B, Maruish ME. User's manual for the SF-36v2 Health Survey. 2nd edn Lincoln. RI: QualityMetric Incorporated. 2007.
23. Hays RD, Hahn H, Marshall G. Use of the SF-36 and other health-related quality of life measures to assess persons with disabilities. Archives of physical medicine and rehabilitation. 2002;83:S4-S9.
24. Von Steinbuechel N, Wilson L, Gibbons H, Muehlan H, Schmidt H, Schmidt S, et al. QOLIBRI overall scale: a brief index of health-related quality of life after traumatic brain injury. Journal of Neurology, Neurosurgery & Psychiatry. 2012;83(11):1041-7.
25. Wilson L, Marsden-Loftus I, Koskinen S, Bakx W, Bullinger M, Formisano R, et al. Interpreting quality of life after brain injury scores: cross-walk with the short form-36. Journal of neurotrauma. 2017;34(1):59-65.
26. McCarthy ML, Dikmen SS, Langlois JA, Selassie AW, Gu JK, Horner MD. Self-reported psychosocial health among adults with traumatic brain injury. Arch Phys Med Rehabil. 2006;87(7):953-61.
27. Warren L, Wrigley JM, Yoels WC, Fine PR. Factors associated with life satisfaction among a sample of persons with neurotrauma. J Rehabil Res Dev. 1996;33:404-8.
28. Stein MB, Jain S, Giacino JT, Levin H, Dikmen S, Nelson LD, et al. Risk of posttraumatic stress disorder and major depression in civilian patients after mild traumatic brain injury: a TRACK-TBI study. JAMA psychiatry. 2019;76(3):249-58.
29. Corrigan JD, Bogner JA, Mysiw WJ, Clinchot D, Fugate L. Life satisfaction after traumatic brain injury. J Head Trauma Rehabil. 2001;16(6):543-55.
30. Teasdale TW, Engberg AW. Subjective well-being and quality of life following traumatic brain injury in adults: A long-term population-based follow-up. Brain inj. 2005;19(12):1041-8.
31. Carroll EL, Manktelow AE, Outtrim JG, Chatfield D, Forsyth F, Hutchinson PJA, et al. Influence of concomitant extracranial injury on functional and cognitive recovery from mild versus moderate to severe traumatic brain injury. Journal of head trauma rehabilitation. 2020;35(6):E513-E23.
32. Steadman-Pare D, Colantonio A, Ratcliff G, Chase S, Vernich L. Factors associated with perceived quality of life many years after traumatic brain injury. J Head Trauma Rehabil. 2001;16(4):330-42.
33. Harrell FE J. rms: Regression Modeling Strategies. R package version 3.4-0. 2012.
34. Grauwmeijer E, Heijenbrok-Kal MH, Ribbers GM. Health-related quality of life 3 years after moderate to severe traumatic brain injury: a prospective cohort study. Archives of Physical Medicine and Rehabilitation. 2014;95(7):1268-76.
35. Mailhan L, Azouvi P, Dazord A. Life satisfaction and disability after severe traumatic brain injury. Brain Injury. 2005;19(4):227-38.
36. Truelle J-L, Koskinen S, Hawthorne G, Sarajuuri J, Formisano R, Von Wild K, et al. Quality of life after traumatic brain injury: the clinical use of the QOLIBRI, a novel disease-specific instrument. Brain Injury. 2010;24(11):1272-91.

37. Wilson L, Horton L, Kunzmann K, Sahakian BJ, Newcombe VFJ, Stamatakis EA, et al. Understanding the relationship between cognitive performance and function in daily life after traumatic brain injury. Journal of Neurology, Neurosurgery & Psychiatry. 2020.
38. Sasse N, Gibbons H, Wilson L, Martinez-Olivera R, Schmidt H, Hasselhorn M, et al. Self-awareness and health-related quality of life after traumatic brain injury. The Journal of head trauma rehabilitation. 2013;28(6):464-72.
39. Cavallo MM, Kay T, Ezrachi O. Problems and changes after traumatic brain injury: Differing perceptions within and between families. Brain injury. 1992;6(4):327-35.
40. Schretlen DJ, Shapiro AM. A quantitative review of the effects of traumatic brain injury on cognitive functioning. International review of psychiatry. 2003;15(4):341-9.
41. Jacob L, Cogné M, Tenovuo O, Røe C, Andelic N, Majdan M, et al. Predictors of access to rehabilitation in the year following traumatic brain injury: a European prospective and multicenter study. Neurorehabilitation and neural repair. 2020;34(9):814-30.
42. Andelic N, Soberg HL, Berntsen S, Sigurdardottir S, Roe C. Self-perceived health care needs and delivery of health care services 5 years after moderate-to-severe traumatic brain injury. Pm&r. 2014;6(11):1013-21.
43. Jennett B, Bond M. Assessment of outcome after severe brain damage: a practical scale. The Lancet. 1975;305(7905):480-4.

9

Chapter 9

Development of Prognostic Models for Health-Related Quality of Life following Traumatic Brain injury

I.R.A. Retel Helmrich, D. van Klaveren, S.A. Dijkland, H.F. Lingsma, S. Polinder, L. Wilson, N. von Steinbuechel, J. van der Naalt, A.I.R. Maas, E.W. Steyerberg, and the CENTER-TBI collaborators

Published in: Quality of Life Research

Abstract

Background: Traumatic brain injury (TBI) is a leading cause of impairments affecting Health-Related Quality of Life (HRQoL). We aimed to identify predictors of, and develop prognostic models for HRQoL following TBI.

Methods: We used data from the Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) Core study, including patients with a clinical diagnosis of TBI and an indication for computed tomography, presenting within 24 hours of injury. The primary outcome measures were the SF-36v2 physical (PCS) and mental (MCS) health component summary scores and the Quality of Life after Traumatic Brain Injury (QOLIBRI) total score six months post-injury. We considered sixteen patient and injury characteristics in linear regression analyses. Model performance was expressed as proportion of variance explained (R^2), and corrected for optimism with bootstrap procedures.

Results: 2666 adult patients completed the HRQoL questionnaires. Most were mild TBI patients (74%). The strongest predictors for PCS were Glasgow Coma Scale, major extracranial injury and pre-injury health status, while MCS and QOLIBRI were mainly related to pre-injury mental health problems, level of education and type of employment. R^2 of the full models was 19% for PCS, 9% for MCS and 13% for the QOLIBRI. In a subset of predominantly patients following mild TBI, including 2-week HRQoL assessment (N=436) improved model performance substantially (R^2 PCS 15% to 37%, MCS 12% to 36%, and QOLIBRI 10% to 48%).

Conclusion: Medical and injury related characteristics are of greatest importance for the prediction of PCS, whereas patient related characteristics are more important for MCS and the QOLIBRI following TBI.

Key words: Prognostic model research; Traumatic Brain Injury; Health Related Quality of Life; SF-36; QOLIBRI

Background

Traumatic brain injury (TBI) is a leading cause of long-term impairments in functional, physical, mental, cognitive, and social domains.[1] These impairments are not restricted to severe cases, but are also known to occur frequently after moderate and mild TBI [2, 3]. Impairments can, for instance, be assessed using functional outcome scales (e.g. Glasgow Outcome Scale (Extended) (GOS(-E)).[4] Although functional measurement scales are useful to portray functional problems, they do not capture the patient's subjective experience of their sequelae and wellbeing in daily life.[5]

Therefore, there has been growing interest in health-related quality of life (HRQoL) in TBI research. HRQoL focuses on an individuals' perception of how a disease and its treatments affect the physical, mental and social aspects of their life.[6] Previous studies confirmed that long-term impairments following TBI affect (HR)QoL.[7-16] To assess HRQoL two types of instruments are available; generic and condition-specific instruments.[6] Generic instruments, such as the Short Form-36 (SF-36),[17] do not take into account diseases or particular conditions and allow comparison with healthy individuals, as well as various health states or conditions. It has been argued that generic HRQoL instruments may not be sensitive enough to detect key issues in TBI, such as cognitive dysfunctions and psychological issues.[6, 18] A TBI-specific instrument, such as the Quality of Life after Traumatic Brain Injury (QOLIBRI),[19, 20] may therefore be complementary.

Outcomes following TBI depend on patient and injury characteristics, mechanisms of trauma, patient response, the social environment, and the quality of care provided.[21-23] Prognostic models predict the outcome of a patient based on characteristics at presentation, and are important to help clinicians provide reliable information to patients and relatives.[24] It would be particularly helpful if poor HRQoL outcomes could be anticipated as these predictions could support clinicians in identifying patients who might benefit from close follow-up and early interventions. Although high-quality and well-validated models exist to predict functional outcomes following moderate and severe TBI,[25] prognostic models for HRQoL following TBI have not been developed yet. Furthermore, efforts have been made to identify predictors of HRQoL following TBI,[11, 12, 14, 26-31] but they are dispersed throughout the literature. Therefore, we aimed to identify predictors of, and develop prognostic models for HRQoL following mild, moderate and severe TBI.

Methods

Study Population

We analyzed patients included in the Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI, version Core 2.1) study. This is a prospective, multicenter, longitudinal, observational study.[32, 33] Data was collected for patients with a clinical diagnosis of TBI and an indication for computed tomography (CT), presenting within 24 hours of injury in one of the 58 participating centers. Participants were recruited from December 2014 to December 2017 in 18 countries across Europe and Israel.

For model development, patients were included if they were aged ≥ 18 years and completed the SF-36v2 or QOLIBRI at six months post-injury.

Data for the CENTER-TBI study was entered on the Quesgen e-CRF (Quesgen Systems Inc, USA), hosted on the International Neuroinformatics Coordinating Facility (INCF) platform and extracted via the INCF Neurobot tool (INCF, Sweden). Informed consent was obtained from all participants according to local and national requirements.

Candidate Predictors

Candidate predictors of HRQoL following TBI were selected based on literature, and included initial severity (Glasgow Come Scale),[12, 26, 27] age,[28] sex,[11, 29, 34] socioeconomic status,[30] social support,[29-31, 34] pre-injury substance abuse,[26, 34] and pre-injury mental health problems (e.g. anxiety, depression).[29, 35] Additionally, major extracranial injury (MEI), injury cause, pre-injury health status, the presence of intracranial traumatic abnormalities, ongoing mental health problems, and two week HRQoL assessment were indicated by experts as potential predictors of HRQoL following TBI.

Ongoing mental health problems were assessed through scores for depression (PHQ9), anxiety (GAD7) and post-traumatic stress disorder (PCL5) at two weeks post-injury. Socioeconomic status was assessed through type of education and type of employment. Social support was assessed through living arrangement. TBI severity was categorized into mild, moderate and severe based on the Glasgow Coma Scale (GCS) at admission. TBI was considered mild in patients with GCS 13-15, moderate in patients with GCS 9-12, and severe in patients with GCS of 3-8.[36] MEI was defined as an Abbreviated Injury Scale (AIS) ≥ 3 on any extracranial domain of the scale.[37] Pre-injury health status was assessed with the American Society of Anesthesiologists - physical status classification system (ASA-PS); patients are categorized as 'normal healthy patient', 'mild systemic disease', 'severe systemic disease', or 'severe systemic disease that is a

constant threat to life'. The categories 'severe systemic disease' and 'severe systemic disease that is constant threat to life' were combined. The presence of intracranial traumatic abnormalities was assessed through the first computed tomography (CT) scan after injury, and indicates whether any of the 12 following abnormalities was present: Mass lesion, hematoma, epidural hematoma, acute or subacute subdural hematoma, subdural collection mixed density, contusion, TAI, traumatic subarachnoid hemorrhage, intraventricular hemorrhage, midline shift or cisternal compression. The candidate predictors were assessed at admission within 24 hours, except for early HRQoL assessment and ongoing mental health problems, which were conducted two weeks post-injury.

Missing predictor values were imputed with 100 iterations with multiple imputation using the *mice* package.[38] All candidate predictors, injury severity score, and HRQoL outcomes between 2 weeks and 12 months were included in the imputation model.

Outcome Assessments

The primary outcomes were the physical (PCS) and mental (MCS) component summary scores from the Short Form-36v2 (SF-36v2) and the Quality of Life after Traumatic Brain Injury (QOLIBRI) total score at six months post-injury. The SF-36v2 is a 36-item patient-reported outcome, which assesses multiple components of HRQoL: PCS; physical functioning, role limitations due to physical health, bodily pain, general health perceptions, vitality, MCS; social functioning, role limitations due to emotional health, and general mental health. Norm-based T-scores (standardized to mean 50 and SD of 10) were calculated for the MCS and PCS.[17]

The QOLIBRI is a 37-item patient-reported outcome, consisting of four subscales assessing satisfaction with aspects of life (cognition, self, daily life and autonomy, and social relationships) and two subscales that concern how bothered the person is by difficulties (emotions, and physical problems).[19, 39]

Data Analyses

Descriptive statistics were presented as medians (Interquartile range [IQR]) or frequencies (percentage) for the predictors and HRQoL data. Differences in patient and injury related characteristics between responders, those who completed the SF-36v2 or QOLIBRI between two weeks and twelve months post-injury, and non-responders were compared using independent sample t tests (continuous) or chi square tests (categorical).

We used linear regression analyses to quantify the relationship between predictors and the SF-36v2 PCS and MCS and the QOLIBRI total score at six months post-injury.

Model performance was expressed as proportion of variance explained (R^2). For the continuous predictors - age and GCS - we assessed nonlinearity with spline functions.

For each outcome, three prognostic models were defined: I) the full model included all candidate predictors; II) the extended model included a reduced set of predictors based on the akaike information criteria (AIC); and III) the core model included the three predictors with the largest partial R^2. We also explored the incremental value of HRQoL assessment and mental health problems at two weeks post-injury for the prediction of the PCS, MCS and QOLIBRI total score. Incremental value was assessed by the difference in R^2 between the model with the additional predictors and the model without the additional predictors. Additionally, we explored the relationship between GCS (3-15) and all other predictors with interaction terms in multivariable analyses. Associations between predictors and outcome measures were presented with estimates of the regression coefficients and their 95% confidence interval (CI).

We assessed model performance through proportion explained variance (R^2), and a bootstrapping procedure to reduce optimistic model performance estimates. Bootstrapping entails drawing random samples with replacement from the development cohort, with sample size equal to that of the original cohort. Model performance was evaluated both in the bootstrap sample and the original cohort and the difference indicated the optimism in performance.[24]

Five sensitivity analyses were performed. First, the models were fitted for the PCS, MCS, and QOLIBRI total score for a subset of patients who completed the questionnaires individually or together with a relative, friend or caregiver, therefore, proxy responses were excluded. Second, the models were fitted for the PCS, MCS and QOLIBRI total score at three months rather than six months post-injury. Third, instead of only selecting patients with available six months outcome, the models were also fitted with additional imputed six months outcome whenever three or twelve months outcomes were available. Fourth, analyses were performed in subgroups of TBI severity – mild versus moderate and severe. Fifth, the models were fitted for impaired SF-36 PCS and MCS (<40) and QOLIBRI total scores (<60).[40]

Analyses were performed with R statistical software 3.6.0 [41]. We used the *rms* package to fit the regression models.[42] Modeling results were reported in accordance with the TRIPOD guidelines.[43]

Results

Study Population

We included 2666 adult patients who completed the SF-36v2 or the QOLIBRI between 2 weeks and 12 months post-injury (Supplementary Figure 1). Patients had a median age of 51 years (IQR = 33-65) (Table 1). More than half (65%) of patients were male, and most (74%) were diagnosed with mild TBI (GCS 13-15). A third (34%) had major extracranial injury. More than half (53%) were employed, and 24% were retired. About 10% had pre-injury mental health problems. Moreover, less than half of the patients (42%) experienced pre-injury comorbid health issues.

Responders and non-responders showed significant differences regarding baseline characteristics (Table 1). Non-responders had a higher median age (47 vs. 51 years), and were more often male (71 vs. 65%) (Table 1). Furthermore, they were more frequently diagnosed with moderate and severe TBI than responders, and had higher median injury severity score (16 vs. 13).

The median PCS, MCS and QOLIBRI total scores increased between three and twelve months post-injury. The largest improvements were observed between three and six months (Figure 1; Supplementary Table 1). PCS showed larger improvements than MCS in mild as well as moderate and severe TBI patients. At six months, 23% of mild and 33% of moderate and severe TBI patients fell within the 'impaired' category on the PCS. On the MCS, 26% of mild and 33% of moderate and severe TBI patients had impaired HRQoL, and on the QOLIBRI 22% of mild and 34% of moderate and severe TBI patients classified as 'impaired' six months post-injury. As expected, patients with moderate and severe TBI had lower median HRQoL scores than patients with mild TBI at every time point. The MCS and QOLIBRI (spearman 0.73) were more strongly related than with PCS (spearman 0.26 with MCS and 0.57 with QOLIBRI; Supplementary Figure 2).

Model Development

For the predictor values most (97%) observations were complete. Of the predictors with the highest percentage missing, 89% and 94% of observations were complete (Table 1).

Physical health component summary score

The strongest predictors of PCS six months after TBI were GCS, MEI, and pre-injury health status (ASA-PS) (Table 2; Figure 2). We found no significant interactions between GCS and the other candidate predictors (p >.05), indicating that predictors of PCS did not differ between patients with mild (GCS ≥ 13), and moderate and severe

Table 1: Patients' demographic and injury characteristics

Characteristics	Responders[a] (n = 2666)	Non-responders[b] (n = 1097)	
			p-value*
Demographics			
Age (18-95) (median, [IQR])	51 [33-65]	47 [30-65]	>.05
% Male sex	65 (1729)	71 (773)	<.05
Living arrangement (N, %)			
Together	2093 (79)	834 (76)	<.05
Missing (%)	3 (0.1)	8 (1)	
Highest level of education			<.001
None or primary school	321 (12)	124 (11)	
Currently in or with diploma/degree oriented program	555 (21)	199 (18)	
Secondary school / High school	820 (31)	305 (28)	
College / University	666 (25)	141 (13)	
Missing (%)	304 (11)	328 (30)	
Employment status			<.001
Yes	1410 (53)	453 (41)	
No	447 (17)	210 (19)	
Retired	643 (24)	243 (22)	
Missing (%)	166 (6)	191 (17)	
Employment type (N, %)			<.001
Working	1410 (53)	453 (41)	
Looking for work, unemployed	145 (5)	74 (7)	
Unable to work/sick leave	70 (3)	39 (4)	
Retired	643 (24)	243 (22)	
Student	190 (7)	74 (7)	
Homemaker	42 (2)	23 (2)	
Missing (%)	166 (6)	191 (18)	
Pre-injury health status			
Pre-injury ASA-PS classification			<.001
Normal healthy patient	1527 (57)	592 (57)	
Mild systemic disease	872 (33)	334 (30)	
Severe systemic disease	233 (9)	115 (11)	
Missing (%)	34 (1)	56 (5)	
History of substance abuse[c]			<.001
Yes	72 (3)	58 (5)	
Missing (%)	43 (2)	59 (5)	
Pre-injury mental health problems[d]			<.001

Table 1: Continued

Characteristics	Responders[a] (n = 2666)	Non-responders[b] (n = 1097)	
			p-value*
Yes	268 (10)	124 (11)	
Missing (%)	43 (2)	59 (5)	
Injury characteristics			
Cause of Injury			<.001
Road traffic accident	1041 (39)	371 (34)	
Incidental fall	1187 (45)	486 (44)	
Other non-intentional injury	239 (9)	84 (8)	
Violence or assault	125 (5)	99 (9)	
Suicide attempt	22 (1)	13 (1)	
Missing (%)	52 (2)	44 (4)	
GCS (3-15)			<.001
Mild (13-15)	1981 (74)	713 (65)	
Moderate/Severe (3-12)	605 (23)	338 (31)	
Missing	80 (3)	46 (4)	
ISS (0-75) (Median, [IQR])	13 [8-25]	16 [9-28]	<.001
Missing (%)	34 (1)	17 (1)	
MEI[e]			>.05
Yes	909 (34)	410 (37)	
Total percentage of observations of baseline characteristics missing	3	7	
Mental health problems two weeks post-injury (N=609)			
Depression (0-27)	5 [1-10]	NA	
Missing (%)	77 (2054)		
Anxiety (0-21)	2 [0-6]	NA	
Missing (%)	77 (2054)		
Post-Traumatic stress disorder (0-72)	9 [3-19]	NA	
Missing (%)	77 (2057)		

* p-values from ANOVA for continuous and chi-square statistics for categorical variables.
[a] Patients <18 years of age (N=158) and non-responders (N=1588) were excluded.
[b] Patients <18 years of age (N=108) and deceased patients (N=491) were excluded.
[c] Patients with a history of substance abuse disorder prior to the injury.
[d] Patients with a history of anxiety, depression, sleep disorders, or schizophrenia prior to the injury.
[e] Patients with an Abbreviated Injury Scale ≥ 3 regarding the following body regions; face, cervical spine, thorax/chest, abdomen/pelvic contents, extremities and pelvic girdle, or external (skin), thus excluding head and neck.
Abbreviations: AIS, Abbreviated Injury Scale; ASA-PS, The American Society of Anesthesiologists - physical status classification system; GCS, Glasgow Coma Scale; ISS, Injury Severity Score; N, Number; MEI, Major Extracranial Injury; SD, Standard Deviation

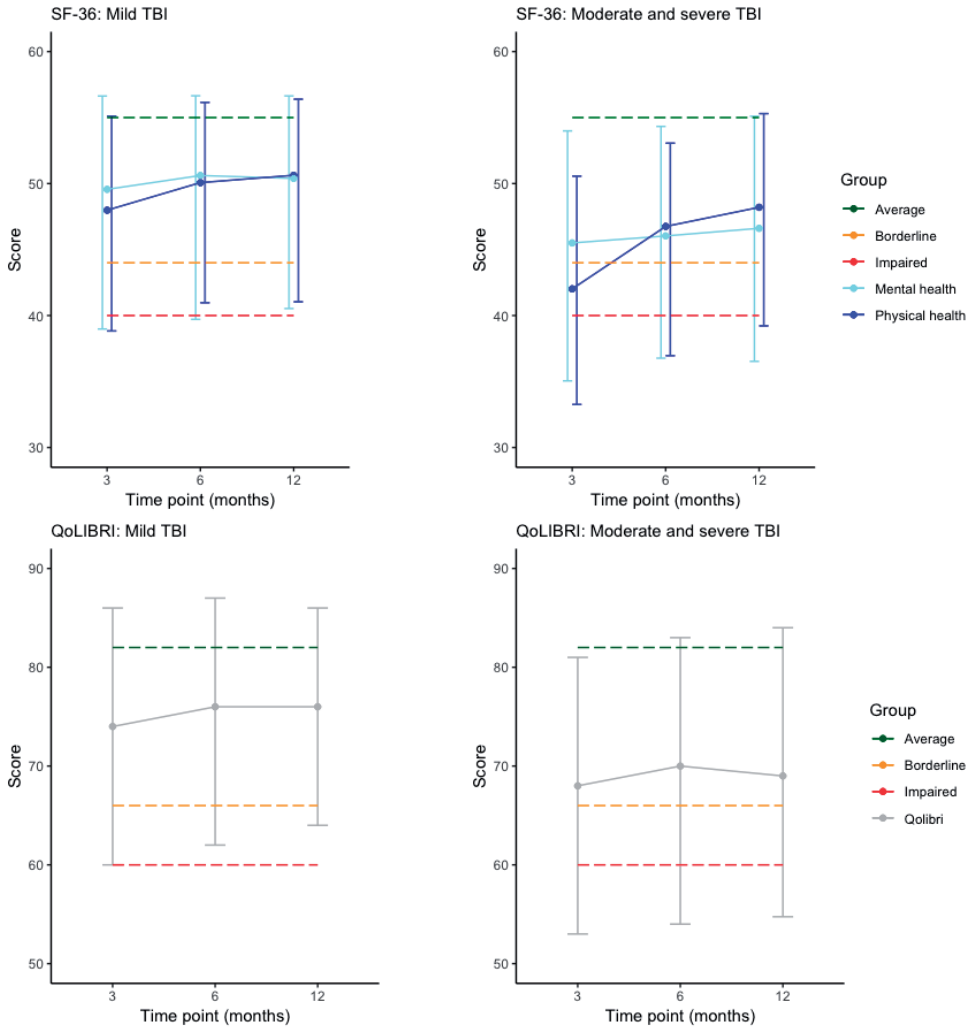


Figure 1: Plots of the median SF-36v2 physical and mental health component summary scores (top) and the Quality of Life after Traumatic Brain Injury (bottom) by time point for mild (left), and moderate and severe TBI (right). For the SF-36v2, scores of 45–55 are considered within the average range (green/upper dotted line), scores of 40–45 are considered borderline (orange/middle dotted line), and scores below 40 (red/lower dotted line) are considered impaired (Ware et al. 2007). For the QOLIBRI, scores of 67–82 are considered within the average range (green/upper dotted line), scores of 60–66 are considered borderline (orange/middle dotted line), and scores below 60 (red/lower dotted line) are considered impaired (Wilson et al. 2017).

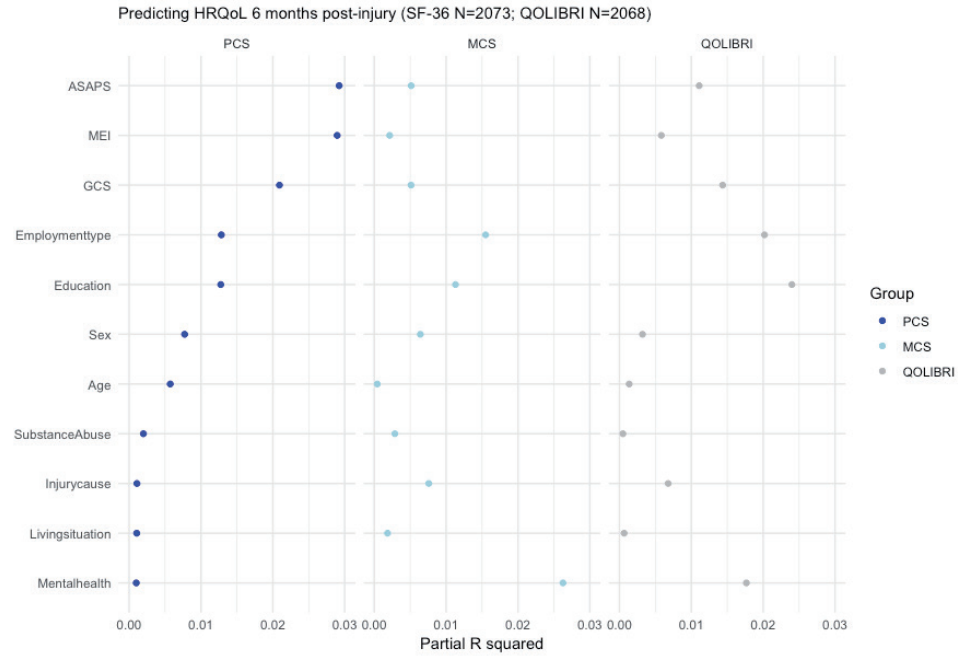


Figure 2: Contribution of predictors to partial explained variance (R^2) of the models for PCS (left), MCS (middle), and QOLIBRI (right). The partial R^2 is calculated as follows: Total R^2 of multivariable model – R^2 multivariable model without individual predictor/Total R^2 of multivariable model without individual predictor = Partial R^2

TBI (GCS ≤ 12). Severe systemic disease had a strong prognostic effect, indicating that patients with severe pre-injury comorbidities had lower PCS six months post-injury (Table 2; Supplementary Figure 3). The model had an R^2 of 11% when the three strongest predictors were considered in the core model. The extended model, also including age, sex, type of employment, and level of education, performed notably better (R^2 = 19%).

Mental health component summary score

The strongest predictors of MCS six months after TBI were pre-injury mental health problems, level of education, and type of employment (Table 3; Figure 2). Again, we found no significant interactions between GCS and the other candidate predictors (p >.05). Patients with a low level of education, as well as those who are unemployed, unable to work, or homemakers had lower MCS six months after injury (Table 3; Supplementary Figure 3). The model had an R^2 of 6% when the three strongest predictors, pre-injury mental health problems, level of education and type of employment, were considered in the core model. The extended model, also including age, employment, education and sex, performed somewhat better (R^2 = 9%).

Table 2: Regression coefficients and 95% confidence intervals for the SF-36v2 physical health component summary score (PCS) with multivariable linear regression analysis. Model performance indicated by explained variance (R^2) and bootstrap validation for each model (N=2073).

PCS	**Core Model**	**Extended Model**	**Full Model**
Constant	46	49	49
Predictors			
GCS	0.35 (0.25, 0.46)	0.38 (0.28, 0.49)	0.39 (0.28, 0.49)
MEI (No[a])			
Yes	-3.7 (-4.6, -2.8)	-4.2 (-5.1,-3.3)	-4.1 (-5.0,-3.1)
ASA-PS (Healthy patient[a])			
Mild systemic disease	-4.0 (-5.0, -3.1)	-2.0 (-3.0,-1.0)	-2.0 (-3.0,-0.96)
Severe systemic disease	-10.0 (-12.0, -8.9)	-7.2 (-8.8,-5.5)	-7.3 (-9.0,-5.7)
Education (College/Uni degree[a])			
Currently in school		-1.7 (-2.9,-0.51)	-1.8 (-3.0,-0.60)
None/Primary school		-4.3 (-5.8,-2.8)	-4.3 (-5.8,-2.8)
Secondary/high school		-1.5 (-2.6,-0.38)	-1.6 (-2.7,-0.45)
Employment (Working[a])			
Homemaker		-4.4 (-8.2,-0.55)	-4.6 (-8.5,-0.81)
Student		0.41 (-1.4, 2.2)	0.45(-1.4, 2.3)
Retired		-1.3 (-2.7, 0.10)	-1.4 (-2.8,-0.06)
Unable to work/sick leave		-6.3 (-8.8, -3.7)	-6.1 (-8.8,-3.5)
Unemployed		-3.2 (-5.1, -1.2)	-3.0 (-5.0, -1.0)
Age (per decade)		-0.73 (-1.0,-0.36)	-0.74 (-1.1, -0.36)
Sex (Male[a])			
Female		-2.1 (-3.0,-1.2)	-2.0 (-2.9, -1.1)
Injury cause (Road traffic[a])			
Incidental fall			0.71 (-0.24,1.7)
Other non-intentional injury			-0.50 (-1.1, 2.1)
Violence/Assault			-0.18 (-2.0, 2.3)
Suicide attempt			-1.4 (-5.8, 2.9)
Pre-injury substance abuse (No[a])			
Yes			3.2 (0.43, 6.0)
Pre-injury mental health problems (No[a])			
Yes			-1.2 (-2.6, 0.26)
Living arrangement (Together[a])			
Alone			-0.87 (-1.9, 0.16)
R^2 development cohort	0.13	0.20	0.21
Optimism	0.01[b]	0.01	0.02
R^2 after bootstrap validation	-	0.19	0.19

Table 2: Continued

Note:
[a] Reference category of categorical variable.
[b] Optimism of the core model is estimated to be similar to that of the extended model.
Core model = Glasgow Coma Scale, Major extracranial injury and pre-injury health status (ASA-PS).
Extended model = Core plus education, employment, age and sex.
Full model = Extended plus injury cause, pre-injury substance abuse, pre-injury mental health problems, and living arrangement.

Quality of Life after Traumatic Brain Injury total score

The strongest predictors of the QOLIBRI total score at six months were type of employment, level of education and pre-injury mental health problems (Table 4; Figure 2), which was similar to the MCS. Again, we found no significant interactions between GCS and the other candidate predictors ($p > .05$). Model performance for the QOLIBRI was intermediate to that of the models for PCS and MCS (R^2 13%, compared to 18% for PCS and 9% for MCS full models) (Table 4).

Early HRQoL assessment, ongoing mental health and intracranial lesions

In a subgroup of predominantly patients following mild TBI (99%), early HRQoL assessment at two weeks (SF36v2 N=432 and QOLIBRI N=434) had substantial incremental value (PCS R^2 37% compared to 15% of the full model without two week PCS; MCS 36% compared to 12% of the full model without two week MCS; QOLIBRI 48% compared to 10% of the full model without two week QOLIBRI) (Figure 3). Similarly, depression, anxiety, and PTSD at two weeks (SF36v2 N=418 and QOLIBRI N=420) had substantial incremental value for the prediction of MCS and the QOLIBRI (MCS R^2 = 35% compared to 11% of the full model without two week depression, anxiety and PTSD; QOLIBRI = 37% compared to 12% of the full model without two week depression, anxiety and PTSD). However, the addition of mental health problems 2 weeks post-injury had limited incremental value for the prediction of PCS (PCS R^2 = 22% compared to 16% of the full model without two week depression, anxiety and PTSD). Further, for the prediction of PCS, MCS and the QOLIBRI, the addition of intracranial traumatic abnormalities (N=1642 and N=1639) had limited to none incremental value (PCS R^2 = 20% compared to 19% of the full model without intracranial traumatic abnormalities; MCS 10% compared to 10%; QOLIBRI 14% compared to 13%).

Sensitivity analyses

Model performance was similar when proxy responses (PCS and MCS N=98, QoLIBRI N=93) were excluded. The full models also performed similarly when three month rather than six month HRQoL was predicted (PCS R^2 20% vs 19% when the

Table 3: Regression coefficients and 95% confidence intervals for the SF-36v2 mental health component summary score (MCS) with multivariable linear regression analysis. Model performance indicated by explained variance (R^2) and bootstrap validation for each model (N=2073).

MCS	**Core Model**	**Extended Model**	**Full Model**
Constant	49	45	44
Predictors			
Pre-injury mental health problems (No[a])			
Yes	-7.5 (-9.2,-5.9)	-6.9 (-8.6, -5.1)	-6.8 (-8.6,-5.1)
Education (College/Uni degree[a])			
Currently in school	-1.7 (-3.2,-0.28)	-1.8 (-3.3,-0.40)	-1.8 (-3.3,-0.39)
None/Primary school	-4.4 (-6.1,-2.6)	-4.3 (-6.1,-2.6)	-4.4 (-6.1,-2.6)
Secondary/high school	-0.96(-2.3, 0.36)	-0.85(-2.2, 0.46)	-0.84(-2.1,0.47)
Employment (Working[a])			
Homemaker	-6.4 (-11.0,-1.9)	-4.4 (-8.9, 0.12)	-4.5 (-9.1, 0.07)
Student	-0.33(-2.3, 1.6)	-0.48(-2.4, 1.5)	-0.31 (-2.5, 1.9)
Retired	2.1 (0.87, 3.3)	2.5 (1.1, 3.8)	2.3 (0.60,4.0)
Unable to work/sick leave	-5.8 (-8.9,-2.7)	-4.5 (-7.6,-1.4)	-4.6 (-7.7,-1.5)
Unemployed	-4.1 (-6.5,-1.7)	-4.0 (-6.4,-1.6)	-4.0 (-6.4,-1.6)
Injury cause (Road traffic[a])			
Incidental fall		2.2 (1.1, 3.4)	2.2 (1.1, 3.3)
Other non-intentional injury		1.2 (-0.67, 3.1)	1.2 (-0.69, 3.1)
Violence or Assault		0.01(-2.6, 2.6)	-0.04 (-2.5, 2.6)
Suicide attempt		4.9 (-0.15,10.0)	4.9 (-0.15,10.0)
GCS		0.22(0.10, 0.34)	0.22(0.09, 0.34)
ASA-PS (Healthy patient[a])			
Mild systemic disease		-0.95(-2.1, 0.21)	-1.0 (-2.2,0.20)
Severe systemic disease		-3.4 (-5.4, -1.5)	-3.5 (-5.5,-1.5)
Pre-injury substance abuse (No[a])			
Yes		-4.4 (-7.7,-1.1)	-4.3 (-7.6,-1.0)
Sex (Male[a])			
Female		-2.1 (-3.2,-1.0)	-2.1 (-3.2,-1.0)
Living arrangement (Together[a])			
Alone		-1.3 (-2.5, 0.06)	-1.3 (-2.5,-0.07)
Mei (No[a])			
Yes		-1.2 (-2.3,-0.15)	-1.2 (-2.3,-0.15)
Age (per decade)			0.08(-0.37,0.53)
R^2 development cohort	0.08	0.11	0.11
R^2 optimism	0.02[b]	0.02	0.02
R^2 after bootstrap validation	-	0.09	0.09

Table 3: Continued

Note:
[a] Reference category of categorical variable.
[b] Optimism of the core model is estimated to be similar to that of the extended model.
Core model = Pre-injury mental health problems, education and employment.
Extended model = Core plus injury cause, GCS, ASA-PS, living arrangement, MEI and sex.
Full model = Extended plus age.

model was fitted for 6 month outcome, respectively; MCS R^2 9% vs 9%; QOLIBRI R^2 14% vs 13%;). Furthermore, the models performed similarly when missing six month HRQoL outcomes (N=462) were imputed for with HRQoL outcomes on three and twelve months (PCS R^2 20% vs 19%, respectively; MCS R^2 9% vs 9%; QOLIBRI R^2 13% vs 13%) (Supplementary Tables 2-4). As expected, the predictive value of GCS diminished when patients were separated based on GCS (Mild ≥ 13, Moderate and Severe ≤ 12) (Supplementary Figure 4). The models were fitted for impaired PCS and MCS (<40) and QOLIBRI total scores (<60). The strongest predictors of impaired PCS were GCS, pre-injury health status and MEI (Supplementary Table 5). For impaired MCS the strongest predictors were pre-injury mental health problems, employment type and level of education (Supplementary Table 6). The

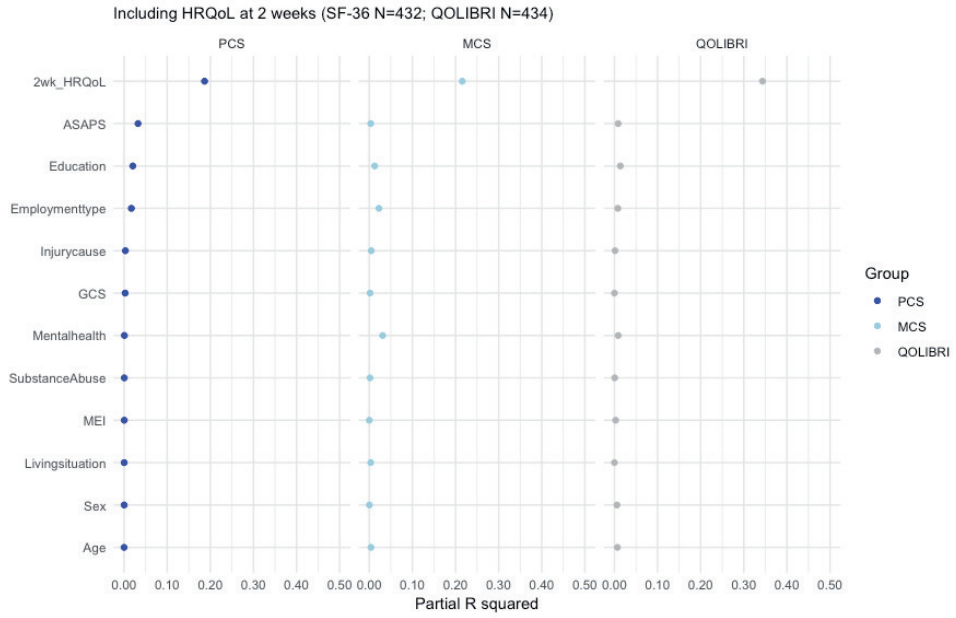


Figure 3: Contribution of predictors to partial explained variance (R^2) of the full models for PCS (left), MCS (middle), and QOLIBRI (right) including early HRQoL assessment at 2 weeks

Table 4: Regression coefficients and 95% confidence intervals for the Quality of Life after Traumatic Brain Injury (QoLIBRI) total score with multivariable linear regression analysis. Model performance indicated by explained variance (R^2) and bootstrap validation for each model (N=2068).

QoLIBRI	**Core Model**	**Extended Model**	**Full Model**
Constant	78	70	73
Predictors			
Pre-injury mental health problems (No[a])			
Yes	-9.8 (-12.0,-7.2)	-9.0 (-12.0, -6.3)	-8.8 (-11.0,-6.2)
Education (College/Uni degree[a])			
Currently in school	-5.1 (-7.3, -2.8)	-5.0 (-7.2,-2.8)	-5.0 (-7.2, -2.8)
None/Primary school	-11.0(-14.0,-8.0)	-10.0(-13.0,-7.6)	-10.0(-13.0,-7.4)
Secondary/high school	-4.8 (-6.9, -2.8)	-4.4 (-6.4, -2.4)	-4.5 (-6.5, -2.5)
Employment (Working[a])			
Homemaker	-12.0(-19.0,-5.6)	-10.0(-17.0,-3.1)	-9.1 (-16.0, -2.2)
Student	-1.6 (-1.5, 4.6)	-1.3 (-1.7, 4.3)	-0.11 (-3.5, 3.2)
Retired	-0.30 (-2.2, 1.6)	-0.47 (-1.6, 2.5)	2.0 (-0.62,4.6)
Unable to work/sick leave	-11.0(-16.0,-6.4)	-9.4 (-14.0,-4.8)	-8.6 (-13.0,-3.9)
Unemployed	-9.4 (-13.0,-5.7)	-9.1 (-13.0,-5.4)	-9.2 (-13.0, -5.5)
Injury cause (Road traffic[a])			
Incidental fall		2.8 (1.1, 4.6)	3.1 (1.4, 4.9)
Other non-intentional injury		3.2 (0.32, 6.0)	3.3 (0.43, 6.1)
Violence or Assault		-1.0 (-5.0, 3.0)	-1.2 (-5.2, 2.8)
Suicide attempt		3.1 (-4.8, 11.0)	3.2 (-4.7, 11.0)
GCS		0.56(0.37, 0.74)	0.57 (0.38,0.76)
ASA-PS (Healthy patient[a])			
Mild systemic disease		-2.4 (-4.2,-0.66)	-1.9 (-3.8, 0.09)
Severe systemic disease		-8.9 (-12.0,-5.8)	-8.1 (-11.0,-5.0)
Pre-injury substance abuse (No[a])			
Yes			-2.9 (-8.3, 2.4)
Sex (Male[a])			
Female		2.4 (0.74, 4.0)	-2.3 (-4.0,-0.69)
Living arrangement (Together[a])			
Alone			-1.2 (-3.1, 0.68)
Mei (No[a])			
Yes		-3.1 (-4.8,-1.4)	-3.2 (-4.9, -1.5)
Age (per decade)		-0.62(-1.3, 0.07)	-0.63 (-1.3,0.06)
R^2 development cohort	0.10	0.15	0.15
R^2 optimism	0.02[b]	0.02	0.02
R^2 after bootstrap validation	-	0.13	0.13

Table 4: Continued

Note:
[a] Reference category of categorical variable.
[b] Optimism of the core model is estimated to be similar to that of the extended model.
Core model = Education, employment type and pre-injury mental health problems.
Extended model = Core plus injury cause, GCS, ASA-PS, sex, MEI, and age.
Full model = Extended plus pre-injury substance abuse, and living arrangement.

strongest predictors of impaired QOLIBRI total score were GCS, level of education, and employment type (Supplementary Table 7).

Model presentation

The proposed models were presented with nomograms (Supplementary Figures 5-7). Prognostic HRQoL scores at six months post-injury can be calculated for individual patients using the formulas (Text box 1; Supplementary Table 8).

Text box 1: Example of calculation of individual SF-36v2 physical (PCS) and mental (MCS) component summary score and the Quality of Life after Traumatic Brain Injury (QoLIBRI) total score at six months post-injury based on the core models.

	Patient characteristics	**PCS score (T-scores)**	**MCS score (T-scores)**	**QoLIBRI score (0-100)**
Constant		46	49	78
GCS	13	0.35 × 13		
MEI	Yes	-3.7 × 1		
ASA-PS	Mild systemic disease	-4.0 × 1		
Pre-injury mental health problems	Yes		-7.5 × 1	-9.8 × 1
Education level	High school		-0.96 × 1	-4.8 × 1
Employment type	Retired		2.1 × 1	-0.30 × 1
Sum score		43	43	63

Discussion

We developed simple and more extended models for predicting Health-Related Quality of Life (HRQoL) six months after traumatic brain injury (TBI), separately for the SF-36v2 physical (PCS) and mental component summary scores (MCS) and the Quality of Life after Traumatic Brain Injury (QoLIBRI) total score. Medical and injury related characteristics were most important for the prediction of PCS, whereas patient related characteristics were more relevant for prediction of MCS and the QoLIBRI. Moderate model performance is indicative for the complexity of predicting HRQoL. Substantial improvement in model performance was achieved by including two-week HRQoL assessment.

Although previously indicated predictors of HRQoL following TBI were also relevant in our study the proportion explained variance (R^2) of the models was moderate. Models that include predictors that move beyond baseline assessment, also known as dynamic or longitudinal predictors, have been proposed to update existing models and potentially improve performance.[21] Prior studies have shown the importance of aspects of current status, including emotional state, for the prediction of HRQoL following TBI.[14, 39, 44] As expected, our study demonstrated that early HRQoL assessment substantially improved model performance in a subset of predominantly mild TBI patients; the R^2 for PCS was 38% compared to 17% of the full model without two week HRQoL; for MCS the increase was to 35% from 12%, and for the QOLIBRI the R^2 increased from 19% to 54%. In our study, HRQoL was highly variable between TBI patients over time, whereas within patients HRQoL scores might be more stable. This could explain the substantial incremental value of two week HRQoL for the prediction of six month HRQoL outcomes. In our study, adherence varied across time points; two week HRQoL assessment was only available in patients that were seen in the Emergency Room (ER) and discharged or in the hospital ward other than the ICU, which almost exclusively comprised mild TBI patients (99%) without MEI (91%). Therefore, the incremental value of early HRQoL assessment can only be generalized to patients following mild TBI. Early after injury, patients might be unable or less inclined to respond to questionnaires. Although patient reported outcomes are increasingly reported in clinical practice, variable or low adherence over time makes early follow-up assessments less feasible to collect, which limits the clinical applicability of dynamic prediction models using patient reported outcomes or assessments. Other longitudinal predictors that can be considered to be included for the prediction of HRQoL following TBI that may be less dependent on patient response are, for instance, biomarkers, duration of hospital stay, and length of coma.

In our study, most patients (74%) classified as mild TBI. More than half (1381/2666, 52%) had intracranial traumatic abnormalities on the initial computed tomography (CT) scan, which might be related to worse long-term outcome and lower HRQoL. In patients following mild TBI, the presence or absence of intracranial traumatic abnormalities is used to differentiate between complicated and uncomplicated mild TBI.[45] A recent study found that although patients after complicated mild TBI reported slightly more post-concussion symptoms compared to those after uncomplicated mild TBI, an abnormality on initial CT was only a weak indicator of these problems after adjusting for baseline covariates (e.g. age, gender, GCS)[46]. However, the relationship between intracranial traumatic abnormalities and HRQoL following TBI has not been examined yet. Our study indicates that when adjusting for patient and injury related characteristics, intracranial traumatic abnormalities had limited to no incremental value for the prediction of HRQoL following TBI. As intracranial traumatic abnormalities are relevant to address the heterogeneity in patients following mild TBI [45], a formal investigation of the relationship between intracranial traumatic abnormalities and HRQoL in a subgroup of patients following mild TBI is warranted. A recent study indicates that the Helsinki CT classification was associated with QoL up to 4 years after TBI [47]. Besides the presence of intracranial traumatic abnormalities, more detailed information such as CT lesion phenotypes, their location, extent and clustering could therefore be considered.

TBI can lead to long-term impairments in functional, physical, mental, cognitive, and social domains. Although median MCS was initially higher than PCS at three months, PCS showed greater improvements between six and twelve months post-injury. This indicates that over time mental health was more strongly affected by TBI. These findings also advocate for a multidimensional outcome assessment of TBI that captures a broad range of difficulties patients may experience, including physical, psychosocial and emotional outcomes. Furthermore, prior studies have shown that patients who sustained TBI, on average, show large HRQoL deficits from full recovery after twelve months when measured by population norms.[6] In our study, post hoc analyses confirmed these findings in mild as well as moderate and severe TBI patients; at twelve months 22% of mild and 27% of moderate and severe TBI patients had impaired PCS scores. Similarly, 24% of mild and 35% of moderate and severe TBI patients had impaired MCS scores, and 21% of mild and 33% of moderate and severe TBI patients had impaired QoLIBRI scores at twelve months. This indicates that a subgroup of patients may experience physical and mental limitations one year after TBI. The pattern of HRQoL scores described in our study also indicates a ceiling effect, which is a prominent issue in TBI outcome studies.[4]

The strongest predictors of the MCS were pre-injury mental health, level of education, and employment. Based on our findings, we can conclude that patient related characteristics are more important for the prediction of MCS than injury related characteristics, such as GCS. In other words, patients' well-being following TBI is more strongly influenced by psychosocial factors than the severity of injury. Furthermore, predictors of functional outcomes differ for patients with mild versus moderate and severe TBI, motivating the development of separate models for these patients.[21] It has been suggested that following moderate and severe TBI, functional outcome is determined by what "the injury brings to the patient" whereas in mild TBI it is determined by what "the patient brings to the injury" [23]. In contrast ,predictors of HRQoL did not significantly differ between patients with mild, and moderate and severe TBI. This might be explained by the fact that HRQoL captures the patient's subjective experience of their wellbeing in daily life, and is therefore likely to be affected by psychological factors and emotional adjustment. Consequently, patient related characteristics (e.g., pre-injury mental health, level of education, and employment) were expected to influence HRQoL and predictor effects to vary less by injury severity.

The combined rate of pre-injury mental health problems (Anxiety, depression, sleep disorders, and schizophrenia) was 10%, which is somewhat lower than pre-injury mental health problems of 19% and 13% for anxiety and depression based on structured diagnostic interviews (Scholten, 2016). Between studies, there is a wide variation in prevalence rates of pre-injury anxiety and depressive disorders. This can be explained by differences in study design, patients' characteristics, definitions, assessment methods, and measures used to assess psychiatric outcomes.

The models for PCS performed better than those predicting MCS and the QOLIBRI total score (R^2 19% compared to 9% and 13% of the full models for MCS and QOLIBRI). Patients' resilience, coping strategies and social support are associated with psychological outcome following TBI.[48-51] Although these psychological processes are typically not assessed in RCTs or observational studies in TBI they have the potential to improve model performance and provide opportunities for focused interventions to improve long-term psychological outcome following TBI. In patients following mild TBI, post-concussion symptoms, relating to a subset of somatic, cognitive, behavioral and emotional symptoms, are negatively associated with HRQoL.[52] Furthermore, cognitive impairments are associated with HRQoL following TBI.[53] Future research should therefore focus on the development of dynamic prediction models for HRQoL following TBI, including resilience, social support, coping, cognitive impairments, and early post-concussion symptoms as (longitudinal) predictors.

The models developed in our study include characteristics that were available at admission and two weeks post-injury. Reliable information about prognosis is of major importance to patients who sustained TBI and their families. For clinicians it would be notoriously difficult, if not impossible, to predict a patient's subjective experience of their sequelae in daily life. Prediction models for HRQoL following TBI have the potential to support clinicians to identify patients at increased risk of experiencing limitations in their daily life, who could then be followed more closely and receive early interventions to alleviate the burden of injury. Before prediction models can be considered for implementation in clinical practice, external validation is required to evaluate their performance in new settings.

Strengths of this study include the use of a longitudinal, prospective observational cohort study (CENTER-TBI). Consequently, we made use of a standardized collection of data, and a well described contemporary cohort of patients. Also, the large sample size of the development cohort allowed for reliable predictions. Another strength is the selection of candidate predictors based on literature and expert knowledge, which is preferred over selection based on data, that may increase the risk for overfitting. The predictors can be easily extracted from patients with standardized questionnaires at admission and early after admission, and are available at the time the model is to be used. Furthermore, we used a generic (SF-36v2) and TBI-specific (QoLIBRI) instrument to assess HRQoL. The SF-36 is validated and most widely used in HRQoL studies and in practice.[6] The proposed models for the SF-36v2 scales can be compared to models for other neurological conditions, such as stroke. Prior research indicates that the QoLIBRI provides additional information to the SF-36.[19]

Several limitations of our study have to be considered. First, candidate predictors were based on literature and expert knowledge. However, among studies, participants, definitions of (HR)QoL, instruments, and time points of HRQoL assessment vary widely.[6] Although prior evidence of predictors is therefore limited our study provides insight in predictors of HRQoL following TBI based on multivariable analysis. Second, living arrangement at admission was considered a proxy of social support and therefore included as a predictor. Social support is associated with psychological outcomes after TBI,[51] but it is typically unmeasured in longitudinal studies. Living arrangement might be related to social support, however, we cannot generalize our findings to the effect of social support on HRQoL following TBI. Third, in our study, non-responders were more frequently diagnosed with moderate/severe TBI than responders. Patients with more severe injury might be unable to respond to questionnaires over time. Furthermore, the SF-36v2 is not suitable for patients with major cognitive impairment or language difficulties, and thus an important subgroup of patients with profound disability is excluded. In the future, options to further improve

adherence rates among TBI patients should be explored. For instance, researchers and clinicians could combine patients' healthcare facility visits with reminders to fill in questionnaires or electronic reminders via smartphone applications.

Conclusion

Whereas prognostic models for functional outcome following TBI typically include medical and injury related characteristics, our results suggest that patient related characteristics contribute to the prediction of HRQoL following TBI. Prediction models for HRQoL have the potential to inform clinicians and patients and their families about prognosis six months after TBI. However, performance of the proposed models was moderate, which reflects the complexity of predicting HRQoL following TBI.

Funding

The research leading to these results was supported by the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 602150 (CENTER-TBI). Additional funding was obtained from the Hannelore Kohl Stiftung (Germany), from OneMind (USA) and from Integra LifeSciences Corporation (USA).

Ethical approval

The CENTER-TBI study (EC grant 602150) has been conducted in accordance with all relevant laws of the EU if directly applicable or of direct effect and all relevant laws of the country where the Recruiting sites were located, including but not limited to, the relevant privacy and data protection laws and regulations (the "Privacy Law"), the relevant laws and regulations on the use of human materials, and all relevant guidance relating to clinical studies from time to time in force including, but not limited to, the ICH Harmonised Tripartite Guideline for Good Clinical Practice (CPMP/ICH/135/95) ("ICH GCP") and the World Medical Association Declaration of Helsinki entitled "Ethical Principles for Medical Research Involving Human Subjects". Informed Consent by the patients and/or the legal representative/next of kin was obtained, accordingly to the local legislations, for all patients recruited in the Core Dataset of CENTER-TBI and documented in the e-CRF. Ethical approval was obtained for each recruiting site. The list of sites, Ethical Committees, approval numbers and approval dates can be found on the website: https://www.center-tbi.eu/project/ethical-approval

Acknowledgements

We would like to thank Daan Nieboer and Benjamin Gravesteijn for their statistical expertise that have supported the analyses performed in our study. Furthermore, we are grateful to all patients and investigators that participated in the CENTER-TBI study.

Conflict of interest

The authors declare that there is no conflict of interest.

Supplementary material

Supplementary material is available at: https://link.springer.com/article/10.1007/s11136-021-02932-z

References

1. Rosenfeld, J.V., Maas, A.I., Bragge, P., Morganti-Kossmann, M.C., Manley, G.T., and Gruen, R.L. (2012). Early management of severe traumatic brain injury. Lancet 9847, 1088-1098.
2. Perry, D.C., Sturm, V.E., Peterson, M.J., Pieper, C.F., Bullock, T., Boeve, B.F., Miller, B.L., Guskiewicz, K.M., Berger, M.S., and Kramer, J.H. (2016). Association of traumatic brain injury with subsequent neurological and psychiatric disease: a meta-analysis. Journal of neurosurgery 2, 511-526.
3. de Koning, M.E., Scheenen, M.E., van der Horn, H.J., Hageman, G., Roks, G., Spikman, J.M., and van der Naalt, J. (2017). Non-hospitalized patients with mild traumatic brain injury: the forgotten minority. Journal of neurotrauma 1, 257-261.
4. Shukla, D., Devi, B.I., and Agrawal, A. (2011). Outcome measures for traumatic brain injury. Clin Neurol Neurosurg 6, 435-441.
5. Nichol, A.D., Higgins, A.M., Gabbe, B.J., Murray, L.J., Cooper, D.J., and Cameron, P.A. (2011). Measuring functional and quality of life outcomes following major head injury: common scales and checklists. Injury 3, 281-287.
6. Polinder, S., Haagsma, J.A., van Klaveren, D., Steyerberg, E.W., and Van Beeck, E.F. (2015). Health-related quality of life after TBI: a systematic review of study design, instruments, measurement properties, and outcome. Popul Health Metr 1, 4.
7. Dikmen, S.S., Machamer, J.E., Powell, J.M., and Temkin, N.R. (2003). Outcome 3 to 5 years after moderate to severe traumatic brain injury. Arch Phys Med Rehabil 10, 1449-1457.
8. Kozlowski, O., Pollez, B., Thevenon, A., Dhellemmes, P., and Rousseaux, M. *Outcome and quality of life after three years in a cohort of patients with severe traumatic brain injury*. in *Annales de readaptation et de medecine physique: revue scientifique de la Societe francaise de reeducation fonctionnelle de readaptation et de medecine physique*. 2002.
9. Svendsen, H.A. and Teasdale, T.W. (2006). The influence of neuropsychological rehabilitation on symptomatology and quality of life following brain injury: A controlled long-term follow-up. Brain Inj 12, 1295-1306.
10. Wood, R.L. and Rutterford, N.A. (2006). Psychosocial adjustment 17 years after severe brain injury. J Neurol Neurosurg Psychiatry 1, 71-73.
11. Scholten, A.C., Haagsma, J.A., Andriessen, T., Vos, P.E., Steyerberg, E.W., Van Beeck, E.F., and Polinder, S. (2015). Health-related quality of life after mild, moderate and severe traumatic brain injury: patterns and predictors of suboptimal functioning during the first year after injury. Injury 4, 616-624.
12. Teasdale, T.W. and Engberg, A.W. (2005). Subjective well-being and quality of life following traumatic brain injury in adults: A long-term population-based follow-up. Brain inj 12, 1041-1048.
13. Andelic, N., Hammergren, N., Bautz-Holter, E., Sveen, U., Brunborg, C., and Røe, C. (2009). Functional outcome and health-related quality of life 10 years after moderate-to-severe traumatic brain injury. Acta Neurologica Scandinavica 1, 16-23.
14. Forslund, M.V., Roe, C., Sigurdardottir, S., and Andelic, N. (2013). Predicting health-related quality of life 2 years after moderate-to-severe traumatic brain injury. Acta Neurologica Scandinavica 4, 220-227.
15. Grauwmeijer, E., Heijenbrok-Kal, M.H., and Ribbers, G.M. (2014). Health-related quality of life 3 years after moderate to severe traumatic brain injury: a prospective cohort study. Archives of Physical Medicine and Rehabilitation 7, 1268-1276.

16. Williamson, M.L.C., Elliott, T.R., Berry, J.W., Underhill, A.T., Stavrinos, D., and Fine, P.R. (2013). Predictors of health-related quality-of-life following traumatic brain injury. Brain Injury 9, 992-999.
17. Maruish, M.E. (2011). User's manual for the SF-36v2 Health Survey, Quality Metric Incorporated.
18. Von Steinbuechel, N., Covic, A., Polinder, S., Kohlmann, T., Cepulyte, U., Poinstingl, H., Backhaus, J., Bakx, W., Bullinger, M., and Christensen, A.-L. (2016). Assessment of health-related quality of life after tbi: Comparison of a disease-specific (qolibri) with a generic (sf-36) instrument. Behav. Neurol.
19. von Steinbüchel, N., Wilson, L., Gibbons, H., Hawthorne, G., Höfer, S., Schmidt, S., Bullinger, M., Maas, A., Neugebauer, E., and Powell, J. (2010). Quality of Life after Brain Injury (QOLIBRI): scale development and metric properties. J Neurotrauma 7, 1167-1185.
20. Truelle, J.-L., Koskinen, S., Hawthorne, G., Sarajuuri, J., Formisano, R., Von Wild, K., Neugebauer, E., Wilson, L., Gibbons, H., and Powell, J. (2010). Quality of life after traumatic brain injury: the clinical use of the QOLIBRI, a novel disease-specific instrument. Brain Injury 11, 1272-1291.
21. Maas, A.I.R., Menon, D.K., Adelson, P.D., Andelic, N., Bell, M.J., Belli, A., Bragge, P., Brazinova, A., Büki, A., and Chesnut, R.M. (2017). Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. Lancet Neurol.
22. Ponsford, J., Willmott, C., Rothwell, A., Cameron, P., Kelly, A.-M., Nelms, R., Curran, C., and Ng, K.I.M. (2000). Factors influencing outcome following mild traumatic brain injury in adults. Journal of the International Neuropsychological Society 5, 568-579.
23. Lingsma, H.F., Yue, J.K., Maas, A.I.R., Steyerberg, E.W., Manley, G.T., including, T.-T.I., Cooper, S.R., Dams-O'Connor, K., Gordon, W.A., and Menon, D.K. (2015). Outcome prediction after mild and complicated mild traumatic brain injury: external validation of existing models and identification of new predictors using the TRACK-TBI pilot study. Journal of neurotrauma 2, 83-94.
24. Steyerberg, E.W. (2008). Clinical prediction models: a practical approach to development, validation, and updating, Springer Science & Business Media: New York.
25. Dijkland, S.A., Foks, K.A., Polinder, S., Dippel, D.W.J., Maas, A., Lingsma, H., and Steyerberg, E.W. (2019). Prognosis in moderate and severe traumatic brain injury: a systematic review of contemporary models and validation studies. J Neurotrauma.
26. Corrigan, J.D., Bogner, J.A., Mysiw, W.J., Clinchot, D., and Fugate, L. (2001). Life satisfaction after traumatic brain injury. J Head Trauma Rehabil 6, 543-555.
27. Klonoff, P.S., Costa, L.D., and Snow, W.G. (1986). Predictors and indicators of quality of life in patients with closed-head injury. J Clin Exp Neuropsychol 5, 469-485.
28. McCarthy, M.L., Dikmen, S.S., Langlois, J.A., Selassie, A.W., Gu, J.K., and Horner, M.D. (2006). Self-reported psychosocial health among adults with traumatic brain injury. Arch Phys Med Rehabil 7, 953-961.
29. Steadman-Pare, D., Colantonio, A., Ratcliff, G., Chase, S., and Vernich, L. (2001). Factors associated with perceived quality of life many years after traumatic brain injury. J Head Trauma Rehabil 4, 330-342.
30. Warren, L., Wrigley, J.M., Yoels, W.C., and Fine, P.R. (1996). Factors associated with life satisfaction among a sample of persons with neurotrauma. J Rehabil Res Dev, 404-408.
31. Webb, C.R., Wrigley, M., Yoels, W., and Fine, P.R. (1995). Explaining quality of life for persons with traumatic brain injuries 2 years after injury. Arch Phys Med Rehabil 12, 1113-1119.

32. Maas, A.I.R., Menon, D.K., Steyerberg, E.W., Citerio, G., Lecky, F., Manley, G.T., Hill, S., Legrand, V., and Sorgner, A. (2014). Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) A Prospective Longitudinal Observational Study. Neurosurgery 1, 67-80.
33. Steyerberg, E.W., Wiegers, E., Sewalt, C., Buki, A., Citerio, G., De Keyser, V., Ercole, A., Kunzmann, K., Lanyon, L., and Lecky, F. (2019). Case-mix, care pathways, and outcomes in patients with traumatic brain injury in CENTER-TBI: a European prospective, multicentre, longitudinal, cohort study. Lancet Neurol 10, 923-934.
34. McCarthy, M.L., Dikmen, S.S., Langlois, J.A., Selassie, A.W., Gu, J.K., and Horner, M.D. (2006). Self-reported psychosocial health among adults with traumatic brain injury. Arch Phys Med Rehabil. 7, 953-961.
35. Vickery, C.D., Gontkovsky, S.T., and Caroselli, J.S. (2005). Self-concept and quality of life following acquired brain injury: A pilot investigation. Brain Inj 9, 657-665.
36. Teasdale, G., Maas, A., Lecky, F., Manley, G., Stocchetti, N., and Murray, G. (2014). The Glasgow Coma Scale at 40 years: standing the test of time. Lancet Neurol 8, 844-854.
37. Van Leeuwen N, L.H., Perel P, Lecky F, Roozenbeek R, Lu J, Shakur H, Weir J, Steyerberg EW, Maas AIR. (2011). Prognostic value of major extracranial injury in traumatic brain injury: an individual patient data meta-analysis in 39 274 patients. Neurosurgery 4, 811-818.
38. Buuren, S.v. and Groothuis-Oudshoorn, K. (2010). mice: Multivariate imputation by chained equations in R. J Stat Softw, 1-68.
39. Von Steinbüchel, N., Wilson, L., Gibbons, H., Hawthorne, G., Höfer, S., Schmidt, S., Bullinger, M., Maas, A., Neugebauer, E., and Powell, J. (2010). Quality of Life after Brain Injury (QOLIBRI): scale validity and correlates of quality of life. Journal of neurotrauma 7, 1157-1165.
40. Wilson, L., Marsden-Loftus, I., Koskinen, S., Bakx, W., Bullinger, M., Formisano, R., Maas, A., Neugebauer, E., Powell, J., and Sarajuuri, J. (2017). Interpreting quality of life after brain injury scores: cross-walk with the short form-36. J Neurotrauma 1, 59-65.
41. Team, R.C. (2013). R: A language and environment for statistical computing.
42. Harrell FE, J. (2012). rms: Regression Modeling Strategies. R package version 3.4-0.
43. Moons, K.G.M., Altman, D.G., Reitsma, J.B., and Collins, G.S. (2015). New guideline for the reporting of studies developing, validating, or updating a multivariable clinical prediction model: the TRIPOD statement. Ann Intern Med 5, 303-305.
44. Morton, M.V. and Wehman, P. (1995). Psychosocial and emotional sequelae of individuals with traumatic brain injury: a literature review and recommendations. Brain inj 1, 81-92.
45. Williams, D.H., Levin, H.S., and Eisenberg, H.M. (1990). Mild head injury classification. Neurosurgery 3, 422-428.
46. Voormolen, D.C., Haagsma, J.A., Polinder, S., Maas, A.I.R., Steyerberg, E.W., Vuleković, P., Sewalt, C.A., Gravesteijn, B.Y., Covic, A., and Andelic, N. (2019). Post-concussion symptoms in complicated vs. uncomplicated mild traumatic brain injury patients at three and six months post-injury: results from the center-tbi study. Journal of clinical medicine 11, 1921.
47. Summaka, M., Zein, H., Elias, E., Naim, I., Fares, Y., and Nasser, Z. (2020). Prediction of quality of life by Helsinki computed tomography scoring system in patients with traumatic brain injury. Brain Injury 9, 1229-1236.
48. Curran, C.A., Ponsford, J.L., and Crowe, S. (2000). Coping strategies and emotional outcome following traumatic brain injury: a comparison with orthopedic patients. J Head Trauma Rehabil 6, 1256-1274.
49. Anson, K. and Ponsford, J. (2006). Coping and emotional adjustment following traumatic brain injury. J Head Trauma Rehabil 3, 248-259.

50. Rainey, E.E., Petrey, L.B., Reynolds, M., Agtarap, S., and Warren, A.M. (2014). Psychological factors predicting outcome after traumatic injury: the role of resilience. Am J Surg 4, 517-523.
51. Tomberg, T., Toomela, A., Pulver, A., and Tikk, A. (2005). Coping strategies, social support, life orientation and health-related quality of life following traumatic brain injury. Brain inj 14, 1181-1190.
52. Voormolen, D.C., Polinder, S., Von Steinbuechel, N., Vos, P.E., Cnossen, M.C., and Haagsma, J.A. (2019). The association between post-concussion symptoms and health-related quality of life in patients with mild traumatic brain injury. Injury 5, 1068-1074.
53. Wilson, L., Horton, L., Kunzmann, K., Sahakian, B.J., Newcombe, V.F.J., Stamatakis, E.A., von Steinbuechel, N., Cunitz, K., Covic, A., and Maas, A. (2020). Understanding the relationship between cognitive performance and function in daily life after traumatic brain injury. Journal of Neurology, Neurosurgery & Psychiatry.

10

Chapter 10

US and Dutch Perspectives About the Use of COVID-19 Clinical Prediction Models: Findings From a Qualitative Analysis

M. Basile, I.R.A. Retel Helmrich, J. Park, J. Polo, J. Rietjens, D. van Klaveren, T. Zanos, J. Nelson, H.F. Lingsma, D.M. Kent, J. Alsma, R.J.C.G. Verdonschot, and N. Hajizadeh

Published in: Medical Decision Making

Abstract

Introduction: Clinical prediction models (CPMs) for coronavirus disease 2019 (COVID-19) may support clinical decision-making, treatment, and communication. However, attitudes about using CPMs for COVID-19 decision-making are unknown.

Methods: Online focus groups and interviews were conducted among healthcare providers, survivors of COVID-19, and surrogates (i.e., loved-ones/surrogate decision-makers) in the United States (US) and the Netherlands (NL). Semi-structured questions explored experiences about clinical decision-making in COVID-19 care and facilitators and barriers for implementing CPMs.

Results*:* In the US, we conducted four online focus groups with (1) providers and (2) surrogates and survivors of COVID-19 between January 2021 and July 2021. In the NL, we conducted three focus groups and four individual interviews, with (1) providers and (2) surrogates and survivors of COVID-19 between May 2021 and July 2021. Providers expressed concern about CPM validity and the belief that patients may interpret CPM predictions as absolute. They described CPMs as potentially useful for resource allocation, triaging, education, and research. Several surrogates and people who had COVID-19 were not given prognostic estimates but believed this information would have supported and influenced their decision-making. A limited number of participants felt the data would not have applied to them and that they or their loved ones may not have survived, as poor prognosis may have suggested withdrawal of treatment.

Conclusions: Many providers had reservations about using CPMs for people with COVID-19 due to concerns about CPM validity and patient-level interpretation of the outcome predictions. However, several people who survived COVID-19 and their surrogates indicated that they would have found this information useful for decision-making. Therefore, information provision may be needed to improve provider-level comfort and patient and surrogate understanding of CPMs.

Key words: COVID-19; Clinical Decision Rules; Prognosis; Implementation Science; Decision Making; Decision Support Model; Decision Support Techniques; Critical Care Outcomes; Critical Care

Introduction

People hospitalized with coronavirus disease 2019 (COVID-19) may require admission to an intensive care unit (ICU), possibly escalating to the need for invasive mechanical ventilation (MV). Individual preferences for ICU admission and/or MV are often influenced by concerns about poor outcomes including prolonged MV and subsequent mortality.[1,2] However, the COVID-19 pandemic has been widely characterized by high degrees of clinical uncertainty in terms of severity of symptoms, disease trajectories, and mortality for those contracting the virus. Additionally, variation in governmental public health responses among countries and surges in COVID-19 cases ('waves') have been significant and overall outcomes have varied both by geographic region and temporally with each wave. Over the course of the pandemic, these factors have, therefore, exacerbated clinical uncertainty among healthcare providers, people with COVID-19, and their loved ones/surrogate decision makers (surrogates). This has resulted in difficulty predicting outcomes and subsequent treatment decisions, particularly for those people admitted to the hospital with COVID-19.

Clinical prediction models (CPMs) have the potential to support providers and people with COVID-19 and their surrogates in medical treatment decision-making and communication about prognosis. Further, given the continuous pressure on healthcare systems, CPMs may also support decision-making in triaging people with COVID-19 in the Emergency Department (ED) for hospital or ICU admission, and discharge. Since the start of the pandemic, several prognostic models have been developed to predict outcomes in people with COVID-19. However, almost all published models were identified as high risk of bias, indicating that their reported predictive performance is most likely overly optimistic.[5] Two CPMs that have been developed to predict outcomes in people suspected of COVID-19 are the Northwell COVID-19 Survival (NOCOS)[6] and the Erasmus Medical Center COVID Outcome Prediction in the Emergency Department (COPE) models.[7] The Northwell COVID-19 Survival (NOCOS) model was developed using data from 13 New York City area hospitals, and the COVID Outcome Prediction in the Emergency Department (COPE) model was developed on data from four hospitals located throughout the Netherlands (NL) (Rotterdam, Zwolle, Eindhoven, Heerlen). Data from the models are based on first-wave data from people who presented to the ED with suspected COVID-19. These models were validated on second-wave data at the same sites and further validated against each other to determine their temporal and geographic transportability. Both are based on readily available predictors in the electronic health record.[8,9] Figure 1 depicts the timeline of the development and validation of the two models.

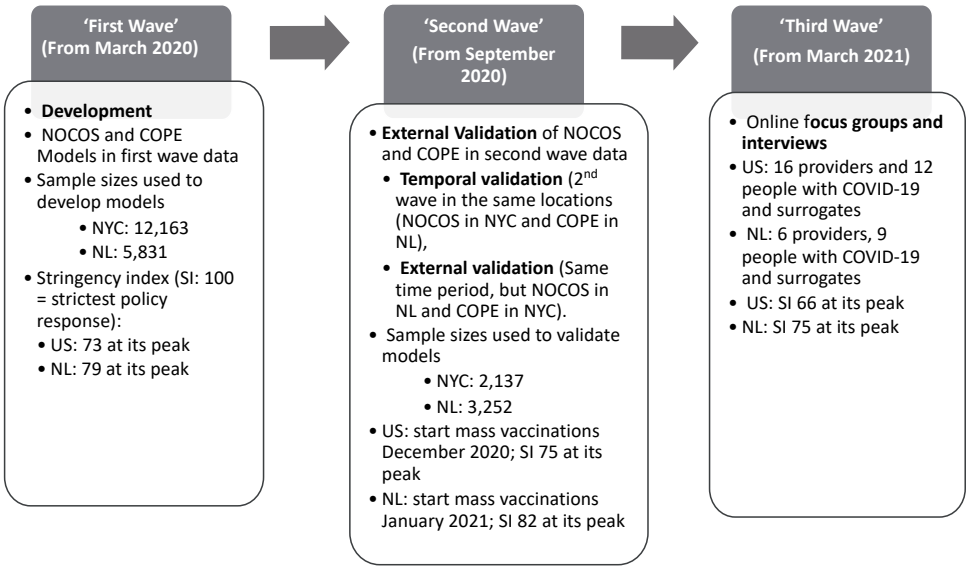


Figure 1: Timeline of Development and Validation of NOCOS and COPE Models

The NOCOS and COPE models provide risk predictions (e.g., risk of mortality expressed as a percentage given clinical characteristics of an individual patient) for people with COVID-19 based on a combination of clinical predictors. Specifically, the clinical predictors used in the NOCOS model are patient's age, oxygen saturation (%), absolute neutrophil count (k/uL), red cell distribution width (%), serum sodium (mmol/l), and serum blood urea nitrogen (mg/DL) and are used to calculate probability of hospital survival. The clinical predictors used in the COPE model are patient's age, respiratory rate (per minute), lactate dehydrogenase (LDH) (U per L), C-reactive protein level (mg per L), serum albumin (g per L), and serum urea (mmol per L) and are used to calculate mortality and ICU admission within 28 days. The models do not explicitly define treatment decisions or make treatment recommendations—i.e., they are not clinical practice guidelines (CPGs) for defining either (a) mortality risk thresholds below which a person with COVID-19 can be sent home from the emergency room or (b) thresholds above which a person should be admitted to the ICU. Rather, they provide evidence-based information to assist in decision making. These models have the potential to support providers, people with COVID-19, and surrogates about decisions concerning hospital and ICU admission for COVID-19 and the use of MV. They also allow providers and health systems to define their own risk thresholds to guide decision-making based on each health system's most up to date protocols and available resources, as well as the patient's own goals of care. Figure 2 depicts the COVID-19 care pathways and intended uses of the models, namely, to support decisions regarding admission to the hospital and ICU or discharge.

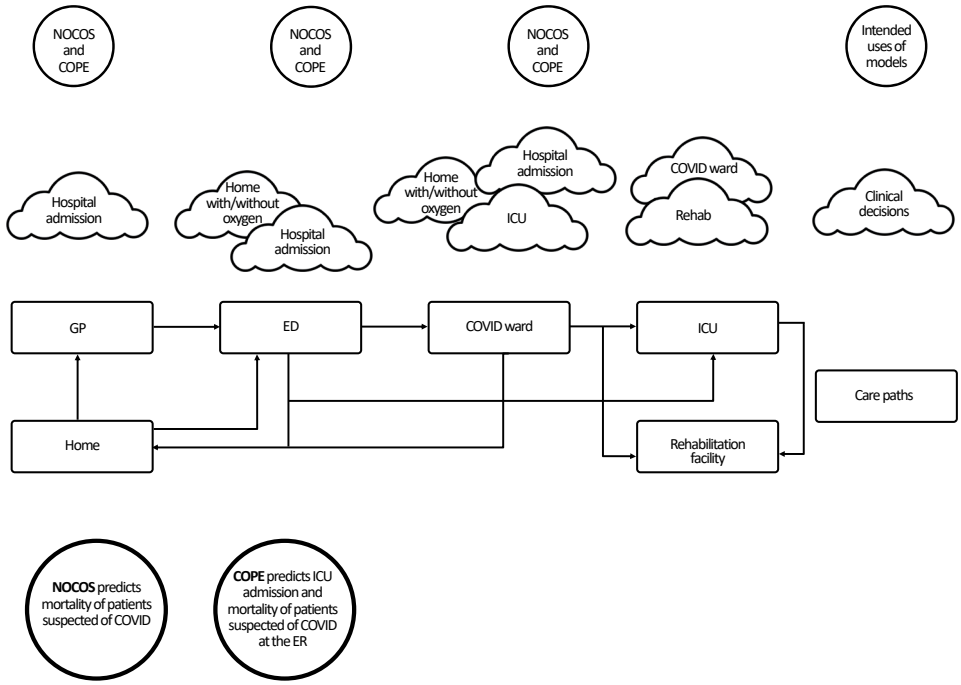


Figure 2: COVID-19 Care Pathways and Intended Uses of NOCOS and COPE Models
Boxes follow possible treatment seeking pathways of an individual at the onset of COVID-19 symptoms. Clouds represent treatment seeking decision points, beginning with a GP consult leading to an ED visit or the decision to go directly to the ED visit with subsequent decision points (e.g., back home, admission to a COVID-19 ward, or to admission to an ICU). Circles represent intended uses of the NOCOS and COPE CPMs.

Before prognostic tools such as NOCOS and COPE can be implemented in clinical practice, we should understand end-user perceptions of CPMs, including those of healthcare providers, people with COVID-19, and their surrogates. It is important to develop an understanding of how these individuals navigate clinical uncertainty, and understand probabilistic data used in CPMs during decision making. Earlier work on CPMs has primarily focused on difficulties in understanding prognosis in clinical practice. Such work has shown that there may be both lay- and provider-level misunderstanding when interpreting probabilistic data generated from the CPMs, which can occur when data is presented in relative, as opposed to absolute terms. For example, Bodemar *et al* and Gigerenzer have both shown that while relative and absolute risks are based on the same data, both providers and lay people may interpret data in a more favorable light when it is presented as relative as opposed to absolute risk, which may ultimately impact decision making.[10,11] Additionally, there may also be provider level discomfort when communicating prognosis to patients and surrogates due to the high

degrees of prognostic uncertainty inherent in applying probabilistic data generated from population-level outcomes to the individual patient.[12]

While earlier work focused primarily on probability, expansion of this earlier work has shown that clinical uncertainty is often multidimensional. Work by Han *et al* led to the development of a taxonomy for categorizing different types of clinical uncertainty. This taxonomy highlights that uncertainty may result from what are broadly categorized as a) the source (probability of a specific outcome, ambiguities related to reliability and credibility of the sources from which the CPMs were developed, and complexities linked to myriad social contextual factors such as prior functional status, post-discharge access to care, or health system resources), b) the issue (i.e., the contexts in which the uncertainty occurs. These include scientific uncertainty (cause of disease, diagnosis, treatment and prognoses), practical uncertainty (i.e., linked to the health system structures and processes), and personal uncertainty which focuses on the impact of the illness on future well-being, quality of life and personal relationships, and c) the locus of the uncertainty (which may reside with the provider, the patient, or both).[13,14]

Finally, a separate focus of existing work has explored CPM objectivity and the extent to which users perceive CPM data to be objective and based on a patient's clinical characteristics. This work has examined how specific clinical variables are selected by the model's developers while leaving out other variables (e.g., using or not using race or ethnicity as variables in a model and the extent to which this may exacerbate racial and ethnic inequalities) and whether CPMs have successfully been able to incorporate wider socio-contextual disparities impacting outcomes.[15-18]

The primary objective of our qualitative study, therefore, was to better understand how the NOCOS and COPE CPMs may be implemented to support providers, people with COVID-19, and surrogates in making critical, patient-centered decisions in COVID-19 care by situating our study within the aforementioned bodies of work on clinical uncertainty, the communication of prognostic data among providers, patients and surrogates, and CPM development and variable selection. To that end, we convened participants to gather information about what and how risk information was integrated in their prior COVID-19 treatment decision-making and to obtain pointed feedback on the NOCOS and COPE CPMs for future use. Because of cultural and healthcare system differences between the United States (US) and the NL,[19,20] we sought to understand perspectives of providers, people who were hospitalized with COVID-19, and surrogates in both countries.

The aims of our study were twofold:

- To understand specific experiences with COVID-19 decision-making among our stakeholder groups—including estimating prognosis and information communicated for decision-making.
- To identify facilitators and barriers for implementation of CPMs in COVID-19 care, including perceptions and attitudes about the usefulness of CPMs for COVID-19 decision-making.

Methods

Study Design

For this qualitative study, we conducted online focus groups and a limited number of one-on-one interviews among Dutch stakeholders who were unavailable for the focus groups. In both the US and the NL, the online focus groups were held separately for a) healthcare providers, b) and people who had COVID-19 and surrogates. The semi-structured focus group guides (Supplemental Appendices A-B) for this study were developed to facilitate a flexible conversational approach with open-ended probes focused on assessing the following:

1. Stakeholder experiences with COVID-19:
   a. Provider's estimating prognosis, communication about prognosis, and decision-making
   b. Information communicated to people with COVID-19 and surrogates about prognosis and decision-making related to COVID-19 treatment-seeking
2. Stakeholder attitudes and beliefs about the NOCOS and COPE CPMs
3. Facilitators and barriers for using the NOCOS and COPE CPMs in COVID-19 care.

The study was approved by Research Ethics Committee of Erasmus Medical Center, Northwell Health's IRB in the Feinstein Institutes for Medical Research (IRB# 20-1017), and the Tufts University Medical Center (IRB#00001044). We followed the Consolidated Criteria for Reporting Qualitative Research (COREQ) guidelines in reporting this study (Supplemental Appendix C). The interviews took place between January 2021 and July 2021.

Participants and Inclusion Criteria

Healthcare providers in both countries, were invited to take part if they had experience caring for people with COVID-19. In the US, people who had been hospitalized for COVID-19 and surrogates of people who had COVID-19 were invited to take part if they were older than 18 years of age and had proficiency in the language in which the

focus groups were conducted (i.e., English or Spanish). In the NL, people who had visited the ED or had been admitted to the hospital (COVID-19 ward or ICU) for COVID-19 and surrogates of people who had COVID-19 were invited to take part if they were older than 18 years of age and had proficiency in the language in which the focus group or interview was conducted (i.e., Dutch). Surrogates could include a partner, parent, child, or other significant other (e.g., roommate). Among surrogates, death of the patient had to be at least three months prior to enrolling in the study. All participants needed access to a device (e.g., laptop, computer, tablet) with a working camera and microphone, and they needed to provide informed consent to participate. The US-based focus groups only invited those who were hospitalized with COVID-19 or were relatives of hospitalized patients, while the NL-based focus groups invited individuals who either visited the ED (i.e., they were not admitted during their ED visit) or were hospitalized for COVID-19 (either in a COVID-19 ward or ICU). All other eligibility criteria as described above were similar across countries.

Recruitment

Healthcare providers were recruited through (clinical) collaborators from the study team at Tufts University Medical Center, Northwell Health, and Erasmus University Medical Center using purposive sampling of COVID-19 health care providers to capture a diverse array of clinical specialties (e.g., critical care physicians, pulmonologists, acute intensivists, geriatricians, nurses, and members of palliative care support teams such as clinicians, hospital chaplains). People who had COVID-19 and surrogates were recruited through online advertisements and hard-copy fliers placed in our pulmonary and respiratory clinics. Surrogates were further identified via our COVID-19 participants already enrolled in the study. Members of the study team contacted by telephone, people who had COVID-19 and surrogates who responded to the advertisements and were interested in participating to further explain the study. If they agreed to participate, the research team sent an information letter and informed consent form, which were returned to the researcher by email (US and NL) or by post (NL only). Individual subjects could withdraw from the study at any time without consequences and without having to state their reasons for withdrawal. Following local norms and common practice for compensation for research study participation, US-based participants received $200 for each session in which they participated (providers had the option of attending two sessions). All Dutch participants received €25, which was not disclosed beforehand.

Data Collection

Focus groups and interviews used a semi-structured interview guide (Supplemental Appendix). All focus groups and interviews were conducted via Microsoft Teams or Zoom. In preparation for the online interviews, a brief online practice session was

held with participants to test the internet connection and familiarize participants with the software that was used. Each session took approximately one hour and was led by an experienced moderator. For the US-based focus groups, MB (a PhD-level medical anthropologist) and NH (a critical care pulmonologist) conducted the two provider focus groups and English-speaking patient/surrogate focus group. JP (a research coordinator and bilingual English/Spanish speaker) conducted the Spanish-speaking patient/surrogate focus group. A bilingual Spanish/English speaking interpreter was present to translate the discussion to the non-Spanish-speaking members of the study team. Additional members of the US and NL research teams were present at each of the sessions. For the NL-based focus groups, IRH (a PhD candidate in public health with a focus on CPMs) and JR (a PhD-level associate professor with expertise in qualitative research) conducted all focus groups with HL present. All focus group moderators were female. All focus groups and interviews were audio-recorded (following introductions and verbal consent for recording), and later transcribed by a professional company with all identifiers removed to maintain confidentiality. Spanish and Dutch focus groups and interviews were also translated to English. Additionally, members of the study team took notes during the focus groups. At the start of each session, the moderators introduced the study team including their academic and or clinical positions, and participants introduced themselves to the group. The moderators gave an overview on CPMs in general and provided details about the development of NOCOS and COPE specifically (including which predictors were included in the model and the prediction (estimate) that the model would give). Participants were shown static PowerPoint images of the 2 CPMs including the specific clinical data points used to calculate prognosis in both models as describe above, and we explained that the models were developed based on data from multiple hospitals.

Data Analysis

Qualitative analysis involved an iterative coding process using open, axial, and selective coding, allowing an inductive exploration of themes and constructs. First, the open and axial coding was done separately for the US- and NL-based focus groups and interviews. The US-based transcripts were coded independently by three researchers and the NL-based interviews by two researchers. Second, the codebooks were compared and similarities and differences between codes for the US- and NL-based focus groups and interviews were examined. Third, the codebooks were merged into one final codebook to allow for the selective coding process (Supplemental Appendix D). Results were reported separately for providers and people who had COVID-19 and surrogates. Additionally, differences between the US and the NL participants were examined. All transcripts were hand-coded (i.e., coding software was not used).

Results

Study Population

In the US, we conducted four online focus groups: two for healthcare providers (N=9), one for English speakers who had been hospitalized for COVID-19 and surrogates (N=12) and one for Spanish speakers who had been hospitalized for COVID-19 and surrogates (N=12) between January 2021 and July 2021 (Table 1). In the NL, we conducted three focus groups and four individual interviews with healthcare providers (N=6) and people who had visited the ED and were subsequently either discharged or admitted to the hospital for COVID-19 and surrogates (N=9) between May 2021 and July 2021 (three Dutch provider-participants and one Dutch surrogate who were recruited but unable to participate in the scheduled focus groups took part in one-on-one interviews using the focus group guides).

Table 1: Participant Demographics

Characteristic	**US** **N (%)**	**NL** **N (%)**
People with COVID-19	N = 5	N=5
Sex:		
Male	2 (40%)	2 (40%)
Female	3 (60%)	3 (60%)
Other	0	0
Race/Ethnicity:*		
White	2 (40%)	5 (100%)
Black	2 (40%)	0
Hispanic	2 (40%)	0
Asian	0	0
Other	0	0
Language of Focus Group		
English	3 (60%)	0
Spanish	2 (40%)	0
Dutch	0	5 (100%)
Admitted to ICU for COVID-19	5 (100%)	1 (20%)
Experienced mechanical ventilation	3 (60%)	1 (20%)
Caregivers of People with COVID-19	N = 7	N=4
Relationship to individual with COVID-19:		
Parent	1 (14.3%)	0
Partner or spouse	0	3 (75%)
Other relative	5 (71.4)	1 (25%)
Other non-relative	1 (14.3)	0
Sex:		
Male	2 (28.6)	2 (50%)
Female	5 (71.4)	2 (50%)
Other	0	0

Table 1: Continued

Characteristic	US N (%)	NL N (%)
Race/Ethnicity:*		
White	3 (42.9)	N/A**
Black	1 (14.3)	
Hispanic	2 (28.6)	
Asian	1 (14.3)	
Other	1 (14.3)	
Declined response	1 (14.3)	
Language of Focus Group:		
English	5 (71.4)	0
Spanish	2 (28.6)	0
Dutch	0	4 (100%)
Cares for someone admitted to ICU for COVID-19	7 (100%)	1 (25%)
Cares for someone admitted who experienced MV for COVID-19	4 (57.1%)	0
Providers by Clinical Specialty*	N = 9	N=6
Geriatric Hospitalists	2 (22.2%)	0
ED Physicians and Surgical/Critical Care Physicians/ICU intensivists	1 (11.1%)	0
Internal Medicine	1 (11.1%)	0
Pulmonary and Critical Care Physicians	6 (66.7%)	1 (16.6)
Palliative Care Physicians	3 (33.3%)	0
Internal Medicine Physician Trainees	1 (11.1%)	1 (16.6)
Emergency department physician in training	0	1 (16.6)
physicians	0	1 (16.6)
Nursing Home Physicians	0	1 (16.6)
ICU nurse and senior researcher	0	1 (16.6)
Pastoral Care/Chaplains	1 (11.1%)	0

*Some US respondents selected/identified with multiple categories for race/ethnicity, and clinical specialty.
** Data on race/ethnicity was not collected from the NL participants.

Summary of Results

US and Dutch Providers

Overall, we found greater similarities than differences among providers in the two countries (Table 2 shows results of qualitative analysis from the provider responses). In the first wave, similarities noted among providers were the high degrees of clinical uncertainty for outcomes and a sense of 'groping in the dark'. In the initial absence of COVID-19–specific outcomes data, providers prognosticated based on intuition ('gut feelings'), gestalt, 'laying eyes,' objective real-time data such as vital signs and the course of illness of the patient (i.e., rapidity of deterioration, days of infection, and severity of illness), and the patients' co-morbidities and prior functional status. Providers used descriptions and general estimates to communicate prognosis to patients

Table 2: Domains Addressed and Themes Identified in Focus Group Interviews with US- and NL-Based Providers

Domains Addressed	Themes	Quotes
Experiences with estimating prognosis and clinical decision-making for COVID-19	• Focus on clinical factors, gut feeling, and gestalt; • No use of CPMs; • Uncertainty.	"We don't use a decision model. But of course, we know what factors are favorable and unfavorable [for outcome in COVID-19 patients]." (Dutch provider) "I know that in that first wave we were really inventing the wheel" (Dutch provider)
Communicating about COVID-19 prognosis	• Prognosis in words instead of numbers; • Short-term prognosis instead of long-term prognosis; • Difficult as surrogates could not be present.	"I would group their family member into a category and tell them that, you know, we're doing our best and based on these vital signs and the lab data it's looking like it's going this direction." (US provider) "Then we want to say we think the chances are very bad if you go to the ICU, we advise you not to go there. We speak more in terms of poor odds than percentages."(Dutch provider)
Possible use of CPMs	*Bedside use:* • Communication with patients/surrogates; • Decision to admit patient to ward or ICU; • Consensus building among providers. *Non-bedside use:* • Education/training; • Risk stratification; • Resource allocation; • Triaging.	*Bedside use:* "If there's something more standardized, you know, then it helps everybody sort of speak that same language. Because I think that confusing messages between physicians can be very difficult for families." (US provider) *Non-bedside use:* "The Department of Public Health, or whoever, if they would know ahead of the time that there are these 100 patients around in the hospitals who are not in the ICU, but they are at a higher risk, they could have looked at things differently, or moved patients around better."(US provider) "For new doctors and the next generation it might be good."(Dutch provider)
Attitudes towards CPMs	*Positive:* • Helpful; • Supportive; • Objective. *Negative:* • Not relevant; • Misleading; • Not patient specific.	*Positive:* "But precisely in that middle area where you may not be sure, then I think it's very good to indeed have that based on a larger database of patient data, yes then it can actually be helpful in your own decision-making." (Dutch provider) "I think that it would be an extra tool that we can get into our decision-making tree." (US provider) "If I'm speaking to families, I think it does give a nice objective measure to use." (US provider) *Negative:* "[Families may think] "Oh, this is a really valid piece of information." But what really does that information tell us?" (US provider) "It's always hard for me to take a prognostic model from a big pool of people and apply it to the person I'm seeing." (US provider) "Much of the work can't just be captured in a score" (Dutch provider)

Table 2: Continued

Domains Addressed	Themes	Quotes
Facilitators for use of CPMs in practice	*Acceptability:* • External and temporal validation; • Impact analysis. *Implementation:* • Linked to treatment decision; • Able to adapt daily based on changes in patient's condition. *Application:* • Accessible through website; • Smartphone application, • Electronic patient record; • Information on how to interpret risk estimate.	*Acceptability:* "It does give more confidence when you say it has been externally validated, including the second wave." (Dutch provider) *Application:* "You could easily embed this into an EMR, or just have either a phone app, or just have an app on the computer, that you could plug the numbers in real quick." (US provider)
Barriers for use of CPMs in practice	*Provider-level:* • Limited knowledge about CPMs; • Score fatigue; • Difficulty interpreting risk estimates. *Model-level:* • Model incomplete (e.g. no comorbidities); Outcomes are of less relevance (e.g. no ICU mortality).	*Provider-level:* "The tricky thing is that many clinicians are not familiar with prediction models. That makes it difficult." (Dutch provider) *Model-level:* "Like I said the model is quite limited […] So in the model, maybe then it gets too complex, but I would put in risk factors like BMI. Does someone have an underlying comorbidity, recently had chemo, recent immune suppressive therapy, an organ transplant? I think you can do more with that."(Dutch provider) "I always worry whether is this a case where something pops out statistically as correlating, but clinically is not relevant?"(US Provider)

Abbreviations used throughout: US (United States), NL (Netherlands), CPMs (clinical prediction models), ICU (intensive care unit), EMR (electronic medical record), BMI (body mass index).

and families rather than numbers—e.g., "things look bad" or "the odds are not good." As stated by one Dutch provider, "I never use numbers or whatever, because if I'm worried then, indeed, I'll just say that 'I am worried' about the patient." In subsequent waves, clinical uncertainty was reduced as patterns in outcomes began to emerge.

Providers suggested that the NOCOS and COPE CPMs could be used at the bedside to build consensus among providers and in non-bedside contexts for triage and resource allocation. Positive feedback on the CPMs was that they appeared objective

and that a tool to augment subjective decision-making was needed. Negative feedback was that the CPMs were not patient specific and did not include variables such as patient baseline functional status or socio-cultural contexts that might heavily influence outcomes. Perceived facilitators to using CPMs included enhancing provider acceptability by providing data on CPM validation and making the CPMs easily accessible within the health system software. Barriers to use were providers' limited knowledge of prediction models (i.e., a provider-level barrier) and provider perceptions that the CPM outcomes used were not relevant to their decision-making; for example, not including likelihood of morbidity if a patient survived MV (i.e., model-level barrier).

The one key difference noted among providers regarded critical care decision points. Among Dutch providers, the most important decision point was in the ED and focused on whether to admit a person with COVID-19 to the ICU. Dutch providers explained that they had specific criteria for ICU eligibility and that, in some cases, a patient may be too sick to survive in the ICU. Here, decisions were typically made between providers and people with COVID-19 directly before an individual was admitted to the ICU and therefore before MV decision-making. In the US, decisions about treatments such as MV would typically be decided among providers, people with COVID-19, and surrogates once the individual was already admitted to the ICU.

US and Dutch Surrogates and People Who Had COVID-19

Participants in both countries described uncertainty regarding decisions about when to go to the hospital and when a high chance of relapse after discharge from the emergency room or hospital would be a concern (Table 3 shows results of qualitative analysis from the patient and surrogate responses). Respondents reported that prognosis was often not explicitly discussed, and that providers tended to use words instead of numbers (supporting providers' own descriptions of how they presented information about prognosis to their patients and surrogates), and that data was rarely referred to— possibly due to a lack of data at the start of the pandemic. References to media representation of COVID-19 came up in each of the focus groups. For example, participants in both countries described hearing about high mortality rates and poor outcomes for people with COVID-19 needing MV. This weighed heavily on their own perceptions of likely outcomes and decision-making. For example, as one person who had COVID-19 requiring MV in the US/Spanish language focus group stated about his prior knowledge of MV, "the only thing I heard was that people who were intubated in New York were all dying, so my feelings about intubation were very bad and I refused intubation twice."

Table 3: Domains Addressed and Themes Identified in Focus Group Interviews with US- and NL-Based Surrogates and People Who Had COVID-19

Domains Addressed	Themes	Quotes
Experiences with clinical decision-making in COVID-19	• Shared decision-making; • Acute care decisions; • Lay knowledge.	*Shared decision-making:* "So, when it came time to be intubated, we knew that my dad had already agreed that he wanted that. So, we did allow that, we had to allow that because that's what his wishes were." (US surrogate, English language) *Acute care decisions:* "I really had no choice myself, so I was brought to the [hospital] and there I was immediately put on the ventilator." (Dutch patient) "[ICU admission] was not in consultation with us. It was immediate because it was necessary." (Dutch surrogate) *Lay knowledge* "At a certain point there was an intervention and it is called faith, and in my religion, that is called a miracle, and that is what happened." (US surrogate, Spanish language)
Communication with Providers about COVID-19 prognosis	• Communication with patient and/or caregivers; • Prognosis often not explicitely discussed; • In words instead of numbers.	"There was never a discussion of probabilities or predictive variables. In my case, I certainly would have valued that." (US patient, English speaking)
Attitudes towards CPMs	*Positive:* • Supportive; • Preference for risk estimates is Individual. *Negative:* • Induce fear and anxiety; • Incorrect/not patient specific; • No added value.	*Positive:* "I think it's nice to be as transparent as possible. And especially if someone is intubated or ventilated and the patient themselves cannot communicate, you really only have the doctor and the nursing staff. So then you prefer to have all available information."(Dutch surrogate) " I do think that [preference for risk estimation] is different from person to person. And as such a doctor will also have to sense that, I think, to whom they can say that and to whom not."(Dutch patient) *Negative:* "Had I known any prediction models, they might have been a little scary and anxiety-provoking."(US patient, English speaking) "I think that the problem is that I am the exception, they gave me very less chances of survival, then I think, according to me, if they had used with me that model, I would not be here."(US patient, Spanish speaking) "I think at the time I was admitted, [risk estimations] wouldn't have mattered to me at all. It would have added absolutely no value." (Dutch patient)

Table 3: Continued

Domains Addressed	Themes	Quotes
Possible use of CPMs	• Guideline for conversations; • To support prognosis in 'words'; • Explain and support treatment decisions.	*Support prognosis in words:* "Yes, it is serious, but how serious, on a scale of one to ten. For me [providing a number], that is kind of pleasant, then I know how to look at it." (Dutch patient) *Explain and support treatment decisions:* "That it [CPM] is kind of part of the decision of whether or not to admit someone. Particularly if there is also a 'code black' or emergency situation or that there are indeed too few beds. Or in the case of transfer for example that in that case it should be part of the decision." (Dutch surrogate) "Sometimes I regret and I think wow, maybe if he [dad] had been with the mask for some more days, maybe he could have recovered, so yes, data and also some assurance with numbers, or any other thing, probabilities, a doctor saying what it is best in order to make a decision." (US surrogate, Spanish language)
Facilitators for use of CPMs in practice	• Understandable language; • Real-world interpretation by clinician (i.e. pertaining to real life).	"Indeed, in normal understandable [language]. I always tell doctors that. I like to hear things I can understand. All that medical language and numbers, I didn't study [medicine]." (Dutch patient)
Barriers for use of CPMs in practice	• Not being able to understand the risk estimates; • Dangerous misunderstanding; • Use of difficult language; • No added value.	"I don't know what would have happened If I had not been able to understand him [the doctor], I don't know.. a bad decision or a misunderstanding. I think that for these medical situations, language is extremely important. A language with less words but that has to be accurate. In a translation you cannot add information, misunderstand information."(US surrogate, Spanish language)

Positive attitudes toward the use of CPMs included (1) that they could provide greater transparency to how the providers were making their decisions and recommendations (i.e., a numeric value to lend support for a provider's recommendation), and (2) that it allowed people with COVID-19 to view their chances compared to others on a visual scale. Negative attitudes included concern that the CPMs may induce fear in the person with COVID-19 or surrogate and that they therefore need to be carefully communicated to select patients. In fact, one participant suggested that the provider should decide whether or not to share this information with the person depending on whether they felt that he/she could handle the information and associated fears: "I think it differs from person to person. I think a doctor will have to have a feel for who

he can say that to and who he can't." Participants suggested using CPMs to explain decision recommendations and to help them visualize the information being communicated. Perceived facilitators to CPM use included ensuring that the content was easily understandable to reduce the chance for misunderstanding. Barriers included the user perception that the predictions may not be accurate, i.e., that some people with COVID-19 may have a different outcome from what the model might estimate.

A key difference among our Dutch and US participants was that most of Dutch people who had COVID-19 were not admitted to the ICU. Therefore, the US participants overall had experienced more severe COVID-19. However, among those US and Dutch participants admitted to the ICU both had experiences in which providers needed to make immediate decisions about MV and did not involve surrogates or COVID-19 patients in decision-making. For example, one Dutch participant who had COVID-19 stated, "I was brought to the hospital and immediately put on a ventilator. I was so sick, they didn't even ask." However, most Dutch participants described consulting their general practitioners (GPs) at the onset of symptoms, and GPs often made house calls to assess them. Typically, people with COVID-19 and their GPs would decide together that the person should then go to the ED where triage would take place. Individuals would be assessed and then either sent home (with oxygen if needed), to a rehabilitation center, a COVID-19 ward, or the ICU. Some Dutch surrogates discussed losing contact with family members once an ICU admission had occurred. While all participants mentioned hearing about COVID-19 in the media, among US participants, references to lay or folk knowledge such as religion were more frequent compared to Dutch participants.

Discussion

The use of CPMs in critical care settings has precedent. Many centers have mortality prediction scores (e.g., the APACHE score)[22,23] incorporated into the electronic health records. Whether these scores are used by frontline providers in the direct care of their patients to inform their decision-making or in their communication with patients and surrogates is unclear. The largest study measuring the effect of providing calculated prognostic estimates to ICU clinicians, conducted 25 years ago, was the Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments (SUPPORT) study.[24] It did not find that the provision of objective (calculator-derived) prognostic estimates to ICU clinicians altered the use of life-supporting technologies or was associated with improved communication or care of critically ill patients—mainly due to provider-level lack of use. Therefore, when engaging our stakeholders in the US and the NL, we aimed to better understand how CPMs might be implemented

to support healthcare providers, people hospitalized with COVID-19, and surrogates in making medical decisions about COVID-19 care in the US and the NL. This includes determining reasons for potential hesitancy of using the COVID-19 CPMs in the ED, ICU, or other clinical settings. We also sought to better understand end-user perceptions about CPMs, how prognostic uncertainly was communicated within the context of COVID-19, and how CPMs may support future communication in these settings.

Prior work on CPM implementation in clinical practice shows multi-level– and multi-stakeholder–based factors (e.g., provider, patient, CPM, and health system) potentially impacting the use and usefulness of CPMs. For example, the non-adoption, abandonment, scale-up, spread, and sustainability (NASSS) framework[25] explains that the implementation of technological health innovations is influenced by factors related to seven domains: the characteristics of the condition or illness (i.e., is it well known or well understood?), the technology, the value proposition (i.e., the extent to which users find it valuable or 'worth using'), adopter system (i.e., the professional staff, patients, and surrogates who will be adopting the technology), the organization, the wider institutional and societal context, and the interactions among all of these domains over time. Taken together, these domains allow consideration of COVID-19 CPM implementation in a wider context—namely, as an illness that was initially not well understood, with an uncertain trajectory, further complicated by the fact that many stakeholders are unfamiliar with CPMs in general and may not see the value of the CPMs in facilitating decision-making. Moskowitz et al. also explored reasons for non-adaption of a well-known prediction model in traumatic brain injury (the IMPACT model) and found that mistrust in underlying data and the belief that presenting numbers derived from statistical methods may mislead patients and relatives were the main barriers to implementation.[26] These are consistent with our findings. Additional prior work which we found to be significant for our analysis includes Han et al.'s taxonomy of clinical uncertainty which provides a language for describing the myriad clinical uncertainties found among our study participants, work on communicating clinical uncertainty among providers, patients and surrogates,[12,27] and studies exploring the impact of variable selection in CPM development.[15-18]

In our study, we found that among participants in both countries, first-wave experiences were characterized by various types of clinical uncertainty that fit within the taxonomy of uncertainty. Providers described an inability to draw on prior anecdotal experiences or clinical outcomes in assessing prognosis for people with COVID-19, which represents *source uncertainty* specifically *ambiguity* due to the absence of credible data available at the time. Therefore, providers from both countries frequently drew on intuition or gut feelings based on direct observations of patients at the bedside

and their clinical trajectories over time. This uncertainty was reduced in subsequent waves as patterns in outcomes began to emerge, although here, too, CPMs were not used. Rather, published outcome studies and personal experience were extrapolated to make prognostic estimates. Providers in both countries felt that CPMs could be useful in aiding decision-making since the CPMs could provide a standardized way of communicating prognosis based on an objective set of criteria. However, both groups were wary of relying solely on CPMs due to the fact that the models are based on the general population and may not reflect the individual characteristics of the people with COVID-19 they were treating, which represents *probabilistic uncertainty*.[13,14] Some providers also pointed out that factors such as baseline functional status should be considered, and quality of life should be predicted in addition to survival. Furthermore, providers stressed that access to resources post-discharge may impact outcomes, both of which represents clinical uncertainty rooted in *complexity* linked to patient-level personal and social contextual factors.

Providers felt that CPMs might be useful in communicating with surrogates and people with COVID-19, perhaps used as part of a larger "toolset" to estimate prognosis and support clinical decisions and communication, instead of as a stand-alone tool as is typically recommended in prediction model studies. However, some providers were hesitant to use models due to concerns that people who had COVID-19 may misunderstand the data or may not understand that the numbers were only estimates and not guaranteed outcomes. For example, several providers worried that patients and families would place too much value on the number given. Some providers were also concerned that surrogates and people who had COVID-19 would find the numbers dehumanizing. While some survivors of COVID-19 in our study pointed out that CPM estimates would likely not have accurately captured their own outcomes (many were survivors of the most severe COVID-19 illness), we found that most people who had COVID-19 or surrogates did not question the accuracy or validity of the data itself. Overall, a greater acceptance of using CPMs compared to providers was notable among surrogates and people who had COVID-19. Many stated that they would have wanted to see or hear about data and estimates in numbers and would have found this information helpful when making treatment decisions. This supports concerns of our providers about how patients and surrogates understand prognostic data and how they might rely on it for decision making. Therefore, an important caveat to the use of CPMs was that the communication of prognosis to surrogates and people with COVID-19 should be tailored to accommodate health literacy levels and as prior work shows, should include discussions about prognostic uncertainty inherent in CPMs.[12]

While we identified fewer barriers to use among surrogates and people who had COVID-19 some stated the information might be scary or 'anxiety provoking,' the

main priority for people who had COVID-19 and surrogates in our study is that data are presented in a way that is easy to understand—both in terms of health literacy levels—and accommodating their cultural and linguistic needs. Subsequent revisions to the CPM platforms can accommodate these suggestions by making the information available in multiple languages as well as providing training to providers for integrating information from the CPMs into their conversations with people who had COVID-19. Additionally, providers must remain aware that in some cases, sharing CPM data may do more harm than good in some individuals who misinterpret data or have their fears amplified when seeing prognostic estimates.

Our focus groups and interviews allowed us to assess facilitators and barriers that may occur when implementing the CPMs in clinical practice. Most providers felt that the CPMs could be easily embedded in the EHR since both NOCOS and COPE are already available through websites, and can be made available on an app, tablet, or desktop workstation. Overall, the biggest barrier among providers was a general hesitancy toward using the CPMs in the first place, due to their attitudes and beliefs about CPMs. As mentioned, providers in our study saw as it as a limitation that the models did not take into account co-morbidities, an individual's prior functional status, or access to resources upon discharge when assessing prognosis. However, because the NOCOS and COPE models were explicitly developed based on quickly and objectively obtainable predictors at presentation to the ED, pre-existing comorbidities or prior functional status were not considered. To address these concerns, providing detailed information on how the models were developed and validated, including the data and variables included and excluded in the models, is vital to obtaining provider-level buy-in for using the CPMs. However, this also underscores the need to balance objective data from the CPMs with wider patient-level socio-contextual factors including race, socio-economic status, and disabilities, which may be left out of CPMs but that indeed impact outcomes,[15-18] and which may result in some individuals being denied care if CPMs are used to determine which patients should be allocated scarce resources. Ensuring clarity on which variables are included, while at the same time offering alternative ways of "contextualizing" the patient during clinical conversations may be a way of further tailoring communication about prognostic uncertainty when using CPMs. It is also vital that providers address lay knowledge (e.g., spiritual beliefs, media representations) and local contexts (e.g., resource shortages) in addition to goals of care, when discussing prognosis as these factors may also be considered by patient and surrogate when making clinical decisions. Finally, CPMs must be continuously updated and maintained to ensure ongoing validity as new data becomes available or when conditions 'on the ground' evolve rapidly, as has been seen with COVID-19 over time. This may further alleviate provider-level concerns about using the CPMs.

Finally, while providers had some concerns about using the CPMs for individual-level prognosis and communication, their suggestions for "non-bedside uses" are useful. Specifically, using the CMPs CPMs to predict community-level, or hospital-level surges would allow for more efficient distribution of resources and triaging of individuals to locations with greater bed availability. Currently, the NOCOS and COPE models do not provide prescriptive risk thresholds, meaning that they do not make treatment recommendations based on a given risk estimate. These risk thresholds would be influenced by availability of resources and social norms. For example, whether or not to admit a person with COVID-19 to the ICU, may also depend on hospital bed capacity and patient preferences for care. Our study suggests that, with careful consideration of recommendations for implementation, the risk predictions provided by NOCOS and COPE can support providers, people with COVID-19, and surrogates when making decisions about hospital or ICU admission.

Strengths and Limitations

Our study is one of the few that provides insight on considerations of potential users of CPMs in two different countries. We obtained multiple perspectives from participants with different experiences and from different settings. Our findings support prior work on clinical uncertainty and may have practical implications which can guide the development, validation, and implementation of CPMs for COVID-19 in the future. Furthermore, we believe that some of the facilitators (e.g., external validation) and barriers (e.g., limited knowledge of CPMs among providers) that were identified apply to CPMs in other fields.

Several limitations of our study should be considered. Firstly, in the NL, we did not include stakeholders from non–Dutch-speaking communities, some of whom may have different concerns about the use of CPMs. Secondly, in our study, 'context' was limited to personal experiences; we do not present the wider socio-cultural contexts that may shape those experiences. The fact that the NL has national-level, universal health care, whereas the US is characterized by a private and more fragmented health system, is well known. While relevant to decision-making and prognosis—particularly regarding access to care and the use of life extending treatment—these considerations were beyond the scope of our study. However, it must be emphasized that while race/ethnicity was not a variable used in either NOCOS or COPE, racial inequalities are known to impact health outcomes. Thirdly, differences in disease severity, lived experiences of people who had COVID-19, and surrogate participants may impact perceptions of the models. For instance, overall, the Dutch patients suffered less severe COVID-19 illness, which in turn likely influenced their feedback. Lastly, the participants represent health systems in Boston, New York City, and Rotterdam—all metropolitan areas that do not reflect the healthcare experiences of areas with more

rural and isolated populations. Future work may explore how CPMs may be combined with lay knowledge and gestalt to tailor shared decision-making conversations, allowing a balance between CPMs on the one hand and other epistemological frameworks on the other. We may also explore the use of CPMs during clinical conversations to obtain first-hand accounts of CPM use, and to measure the impact of actual use on decision making and clinical outcomes.

Conclusions

Despite differences in healthcare systems and national-level public health programs, we saw more similarities than differences between our US and Dutch stakeholders. While providers had reservations about using CPMs for people who had COVID-19 due to concerns about CPM validity and patient-level interpretation of the data, surrogates and people who had COVID-19 indicated that they would have found this information useful for decision-making provided that the information is carefully and possibly selectively communicated. Future studies must develop and test high-quality strategies for communicating prognostic uncertainty and to enhance shared decision-making using CPMs, and they must continue to measure acceptability for surrogate decision-makers and people with COVID-19 via ongoing stakeholder engagement. This may help increase both provider-level comfort and patient and surrogate decision-maker understanding of CPMs.

References

1. Richardson S, Hirsch JS, Narasimhan M, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. *Jama.* 2020;323(20):2052-2059.
2. Cazeau N. Social Isolation: Managing Psychological Distress in Hospitalized Patients During the COVID-19 Pandemic. *Clinical journal of oncology nursing.* 2020;24(5):472-474.
3. Nkengasong J, Iwasaki A, Victora C, et al. The Global Response to the COVID-19 Pandemic. *Med (New York, NY).* 2020;1(1):3-8.
4. World Health Organization. Coronavirus disease (COVID-19) pandemic. 2021; https://www.who.int/emergencies/diseases/novel-coronavirus-2019.
5. Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ (Clinical research ed).* 2020;369:m1328.
6. Levy TJ, Richardson S, Coppa K, et al. Development and Validation of a Survival Calculator for Hospitalized Patients with COVID-19. *medRxiv : the preprint server for health sciences.* 2020.
7. van Klaveren D, Rekkas A, Alsma J, et al. COVID outcome prediction in the emergency department (COPE): using retrospective Dutch hospital data to develop simple and valid models for predicting mortality and need for intensive care unit admission in patients who present at the emergency department with suspected COVID-19. *BMJ open.* 2021;11(9):e051468.
8. Levy TJ, Coppa K, Cang J, et al. Development and validation of self-monitoring auto-updating prognostic models of survival for hospitalized COVID-19 patients. *Nature communications.* 2022;13(1):6812.
9. van Klaveren D, Zanos TP, Nelson J, et al. Prognostic models for COVID-19 needed updating to warrant transportability over time and space. *BMC medicine.* 2022;20(1):456.
10. Bodemer N, Meder B, Gigerenzer G. Communicating Relative Risk Changes with Baseline Risk: Presentation Format and Numeracy Matter. *Medical decision making : an international journal of the Society for Medical Decision Making.* 2014;34(5):615-626.
11. Gigerenzer G. Making sense of health statistics. *Bulletin of the World Health Organization.* 2009;87(8):567.
12. Anderson RJ, Stone PC, Low JTS, Bloch S. Managing uncertainty and references to time in prognostic conversations with family members at the end of life: A conversation analytic study. *Palliative medicine.* 2020;34(7):896-905.
13. Han PK, Klein WM, Arora NK. Varieties of uncertainty in health care: a conceptual taxonomy. *Medical decision making : an international journal of the Society for Medical Decision Making.* 2011;31(6):828-838.
14. Kalke K, Studd H, Scherr CL. The communication of uncertainty in health: A scoping review. *Patient education and counseling.* 2021;104(8):1945-1961.
15. Adams R. Urgent care: disability, pandemic, and the value of a life. *The Lancet.* 2022;399(10323):430-431.
16. Hathaway B. Triage standards during pandemic may worsen racial disparities in treatment. 2021; https://news.yale.edu/2021/09/20/triage-standards-during-pandemic-may-worsen-racial-disparities-treatment#:~:text=Triage%20standards%20during%20pandemic%20may%20worsen%20racial%20disparities%20in%20treatment,-By%20Bill%20Hathaway&text=During%20the%20worst%20of%20the,receive%20scarce%20health%20care%20resources.
17. Schmidt H, Roberts DE, Eneanya ND. Rationing, racism and justice: advancing the debate around 'colourblind' COVID-19 ventilator allocation. 2022;48(2):126-130.

18. Vyas DA, Eisenstein LG, Jones DS. Hidden in Plain Sight - Reconsidering the Use of Race Correction in Clinical Algorithms. *The New England journal of medicine.* 2020;383(9):874-882.
19. Blumenthal D, Fowler EJ, Abrams M, Collins SR. Covid-19 - Implications for the Health Care System. *The New England journal of medicine.* 2020;383(15):1483-1488.
20. Aleva FE, van Mourik L, Broeders M, Paling AJ, de Jager CPC. COVID-19 in critically ill patients in North Brabant, the Netherlands: Patient characteristics and outcomes. *Journal of critical care.* 2020;60:111-115.
21. Saldana J. *The Coding Manual for Qualitative Researchers, 3rd Edition.* SAGE Publications Ltd; 2015.
22. Knaus W, Wagner D, Draper E. APACHE III study design: analytic plan for evaluation of severity and outcome in intensive care unit patients. Development of APACHE. *Critical care medicine.* 1989;17(12 Pt 2):S181-185.
23. Wheeler MM. APACHE: an evaluation. *Critical care nursing quarterly.* 2009;32(1):46-48.
24. Connors AF, Jr, Dawson NV, Desbiens NA, et al. A Controlled Trial to Improve Care for Seriously III Hospitalized Patients: The Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments (SUPPORT). *Jama.* 1995;274(20):1591-1598.
25. Greenhalgh T, Wherton J, Papoutsi C, et al. Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies. *Journal of medical Internet research.* 2017;19(11):e367.
26. Moskowitz J, Quinn T, Khan MW, et al. Should We Use the IMPACT-Model for the Outcome Prognostication of TBI Patients? A Qualitative Study Assessing Physicians' Perceptions. *MDM policy & practice.* 2018;3(1):2381468318757987.
27. Simpkin AL, Armstrong KA. Communicating Uncertainty: a Narrative Review and Framework for Future Research. *Journal of general internal medicine.* 2019;34(11):2586-2591.

11

Chapter 11

General discussion

The overall aim of this thesis was to increase our knowledge of prediction of outcome in acute care. Methodological aspects and their importance in prognostic research were reviewed and various applications of prognostic research were presented. We addressed four research questions (Text box 1). This final chapter provides a summary and discussion of the main findings, followed by recommendations for future research and clinical practice.

Text box 1: Main research findings per research question

What methodological aspects are of key importance in prognostic research?

We found that methodological quality of the model development study is related to performance at external validation. Therefore, methodological aspects, such as sample size, selection of predictors, and internal validation, should be considered carefully. Prognostic research can progress when methodological recommendations are followed, and when reporting guidelines are adhered to. Collaboration between research groups provide opportunities to develop, update and validate prognostic models.

To what extent can we predict functional outcome and Health-Related Quality of Life after TBI in contemporary patients?

We found that the IMPACT and CRASH models adequately identify patients at high risk for mortality or unfavorable outcome after TBI. However, calibration is variable between settings, which reflects heterogeneity in the reliability of predictions.
We developed models for the prediction of Health-Related Quality of Life (HRQoL) following TBI, which showed that HRQoL is challenging to predict using baseline predictors. Our results imply that repeated HRQoL assessments can further improve the prediction of HRQoL: including 2-week HRQoL assessment to baseline predictors improved model performance substantially.

Can blood-based biomarkers further improve prediction of functional outcome following TBI?

We found that all examined serum biomarkers – UCH-L1, S100B, GFAP, NFL, t-tau, and NSE - obtained within 24h after injury, improved the prognostic value for functional outcome, over patient's demographic, clinical and radiological characteristics.

How do health care providers, patients and surrogate decision makers perceive the use of prediction models to support clinical decision-making?

We found that health care providers had concerns about using prognostic models for medical decision making at the bedside. They felt that prediction models may have a greater value if used for resource allocation, triage, or educational purposes. In contrast, patients and surrogate decision makers felt that prediction models would have been informative and may have supported them in making COVID-19 treatment decisions. We found that factors related to the health care provider, prediction model, and patient, might influence the use of prognostic models in practice.

Part I: Methodological aspects

Models combining multiple patient characteristics to predict the risk of an outcome for an individual patient have the potential to support health care providers and patients in conversations about prognosis and in making medical decisions.[2] As these decisions may have far reaching consequences, it is crucial that predictions are reliable and accurate.[3] However, several reviews showed shortcomings in model development.[9, 12, 13] Typically, limitations are found in the design and statistical analyses, which make models at risk of overfitting. Overfitting relates to the notion of asking too much from the available data, which will result in overly optimistic estimates of model performance.[47] When the model is overfitted, results are not valid in underlying or related populations. Consequently, the model may predict poorly, with severe drawbacks when the model is applied in practice: it does not separate low from high-risk patients (poor discrimination), and may give unreliable, or even misleading risk estimates (poor calibration). We found that lower methodological quality of model development studies is related to poorer model performance at external validation in the field of TBI (**Chapter 5**). A large-scale validation study of a short form based on the PROBAST in the field of cardiovascular disease showed that higher risk of bias was associated with poorer discrimination. [8]

Performing high quality prognostic research requires general methodological and specific statistical knowledge. Chapter 3-5 were written for audiences from different fields, and consequently differ in length, format, and use of technical language. By doing so, we aim to give researchers and health care providers from different educational backgrounds a better understanding of (the importance of) methodological aspects in prognostic research. We highlight guidelines and resources that can be consulted. In **Chapter 2-5** we used different sources, including literature (**Chapter 2 and 3**), reporting checklists (**Chapter 4**) and a tool to assess risk of bias of prognostic studies (**Chapter 5**), to provide recommendations for methodological aspects that should be carefully considered in prognostic research (Text box 2). These sources all emphasize the importance of transparent reporting and the consideration of the study design and statistical analysis.[3, 10, 48, 49]

In Text box 3 we provide an overview of methodological challenges and opportunities in prognostic research. Prediction modeling studies often suffer from incomplete reporting,[9] which might indicate that specific methodological aspects were not considered. Reporting checklist such as TRIPOD and REMARK can be consulted to avoid incomplete reporting.[10, 48] Similarly, quality checklists such as the PROBAST can inform investigators on what should be reported in prognostic model studies.[49] A short form based on the PROBAST, consisting of 8/20 items, was recently validated

Text box 2: Sources discussed in this thesis that offer methodological guidelines

First author, year of publication	Name or abbreviation	Aim	Important aspects
Collins, 2015	TRIPOD	Checklist to provide guidelines for the reporting of prediction model studies	Transparent reporting
Altman, 2012	REMARK	Checklist to provide guidelines for the reporting of prognostic factor studies	Transparent reporting
Wolff, 2019	PROBAST	Tool to assess the risk of bias and applicability of prediction model studies	Includes 20 items on participant selection, study design, predictors, outcome and statistical analysis
Steyerberg, 2019*	Clinical prediction models	Book that provides a practical checklist with seven steps that need to be considered for development of a valid prediction model	Missing values Estimation Selection Presentation Internal validation External validation

Note: * the first edition of the book was published in 2009

and could distinguish well between high and low RoB.[8] In our study, the overall judgment on the short form was consistent with the original PROBAST for almost all studies (**Chapter 5**).

External validation is often lacking and can therefore be seen as an opportunity for the improvement of methodological quality of prognostic models in TBI.[50] We emphasize the importance of external validation, preferably across a range of settings, before the use of a model can be considered in practice. External validation can be facilitated when research is performed in collaboration among multiple centers, countries, or research groups. To facilitate collaboration, standardizing names, definitions and measurement of variables is crucial and can be done using guidelines for common data elements in TBI.[51] A recent example of collaborative efforts is a collaboration between two large longitudinal studies in TBI, in which CT results were evaluated in data from TRACK-TBI and findings were externally validated in data from CENTER-TBI.[52] Collaboration provides opportunities for internal and external validation. Furthermore, collaboration results in a larger sample size, and discourages development of superfluous models as it facilitates updating, adjusting, and recalibrating existing models.

Text box 3: Overview of methodological challenges and opportunities in prognostic research

Challenges	Opportunities
General	
Incomplete reporting	Consult literature, reporting guidelines and quality checklists
Model development	
Insufficient sample size	Collaborative efforts; context-driven selection of predictors
Inappropriate handling of missing data or complete case analysis	Use of multiple imputation methods
Selection of predictors based on univariate analysis or stepwise selection procedures	Shrinkage and penalization in multivariable analysis
Full model equation is not presented for validation	Report full model equation
Internal validation	
Lacking or inefficient	Use of bootstrap resampling or cross-validation for efficient validation
External validation	
Lacking	Validation of models in cohort other than development cohort through collaborative research
No calibration reported	Report calibration intercept and slope

Prior studies describe that calibration, the agreement between observed and predicted outcomes, is reported less often than discrimination.[13, 53] Similarly, a number of the external validation studies in our study did not assess model performance in terms of calibration (**Chapter 5**). When reported, calibration was assessed with the Hosmer-Lemeshow goodness-of-fit test or shown graphically with a calibration plot. To be able to compare model performance between validation studies, reporting the calibration intercept and slope is preferred.

Part II: Applications

TBI can lead to long-term impairments in functional, physical, mental, cognitive, and social domains.[21, 28, 54-58] Impairments can be assessed using functional outcome scales (e.g. Glasgow Outcome Scale (Extended)),[59] and patient-reported outcome measures that focus on an individuals' perception of how a disease and its treatments affect the physical, mental and social aspects of their life (e.g. SF-36 and QOLIBRI).[32, 34, 60]

Prediction of functional outcome

External validation provides information on the models' generalizability and (geographic or temporal) transportability, that is, how the model performs in new patients and settings.[4] External validation and updating of established models is generally preferred over the development of new models.[2] This is especially true in the field of TBI, in which the main predictors of outcome, including age and initial injury severity, are known (**Chapter 2**).[13] In **Chapter 6** we performed detailed evaluations of the external validity of the IMPACT and CRASH prognostic models in a contemporary cohort of patients across Europe. The IMPACT and CRASH models were developed on large cohorts based on relatively historical data,[37, 38] while the epidemiology and survival rates of TBI have changed considerably over the last decades.[15, 19-23] Model performance at external validation can be influenced by several factors, including the methodological quality of the model development study (**Chapter 5**), and study characteristics such as the patient selection.[61] We found that the IMPACT and CRASH models showed good discriminative ability, which improved modestly with the addition of CT variables. There were substantial differences between observed and predicted outcome risk (calibration), specifically for the CRASH CT model. This is consistent with findings from a systematic review that concluded that performance of prognostic models for moderate and severe TBI is highly variable across different settings.[13]

The variable calibration of the models in our and prior studies reflects heterogeneity in reliability of predictions. This motivates continuous validation and updating if clinical implementation is pursued.[13] To update prediction models, several techniques of varying extensiveness are available, including recalibration, and incorporation of novel predictors.[62, 63] A collaborative effort was set up to update the IMPACT prognostic models to setting-specific circumstances: The MoreIMPACT project. The aim of this ongoing project is to bring multiple cohorts together with patients from Europe, the UK, the US, South America, and Australia. The results of this study can be used to provide a web-based calculator for setting-specific IMPACT prognostic models, which may allow for better management of individual patients than provided by a single global model.

The inclusion of novel predictors, such as blood-based biomarkers, has been proposed to further improve prediction and prognostic models for TBI.[15] Biomarkers, such as S100 calcium-binding protein B (S100B), glial fibrillary acidic protein (GFAP), and ubiquitin C-terminal hydrolase L1 (UCH-L1) have received much attention for their role in diagnosing mild TBI and triaging patients for computed tomography (CT) scanning of the head.[64] S100B has been implemented in the Scandinavian TBI Guidelines, and based on results of the ALERT-TBI study the combination of GFAP

and UCH-L1 was approved by the FDA as a diagnostic test in patients suspected of mild TBI.[65, 66] In addition to the diagnostic role of biomarkers in TBI evidence indicates the potential for a prognostic role.[67-74] However, most studies have focused on the unadjusted prognostic effect of biomarkers rather than estimating their value over and above established prognostic factors, which is considered essential.[75] In **Chapter 7**, we examined the incremental prognostic value of serum biomarkers, independent of patient's demographic, clinical and radiological characteristics, for prediction of six-month GOSE following TBI. All examined serum biomarkers – UCH-L1, S100B, GFAP, NFL, t-tau, and NSE - obtained within 24h after injury, improved the prognostic value for functional outcome. We found that UCH-L1 had the greatest incremental prognostic value, and thus biomarkers that have greatest incremental prognostic value (UCH-L1) are different to those that have greatest diagnostic value (such as GFAP) in TBI. We examined the incremental value of six serum biomarkers that have been studied most extensively in recent studies, both in isolation and in combination (including the specific combination of GFAP and UCH-L1, which has specific diagnostic value). Combining all six biomarkers resulted in small further increments in C-statistic and R^2, compared to the best performing individual biomarkers separately. Adding biomarkers to the IMPACT and CRASH models resulted in an R^2 up to 45% and 46% for mortality and 37% and 45% for unfavorable outcome, respectively. Therefore, these biomarkers should be considered in updating existing prognostic models or when developing new models for functional outcome after TBI. In the CENTER-TBI study the time of biomarker sampling is widely varying and typically late. Serial sampling of biomarkers has revealed different temporal trajectories.[76] Therefore, future research should consider mixed model approaches (also known as dynamic prediction models) for the prediction of functional outcome following TBI including repeated measures of serum biomarkers.

Prediction of HRQoL

In this thesis, we examined HRQoL following TBI in contemporary patients across Europe. We found that individuals' perceptions of well-being are often discordant with their objective functioning following TBI (**Chapter 8**). Over time, physical health showed greater improvements between six- and twelve-months post-injury than mental health (**Chapter 9**). This indicates that over time mental health was more strongly affected by TBI. These findings advocate for a multidimensional outcome assessment of TBI that captures a broad range of difficulties that patients may experience. Hereby, it is important to assess wellbeing using patient reported outcome measures (PROMs), besides patients' objective functioning. PROMs incorporate patient perception on physical and mental wellbeing. In practice, the systematic use of information from PROMs leads to better communication and decision-making between health care providers and patients, and patient's satisfaction with care.[77-79]

In TBI, there is an increasing emphasis on assessing multi-dimensional outcomes.[15] However, many overlapping assessment tools are available, without clear guidelines on the choice of outcomes.[26, 32] A CENTER-TBI study showed that GOSE identified impairment in 59-61%, the SF-12v2 in 28% and the Trail Making Test Part A in 19% of patients.[80]

We found that for patients following mild TBI, the lowest mean score for mental health was reported for those with lower moderate disability (GOSE 5) (**Chapter 8**). Our results thus suggest that patients with lower moderate disability living in the community should receive additional support, rehabilitation and interventions. As suggested by a prior study,[81] our results indicate that patients with moderate disability might be less satisfied with their social support and were less likely to receive rehabilitation. Generally, access to rehabilitation services is more likely among patients following moderate and severe TBI and patients with severe disability compared to less severely injured and disabled patients.[82] Consequently, patients after less severe TBI report more unmet rehabilitation needs than those following severe TBI.[83]

In **Chapter 8** we aimed to examine the relationship between disability assessed with the GOSE and HRQoL measured with the SF-12v2 mental health component summary score (MCS) and the QOLIBRI-OS following TBI, while taking variation in personal, injury-related, and environment factors into account. GOSE, personal, injury-related and environment factors explained a limited amount of variance in HRQoL (up to 29%). As the majority of variance remained unexplained, future research should consider the effect of coping, resilience, adaptation, and cognitive impairments on HRQoL following TBI. Another CENTER-TBI study found that MCS scores generally decreased with increasing cognitive impairment, and reached a plateau in the severely disabled group.[84] In the CENTER-TBI study, data on cognitive impairments in severely disabled patients (GOSE 3-4) is too limited to allow us to examine the effect of cognitive impairments on HRQoL in these patients.

Reliable information about prognosis is of major importance to patients who sustained TBI and their families. However, for clinicians it would be notoriously difficult, if not impossible, to predict a patient's subjective experience of their wellbeing. Therefore, models predicting HRQoL have the potential to support clinicians to identify patients at increased risk of experiencing limitations in their daily life, who could then be followed more closely and receive early interventions to alleviate the burden of injury. In **Chapter 9**, we aimed to identify predictors of, and develop prognostic models for the prediction of HRQoL after TBI. We developed simple and more extended models for predicting HRQoL six months after TBI, separately for the SF-36v2 PCS and MCS and the QoLIBRI total score. Medical and injury related characteristics were

most important for the prediction of the PCS, whereas patient related characteristics were more relevant for prediction of the MCS and the QoLIBRI. The performance of models to predict HRQoL was limited: the proportion of variance explained of the full models was 19% for PCS, 9% for MCS and 13% for the QOLIBRI. Patients' resilience, coping strategies, social support, cognitive impairments, and post-concussive symptoms are associated with HRQoL following TBI, and might further improve prediction of HRQoL.[84-87]

Considering the variation in HRQoL between patients, we assumed that moderate model performance was an indication of the complexity of predicting HRQoL. Furthermore, we hypothesized that the use of solely baseline predictors was suboptimal to predict HRQoL after TBI. In a subset of predominantly patients following mild TBI, including 2-week HRQoL assessment (N=436) improved model performance substantially (R^2 PCS 15% to 37%, MCS 12% to 36%, and QOLIBRI 10% to 48%) (**Chapter 9**). Prior studies have shown the importance of aspects of current status, including emotional state, for the prediction of HRQoL following TBI.[88-90] Similarly, in a study by Haagsma et al., (submitted) repeated HRQoL assessments improved the discriminative ability of models predicting HRQoL using the EQ-5D and the EQ-VAS in patients following trauma. The incorporation of predictors that capture information beyond hospital admission, for instance at 2 weeks post-injury, is known as dynamic prediction. Studies showed that dynamic prediction in moderate and severe TBI resulted in variable improvement in model performance,[13] whereas studies on prognosis following mild TBI showed that symptoms days or weeks after the sustained injury are strong predictors of long-term outcomes.[39, 91]

In the CENTER-TBI study, adherence varied across time points; two week HRQoL assessment was only available in patients that were seen in the Emergency Room (ER) and discharged or in the hospital ward other than the ICU, which in our study almost exclusively comprised mild TBI patients (99%) without MEI (91%) (**Chapter 9**). The incremental value of early HRQoL assessment can therefore only be generalized to patients following mild TBI. Patients might be unable or less inclined to respond to questionnaires early after injury. Although patient reported outcomes are increasingly reported in clinical practice,[77] variable or low adherence over time limits the clinical applicability of dynamic prediction using patient reported outcomes or assessments. Other longitudinal predictors of outcome following TBI, such as blood-based biomarkers, might be less dependent on patient response.

Implementation of prognostic models in clinical practice

Prognostic models can be used in research and clinical practice.[2] Many centers have for instance, mortality prediction models, such as the APACHE score,[92] incorporated

into the electronic health records. Despite extensive research on prognosis and the development and validation of prognostic models in TBI, high-quality and well validated models are not routinely used in clinical practice. Studies exploring the reasons behind the scarcity of the implementation of prognostic models for TBI and other medical conditions are lacking. The few studies that have explored reasons for lack of implementation found several barriers related to the clinician, model and condition, including lack of awareness of established models, mistrust in prognostic estimates for individual patients, missing of relevant predictors or outcomes, and a heterogeneous disease course.[93-95] Prior studies have explored the perspective of clinicians whereas the perspective of other end-users (e.g. patients and their relatives) has been overlooked.

In **Chapter 10,** qualitative analyses were used to explore considerations of health care providers, patients, and surrogate decision makers (e.g. relatives and caregivers) about the use of prediction models to support clinical decision-making in COVID-19 care. We conducted focus groups and individual interviews with health care providers, patients and surrogate decision makers in the US and the Netherlands. Participants were introduced to two prognostic models predicting outcome after COVID-19 and were asked about their considerations regarding the use of these models in practice. Providers had reservations about using prognostic models for COVID-19 patients at the bedside due to concerns about data accuracy and patient-level interpretation of the data. They felt that prediction models may have a greater value if used for resource allocation, triage, or educational purposes. However, patients and relatives indicated that they would have found this information useful for decision making. We found factors related to the health care provider, model, and patient, that might influence the use of prognostic models in practice (Text box 4). This is consistent with the non-adoption, abandonment, scale-up, spread, and sustainability (NASSS) framework, which explains that the implementation of technological health innovations is influenced by factors related to several domains, including the technology (e.g. prediction model) and the adopter (e.g. health care provider and patient).[96] Our results also showed that patients need help putting risk estimates into context and to understand the results. Knowledge about prognostic models among health care providers should be increased to facilitate implementation.

Based on the NASSS complexity assessment tool the implementation of prediction models for heterogenous conditions such as COVID-19 would have several complexities (related to the condition, technology and intended adopter), making successful long-term adoption less feasible if these complexities are not properly managed.[96] Prior studies suggest that models that affect diagnostic or therapeutic decisions are more likely to be adopted in clinical practice. Currently, the NOCOS and COPE models, that were introduced to participants in our study, do not provide prescriptive/

Text box 4: Barriers and facilitators of implementation of prognostic models

	Barriers	Facilitators	Practical implications
Health care provider	Limited knowledge and awareness of prediction models Concerns about data accuracy		Increase knowledge and familiarity among health care providers
Prediction model	Limited relevance Relevant predictors missing	Externally validated prediction models Embeded in Electronic Patient Records	Develop models in consultation with health care providers
Patient and surrogate decision makers		Easy to understand	Design patient information materials with patients and surrogate decision-makers

proscriptive risk thresholds.[46] This means that they do not make treatment recommendations based on a given risk estimate. These risk thresholds would depend on a trade-off between benefits and harms of hospital or ICU admission and will be influenced by availability of resources and social norms. Nevertheless, our study suggests that, the NOCOS and COPE models can support providers, patients with COVID-19, and surrogate decision-makers when making decisions about hospital or ICU admission (**Chapter 10**). In our study, we obtained multiple perspective from participants with different experiences, including health care providers and patients with varying levels of illness severity. Our findings can be used to guide the development and implementation of prediction models for COVID-19 in the future. Furthermore, facilitators and barriers identified in our study can be generalized to prediction models in other fields, including TBI. Implementation and model presentation should be considered early on, for instance, at model development. As facilitators and barriers differ between physicians and patients it is important to include different stakeholders and address their needs.

Limitations

Prognostic research is preferably conducted on data from observational cohort studies, while data from randomized controlled trials can also be used for secondary analysis.[3] In this thesis we mainly used data from the CENTER-TBI study: a large international, multicenter observational study.[21, 41] The richness of CENTER-TBI data allowed us to describe and predict functional outcome using the Glasgow Outcome Scale Extended (GOSE) and Health-Related Quality of Life (HRQoL) using the SF-36v2/SF-12v2 and the QoLIBRI-OS in TBI patients of all severities. The CENTER-TBI study includes mostly patients following mild TBI, reflecting contemporary clinical

practice. In the CENTER-TBI study, patients with lower functional outcome on the GOSE and lower HRQoL were less likely to complete the questionnaires, potentially resulting in a response bias. Furthermore, the SF-36v2/SF-12v2 is not suitable for patients with major cognitive impairment or language difficulties. Thus, in this thesis the most severely disabled patients are likely not represented.

Recommendations for future research and clinical practice

Figure 1 provides an overview of recommendations for future research and clinical practice.

Overall conclusions

The overall aim of this thesis was to increase our knowledge on prediction of outcome in acute care by exploring methodological aspects and applications of prognostic research.

Although many prognostic models are available to predict functional outcome after TBI, the methodological quality is often suboptimal. As poor methodological quality is related to poorer performance in new patients and settings, it is crucial to pursuit adequate methodological quality. Several sources, including books, reporting checklists, and guidelines all emphasize the importance of transparent reporting and careful consideration of the study design and statistical analysis when developing a prediction model. Besides, we emphasize the importance of external validation, preferably across a range of settings, before the use of a model can be considered in practice. External validation can be facilitated when research is performed in collaboration among multiple centers, countries, or research groups.

Novel predictors, such as blood-based biomarkers and repeated HRQoL assessments, can further improve prediction of functional outcome and HRQoL after TBI, over known baseline demographic, clinical and radiology characteristics. Biomarkers can be incorporated in established prognostic models for functional outcome. As the prediction of HRQoL using baseline predictors was found to be challenging, dynamic prediction can be utilized which incorporates repeated HRQoL assessments, to better predict HRQoL after TBI.

Even with high-quality and extensively validated prediction models available, only few models are adopted in clinical practice. Based on qualitative findings, we conclude that multi-level factors, related to the end-users, model, and condition, might influence the use of prognostic models in practice. When clinical use is pursued, these facilitators and barriers should be considered while developing, updating and implementing a prediction model.

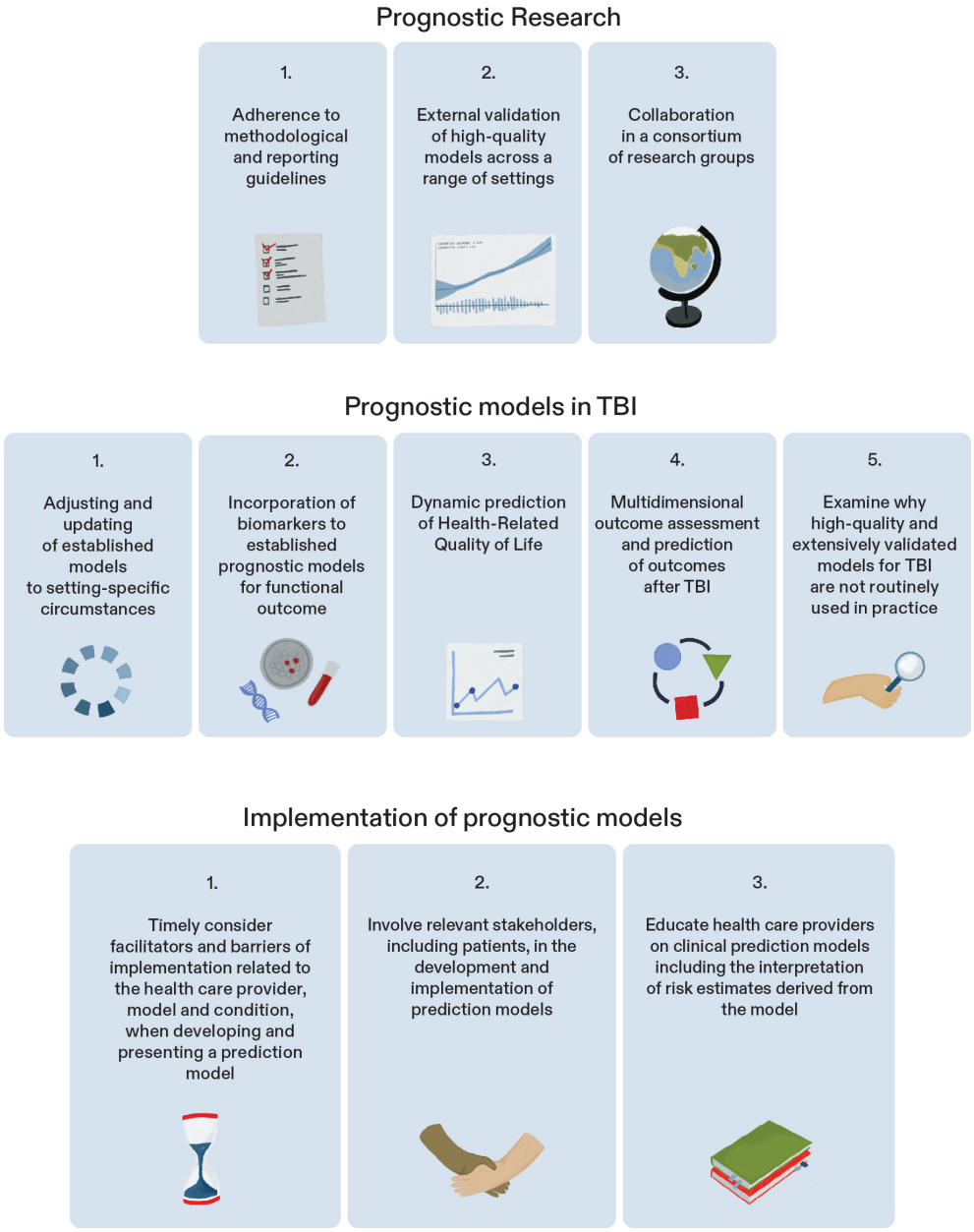


Figure 1: Recommendations for future research and applications of prognostic models in general, prognostic models in TBI and the implementation of prognostic models, based on this thesis.

References

1. Moons K G M, Altman D G, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *Bmj*. 2009;**338**.
2. Steyerberg E W. Clinical prediction models: a practical approach to development, validation, and updating. New York: Springer Science & Business Media; 2019.
3. Dijkland S A, Foks K A, Polinder S, et al. Prognosis in moderate and severe traumatic brain injury: a systematic review of contemporary models and validation studies. *J Neurotrauma*. 2019.
4. Perel P, Edwards P, Wentz R, Roberts I. Systematic review of prognostic models in traumatic brain injury. *BMC medical informatics and decision making*. 2006;**6**(1):38.
5. Bouwmeester W, Zuithoff N P A, Mallett S, et al. Reporting and methods in clinical prediction research: a systematic review. *PLoS Med*. 2012;**9**(5):e1001221.
6. Babyak M A. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosom Med*. 2004;**66**(3):411-21.
7. Venema E, Wessler B S, Paulus J K, et al. Large-scale validation of the Prediction model Risk Of Bias ASsessment Tool (PROBAST) using a short form: high risk of bias models show poorer discrimination. *medRxiv*. 2021.
8. Collins G S, Reitsma J B, Altman D G, Moons K G M. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMC med*. 2015;**13**(1):1.
9. Altman D G, McShane L M, Sauerbrei W, Taube S E. Reporting recommendations for tumor marker prognostic studies (REMARK): explanation and elaboration. *BMC med*. 2012;**10**(1):51.
10. Wolff R F, Moons K G M, Riley R D, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med*. 2019;**170**(1):51-8.
11. Dijkland S, Retel Helmrich I, Steyerberg E. Validation of prognostic models: challenges and opportunities. *Journal of Emergency and Critical Care Medicine*. 2018;**2**(91):1-4.
12. Maas A I, Harrison-Felix C L, Menon D, et al. Common data elements for traumatic brain injury: recommendations from the interagency working group on demographics and clinical assessment. *Archives of physical medicine and rehabilitation*. 2010;**91**(11):1641-9.
13. Yuh E L, Jain S, Sun X, et al. Pathological computed tomography features associated with adverse outcomes after mild traumatic brain injury: a TRACK-TBI study with external validation in CENTER-TBI. *JAMA neurology*. 2021;**78**(9):1137-48.
14. Van Calster B, McLernon D J, Van Smeden M, Wynants L, Steyerberg E W. Calibration: the Achilles heel of predictive analytics. *BMC medicine*. 2019;**17**(1):1-7.
15. Wilson L, Stewart W, Dams-O'Connor K, et al. The chronic and evolving neurological consequences of traumatic brain injury. *The Lancet Neurology*. 2017;**16**(10):813-25.
16. Scholten A C, Haagsma J A, Andriessen T, et al. Health-related quality of life after mild, moderate and severe traumatic brain injury: patterns and predictors of suboptimal functioning during the first year after injury. *Injury*. 2015;**46**(4):616-24.
17. Steyerberg E W, Wiegers E, Sewalt C, et al. Case-mix, care pathways, and outcomes in patients with traumatic brain injury in CENTER-TBI: a European prospective, multicentre, longitudinal, cohort study. *Lancet Neurol*. 2019;**18**(10):923-34.
18. Voormolen D C, Zeldovich M, Haagsma J A, et al. Outcomes after complicated and uncomplicated mild traumatic brain injury at three-and six-months post-injury: results from the CENTER-TBI study. *Journal of clinical medicine*. 2020;**9**(5):1525.
19. Teasdale T W, Engberg A W. Subjective well-being and quality of life following traumatic brain injury in adults: A long-term population-based follow-up. *Brain inj*. 2005;**19**(12):1041-8.

20. Corrigan J D, Bogner J A, Mysiw W J, Clinchot D, Fugate L. Life satisfaction after traumatic brain injury. *J Head Trauma Rehabil*. 2001;**16**(6):543-55.
21. Grauwmeijer E, Heijenbrok-Kal M H, Ribbers G M. Health-related quality of life 3 years after moderate to severe traumatic brain injury: a prospective cohort study. *Archives of Physical Medicine and Rehabilitation*. 2014;**95**(7):1268-76.
22. Wilson L, Boase K, Nelson L D, et al. A manual for the glasgow outcome scale-extended interview. *Journal of neurotrauma*. 2021;**38**(17):2435-46.
23. Polinder S, Haagsma J A, van Klaveren D, Steyerberg E W, Van Beeck E F. Health-related quality of life after TBI: a systematic review of study design, instruments, measurement properties, and outcome. *Popul Health Metr*. 2015;**13**(1):4.
24. Wilde E A, Whiteneck G G, Bogner J, et al. Recommendations for the use of common outcome measures in traumatic brain injury research. *Archives of physical medicine and rehabilitation*. 2010;**91**(11):1650-60. e17.
25. Von Steinbuechel N, Covic A, Polinder S, et al. Assessment of health-related quality of life after tbi: Comparison of a disease-specific (qolibri) with a generic (sf-36) instrument. *Behav Neurol*. 2016;**2016**.
26. Steyerberg E W, Harrell F E. Prediction models need appropriate internal, internal–external, and external validation. *J Clin Epidemiol*. 2016;**69**:245-7.
27. Steyerberg E W, Mushkudiani N, Perel P, et al. Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. *PLoS Med*. 2008;**5**(8):e165.
28. Collaborators M R C C T. Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients. *Bmj*. 2008;**336**(7641):425-9.
29. Maas A I R, Menon D K, Adelson P D, et al. Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. *Lancet Neurol*. 2017.
30. Roozenbeek B, Maas A I R, Menon D K. Changing patterns in the epidemiology of traumatic brain injury. *Nature Reviews Neurology*. 2013;**9**(4):231-6.
31. Peeters W, van den Brande R, Polinder S, et al. Epidemiology of traumatic brain injury in Europe. *Acta neurochirurgica*. 2015;**157**(10):1683-96.
32. Mollayeva T, Mollayeva S, Colantonio A. Traumatic brain injury: sex, gender and intersecting vulnerabilities. *Nature Reviews Neurology*. 2018;**14**(12):711-22.
33. Gardner R C, Dams-O'Connor K, Morrissey M R, Manley G T. Geriatric traumatic brain injury: epidemiology, outcomes, knowledge gaps, and future directions. *Journal of neurotrauma*. 2018;**35**(7):889-906.
34. Damen J A A G, Debray T P A, Pajouheshnia R, et al. Empirical evidence of the impact of study characteristics on the performance of prediction models: a meta-epidemiological study. *BMJ open*. 2019;**9**(4):e026160.
35. Vergouwe Y, Nieboer D, Oostenbrink R, et al. A closed testing procedure to select an appropriate method for updating prediction models. *Stat Med*. 2017;**36**(28):4529-39.
36. Su T-L, Jaki T, Hickey G L, Buchan I, Sperrin M. A review of statistical updating methods for clinical prediction models. *Statistical methods in medical research*. 2018;**27**(1):185-97.
37. Mondello S, Sorinola A, Czeiter E, et al. Blood-based protein biomarkers for the management of traumatic brain injuries in adults presenting to emergency departments with mild brain injury: a living systematic review and meta-analysis. *J Neurotrauma*. 2021;**38**(8):1086-106.
38. Bazarian J J, Biberthaler P, Welch R D, et al. Serum GFAP and UCH-L1 for prediction of absence of intracranial injuries on head CT (ALERT-TBI): a multicentre observational study. *Lancet Neurol*. 2018;**17**(9):782-9.

39. Undén J, Ingebrigtsen T, Romner B. Scandinavian guidelines for initial management of minimal, mild and moderate head injuries in adults: an evidence and consensus-based update. *BMC medicine*. 2013;**11**(1):1-14.
40. Vos P E, Jacobs B, Andriessen T, et al. GFAP and S100B are biomarkers of traumatic brain injury: an observational cohort study. *Neurology*. 2010;**75**(20):1786-93.
41. Zhang Z, Zhang L, Dong X, et al. Comparison of the performances of copeptin and multiple biomarkers in long-term prognosis of severe traumatic brain injury. *Peptides*. 2014;**60**:13-7.
42. Di Battista A P, Buonora J E, Rhind S G, et al. Blood biomarkers in moderate-to-severe traumatic brain injury: potential utility of a multi-marker approach in characterizing outcome. *Front Neurol*. 2015;**6**:110.
43. Gradisek P, Carrara G, Antiga L, et al. Prognostic value of a combination of circulating biomarkers in critically ill patients with traumatic brain injury: results from the European CREACTIVE study. *J Neurotrauma*. 2021;**38**(19):2667-76.
44. Graham N S N, Zimmerman K A, Moro F, et al. Axonal marker neurofilament light predicts long-term outcomes and progressive neurodegeneration after traumatic brain injury. *Sci Transl Med*. 2021;**13**(613):eabg9922.
45. Mondello S, Papa L, Buki A, et al. Neuronal and glial markers are differently associated with computed tomography findings and outcome in patients with severe traumatic brain injury: a case control study. *Crit Care*. 2011;**15**(3):1-10.
46. Takala R S K, Posti J P, Runtti H, et al. Glial fibrillary acidic protein and ubiquitin C-terminal hydrolase-L1 as outcome predictors in traumatic brain injury. *World Neurosurg*. 2016;**87**:8-20.
47. Dadas A, Washington J, Diaz-Arrastia R, Janigro D. Biomarkers in traumatic brain injury (TBI): a review. *Neuropsychiatric disease and treatment*. 2018;**14**:2989.
48. Huie J R, Mondello S, Lindsell C J, et al. Biomarkers for traumatic brain injury: data standards and statistical considerations. *J Neurotrauma*. 2020.
49. Thelin E P, Zeiler F A, Ercole A, et al. Serial sampling of serum protein biomarkers for monitoring human traumatic brain injury dynamics: a systematic review. *Front Neurol*. 2017;**8**:300.
50. Nelson E C, Eftimovska E, Lind C, Hager A, Wasson J H, Lindblad S. Patient reported outcome measures in practice. *Bmj*. 2015;**350**.
51. Øvretveit J, Zubkoff L, Nelson E C, Frampton S, Knudsen J L, Zimlichman E. Using patient-reported outcome measurement to improve patient care. *International Journal for Quality in Health Care*. 2017;**29**(6):874-9.
52. Rivera S C, Kyte D G, Aiyegbusi O L, Slade A L, McMullan C, Calvert M J. The impact of patient-reported outcome (PRO) data from clinical trials: a systematic review and critical analysis. *Health and quality of life outcomes*. 2019;**17**(1):1-19.
53. Shukla D, Devi B I, Agrawal A. Outcome measures for traumatic brain injury. *Clin Neurol Neurosurg*. 2011;**113**(6):435-41.
54. Wilson L, Horton L, Polinder S, et al. Tailoring Multi-Dimensional Outcomes to Level of Functional Recovery after Traumatic Brain Injury. *Journal of neurotrauma*. 2022.
55. Mailhan L, Azouvi P, Dazord A. Life satisfaction and disability after severe traumatic brain injury. *Brain Injury*. 2005;**19**(4):227-38.
56. Jacob L, Cogné M, Tenovuo O, et al. Predictors of access to rehabilitation in the year following traumatic brain injury: a European prospective and multicenter study. *Neurorehabilitation and neural repair*. 2020;**34**(9):814-30.
57. Andelic N, Soberg H L, Berntsen S, Sigurdardottir S, Roe C. Self-perceived health care needs and delivery of health care services 5 years after moderate-to-severe traumatic brain injury. *Pm&r*. 2014;**6**(11):1013-21.

58. Wilson L, Horton L, Kunzmann K, et al. Understanding the relationship between cognitive performance and function in daily life after traumatic brain injury. *Journal of Neurology, Neurosurgery & Psychiatry*. 2020.
59. Voormolen D C, Polinder S, Von Steinbuechel N, Vos P E, Cnossen M C, Haagsma J A. The association between post-concussion symptoms and health-related quality of life in patients with mild traumatic brain injury. *Injury*. 2019;**50**(5):1068-74.
60. Tomberg T, Toomela A, Pulver A, Tikk A. Coping strategies, social support, life orientation and health-related quality of life following traumatic brain injury. *Brain Injury*. 2005;**19**(14):1181-90.
61. Rainey E E, Petrey L B, Reynolds M, Agtarap S, Warren A M. Psychological factors predicting outcome after traumatic injury: the role of resilience. *The American Journal of Surgery*. 2014;**208**(4):517-23.
62. Morton M V, Wehman P. Psychosocial and emotional sequelae of individuals with traumatic brain injury: a literature review and recommendations. *Brain inj*. 1995;**9**(1):81-92.
63. Von Steinbüchel N, Wilson L, Gibbons H, et al. Quality of Life after Brain Injury (QOLIBRI): scale validity and correlates of quality of life. *Journal of neurotrauma*. 2010;**27**(7):1157-65.
64. Forslund M V, Roe C, Sigurdardottir S, Andelic N. Predicting health-related quality of life 2 years after moderate-to-severe traumatic brain injury. *Acta Neurologica Scandinavica*. 2013;**128**(4):220-7.
65. van der Naalt J, Timmerman M E, de Koning M E, et al. Early predictors of outcome after mild traumatic brain injury (UPFRONT): an observational cohort study. *The Lancet Neurology*. 2017;**16**(7):532-40.
66. Silverberg N D, Gardner A J, Brubacher J R, Panenka W J, Li J J, Iverson G L. Systematic review of multivariable prognostic models for mild traumatic brain injury. *Journal of neurotrauma*. 2015;**32**(8):517-26.
67. Knaus W A, Draper E A, Wagner D P, Zimmerman J E. APACHE II: a severity of disease classification system. *Critical care medicine*. 1985;**13**(10):818-29.
68. Moskowitz J, Quinn T, Khan M W, et al. Should we use the IMPACT-model for the outcome prognostication of TBI patients? A qualitative study assessing physicians' perceptions. *MDM policy & practice*. 2018;**3**(1):2381468318757987.
69. Kappen T H, Van Loon K, Kappen M A M, et al. Barriers and facilitators perceived by physicians when using prediction models in practice. *Journal of clinical epidemiology*. 2016;**70**:136-45.
70. Letsinger J, Rommel C, Hirschi R, Nirula R, Hawryluk G W J. The aggressiveness of neurotrauma practitioners and the influence of the IMPACT prognostic calculator. *PLOS one*. 2017;**12**(8):e0183552.
71. Greenhalgh T, Wherton J, Papoutsi C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. *Journal of medical Internet research*. 2017;**19**(11):e367.
72. van Klaveren D, Rekkas A, Alsma J, et al. COVID Outcome Prediction in the Emergency Department (COPE). *medRxiv*. 2021:2020.12. 30.20249023.
73. Maas A I R, Menon D K, Steyerberg E W, et al. Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) A Prospective Longitudinal Observational Study. *J Neurosurg*. 2014;**76**(1):67-80.

Summary

Zie pagina 335-340 voor de Nederlandse samenvatting en pagina 341-345 voor de lekensamenvatting.

Introduction

Over the past 4 years, I have studied prediction of outcomes for patients in acute care. A prediction of the risk of future conditions, such as a patient's health or wellbeing, is called a prognosis. In clinical practice, health care providers frequently aim to predict a future outcome of an individual patient. The ability to accurately predict a patient's outcome is important and has several purposes. A prognosis can be used in communication with patients and relatives, it can support clinical decisions, and it can be used for risk stratification in research and for quality-of-care assessments.

In the hospital, acute care services are provided to a patient with a severe illness or condition. Patients are, for instance, treated briefly for a severe illness or condition that resulted from a disease or trauma at the emergency department or in the intensive care unit. A considerable proportion of this thesis came into being during the COVID-19 pandemic, in which research on COVID-19 emerged rapidly and took priority. Therefore, we will predominantly focus on traumatic brain injury, while also including results from a study on COVID-19 care.

Traumatic brain injury (TBI) can be defined as an injury to the brain induced by an external force and is a major health concern with over 50,000,000 new cases reported globally every year. There has been considerable interest in prognosis following TBI. TBI is said to be one of the most heterogeneous neurological conditions, which makes the prediction of outcome challenging. It is important to identify patients who are at high risk of mortality or long-term consequences. Accurate and reliable prognostic models for outcome prediction after TBI have the potential to support health care providers and patients in making clinical decisions. Improving prognostication has been considered critical by health care providers, researchers, and patients and caregivers alike.

Aim of this thesis

The overall aim of this thesis is to increase our knowledge of prediction of outcome in acute care by exploring methodological aspects (Part I) and applications (Part II) of prognostic research. In this thesis we mainly made use of data from the Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) core study. The CENTER-TBI study is a large prospective observational cohort study including 4509 patients with mild, moderate and severe TBI.

Part I: Methodological aspects

Main results

Part I of this thesis focuses on the methodological aspects of prognostic research. **Chapter 2** describes prediction of outcome following TBI, including prognostic factors and established prognostic models, in detail. **Chapter 3** gives a concise overview of the steps and considerations in prognostic research and introduces the reader to the concept of overfitting, illustrated by two established models in the field of Physiotherapy. In **Chapter 4**, we offer methodological recommendations for prognostic research in TBI using the REMARK and TRIPOD reporting guidelines. In **Chapter 5**, we assessed the methodological quality of model development studies using the PROBAST and examined the relation between methodological quality and model performance at external validation. Of the ten included model development studies, four models were found to have low risk of bias (RoB). On average, the change in discriminative ability was positive in validations of 'low' RoB models meaning that the models performed better at external validation. Conversely, the change in discriminative ability was negative for 'high' RoB models, which means that the models performed worse at external validation. We concluded that lower methodological quality at model development was associated with poorer model performance at external validation.

Discussion

Although many prognostic models are available to predict functional outcome after TBI, the methodological quality is typically suboptimal. As we found that poor methodological quality is related to poorer performance in new patients and settings, it is crucial to pursuit adequate methodological quality. Prognostic research can be improved if key methodological principles are adhered to, which would result in models of higher methodological quality with better model performance in new patients and settings. To ensure that models are of good quality, it is important to consult existing checklists and guidelines. These sources provide information on methodological aspects that should carefully be considered. In Chapters 2-4, we emphasize the importance of transparent reporting and the consideration of the study design and statistical analysis. Furthermore, we stress that a model should be externally validated, preferably across a range of settings, before the use of a model can be considered in practice. External validation can be facilitated when research is performed in collaboration among multiple centers, countries, or research groups.

Part II: Applications

Main results

In Part II we investigate several applications, including external validation, development, and implementation of prognostic models. First, we externally validate established models for predicting mortality and unfavorable outcome after moderate and severe TBI in **Chapter 6.** These models (the IMPACT and CRASH), which were developed on historical data, also performed well in contemporary patients. The models showed good discriminative ability, which improved modestly with the addition of CT variables. There were substantial differences between observed and predicted outcome risk (calibration), specifically for the CRASH CT model. Second, we examined the incremental prognostic value of serum biomarkers over demographic, clinical and radiological characteristics and over established prognostic models for the prediction of functional outcome after TBI in **Chapter 7**. All examined serum biomarkers – UCH-L1, S100B, GFAP, NFL, t-tau, and NSE - obtained within 24h after injury, improved the prognostic value for functional outcome. Combining all six biomarkers resulted in small further increments in C-statistic and R^2, compared to the best performing individual biomarkers separately. We found that UCH-L1 had the greatest incremental prognostic value. Adding biomarkers to the IMPACT and CRASH models resulted in an R^2 up to 45% and 46% for mortality and 37% and 45% for unfavorable outcome, respectively.

The relationship between disability and wellbeing following TBI is examined in **Chapter 8**. Functional outcome assessed with the GOSE, personal, injury-related and environment factors explained a limited amount of variance in HRQoL (up to 29%). In **Chapter 9**, we aimed to identify predictors of, and develop prognostic models for the prediction of Health-Related Quality of Life after TBI, separately for the SF-36v2 PCS and MCS and the QoLIBRI total score. Medical and injury related characteristics were most important for the prediction of the PCS, whereas patient related characteristics were more relevant for prediction of the MCS and the QOLIBRI. The performance of models to predict HRQoL was limited: the proportion of variance explained of the full models was 19% for PCS, 9% for MCS and 13% for the QOLIBRI. In a subset of predominantly patients following mild TBI, including 2-week HRQoL assessment (N=436) improved model performance substantially (R^2 PCS 15% to 37%, MCS 12% to 36%, and QOLIBRI 10% to 48%).

In **Chapter 10,** qualitative analyses are used to explore considerations of health care providers, patients, and surrogate decision makers (e.g. relatives and caregivers) about the use of prediction models to support clinical decision-making in COVID-19 care. Providers had reservations about using prognostic models for COVID-19 patients due

to concerns about data accuracy and patient-level interpretation of the data. However, patients and relatives indicated that they would have found this information useful for decision making. We found factors, related to the health care provider, model, and patient, that might influence the use of prognostic models in practice.

Discussion

As established models may become outdated, it is important to validate these models in new patients and settings. External validation studies show whether the model can also be used in new patients and settings. External validation can be facilitated when researchers from different countries or institutes collaborate and share research data. If the model doesn't perform adequately, it should be adjusted or updated. The variable calibration of the IMPACT and CRASH models reflects heterogeneity in reliability of predictions, which motivates continuous validation and updating when clinical implementation is pursued.

Prognostic models for functional outcome and HRQoL after TBI can be further improved by incorporating additional information. To further improve models for mortality and unfavorable outcome, biomarkers can be incorporated. Biomarkers provide additional information to patient data such as age and severity of symptoms and have incremental prognostic value. Our research shows that the daily functioning of the patient is insufficient to predict HRQoL after TBI. Therefore, to get a complete picture of consequences after TBI, it is important to measure different outcomes, including functional outcome, quality of life, and cognitive functioning, in these patients. We found that the use of solely baseline predictors was suboptimal to predict HRQoL after TBI, and models could be improved by the addition of repeated HRQoL assessments. Therefore, models that predict quality of life can be improved by using information about a patient's quality of life after admission.

Models are generally developed to provide health care providers, patients, and their relatives with information. This information can then be used in conversations between health care providers and patients, and to support treatment decisions. To ensure that models are suitable for use in clinical practice, it is important to involve stakeholders. Healthcare providers need to be educated on prognostic research so that they have sufficient knowledge on how to use prediction models in practice. Future research should also investigate reasons why prediction models for traumatic brain injury patients are often not used in practice.

Samenvatting

Zie pagina 330-334 voor de Engelstalige samenvatting en pagina 341-345 voor de lekensamenvatting.

Inleiding

De afgelopen 4 jaar heb ik mij verdiept in het voorspellen van uitkomsten voor patiënten in de acute zorg. Een voorspelling van het risico op een toekomstige resultaat of verloop van een ziekte, zoals de gezondheid of het welzijn van een patiënt, wordt een prognose genoemd. In de klinische praktijk proberen zorgverleners vaak een toekomstige uitkomst van een individuele patiënt te voorspellen. Het vermogen om de uitkomst van een patiënt nauwkeurig te voorspellen is belangrijk en heeft verschillende doelen. Een prognose kan worden gebruikt in de communicatie met patiënten en familieleden, het kan klinische beslissingen ondersteunen, en het kan worden gebruikt voor risicostratificatie in onderzoek en kwaliteitsbeoordelingen van de zorg.

In het ziekenhuis wordt acute zorg verleend aan patiënten met een ernstige ziekte of aandoening. Patiënten worden bijvoorbeeld kort behandeld voor ziekte of trauma op de spoedeisende hulp of op de intensive care afdeling. Een aanzienlijk deel van dit proefschrift is ontstaan tijdens de COVID-19 pandemie, waarin het onderzoek naar COVID-19 snel opkwam en prioriteit kreeg. Daarom zullen we ons in dit proefschrift naast traumatisch hersenletsel ook richten op COVID-19.

Traumatisch hersenletsel kan worden gedefinieerd als letsel aan de hersenen veroorzaakt door een externe kracht en is een belangrijk gezondheidsprobleem met meer dan 50.000.000 nieuwe gevallen per jaar wereldwijd. Er is veel belangstelling voor de prognose na traumatisch hersenletsel. Traumatisch hersenletsel zou een van de meest heterogene neurologische aandoeningen zijn, waardoor het voorspellen van de uitkomst na traumatisch hersenletsel een uitdaging vormt. Het is belangrijk patiënten te identificeren die een hoog risico lopen op sterfte of lange termijn gevolgen. Nauwkeurige en betrouwbare prognostische modellen voor het voorspellen van de uitkomst na traumatisch hersenletsel hebben de potentie om zorgverleners en patiënten te ondersteunen bij het nemen van klinische beslissingen. Het verbeteren van de prognose wordt door zowel zorgverleners, onderzoekers, en patiënten als cruciaal beschouwd.

Doel van dit proefschrift

Het doel van dit proefschrift is het vergroten van onze kennis over het voorspellen van uitkomsten in de acute zorg door het verkennen van methodologische aspecten

(deel 1) en toepassingen (deel 2) van prognose onderzoek. In dit proefschrift hebben we voornamelijk gebruik gemaakt van gegevens uit de Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) kernstudie. De CENTER-TBI studie is een grote prospectieve observationele cohortstudie met 4509 patiënten met mild, matig en ernstig traumatisch hersenletsel.

Deel I: Methodologische aspecten

Belangrijkste resultaten

Deel I van dit proefschrift richt zich op methodologische aspecten van prognose onderzoek. **Hoofdstuk 2** beschrijft in detail de voorspelling van de uitkomst na traumatisch hersenletsel, inclusief prognostische factoren en bestaande prognostische modellen. **Hoofdstuk 3** geeft een beknopt overzicht van de stappen en overwegingen in prognose onderzoek en introduceert het concept van 'overfitting', geïllustreerd door twee bestaande modellen in de Fysiotherapie. In **hoofdstuk 4** bieden wij methodologische aanbevelingen voor prognose onderzoek in traumatisch hersenletsel aan de hand van de REMARK en TRIPOD-richtlijnen. In **hoofdstuk 5** hebben we de methodologische kwaliteit van modelontwikkelingsstudies beoordeeld met behulp van de PROBAST en de relatie tussen methodologische kwaliteit en prestaties van het model bij externe validatie onderzocht. Van de tien geïncludeerde modelontwikkelingsstudies bleken vier modellen een laag risico op *bias* (RoB) te hebben. Gemiddeld was de verandering in discriminerend vermogen positief bij validaties van modellen met een 'lage' RoB, wat betekent dat de modellen beter presteerden bij externe validatie. Omgekeerd was de verandering in discriminerend vermogen negatief voor modellen met een "hoge" RoB, wat betekent dat de modellen slechter presteerden bij externe validatie. Wij concludeerden dat een lagere methodologische kwaliteit bij de modelontwikkeling gepaard ging met slechtere modelprestaties bij externe validatie.

Discussie

Hoewel er veel prognostische modellen beschikbaar zijn om de functionele uitkomst na traumatisch hersenletsel te voorspellen, is de methodologische kwaliteit doorgaans suboptimaal. Aangezien wij vonden dat slechte methodologische kwaliteit samenhangt met slechtere prestaties in nieuwe patiënten en settings, is het van cruciaal belang om adequate methodologische kwaliteit na te streven. Prognose onderzoek kan worden verbeterd als belangrijke methodologische richtlijnen worden nageleefd, wat zou resulteren in modellen van hogere methodologische kwaliteit met betere modelprestaties bij nieuwe patiënten en in nieuwe settings. Om ervoor te zorgen dat de modellen van goede kwaliteit zijn, is het belangrijk bestaande checklists en richtlijnen te raadplegen. Deze bronnen bieden informatie over methodologische aspecten die zorgvuldig moeten worden overwogen. In de hoofdstukken 2-4 benadrukken wij het

belang van transparante rapportage en de overweging van de onderzoeksopzet en de statistische analyse. Daarnaast benadrukken wij dat een model extern moet worden gevalideerd, bij voorkeur in verschillende settings, voordat het gebruik ervan in de praktijk kan worden overwogen. Externe validatie kan worden vergemakkelijkt wanneer het onderzoek wordt uitgevoerd in samenwerking tussen meerdere centra, landen of onderzoeksgroepen.

Deel II: Toepassingen

Belangrijkste resultaten

In deel II onderzoeken we verschillende toepassingen, waaronder externe validatie, ontwikkeling en implementatie, van prognostische modellen. Ten eerste valideren we bestaande modellen voor het voorspellen van sterfte en ongunstige uitkomst na matige en ernstige traumatisch hersenletsel in **hoofdstuk 6**. Deze modellen (de IMPACT en CRASH), die werden ontwikkeld op basis van relatief oude data, presteerden ook goed voor hedendaagse patiënten. De modellen vertoonden een goed discriminerend vermogen, dat licht verbeterde door toevoeging van CT-variabelen. Er waren aanzienlijke verschillen tussen waargenomen en voorspelde uitkomsten (kalibratie), met name voor het CRASH CT-model. Ten tweede onderzochten wij in **hoofdstuk 7** de incrementele prognostische waarde van serum biomarkers ten opzichte van demografische, klinische en radiologische kenmerken en ten opzichte van bestaande prognostische modellen voor de voorspelling van functionele uitkomst na traumatisch hersenletsel. Alle onderzochte serum biomarkers - UCH-L1, S100B, GFAP, NFL, t-tau, en NSE - verkregen binnen 24 uur na het letsel, verbeterden de prognostische waarde voor functionele uitkomst. Het combineren van alle zes biomarkers resulteerde in kleine verdere toenames in C-statistiek en R^2, vergeleken met de best presterende individuele biomarkers afzonderlijk. Wij vonden dat UCH-L1 de grootste incrementele prognostische waarde had. Het toevoegen van biomarkers aan de IMPACT- en CRASH-modellen resulteerde in een R^2 tot 45% en 46% voor mortaliteit en 37% en 45% voor ongunstige uitkomst, respectievelijk.

De relatie tussen functioneren en welzijn na traumatisch hersenletsel wordt onderzocht in **hoofdstuk 8**. Functionele uitkomst (op basis van de GOSE), persoonlijke, letselgerelateerde en omgevingsfactoren verklaarden een beperkte hoeveelheid variantie in gezondheid gerelateerde kwaliteit van leven (tot 29%). In **hoofdstuk 9** beoogden wij voorspellers van en prognostische modellen te ontwikkelen voor de voorspelling van gezondheid gerelateerde kwaliteit van leven na traumatisch hersenletsel, afzonderlijk voor de SF-36v2 PCS en MCS en de QOLIBRI totaalscore. Medische en letsel gerelateerde kenmerken waren het belangrijkst voor de voorspelling van de PCS, terwijl patiënt gerelateerde kenmerken relevanter waren voor de voorspelling van de MCS

en de QOLIBRI. De prestatie van de modellen om gezondheid gerelateerde kwaliteit van leven te voorspellen was beperkt: het verklaarde deel van de variantie van de volledige modellen was 19% voor PCS, 9% voor MCS en 13% voor de QOLIBRI. In een subset van voornamelijk patiënten na mild traumatisch hersenletsel verbeterde het opnemen van gezondheid gerelateerde kwaliteit van leven op 2 weken na opname (N=436) de modelprestaties aanzienlijk (R^2 PCS 15% naar 37%, MCS 12% naar 36%, en QOLIBRI 10% naar 48%).

In **hoofdstuk 10** worden kwalitatieve methoden gebruikt om de overwegingen van zorgverleners, patiënten en naasten (bv. familieleden) over het gebruik van voorspelmodellen ter ondersteuning van de klinische besluitvorming in de COVID-19-zorg te onderzoeken. Zorgverleners hadden bedenkingen bij het gebruik van voorspelmodellen voor COVID-19 patiënten wegens bezorgdheid over de nauwkeurigheid van de gegevens en de interpretatie van de gegevens voor individuele patiënten. Patiënten en naasten gaven echter aan dat zij deze informatie nuttig zouden hebben gevonden voor de besluitvorming. Wij vonden factoren gerelateerd aan de zorgverlener, het model en de patiënt, die het gebruik van prognostische modellen in de praktijk zouden kunnen beïnvloeden.

Discussie

Aangezien bestaande modellen verouderd kunnen raken, is het belangrijk deze modellen te valideren in nieuwe patiënten en settings. Externe validatiestudies laten zien of het model ook in nieuwe patiënten en settings gebruikt kan worden. Externe validatie kan worden vergemakkelijkt wanneer onderzoekers uit verschillende landen of instituten samenwerken en onderzoeksgegevens delen. Als het model onvoldoende presteert, moet het worden aangepast of geupdate. De variabele kalibratie van de IMPACT- en CRASH-modellen weerspiegelt de heterogeniteit in de betrouwbaarheid van de voorspellingen, hetgeen motiveert voortdurende validatie en aanpassingen wanneer klinische implementatie wordt nagestreefd.

Voorspelmodellen voor functionele uitkomst en gezondheid gerelateerde kwaliteit van leven na traumatisch hersenletsel kunnen verder worden verbeterd door aanvullende informatie op te nemen. Om de modellen voor mortaliteit en ongunstige uitkomst verder te verbeteren, kunnen biomarkers worden opgenomen. Biomarkers geven aanvullende informatie naast patiëntgegevens zoals leeftijd en ernst van de symptomen en hebben incrementele prognostische waarde. Uit ons onderzoek blijkt dat het dagelijks functioneren van de patiënt onvoldoende is om gezondheid gerelateerde kwaliteit van leven na traumatisch hersenletsel te voorspellen. Om een volledig beeld te krijgen van de gevolgen na traumatisch hersenletsel is het daarom belangrijk om bij deze patiënten verschillende uitkomsten te meten, waaronder dagelijks functioneren, kwaliteit van

leven en cognitief functioneren. Wij zagen dat als we de kwaliteit van leven van een patiënt na opname ook wisten, we beter in staat waren om kwaliteit van leven op een later moment na het ongeval te voorspellen. Daarom kunnen modellen die kwaliteit van leven voorspellen worden verbeterd door informatie over de kwaliteit van leven van een patiënt na opname te gebruiken.

Modellen worden doorgaans ontwikkeld om zorgverleners, patiënten en hun naasten van informatie te voorzien. Deze informatie kan dan worden gebruikt in gesprekken tussen zorgverleners en patiënten, en ter ondersteuning van beslissingen over de behandeling. Om ervoor te zorgen dat de modellen geschikt zijn voor gebruik in de klinische praktijk, is het belangrijk belanghebbenden te betrekken. Zorgverleners moeten worden voorgelicht over Prognose onderzoek, zodat zij voldoende kennis hebben over het gebruik van predictiemodellen in de praktijk. In toekomstig onderzoek moet ook worden nagegaan waarom voorspellingsmodellen voor patiënten met traumatisch hersenletsel vaak niet in de praktijk worden gebruikt.

Lekensamenvatting

Inleiding

De afgelopen 4 jaar heb ik onderzoek gedaan naar het voorspellen van uitkomsten voor patiënten in de acute zorg. Zo'n voorspelling over de gezondheid of welzijn van een patiënt noemen we ook wel een prognose. In de acute zorg worden patiënten vaak kort behandeld, bijvoorbeeld op de spoedeisende hulp of Intensive Care, voor een ernstig letsel of aandoening. Zo keek ik met name naar patiënten die traumatisch hersenletsel hadden opgelopen, bijvoorbeeld na een verkeersongeluk of val. Ook onderzocht ik patiënten die in het ziekenhuis lagen na besmetting met het Coronavirus.

Om te kunnen voorspellen hoe de gezondheid of welzijn van een patiënt na bijvoorbeeld hersenletsel zal zijn, ontwikkelen we modellen. Een model kun je zien als een rekenmachine waar je verschillende gegevens van de patiënt in stopt. Het resultaat is dan de kans op een bepaalde uitkomst. Zo'n model zegt dan bijvoorbeeld: "Op basis van de leeftijd, en ernst van de klachten van deze patiënt, is de kans op overlijden 6 maanden na traumatisch hersenletsel 10%". Vervolgens is het aan de zorgverlener om te bepalen wat ze met deze informatie doen. Zo kan de zorgverlener er bijvoorbeeld voor kiezen om deze informatie te bespreken met de patiënt en zijn/haar familie. Of om deze informatie te gebruiken om een keuze te maken over de behandeling.

In mijn onderzoek richtte ik me op het vergroten van onze kennis over het voorspellen van uitkomsten in de acute zorg. Hierbij keek ik welke onderdelen ('methodologische aspecten') belangrijk zijn om te zorgen dat modellen van goede kwaliteit zijn. Daarnaast heb ik bestaande modellen getest voor nieuwe patiënten, nieuwe modellen ontwikkeld, en onderzoek gedaan naar het mogelijke gebruik van voorspelmodellen.

Belangrijkste resultaten

Omdat er behandelbeslissingen genomen kunnen worden op basis van een voorspelmodel, is het belangrijk dat deze modellen betrouwbaar en precies zijn. Maar uit ons onderzoek blijkt dat de kwaliteit van de modellen vaak niet goed is (**Hoofdstuk 5**). Een gevolg hiervan is dat de voorspelling van het model niet precies of zelfs onjuist is als we dit testen voor nieuwe patiënten (**Hoofdstuk 5**). Om te zorgen dat de kwaliteit van een model goed is kunnen bestaande 'checklists' of richtlijnen worden gebruikt (**Hoofdstuk 2-5**). Deze 'checklists' en richtlijnen kunnen gebruikt worden als stappenplan voor het maken en presenteren van een model.

Allereerst heb ik modellen die eerder zijn ontwikkeld voor het voorspellen van sterfte en herstel na traumatisch hersenletsel getest voor nieuwe patiënten (**Hoofdstuk 6**). Deze modellen (met de naam IMPACT en CRASH), die zijn ontwikkeld op verou-

derde data, bleken het ook goed te doen voor patiënten van nu. Maar de modellen waren soms te negatief over de uitkomst van een patiënt. Als men de modellen wil gebruiken wordt daarom aangeraden om het model eerst aan te passen zodat het geschikt is voor nieuwe patiënten. Op basis van dit onderzoek blijkt ook dat we met een klein aantal patiëntengegevens (leeftijd, en ernst van de klachten) sterfte en herstel na hersenletsel al goed kunnen voorspellen. Extra informatie op basis van bijvoorbeeld een hersenscan (CT-scan) voegen redelijk weinig informatie toe.

Vervolgens keek ik hoe we voorspellingen voor sterfte en herstel van patiënten na traumatisch hersenletsel kunnen verbeteren (**Hoofdstuk 7**). Hierbij onderzochten we de toegevoegde waarde van stofjes die gemeten worden in het bloed ('biomarkers'). Deze stofjes komen vrij na hersenletsel en zijn een graadmeter voor de ernst van de klachten van een patiënt. Uit het onderzoek blijkt dat als je naast patiëntengegevens zoals leeftijd ook de 'biomarkers' weet, je sterfte en herstel beter kunt voorspellen. Ook de eerder onderzochte IMPACT en CRASH modellen verbeterde als we 'biomarkers' toevoegde.

Traumatisch hersenletsel heeft vaak ernstige gevolgen voor het dagelijks leven van een patiënt. Het kan zijn dat een patiënt na het oplopen van hersenletsel minder zelfstandig is: hij/zij kan bijvoorbeeld niet meer werken, is minder mobiel, en heeft hulp nodig in het huishouden. Vaak wordt gedacht dat dit betekent dat een patiënt ook een slechte kwaliteit van leven heeft. Tegen deze verwachting in zien we in ons onderzoek dat er patiënten zijn met beperkingen in het dagelijks leven én een goede kwaliteit van leven (**Hoofdstuk 8**). Wij concluderen dat deze tegenstelling vaak voorkomt na een traumatisch hersenletsel, met name in patiënten die na het ongeval 'mild' letsel hadden. Het dagelijks functioneren van de patiënt is dus niet voldoende om kwaliteit van leven te voorspellen na hersenletsel.

Vervolgens ontwikkelde ik modellen om de kwaliteit van leven na traumatisch hersenletsel te voorspellen (**Hoofstuk 9**). Voor kwaliteit van leven zijn namelijk geen modellen beschikbaar. Daarnaast is het belangrijk om patiënten te identificeren die een hoger risico lopen op een lage kwaliteit van leven na hersenletsel en dus meer ondersteuning nodig zullen hebben. Uit het onderzoek blijkt dat het maken van deze voorspelling erg lastig was als we alleen patiëntengegevens hadden die bekend waren bij de ziekenhuisopname. Denk hierbij bijvoorbeeld aan het opleidingsniveau, de mentale gezondheid en wat voor werk de patiënt doet. Als we de kwaliteit van leven van een patiënt na opname ook wisten, dan waren we beter in staat om kwaliteit van leven op een later moment na hersenletsel te voorspellen.

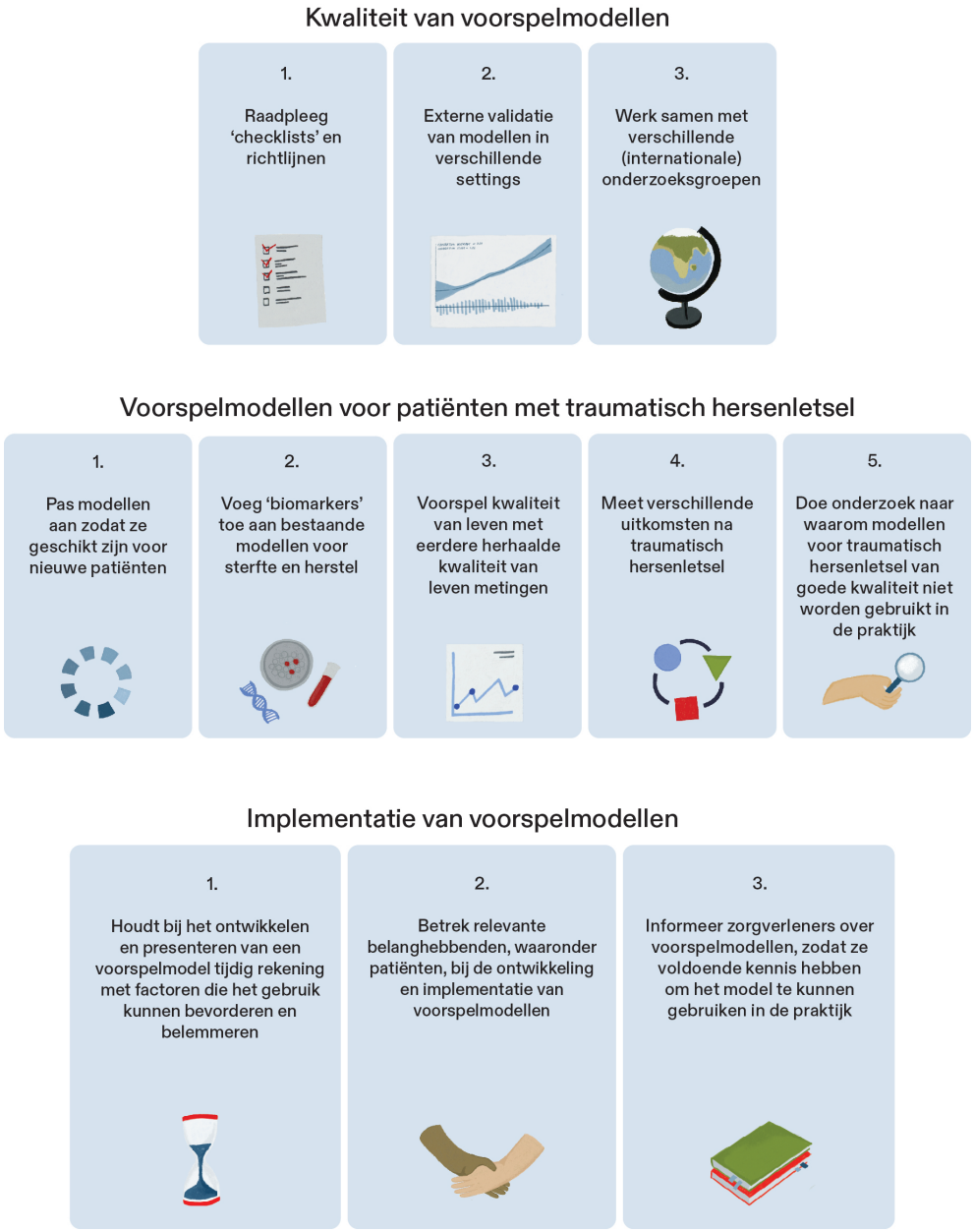
In het begin van de Coronapandemie was het onduidelijk welke patiënten na opname in het ziekenhuis een grote kans hadden om te overlijden. Modellen kunnen zorgverleners en patiënten helpen door het voorspellen van sterfte na opname in het ziekenhuis met het Coronavirus. Mijn collega's in Nederland en Amerika hebben daarom zulke voorspelmodellen ontwikkeld. Samen met collega's ging ik in gesprek met zorgverleners, patiënten die waren opgenomen in het ziekenhuis voor Corona en hun naasten. Wij vroegen hen of zij deze modellen willen gebruiken in de behandelkamer én waarom dan wel of niet (**Hoofdstuk 10**). Ondanks dat er veel onderzoek wordt gedaan naar voorspelmodellen worden de meeste modellen uiteindelijk niet gebruikt in de praktijk. Uit de gesprekken met zorgverleners, patiënten en hun naasten blijkt dat daar een aantal redenen voor zijn (**Hoofdstuk 10**). Zorgverleners geven bijvoorbeeld aan dat ze weinig kennis hebben over voorspelmodellen en ze daarom niet zouden gebruiken. Soms wantrouwen ze de modellen omdat ze niet goed weten hoe het model is ontwikkeld. Patiënten en hun familieleden vinden een voorspelling op basis van een model nuttig. Maar dan moet de zorgverlener het model wel in begrijpelijke taal kunnen uitleggen.

Aanbevelingen voor onderzoek en de praktijk

Op basis van het onderzoek dat ik in de afgelopen 4 jaar heb gedaan kunnen we aanbevelingen doen voor toekomstig onderzoek en de praktijk (Figuur 1).

Om te zorgen dat modellen van goede kwaliteit zijn, is het belangrijk om bij de ontwikkeling ervan bestaande 'checklists' en richtlijnen te raadplegen. Deze bronnen geven informatie over onderdelen waar je op moet letten en kunnen dienen als een stappenplan. Het is bijvoorbeeld belangrijk dat er voldoende patiënten deelnemen aan het onderzoek en dat een beperkt aantal patiëntengegevens wordt gebruikt in het model. Omdat een model verouderd kan zijn of ongeschikt voor bepaalde patiënten is het belangrijk om bestaande modellen te testen voor nieuwe patiënten. Bijvoorbeeld patiënten met een hogere leeftijd of die wonen in een ander land. Dit laat zien of het model voor deze nieuwe patiënten ook gebruikt kan worden. Wanneer onderzoekers van verschillende landen of instituten samenwerken is het makkelijker om een model voor nieuwe patiënten te testen door het delen van onderzoeksdata. Als het model het niet goed doet kan het worden aangepast zodat het wel geschikt is voor nieuwe patiënten.

Om modellen voor sterfte en herstel verder te verbeteren kunnen 'biomarkers' worden toegevoegd. Deze stofjes gemeten in het bloed geven extra informatie naast patiëntengegevens zoals leeftijd en ernst van de klachten, en helpen om een betere voorspelling te doen. Daarnaast blijkt uit ons onderzoek dat het dagelijks functioneren van de patiënt niet voldoende is om kwaliteit van leven te voorspellen na hersenletsel. Om



Figuur 1: Aanbevelingen voor toekomstig onderzoek en de praktijk

een volledig beeld te krijgen van gevolgen na traumatisch hersenletsel is het daarom belangrijk om verschillende uitkomsten te meten, waaronder dagelijks functioneren, kwaliteit van leven en cognitief functioneren, in deze patiënten. Als we de kwaliteit van leven van een patiënt na opname ook wisten, dan waren we beter in staat om kwaliteit van leven op een later moment na het ongeval te voorspellen. Daarom kunnen modellen die kwaliteit van leven voorspellen worden verbeterd door informatie over de kwaliteit van leven van een patiënt na opname te gebruiken.

Modellen worden ontwikkeld om zorgverleners, patiënten en naasten meer informatie te geven. Deze informatie kan vervolgens gebruikt worden in gesprekken tussen zorgverleners, patiënten en hun naasten, of om behandelbeslissingen te maken. Om te zorgen dat modellen geschikt zijn om te gebruiken in de praktijk is het belangrijk om belanghebbende te betrekken. Ook moeten zorgverleners worden opgeleid zodat ze voldoende kennis hebben om voorspelmodellen te gebruiken in de behandelkamer. In de toekomst moet ook onderzoek gedaan worden naar redenen waarom voorspelmodellen voor traumatisch hersenletsel patiënten vaak niet worden gebruikt in de praktijk.

Dankwoord

In 2018 begon ik als junior onderzoeker op de afdeling Maatschappelijke Gezondheidszorg in het Erasmus MC. Tijdens het afronden van mijn Masterprogramma in Maastricht had ik mijn zinnen gezet op een baan in het onderzoek, en dan specifiek het Erasmus MC. Het afronden van mijn proefschrift en alles wat ik tijdens dit proces geleerd heb is met name te danken aan de mensen om me heen.

Mijn promotieteam

Professor Lingsma, **Hester**, ik had mijn allereerste sollicitatiegesprek met jou. Wat gehaast kwam je de keuken op de 23ste verdieping binnen gerend, morste je een halve kop koffie en bood je je excuses aan dat het gesprek wat later zou beginnen. Ondanks de zenuwen voelde ik me snel bij jou op mijn gemak. Jij en Maryse stelden vragen en bij ieder antwoord knikte je aanmoedigend, stelde je kritische vragen, en stemde je vervolgens enthousiast in. De initiële chaos, de manier waarop je anderen direct op hun gemak stelt en jouw aanmoedigende geknik afgewisseld met scherpe vragen waren typisch Hester leerde ik later. Je bent een gewaardeerde collega, sparringpartner en coauteur. Ik waardeer onze gesprekken over politiek in de academische wereld, carrièreadvies, en dat je me het gevoel gaf dat je altijd achter me stond. Van jou leerde ik het belang van samenwerken en het onderhouden van (inter-)nationale contacten. Keer op keer drukte je me op het hart dat ik vooral moest focussen op waar ik goed in ben en niet zozeer op de dingen die me minder goed afgaan, en aan dit advies denk ik vaak terug.

Dr. Van Klaveren, **David**. Wij leerden elkaar ongeveer een jaar na aanvang van mijn aanstelling bij MGZ kennen. In het begin kon ik je wat lastig peilen en, gaf je later toe, jij mij. Wekelijks bespraken we de voortgang en al mijn vragen. 'Thank god for David' zuchtte Ana en ik vaak na lastige gesprekken over data en modellen. We wisten dat we bij errors in R, lastige statistiek of een wollig manuscript altijd bij jou terecht konden. Ondanks dat we best verschillend zijn, bouwde we snel een vertrouwensband op. Tijdens onze overleggen en wandelingen hadden we het ook veel over thuis, jouw gezin, en al mijn huisdieren. Hoewel ik veel aan mezelf twijfelde leek jij dat geen moment te doen. David, dank voor alles, maar nog het meest voor jouw eindeloze vertrouwen.

Professor Steyerberg, **Ewout**, vanaf dag 1 realiseerde ik me hoeveel ik van jou kan leren. Je kwam vaak met voorstellen voor artikelen of samenwerkingen. "Dat kan jij! Je bent toch hartstikke goed!" riep je als ik niet meteen enthousiast reageerde. Je beschreef mijn wetenschappelijke interesses soms als 'soft' en 'psychologisch' (en klonk daar tot mijn verbazing niet erg enthousiast over), maar steunde uiteindelijk

altijd mijn ambities. Na onze overleggen op vrijdagen in het NA-gebouw ging ik met een lange to-do lijst, waardevol commentaar, en een hoofd vol vragen aan de slag. Ik waardeer je kritische blik, oplossingsgerichtheid en de manier waarop je je inzet om voor anderen kansen te creëren.

Het CENTER-TBI project

Mijn dank gaat uit naar het CENTER-TBI project en alle betrokkenen. Hier kreeg ik de mogelijkheid om me als onderzoeker te ontwikkelen en kreeg ik een uniek kijkje achter de schermen bij een groot, complex, internationaal onderzoeksproject. In het bijzonder dank aan professor Andrew Maas, professor David Menon, Veronique de Keizer, en professor Lindsay Wilson voor de samenwerking en ondersteuning.

Professor Maas, **Andrew**, een e-mail van jou opende ik met gemengde gevoelens: Aan de ene kant keek ik op tegen de (zeer hoge) kans dat ik al mijn analyses opnieuw moest doen, of in ieder geval sterk zou worden aangeraden om een aanzienlijk aantal subgroep en sensitiviteitsanalyses toe te voegen. Aan de andere kant was er de zekerheid dat na het verwerken van jouw scherpe feedback, inclusief 'English language check', het manuscript altijd naar een hoger niveau steeg. Dank voor jouw kritische blik, oog voor detail en dat je ieder uur van de dag beschikbaar was voor vragen.

Professor Wilson, **Lindsay**, thank you for supporting my interest in the 'disability paradox' and wellbeing of TBI patients. Me and the manuscript benefited a lot from your critical appraisal, kind and helpful comments and our in-depth discussions over Skype. You were one of my most valued co-authors (if I am allowed to say so) as I treasure your insights, attention to detail and you were always first to reply to my manuscript with extensive feedback. I appreciate the opportunities we had to work together, and I wish I could have visited you at the University of Stirling, Scotland.

De maatschappelijke gezondheidszorg afdeling

Dank aan al mijn collega's op de afdeling Maatschappelijke Gezondheidszorg (MGZ). In mijn begindagen bij MGZ voelde ik me erg welkom door het groot aantal junioren. Ik was onder de indruk van hoe betrokken en geïnteresseerd collega's waren in elkaars onderzoek. De afdeling voelt, ondanks het grote aantal onderzoekers en de verschillende expertise gebieden, als één team. **Simone**, dank voor al jouw hulp bij mijn eerste artikel en het wegwijs maken op de afdeling. Ook wil ik mijn CMB collega's, en in het bijzonder **Eveline, Jilske, Arvind, Rana, Ursula, Caroline, Nikki, Shannon, Margrietha, Ellen, Jasper en Jordi**, bedanken voor jullie interesse in mijn werk, alle emotionele support en aanmoedigingen, en natuurlijk de vele borrels en etentjes (waar ik zelfs na mijn vertrek nog bij mag zijn!).

Bij MGZ kreeg ik de kans om met collega's van verschillende secties samen te werken, waaronder met **dr. Suzanne Polinder, dr. Juanita Haagsma en dr. Judith Rietjens**. **Suzanne**, dank voor jouw feedback en ondersteuning. Met name in mijn begindagen bij MGZ en bij het schrijven van mijn eerste CENTER-TBI artikel hebben we veel gesproken over onderzoek naar kwaliteit van leven van patiënten met traumatisch hersenletsel. **Juanita**, in mijn laatste jaar bij MGZ kreeg ik de mogelijkheid jou te ondersteunen bij analyses naar het voorspellen van kwaliteit van leven bij patiënten in de BIOS studie. Je gaf me altijd het gevoel dat je mijn bijdrage waardeerde en ik keek iedere keer weer uit naar onze overleggen. **Judith**, ik was heel dankbaar toen ik de kans kreeg om met jou samen te werken. Je ondersteunde me bij het opzetten van kwalitatief onderzoek en we hadden het regelmatig over hoe we effectief samen kunnen werken binnen een internationaal team. In deze fase waar ik veel nadacht over wat ik na mijn PhD zou willen doen was het waardevol om met jou te sparren over mogelijke vervolg stappen.

In het bijzonder wil ik **Judith Spek**, **Kees Noordsij-Waagenaar**, **Olaf Donkervoort**, **Petra de Vries**, en **Kai Rock Ho** bedanken voor hun eindeloze geduld en ondersteuning bij administratieve, organisatorische en IT-gerelateerde vragen.

During the COVID pandemic I had the pleasure to work with the team from Tufts medical center and Northwell Health, including **dr. Negin Hajizadeh, dr. Melissa Basile, dr. Theo Zanos, Jinny Park and Professor David Kent**. Professor Kent, **David**, thank you for welcoming me into your team, of which you made me feel like a valued member from the start. I enjoyed our weekly discussions about the COVID-19 models that DvK and Theo have developed and validated. **Jinny**, thank you for your support and diligent agenda and minutes. As most meetings started with Professor Kent calling out your name, it was perfectly clear that without you we would have been utterly lost. **Melissa**, I very much enjoyed our weekly meetings and in particular our in-depth discussion of the analyses and results of the focus groups. Thank you for being patient and supportive and for giving me many opportunities to learn. **Negin**, thank you for sharing your knowledge about the US health system and COVID-19 care. I appreciated your critical questions and the positive energy you brought to each meeting.

De Klinische Genetica afdeling en de DNA dialogen

Ook wil ik mijn 'nieuwe' collega's op de afdeling Klinische Genetica van het Erasmus MC en alle betrokkenen bij het consortium de DNA dialogen, en in het bijzonder, **Sam Riedijk, Diewertje Houtman, Boy Vijlbrief, Wendy Geuverink en Jeanne Arnold**, bedanken. Vanaf mijn allereerste werkdag geven jullie mij het gevoel dat ik een gewaardeerd lid ben van het team. Ik ben onwijs dankbaar dat ik de mogelijkheid

krijg me met jullie hulp verder te ontwikkelen als projectmanager en onderzoeker. Jullie inspireren me iedere dag om dingen net iets anders te doen en met andere ogen naar onderzoek en de academische wereld te kijken.

Mijn paranimfen

Dear Ana and Ernest, you are both colleagues that turned into close friends fast. I am very thankful that you have been by my side every step of the way. Without you I wouldn't have held it possible to finish this thesis.

Ana, the first time we met was on the 15th of August 2018 in the hallway of the NA-building. I remember thinking you were sweet, shy, even quiet, at first: Oh, little did I know. I cherish our long talks (which have continued via voice messages now you moved to Canada), dinner dates, intense workouts, relaxing yoga sessions, and how you always managed to make me reach my 10.000 steps a day, easily. I appreciate the diligent and critical way you analyzed CENTER-TBI data, your adventures spirit inspired me to take more trips and balance work and life better, and I learned a lot from our discussions about work, career and life. I miss you every day and I cannot wait for you to move back to the Netherlands.

Ernest, wanneer we ons samen in een kantoorruimte bevonden wist je; vandaag zal geen productieve dag worden. Wel vond ik dit stiekem de leukste dagen. David herkende onze gezamenlijke aanwezigheid aan 'gegiechel' (zijn woorden, en zeker niet de manier waarop ik onze inhoudelijke professionele discussies zou omschrijven). Als we het over werk hadden stelden we samen belangrijke mails op en bespraken we de politiek van de academische wereld. Ik neem graag voorbeeld aan jouw assertiviteit, gezonde dosis zelfvertrouwen en positieve instelling en ik ben dankbaar dat je in mijn leven bent.

Dana, I mention you here because you feel like my third paranimf. I remember that Ana and I were asked to welcome you on your first day at the CMB section and we were so excited to have someone new join our team. A few years later I am grateful that we got to spend time together (despite COVID, both of us only rarely making it into the office, and you often spontaneously leaving the country), and that we got to know each other. I really value our walks (with and without Yuki) around the Kralingse plas, and our long talks. During presentations I could always count on seeing your face in the crowd smiling at me encouragingly! Thank you for always making time for me and your endless support.

Mijn vrienden en familie

De cover en de illustraties die te vinden zijn in de introductie en discussie van mijn proefschrift zijn geïllustreerd door Veerle van Herk. **Veerle**, ik vond het onwijs waardevol om samen met jou te sparren over de kern van mijn werk en wat ik graag over zou willen brengen. We hebben met name veel gesproken over de cover in een periode waarin ik stiekem helemaal niet met mijn proefschrift bezig wilde zijn. Onze gesprekken gaven mij nieuwe energie om het af te ronden. Ik was onder de indruk van hoe je mijn gedachten in beeld hebt weten te vangen. Jij bracht mijn onderzoek tot leven en hier ben ik je heel dankbaar voor.

Daarnaast zijn er een hele hoop lieve mensen in mijn leven bij wie ik in de eerste plaats dochter, zus, nichtje, vriendin, en partner ben. In het bijzonder wil ik **Ilse, Jelena, Linet, Nienke** en **Ellis** bedanken. Jullie zijn, ieder op jullie eigen manier, een belangrijk en onmisbaar onderdeel van mijn leven. **Ilse**, dank voor al onze gesprekken, jouw openheid en natuurlijk carrièreadvies. Je helpt me om dichter bij mezelf te blijven en mijn kwetsbaarheid te omarmen. **Jelena**, dank voor jouw eindeloze support (sinds groep 4!) en dat je altijd het beste in me ziet, ook als ik dat zelf even niet meer zie. **Nienke**, dank voor jouw vriendschap (die na een intermezzo voor mij nog waardevoller is), de vele koppen thee, en dat we altijd alles kunnen bespreken. **Linet**, thank you for everything; for only being one phone call away, for being my most trusted advisor, and being there every step of the way. **Ellis**, jouw vertrouwen in mij is altijd grenzeloos geweest en dit is volledig wederzijds. Dank dat je me eraan herinnerd dat ik volledig mezelf mag zijn (goed voorbeeld doet volgen!).

Lieve mama, **Anita**, bij iedere (kleine of grote) beslissing in mijn leven denk ik allereerst aan jou. Jij leerde me om mijn eigen plan te trekken, te zorgen voor de mensen om me heen, maar ook jezelf niet te vergeten. Ik ben heel dankbaar dat ik jou en Piet heb om op te bouwen.

Lieve papa, **John**, met jouw bodemloze energie en enthousiasme moedig je me aan om met optimisme naar het leven te kijken. Jij staat altijd open voor nieuwe ervaringen van nieuwe restaurants en voorstellingen tot stedentrips. Dank dat jij en Cathy mij meenemen naar wat achteraf hele bijzondere ervaringen blijken te zijn die ik niet had willen missen.

Nino en Cas, ik ben zo trots om jullie 'grote' zus te zijn. Jullie leren me anders naar het leven te kijken. In de praktijk komt dit erop neer dat ik dit vaak wat minder serieus mag nemen en uitdagingen aan moet gaan. Ik volg hierin (soms met tegenzin) jullie voorbeeld.

Olaf, mijn persoonlijke chef-kok, nummer één cheerleader, grote liefde en mijn thuis. Over de telefoon op station Tilburg overtuigde jij mij om 'ja' te zeggen op een baan als junior onderzoeker (wat uiteindelijk toch gewoon een PhD bleek te zijn). Zonder jou was dit proefschrift er nooit gekomen, mogelijk omdat ik niet het zelfvertrouwen had gehad om ervoor te gaan of omdat ik was omgekomen van de honger. Ik ben je heel dankbaar, maar nog het meest omdat ik jou heb om bij thuis te komen.

List of publications

In this thesis

- **Retel Helmrich, I. R. A.,** Steyerberg, E. W., Maas, A. I. R., Lingsma, H. F. (2022). Sequelae and Outcome in Traumatic Brain Surgery: Prognosis after Traumatic Brain Injury. In Winn, H. R. (Ed.). *Youmans and Winn Neurological Surgery E-Book*. Elsevier Health Sciences.
- **Retel Helmrich, I. R. A.,** van Klaveren, D., & Steyerberg, E. (2019). Research Note: Prognostic model research: overfitting, validation and application. *Journal of physiotherapy*.
- **Retel Helmrich, I. R. A.,** Lingsma, H. F., Turgeon, A. F., Yamal, J. M., & Steyerberg, E. W. (2020). Prognostic Research in Traumatic Brain Injury: Markers, Modeling, and Methodological Principles. *Journal of neurotrauma*.
- Dijkland, S. A., **Retel Helmrich, I. R. A.,** Nieboer, D., Van Der Jagt, M., Dippel, D. W., Menon, D. K., ... & CENTER-TBI Participants and Investigators. (2021). Outcome prediction after moderate and severe traumatic brain injury: external validation of two established prognostic models in 1742 European patients. *Journal of Neurotrauma*, 38(10), 1377-1388.
- **Retel Helmrich, I. R. A.,** van Klaveren, D., Dijkland, S. A., Lingsma, H. F., Polinder, S., Wilson, L., ... & Steyerberg, E. W. (2021). Development of prognostic models for Health-Related Quality of Life following traumatic brain injury. *Quality of life research*, 1-21.
- **Retel Helmrich, I. R. A.,** Mikolić, A., Kent, D. M., Lingsma, H. F., Wynants, L., Steyerberg, E. W., & van Klaveren, D. (2022). Does poor methodological quality of prediction modeling studies translate to poor model performance? An illustration in traumatic brain injury. *Diagnostic and prognostic research*, *6*(1), 1-12.
- **Retel Helmrich, I. R. A.,** van Klaveren, D., Andelic, N., Lingsma, H., Maas, A., Menon, D., ... & Wilson, L. (2022). Discrepancy between disability and reported well-being after traumatic brain injury. *Journal of Neurology, Neurosurgery & Psychiatry*.
- **Retel Helmrich, I. R. A.,** Czeiter, E., Amrein, K., Büki, A., Lingsma, H. F., Menon, D. K., ... & CENTER-TBI participants and investigators. (2022). Incremental prognostic value of acute serum biomarkers for functional outcome after traumatic brain injury (CENTER-TBI): an observational cohort study. *The Lancet Neurology*, *21*(9), 792-802.
- Basile, M. J., **Retel Helmrich, I. R. A.,** Park, J. G., Polo, J. Rietjens, D. van Klaveren, T. Zanos, J. Nelson, H.F. Lingsma, D.M. Kent, J. Alsma, R.J.C.G. Verdonschot, and N. Hajizadeh. (2023). US and Dutch Perspectives on the Use of COVID-19 Clinical Prediction Models: Findings from a Qualitative Analysis. *Medical Decision Making*, 0272989X231152852.

Other publications

- Dijkland, S., **Retel Helmrich, I. R. A.,** & Steyerberg, E. (2018). Validation of prognostic models: challenges and opportunities. *Journal of Emergency and Critical Care Medicine*, *2*(91), 1-4.
- Mikolić, A., Polinder, S., **Retel Helmrich, I. R. A.,** Haagsma, J.A., & Cnossen, M.C. (2019). Treatment for posttraumatic stress disorder in patients with a history of traumatic brain injury: a systematic review. *Clinical psychology review*, *73*, 101776.
- Gravesteijn, B., Krijkamp, E., Busschbach, J., Geleijnse, G., **Retel Helmrich, I. R. A.**, Bruinsma, S., ... & van Vledder, M. (2021). Minimizing population health loss in times of scarce surgical capacity during the coronavirus disease 2019 crisis and beyond: a modeling study. *Value in Health*, *24*(5), 648-657.
- Mikolić, A., Polinder, S., Steyerberg, E. W., **Retel Helmrich, I. R. A.**, Giacino, J. T., Maas, A. I., ... & Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) Study Participants and Investigators. (2021). Prediction of global functional outcome and post-concussive symptoms after mild traumatic brain injury: external validation of prognostic models in the Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) study. *Journal of neurotrauma*, *38*(2), 196-209.
- Van der Vlegel, M., Mikolić, A., Hee, Q.L., Kaplan, Z. R., **Retel Helmrich, I. R. A.,** van Veen, E., ... & Dulière, G.L. (2022). Health care utilization and outcomes in older adults after Traumatic Brain Injury: a CENTER-TBI study. *Injury*.
- Thomas, I., Dickens, A. M., Posti, J. P., Czeiter, E., Duberg, D., Sinioja, T., ... & Orešič, M. (2022). Serum metabolome associated with severity of acute traumatic brain injury. *Nature communications*, *13*(1), 1-15.
- Gravesteijn, B., Krijkamp, E., van Alphen, A., Busschbach, J., Geleijnse, G., **Retel Helmrich, I. R. A**., ... & de Jong, R.B. (2022). Update on a Model to Minimize Population Health Loss in Times of Scarce Surgical Capacity During the COVID-19 Crisis and Beyond. *Value in Health*, *25*(5), 785-786.
- Van Klaveren, D., Zanos, T. P., Nelson, J., Levy, T. J., Park, J. G., **Retel Helmrich, I. R. A**., ... & Kent, D. M. (2022). Prognostic models for COVID-19 needed updating to warrant transportability over time and space. *BMC medicine*, *20*(1), 1-11.
- Wu, X., Sun, Y., Xu, X., Steyerberg, E., **Retel Helmrich, I. R. A.,** Lecky, F., ... & Jiang, J. (2023). Mortality prediction in severe traumatic brain injury using traditional and machine learning algorithms. *Journal of Neurotrauma*.

PhD portfolio

Activity	Period	EC's
Courses on general academic skills		
Scientific writing - Department of Public Health, Erasmus MC	2020	**1.0**
Scientific integrity – Erasmus MC	2020	**0.3**
NIHES courses		
Intermediate course in R	2019	**1.4**
Quality of Life Measurement	2019	**0.9**
Advanced topics in decision-making in medicine	2020	**2.4**
Advanced analysis of prognosis studies	2020	**0.9**
NIHES summer program: Advances in clinical epidemiology, causal inference, and social epidemiology	2020	**2.1**
Seminars and workshops		
Research seminars, Dep. Public Health, Erasmus MC, Rotterdam, the Netherlands	2018-2022	**3.0**
Research meetings Medical Decision Making, Dep. Public Health, Erasmus MC, Rotterdam, The Netherlands	2018-2022	**1.0**
Data Analytics in CENTER-TBI, Erasmus MC, Rotterdam, The Netherlands	2018	**0.3**
Smarter choices for better health, Erasmus University, Rotterdam, The Netherlands	2018	**0.5**
Data Analytics in CENTER-TBI, University of Antwerp, Antwerp, Belgium	2019	**0.3**
CENTER-TBI General Assembly, University of Antwerp, Antwerp, Belgium	2019	**0.3**
Evolutie of Revolutie, The Hague, The Netherlands	2019	**0.5**
International Initiative for Traumatic Brain Injury Research (InTBIR) Annual Meeting, Washington, United States	October 2019	**2.0**
CQM symposium Prediction to Personalized Medicine, Rotterdam, the Netherlands.	2020	**0.5**
Several meetings of the RIOT science club	2020	**0.2**
Internal neurotrauma symposium (INTS), online	February 2021	**0.1**
CENTER-TBI close-out meeting, online	February 2021	**0.3**
CEPHIR webinar "Natuurlijke experimenten en de volksgezondheid: onbenutte kansen voor preventie?"	June 2021	**0.1**
Presentations at conferences and seminars		
Development of Prognostic Models for Health-Related Quality of Life in Patients with Traumatic Brain Injury, IBIA, Toronto, Canada. (Poster presentation)	March 2019	**1.0**

Activity	Period	EC's
Outcome prediction following moderate and severe TBI: external validation of two established prognostic models in the CENTER-TBI study, RCEM Academic trainees day, UK. (Oral presentation)	January 2021	**0.3**
Sequelae and Outcome in Traumatic Brain Surgery: Prognosis after Traumatic Brain Injury. (Oral presentation)	March 2021	**0.3**
Several presentations for Research meetings Medical Decision Making, Dep. Public Health, Erasmus MC, Rotterdam, The Netherlands	2018-2022	**1.0**
Teaching		
Supervisor medical students' community projects	2018-2019	**1.0**
Evaluation of medical students' bachelor theses	2019-2020	**1.5**
Statistical and methodological advisor for medical students at the Erasmus MC	2020-2021	**0.5**
Teaching assistant CEO2 Prognosis	2021	**0.5**
Teaching assistant Advanced analysis of prognosis studies	2021	**2.0**
Other activities		
Peer reviewer for several international journals	2020-2022	**0.6**
Methodological advisor for international research projects	2021	**0.5**
Committees		
Chair and committee member of the Junior Representatives Committee (JVO)	2019-2020	**3.0**
Taskforces		
Improve flexworking	2019	**0.5**
Overload, stress and mental health	2021	**0.5**
Social media		
Twitter editor for the official twitter accounts of the Public Health department (Erasmus MC) and the CENTER-TBI study	2020-2022	**2.0**

About the author

Isabel Retel Helmrich was born in Rotterdam on 10 July 1995 at the Erasmus MC Sophia Children's hospital in Rotterdam, the Netherlands. She finished her A-levels in 2013 at the SG Helinium in Hellevoetsluis. Given her interest in human behaviour and communication she started studying Communication sciences at the university of Amsterdam. She finished her first year but decided to discontinue her studies. She enrolled in Psychology at the Erasmus University in Rotterdam instead to delve more deeply into the human psych and behaviour. During this time, she took part in several committees at the study association, and informed prospective students about the Psychology program at the Erasmus University as a student ambassador. After following a minor in Medical Psychology at the Erasmus University Medical Center she wrote her Bachelor thesis on informed decision making in the field of prenatal screening at the clinical genetics department.

Isabel pursued her interest in medical psychology and health behaviour and moved to Maastricht to obtain her master's degree in health education and promotion at the faculty of Health Sciences at the University of Maastricht. She enjoyed working as a research assistant at the clinical genetics department at the Maastricht UMC+. Here she wrote her master thesis on the effectiveness of a decision aid to support reproductive decision making of persons with a hereditary predisposition to cancer and their partners.

In August 2018, she moved back to Rotterdam and started working as a junior researcher and PhD candidate at the Department of Public Health in the field of medical decision making. Her work predominantly focused on the prediction of outcomes for individual patients in acute care within the European CENTER-TBI project, which has resulted in this thesis. During the COVID-19 pandemic, she conducted research on the potential implementation of prediction models in COVID-19 care together with colleagues from Tufts Medical center and Northwell health in the US.

During her PhD Isabel experienced the pressure and insecurities of working in academia firsthand. This led her to push for mental health resources and agency of junior researchers during her time as chair and committee member of the JVO (Committee for Junior Researchers at the department of public health). She also contributed to

NIHES courses as a teaching assistant and supported students as a methodological and statistical advisor.

From September 2022 on, she has continued her career in academia at the clinical genetics department at the Erasmus University Medical Center. Her work is focused on public deliberation on DNA-therapies, combining her interests in medical psychology, shared decision making and implementation of technologies in health care. She simultaneously works as a freelance thesis supervisor for bachelor and master students and is a mentor for youths at a non-profit organization in Rotterdam.

Isabel lives in Rotterdam with her long-term partner, their two cats and dog.