DESIGN AND ANALYSIS OF RANDOMIZED CONTROLLED TRIALS IN TRAUMATIC BRAIN INJURY

BOB ROOZENBEEK

DESIGN AND ANALYSIS OF RANDOMIZED CONTROLLED TRIALS IN TRAUMATIC BRAIN INJURY

BOB ROOZENBEEK

The research described in this thesis was conducted at the Department of Public Health, Erasmus MC Rotterdam, The Netherlands, and the Department of Neurosurgery, Antwerp University Hospital, Belgium and is the result of a “joint PhD program” of Erasmus University Rotterdam (The Netherlands) and Antwerp University (Belgium).

The research project was financially supported by the U.S. National Institutes of Health, project “Clinical Trial Design and Analysis in TBI”, R01 NS-042691.

Printing of this thesis was kindly supported by:
Erasmus University Rotterdam
Department of Public Health, Erasmus MC, Rotterdam

ISBN: 978-90-5335-594-7
Cover design: Amanda Nedermeijer – Media & Design, Rotterdam, The Netherlands
Layout: Simone Vinke, Ridderprint BV, Ridderkerk, The Netherlands
Printing: Ridderprint BV, Ridderkerk, The Netherlands

Design and Analysis of Randomized Controlled Trials in Traumatic Brain Injury

Ontwerp en analyse van gerandomiseerde klinische studies
voor traumatisch hersenletsel

Proefschrift

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

Prof.dr. H.G. Schmidt

en volgens besluit van het College voor Promoties.

In het kader van een dubbeldoctoraat wordt dit proefschrift
tevens voorgelegd tot het behalen van de graad van
doctor in de medische wetenschappen
aan de Universiteit Antwerpen.

De openbare verdediging zal plaatsvinden op
donderdag 15 november 2012 om 13:30 uur
te Rotterdam

door

Bob Roozenbeek

geboren te Alkmaar

ERASMUS UNIVERSITEIT ROTTERDAM

Universiteit Antwerpen

Promotiecommissie Erasmus Universiteit Rotterdam

Promotoren: Prof.dr. A.I.R. Maas
Prof.dr. E.W. Steyerberg

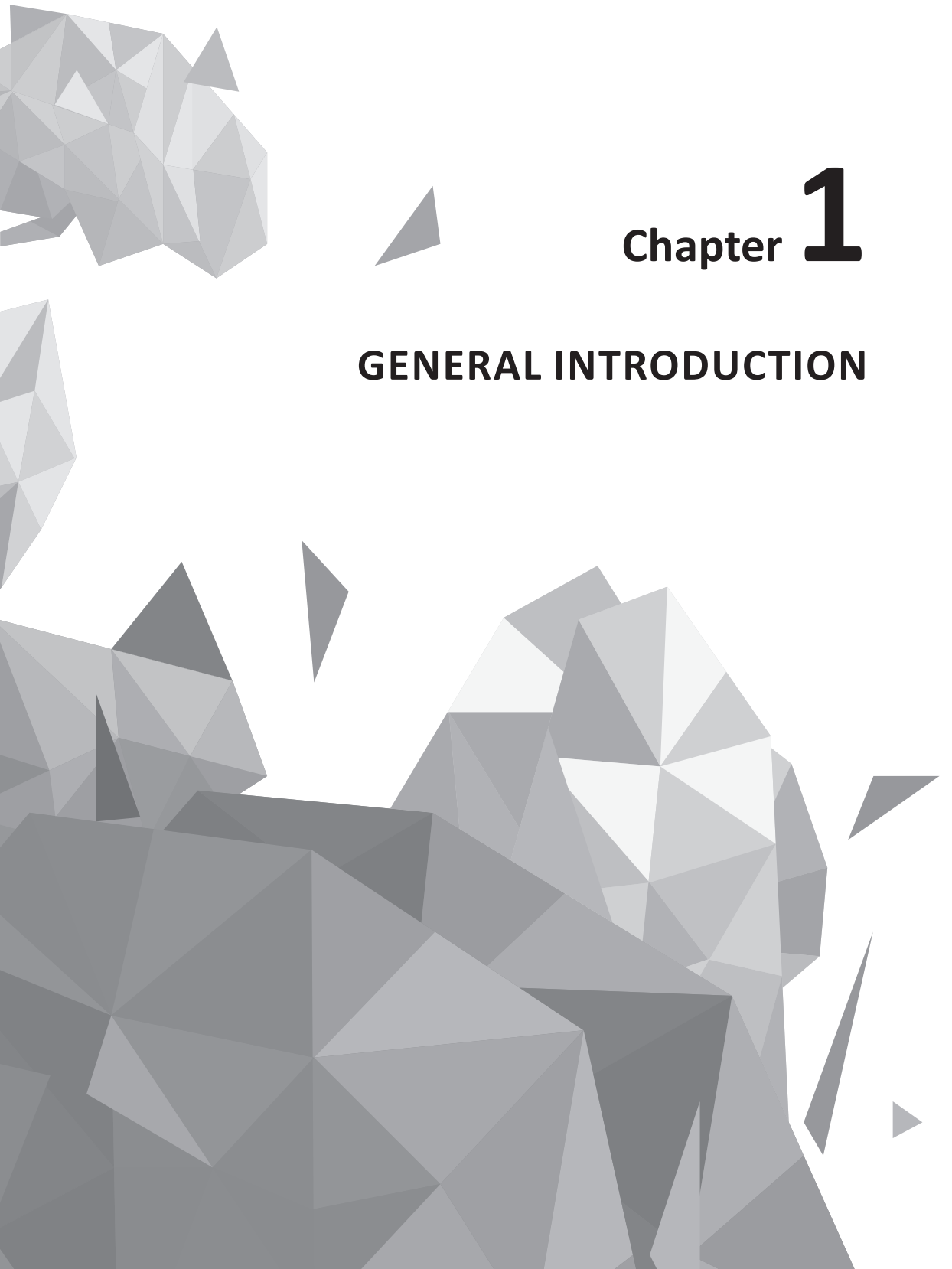
Overige leden: Prof.dr. D.W.J. Dippel
Prof.dr. M.G.M. Hunink
Prof.dr. P.G. Jorens

Doctoraatscommissie Universiteit Antwerpen

Promotoren: Prof.dr. A.I.R. Maas
Prof.dr. E.W. Steyerberg

Overige leden: Prof.dr. P.G. Jorens
Prof.dr. P.M. Parizel
Prof.dr. J. Weyler

CONTENTS



Chapter 1

GENERAL INTRODUCTION

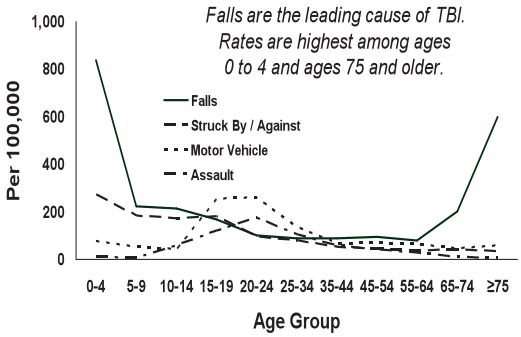
Traumatic Brain Injury

Incidence, age distribution and cause

Traumatic Brain Injury (TBI) is a serious public health problem (1, 2). It is a leading cause of death and disability among young adults in developed countries. The overall incidence of hospitalized TBI in Europe is estimated at 235 per 100,000 per year, although substantial variation exists between the different European countries (3). The TBI related mortality rate is 15.4 per 100,000 per year in Europe. Since standardized epidemiological monitoring of TBI is only seldom performed and most studies are based on retrospective analyses, these numbers probably underestimate the "real" incidence of TBI.

The age distribution in TBI populations is changing over time. Formerly, the majority of TBI patients were young male adults who were injured by traffic accidents or violence. In the past years the proportion of elderly patients is increasing (4). This shift toward older patients has resulted in falls now to be the leading cause of TBI (Figure 1.1) (5). The predominant type of pathology is changed as well: falls lead to more hemorrhagic contusions. Especially among the elderly, with more pre-existing co-morbidities and higher incidence of the use of anticoagulant medication and platelet aggregation inhibitors, this type of pathology may result in serious morbidity and mortality.

Figure 1.1. Estimated average annual rates of TBI by age and external cause in the United States (2002-2006). From: U.S. Centers for Disease Control and Prevention. From: Faul *et al*, 2010 (5).



Definition and classification

TBI is defined as an alteration in brain function, or other evidence of brain pathology, caused by an external force (6). The diagnosis is established on the basis of clinical symptoms: for example, the presence of any documented loss of consciousness, amnesia (retrograde or post-traumatic) or focal neurologic deficits.

TBI can be classified by mechanism, by clinical severity, by abnormalities on neuro-

imaging and by baseline prognostic risk. A mechanism-based classification divides TBI into closed (or non-penetrating, blunt), penetrating or blast injury (2). These types of TBI all require a different therapeutic approach and have a different prognosis. In this thesis, we will focus on closed TBI.

TBI can also be graded by clinical severity. The Glasgow Coma Scale (GCS) is the worldwide standard scale for scoring the clinical severity of TBI and it comprises three tests: eye, motor and verbal responses (7); Table 1.1). The three values separately as well as their sum are considered. The sum score can vary between 3 and 15 and can be used to classify TBI as mild (≥ 13), moderate (12-9) or severe (≤ 8). It is important to recognize that the GCS score can be confounded by intoxication, sedation or pharmacologically induced paralysis. In most patients with mild traumatic brain injury ('concussion') full neurological recovery occurs, although persisting complaints of problems with concentration and memory do remain in a proportion of these patients. Patients with moderate and severe TBI as well as any patient with structural abnormalities on the CT scan need to be admitted to a hospital and monitored closely, since they have a high risk of developing neurological complications leading to permanent neurologic deficits or even death.

To classify TBI based on neuro-imaging, the gold standard that is used in the acute phase is computed tomography (CT) scanning. The most frequently used classification to score structural abnormalities on the head CT scan is the "Marshall CT Classification" (8; Table 1.2).

Finally, a more novel method of classifying TBI is based on the individual patient's prognostic risk. Prognostic models combining different patient characteristics scored on admission to the hospital can be used to determine each patient's risk of developing a particular functional outcome. This risk can be used for risk-adjustment in clinical trials and to compare datasets in quality of care research. Two sets of well-designed models are currently available for prediction of outcome after TBI (9, 10). Both will be discussed more extensively later in this thesis.

Table 1.1. The Glasgow Coma Scale

Eye opening	Motor response	Verbal response
1= none	1= none	1= none
2= to pain	2= extension	2= incomprehensible
3= to speech	3= abnormal flexion	3= inappropriate
4= spontaneous	4= withdrawal to pain	4= confused
	5= localizes pain	5= oriented
	6= obeys commands	

Table 1.2. The Marshall CT Classification

Category	Definition
Diffuse injury I (no visible pathology)	No visible intracranial pathology seen on CT scan
Diffuse injury II	Cisterns are present with midline shift of 0–5 mm and/or lesions densities present; no high or mixed density lesion >25 cm^3 may include bone fragments and foreign bodies
Diffuse injury III (swelling)	Cisterns compressed or absent with midline shift of 0–5mm; no high or mixed density lesion >25 mm
Diffuse injury IV (shift)	Midline shift >5 mm; no high or mixed density lesion >25 cm^3
Evacuated mass lesion	Any lesion surgically evacuated
Non-evacuated mass lesion	High or mixed density lesion >25 cm^3; not surgically evacuated

Pathophysiology

Primary injury

The primary brain injury is caused at the moment of trauma, by direct impact to the brain or by an acceleration–deceleration mechanism. It results in brain tissue and blood vessels being stretched, compressed or ruptured. Primary brain injuries can be focal or diffuse. Focal injuries include intra-axial cerebral contusions and extra-axial intracranial hematomas. A contusion is a bruise of the brain parenchyma, frequently associated with multiple microhemorrhages. Contusions especially occur in regions near sharp edges of the inner skull (e.g. in the basal frontal and temporal lobes), most frequently as a "coup – contrecoup" injury. The coup injury occurs in the region of the direct impact, while the contrecoup injury is located on the side opposite to the impact. Major controversies exist in the management of contusions, particularly with regard to the indications and timing of surgical treatment.

The main discussion is whether early surgery should be preferred with the intent of preventing deterioration (but at a certain risk) or of delaying intervention until deterioration has occurred (when it may be uncertain that the patient can still recover). The different approaches are currently being compared in a multicenter randomized controlled trial ("Surgical Trial in Traumatic Intracerebral Hemorrhage" (STITCH)).

The extra-axial intracranial hematomas include the epidural hematomas and the subdural hematomas.

An epidural hematoma (EDH) occurs between the skull and the dura mater. It mostly results from a high-pressure arterial bleeding and has a convex or lens shape on CT scan, because its expansion stops at the sutures of the skull bones. An EDH is an absolute neurosurgical emergency and has to be immediately evacuated by a craniotomy.

Subdural hematomas (SDH) are located between the dura mater and the arachnoid

membrane around the brain. It often originates from ruptured bridging veins and occurs more often in older patients using anticoagulant medication. A SDH can also result in intracranial hypertension, with symptoms as a coma and/or focal neurologic deficits. These symptoms can show up immediately or in minutes or hours, but may also be delayed as in weeks. In the latter case, the patient has a *chronic* subdural hematoma (as opposed to an *acute* subdural hematoma). On CT scan, subdural hematomas have a concave aspect. The density of the hematoma is indicative for its age: hyperdens hematomas are acute and iso- or hypodens hematomas are subacute or chronic (11). Acute SDHs need to be evacuated by a craniotomy (12), while making burrholes in the skull can drain chronic SDHs.

In contrast to the focal injuries, TBI can be diffuse as well. As a result of rotational forces or severe deceleration trauma, the axons in the brain may be injured. This traumatic injury of the brain's white matter is called diffuse axonal injury (DAI) or traumatic axonal injury (TAI).

Although classic theories suggest mechanical disruption of the axons at the time of the injury, research in the last decades found out that this is not the case for the majority of the injured axons. Rather, TAI is now considered to be a continuous process that is initiated by tensile forces induced by the injury. This results in focal axonal permeability, which leads to calcium influx inducing a pathway of pathological processes ultimately resulting in axonal failure and disconnection (13). TAI must be suspected in patients whose brain CT scan appears to be normal, but who have severe clinical symptoms such as coma. TAI cannot be visualized well by CT scanning, but modern MRI techniques make it possible to visualize these white matter injuries. Susceptibility weighted images (SWI) or T2*-weighted images show the micro-hemorrhages resulting from disruption of the axons, and diffusion tensor images (DTI) can depict white matter fiber tract injuries in the brain (11).

Secondary injury

In contrast to most other trauma patient populations, a substantial proportion of patients with TBI clinically deteriorate after being hospitalized. This deterioration may be caused by complex processes that result in so-called secondary brain injury. Secondary brain injury is induced by a cascade of different pathophysiologic response mechanisms, triggered by the initial injury. These mechanisms include brain edema, raised intracranial pressure and subsequent decreased cerebral perfusion leading to cerebral ischemia. The secondary brain injury can be even worsened further by systemic second insults, such as hypoxia and hypotension. The principal concept is that secondary brain injury can be prevented.

This prevention of further injury to the brain is the basis of all TBI treatments. Many pharmacological agents aiming at protection of the brain against secondary injury have been proposed in the past decades, such as corticosteroids, free radical scavengers, *N*-methyl-*D*-

aspartate antagonists and calcium channel blockers. Regrettably, none of these agents has proven to be effective (14, 15).

Heterogeneity in TBI populations: a methodological challenge for randomized controlled trials

As outlined previously, TBI is not one single disease entity but it includes a heterogeneous and complex spectrum of pathology, ranging from diffuse axonal injury and focal contusions to epidural hematoma. The population of TBI patients is also extremely heterogeneous in terms of clinical severity and baseline prognostic risk. Furthermore, many TBI patients also suffer from extracranial injuries and second insults, both affecting prognosis. It is widely recognized that this complex heterogeneity is one of the main problems related to the failure of clinical Phase III trials to convincingly demonstrate an overall benefit of any therapeutic intervention in TBI in the past (14- 16). These problems relate for example to the potential imbalances that are induced between trial arms despite proper randomization. Further, this heterogeneity makes that the spectrum of outcomes after TBI is very wide, i.e. ranging from full recovery to death. In clinical trials, functional outcome is commonly measured with the Glasgow Outcome Scale (17, Table 1.3), but the dichotomized way in which the GOS is analyzed seems inefficient. Different concepts for dealing with heterogeneity in TBI trials have been proposed in the past (18, 19, 20), but these approaches were never combined nor were their separate effects compared.

Table 1.3. The Glasgow Outcome Scale.

Category	Label and definition	Dichotomization
1	**Dead** Mortality from any cause	Unfavorable outcome
2	**Vegetative State** Unable to interact with environment, unresponsive	
3	**Severe Disability** Conscious but dependent	
4	**Moderate Disability** Independent but disabled	Favorable outcome
5	**Good Recovery** Return to normal occupation and social activities, may have minor residual deficits	

The IMPACT Study

The IMPACT Study Group was initiated in 2003 as a collaborative venture supported by the US National Institutes of Health (21). The study group includes clinical, epidemiological and statistical investigators from the Antwerp University Hospital, Belgium; the Erasmus MC, University Medical Center Rotterdam, the Netherlands; the University of Edinburgh, Scotland and the Medical College of Virginia Commonwealth University in Richmond, Virginia, US. The IMPACT investigators were granted access to individual patient data of initially eight randomized control trials and three observational studies including a total of 9205 patients (22). During the continuation funding period (2007-2011) the number of studies was expanded providing access to data of over 40,000 patients. Relevant variables from the individual studies were extracted and merged to form a culture medium for exploring concepts to improve the design of clinical trials in TBI. The focus was on methodological approaches for dealing with the heterogeneity inherent to the TBI population. Identification of robust covariates and the development of prognostic models formed the cornerstone for explorations on how best to deal with the heterogeneity inherent to TBI populations (9, 23, 24, 25, 26, 27, 28, 29, 30). The analyses described in this thesis were performed as part of the IMPACT Study.

Aim of this thesis

The **aim** of this thesis is to investigate which adaptations in design and analysis of randomized controlled TBI trials may be best implemented to deal with the heterogeneity of TBI study populations in order to increase chances for detecting clinically relevant treatment effects in future trials for therapeutic interventions for TBI.

The following **specific questions** will be answered:

I. Which problems related to trial design and analysis have contributed to the failure of most randomized controlled TBI trials in the past?
II. What is the most efficient approach to selection of patients for randomized controlled TBI trials?
III. What is the influence of covariate adjustment on statistical power in randomized controlled TBI trials?
IV. What is the influence of ordinal rather than dichotomous analysis of the Glasgow Outcome Scale as primary outcome measure in randomized controlled TBI trials?
V. Are the currently available prognostic models for prediction of outcome after TBI externally valid and can they be used to make design and analysis of randomized controlled TBI trials more efficient?

References

1. Ghajar J. **Traumatic brain injury.** *Lancet*. 2000 Sep 9;356(9233):923-9.
2. Maas AI, Stocchetti N, Bullock R. **Moderate and severe traumatic brain injury in adults.** *Lancet Neurol.* 2008; 7(8): 728-41.
3. Tagliaferri F, Compagnone C, Korsic M, Servadei F, Kraus J. **A systematic review of brain injury epidemiology in Europe.** *Acta Neurochir (Wien)*. 2006; 148(3): 255-68.
4. Stocchetti N, Paternò R, Citerio G, Beretta L, Colombo A. **Traumatic Brain Injury in an Aging Population.** *J Neurotrauma.* 2012. [Epub ahead of print].
5. Faul M, Xu L, Wald MM, Coronado VG. **Traumatic Brain Injury in the United States: Emergency Department Visits, Hospitalizations and Deaths 2002 – 2006.** Atlanta (GA): Centers for Disease Control and Prevention, National Center for Injury Prevention and Control; 2010.
6. Menon DK, Schwab K, Wright DW, Maas AI; Demographics and Clinical Assessment Working Group of the International and Interagency Initiative toward Common Data Elements for Research on Traumatic Brain Injury and Psychological Health. **Position statement: definition of traumatic brain injury.** *Arch Phys Med Rehabil.* 2010; 91(11):1637-40.
7. Teasdale G, Jennett B. **Assessment of coma and impaired consciousness. A practical scale.** *Lancet*. 1974; 2(7872): 81-4.
8. Marshall LF, Bowers S, Klauber MR, et al. **A new classification of head injury based on computerized tomography.** *J Neurosurg* 1991; 75: (suppl): S14-20.
9. Steyerberg EW, Mushkudiani N, Perel P, Butcher I, Lu J, McHugh GS, Murray GD, Marmarou A, Roberts I, Habbema JD, Maas AI. **Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics.** *PLoS Med.* 2008; 5(8): e165.
10. MRC CRASH Trial Collaborators, Perel P, Arango M, Clayton T, Edwards P, Komolafe E, Poccock S, Roberts I, Shakur H, Steyerberg E, Yutthakasemsunt S. **Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients.** *BMJ.* 2008; 336(7641): 425-9.
11. Parizel PM, Van Goethem JW, Ozsarlak O, Maes M, Phillips CD. **New developments in the neuroradiological diagnosis of craniocerebral trauma.** *Eur Radiol.* 2005; 15(3): 569-81.
12. Bullock MR, Chesnut R, Ghajar J, et al., Surgical Management of Traumatic Brain Injury Author Group, **Guidelines for the surgical management of traumatic brain injury**. *Neurosurg*. 2006; 58(3 Suppl.).
13. Büki A, Povlishock JT. **All roads lead to disconnection? Traumatic axonal injury revisited.** *Acta Neurochir (Wien).* 2006; 148(2): 181-93.
14. Maas AI, Steyerberg EW, Murray GD, Bullock R, Baethmann A, Marshall LF, Teasdale GM. **Why have recent trials of neuroprotective agents in head injury failed to show convincing efficacy? A pragmatic analysis and theoretical considerations.** *Neurosurg*. 1999; 44(6): 1286-98.
15. Bullock MR, Lyeth BG, Muizelaar JP. **Current status of neuroprotection trials for traumatic brain injury: lessons from animal models and clinical studies.** *Neurosurg*. 1999; 45(2): 207-17.
16. Narayan RK, Michel ME, Ansell B, Baethmann A, Biegon A, Bracken MB, Bullock MR, Choi SC, Clifton GL, Contant CF, Coplin WM, Dietrich WD, Ghajar J, Grady SM, Grossman RG, Hall ED, Heetderks W, Hovda DA, Jallo J, Katz RL, Knoller N, Kochanek PM, Maas AI, Majde J, Marion DW, Marmarou A, Marshall LF, McIntosh TK, Miller E, Mohberg N, Muizelaar JP, Pitts LH, Quinn P, Riesenfeld G, Robertson CS, Strauss KI, Teasdale G, Temkin N, Tuma R, Wade C, Walker MD, Weinrich M, Whyte J, Wilberger J, Young AB, Yurkewicz L. **Clinical trials in head injury.** *J Neurotrauma*. 2002; 19(5): 503-57.
17. Jennett B, Bond M. **Assessment of outcome after severe brain damage.** *Lancet.* 1975; 1(7905): 480–484.
18. Machado SG, Murray GD, Teasdale GM. **Evaluation of designs for clinical trials of neuroprotective agents in head injury.** *J Neurotrauma.* 1999; 16(12): 1131-8.
19. Murray GD, Barer D, Choi S, Fernandes H, Gregson B, Lees KR, Maas AI, Marmarou A, Mendelow AD, Steyerberg EW, Taylor GS, Teasdale GM, Weir CJ. **Design and analysis of phase III trials with ordered outcome scales: the concept of the sliding dichotomy.** *J Neurotrauma.* 2005; 22(5): 511-7.
20. Hernández AV, Steyerberg EW, Butcher I, Mushkudiani N, Taylor GS, Murray GD, Marmarou A, Choi SC, Lu J, Habbema JD, Maas AI. **Adjustment for strong predictors of outcome in traumatic brain injury trials: 25% reduction in sample size requirements in the IMPACT study.** *J Neurotrauma.* 2006; 23(9): 1295-303.

21. Maas AI, Marmarou A, Murray GD, Teasdale SG, Steyerberg EW. **Prognosis and clinical trial design in traumatic brain injury: the IMPACT study.** *J Neurotrauma.* 2007; 24(2): 232-8.

22. Marmarou A, Lu J, Butcher I, McHugh GS, Mushkudiani NA, Murray GD, Steyerberg EW, Maas AI. **IMPACT database of traumatic brain injury: design and description.** *J Neurotrauma.* 2007; 24(2): 239-50.

23. Mushkudiani NA, Engel DC, Steyerberg EW, Butcher I, Lu J, Marmarou A, Slieker F, McHugh GS, Murray GD, Maas AI. **Prognostic value of demographic characteristics in traumatic brain injury: results from the IMPACT study. J** Neurotrauma. 2007; 24(2): 259-69.

24. Butcher I, McHugh GS, Lu J, Steyerberg EW, Hernández AV, Mushkudiani N, Maas AI, Marmarou A, Murray GD. **Prognostic value of cause of injury in traumatic brain injury: results from the IMPACT study.** *J Neurotrauma.* 2007; 24(2): 281-6.

25. McHugh GS, Engel DC, Butcher I, Steyerberg EW, Lu J, Mushkudiani N, Hernández AV, Marmarou A, Maas AI, Murray GD. **Prognostic value of secondary insults in traumatic brain injury: results from the IMPACT study.** *J Neurotrauma.* 2007; 24(2): 287-93.

26. Maas AI, Steyerberg EW, Butcher I, Dammers R, Lu J, Marmarou A, Mushkudiani NA, McHugh GS, Murray GD. **Prognostic value of computerized tomography scan characteristics in traumatic brain injury: results from the IMPACT study.** *J Neurotrauma.* 2007; 24(2): 303-14.

27. Van Beek JG, Mushkudiani NA, Steyerberg EW, Butcher I, McHugh GS, Lu J, Marmarou A, Murray GD, Maas AI. **Prognostic value of admission laboratory parameters in traumatic brain injury: results from the IMPACT study.** *J Neurotrauma*. 2007; 24(2): 315-28.

28. Butcher I, Maas AI, Lu J, Marmarou A, Murray GD, Mushkudiani NA, McHugh GS, Steyerberg EW. **Prognostic value of admission blood pressure in traumatic brain injury: results from the IMPACT study.** *J Neurotrauma*. 2007; 24(2): 294-302.

29. Marmarou A, Lu J, Butcher I, McHugh GS, Murray GD, Steyerberg EW, Mushkudiani NA, Choi S, Maas AI. **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(2): 270-80.

30. Murray GD, Butcher I, McHugh GS, Lu J, Mushkudiani NA, Maas AI, Marmarou A, Steyerberg EW. **Multivariable prognostic analysis in traumatic brain injury: results from the IMPACT study.** *J Neurotrauma.* 2007; 24(2): 329-37.

Chapter 2

CLINICAL TRIALS IN TRAUMATIC BRAIN INJURY: PAST EXPERIENCE AND CURRENT DEVELOPMENTS

Andrew IR Maas
Bob Roozenbeek
Geoffrey T Manley

Neurotherapeutics 2010; 7(1): 115-26

Abstract

In this manuscript we review past and current experience in clinical trials of traumatic brain injury (TBI), discuss limitations and challenges, and summarize current directions. The focus is on severe and moderate TBI. A systematic literature search over the years 1980-2009 revealed 27 large Phase III trials in TBI; we were aware of a further 6 unpublished trials. Analysis of these 33 trials yielded interesting observations:

- There was a peak incidence of trial initiations occurred in the mid 1990s with a sharp decline over the period 2000-2004
- Most trials reporting a significant treatment effect were studies on a therapeutic strategy (e.g. decompressive craniectomy, hypothermia), and concerned single center studies.
- Increasingly, studies have been shifting toward the Far East.

The currently existing trial registries permit insight into ongoing or recently conducted trials. Compared to the past decade, the number of studies on neuroprotective agents taken forward into efficacy oriented studies is low. In contrast, the number of studies on therapeutic strategies appears to be increasing again.

The disappointing results in trials on neuroprotective agents in TBI have led to a critical reappraisal of clinical trial methodology. This has resulted in recommendations for preclinical workup and triggered extensive analysis on approaches to improve the design and analysis of clinical trials in TBI. An interagency initiative towards standardization on selection and coding of data elements across the broad spectrum of TBI is ongoing, will facilitate comparison of research findings across studies, and encourage high quality meta-analysis of individual patient data in the future.

Introduction

Randomized controlled trials (RCT) are considered the gold standard for proving the efficacy of new treatments. In the field of Traumatic Brain Injury (TBI), however, not a single multicenter phase III RCT on neuroprotective agents has convincingly shown benefit (1,2,3). As a consequence, a sense of disappointment prevails: basic science researchers are increasingly frustrated that their hard and promising work performed under clean experimental conditions does not seem to transfer into the 'dirty' clinical situation and there is some reluctance by pharmaceutical companies to embark on a new high cost venture in TBI. Trials conducted so far have however not definitively established the lack of possible benefit. To the contrary, various trials have shown a trend towards efficacy and this offers hope. Basic research has greatly advanced our knowledge of what happens in the brain after TBI, offering opportunities to limit processes involved in the development of secondary brain damage. Many experimental studies on a multitude of agents have shown encouraging results, demonstrating efficacy of different agents on histologic endpoints, intracranial pressure, brain edema, cerebral blood flow, metabolism and on functional outcome. Translating advances from basic research into clinical benefit has proven complex.

In the experimental situation, studies are performed under tightly controlled conditions. TBI in the clinical situation, however, includes a complex spectrum of pathologies, often superimposed with systemic second insults (e.g. hypoxia and/or hypotension). Approaches to treatment may vary and uncertainty exists as to whether pathophysiologic processes targeted are indeed active in individual patients and if so, at what time after injury. It is here that more basic and clinical research is needed before targeting therapy to the individual needs of a patient can become a clinical reality. TBI populations are also heterogeneous in terms of clinical severity and baseline prognostic risk. This heterogeneity poses complex methodological challenges. It is here that methodological approaches may be optimized to increase statistical power. The challenge to demonstrate benefit of a novel agent, or treatment strategy in TBI is great, but rewards will be correspondingly high.

TBI is a field with one of the greatest unmet needs in medicine and public health (4). Not only is it a major cause of death and disability, incurring great personal suffering to victims and relatives, but it also leads to huge direct and indirect costs to society. In the US, the annual burden of TBI has been estimated at over $ 60 billion (5). Globally, the incidence is increasing, and TBI resulting from blast injuries is being increasingly recognized in military personnel returning from Middle East conflicts. In the Western world the epidemiology of TBI is slowly changing with a relative increase in middle aged and elderly patients injured by falls, resulting typically in contusional brain damage (4). Contusional brain injury is different from diffuse axonal injury, with a much larger component of inflammatory response and is often characterized by lesion progression. Approximately 25 to 45% of cerebral contusions

will enlarge significantly (6,7) and even higher occurrences are reported if the initial CT is performed within 2 hours of injury (8). The more frequent use of anticoagulant medication and platelet aggregation inhibitors in older patients may further increase the risk of lesion progression. The frequent progression of contusive brain injury indicates that this may well constitute a subpopulation of TBI more likely to benefit from neuroprotective agents by limiting processes involved in secondary brain damage. Other mechanisms, and consequently different approaches may be more relevant in patients with diffuse axonal injury, and we emphasize that neuroprotection in a more broad sense also includes strategies and therapies, aimed at promoting regeneration or replacement of lost neuronal and glial cells, neuronal circuits and stimulation of neuroplasticity.

The aim of this manuscript is to review past and current experience in clinical trials in TBI, to discuss limitations and challenges which have been recognized, and to summarize current directions. The focus is on studies on patients with severe and moderate TBI.

Clinical trials in TBI: past experience

The history of multicenter randomized controlled trials in TBI really started in the mid 1980s. Major pillars of clinical research, such as the Glasgow Coma Scale, the Glasgow Outcome Scale and CT scanning had only been introduced in the 1970s. A further boost towards trials resulted from advances in basic research with the identification of various neuroprotective agents. Prior to 1980, the majority of trials reported consisted of single center studies, evaluating treatment results often in respect to historical controls. In this period, a particular interest focused on the efficacy of steroids. These trials were primarily initiated by investigators from scientific interest; in contrast, multicenter trials on neuroprotective agents as from the mid 1980s on, were largely initiated by pharmaceutical companies. The collaboration between investigators and pharmaceutical companies proved beneficial to both sides: Investigators contributed disease-specific expertise, and the participation in trials greatly facilitated international contact and collaboration. Indirectly, this has probably led to improved quality of care.

We conducted a systematic literature search on clinical trials in TBI over the period 1980-2009. We included studies meeting the following criteria:

- prospective, parallel groups, phase III RCT with random assignment to either a new medication/intervention or placebo/best intervention available
- targeting patients with closed, moderate to severe TBI (e.g. GCS ≤ 12)
- with acute presentation (< 24 hours between injury and treatment)
- primary outcome expressed as the Glasgow Outcome Score (GOS ≥ 3 months)
- with a report in English language.

We excluded:
- Phase II RCTs
- studies focusing on mild TBI or on chronic TBI treatments (e.g. rehabilitation)
- studies on pediatric TBI
- studies using a different primary outcome measure than the GOS.

We selected studies focusing on adults (> 15 years) and studies including at least 100 patients (minimum 50 per arm). In total, 27 met these criteria. Collectively, we were aware of another six unpublished large Phase III trials in TBI, but of which we had sufficient knowledge to present details. These studies were therefore added to the selection, resulting in a total of 33 studies. A complete overview of the studies is presented in table 2.1.

Table 2.1. Overview

Publication (Funding)	Agent/Intervention (Mechanism)	Centers	Study population	N	Year of study	Status	Result
Braakman et al 1983[9] (Inv. initiated)	High dose dexamethasone (Various processes)	2	Comatose patients after non-missile TBI	161	1978-1981	Completed	No sign. Tx effect
Dearden et al 1986[10]	Dexamethasone (Various processes)	1	Severe head injury	130	1980-1983	Completed	No sign. Tx effect
Grumme et al 1995[11] (Inv. initiated)	Triamcinolone (Various processes)	9	Severe head injury, not further defined	396	1985-1990	Completed	No sign. Tx effect
Bailey et al 1991[12] (Bayer - HIT I)	Nimodipine (Ca- mediated damage)	6	Not obeying commands	351	1987-1989	Completed	No sign. Tx effect
Eur study group,1994[13] (Bayer - HIT II)	Nimodipine (Ca- mediated damage)	21	Not obeying commands	852	1989-1991	Completed	No significant effect in overall population
Rockswold et al 1992[14] (Inv. initiated)	Hyperbaric oxygen (Cerebral ischaemia)	1	GCS ≤ 9	168	1983-1989	Completed	No significant treatment effect
Wolf et al 1993[15] (NIH:12587)	Tromethamine (THAM) (Cerebral acidosis)	2	GCS ≤ 8	149	1988-1989	Completed	No overall treatment effect
Gaab et al 1994[16] (Inv. initiated)	Dexamethasone (Various processes)	10	GCS ≤ 13	300	1986-1989	Completed	No sign. Tx effect
Unpubl:Tiriliazad-Domestic (Upjohn)	Tirilazad (Lipid peroxidation)	36	GCS ≤ 8: 72% GCS 9-12: 28%	1155	1991-1994	Terminated	No sign. Tx effect reported
Marshall et al 1998:[17] Tirilazad-International (Upjohn)	Tirilazad (Lipid peroxidation)	50	GCS ≤ 8: 85% GCS 9-12: 15%	1120	1992-1994	Completed	No sign. Tx effect
Young et al 1996[18] (Sanofi-Winthrop)	PEGSOD (Free radical damage)	29	GCS ≤ 8	1562	1993-1995	Completed	No sign. Tx effect
Unplublished (SyntheLabo)	Eliprodil (Glutamate exitotoxicity)	20+	GCS 4-8	452	1993-1995	Completed	No sign. Tx effect
Harders et al 1996[19] (Bayer - HIT III)	Nimodipine (Ca- mediated damage)	21	tSAH	123	1994	Completed	Significant reduction in unfavourable outcome

Table 2.1. Overview *(Continued)*

Publication (Funding)	Agent/Intervention (Mechanism)	Centers	Study population	N	Year of study	Status	Result
Robertson et al 1999[20] (Inv. initiated, NIH NS27616)	CBF vs. ICP directed management (Cerebral ischaemia)	1	Motor score ≤ 5	189	1994-1997	Completed	No difference in neurologic outcome. Decrease in episodes of jugular desaturation
Morris et al 1999[21] (Ciba-Geigy,Novartis)	Selfotel (Glutamate exitotoxicity)	95	GCS 4-8	693	1994-1996	Terminated	No sign. Tx effect
Clifton et al 2001[22] (Inv. initiated, NIH NS 32786)	Hypothermia - NABIS (Various processes)	11	GCS 3-8 Motor score 1-5	392	1994-1998	Halted	No effects on outcome. Reduced incidence of ICP>30
Unpublished (Sandoz, Novartis)	D-CPP-ene - Saphir (Glutamate exitotoxicity)	51	Not obeying commands, ≥ one reactive pupil	924	1995-1997	Completed	No sign. Tx effect
Marmarou et al 1999[23] (SmithKlineBeecham/ Cortech Inc.)	Bradycor/CP-0127 (Bradykinine antagonist)	31	GCS 3-8	139	1996	Terminated	12% improvement in favourable outcome (p=0.26)
Unpublished (CambridgeNeuroscience)	Cerestat/Aptiganel (Glutamate exitotoxicity)	38	GCS 4-8 GCS 3 if pupils reactive	532	1996-1997	Terminated	No sign. Tx effect
Unpublished (Parke Davis)	SNX-111 (Glutamate exitotoxicity)	? (multi-center)	GCS 4-8	237	1997-1998	Terminated	Higher mortality
Unpublished (Bayer - HIT IV)	Nimodipine (Ca- mediated damage)	36	GCS<15 + tSAH	592	1997-1999	Completed	No significant effect
Yurkewich et al 2005[24] (Pfizer)	Traxoprodil (CP-101606) (Ca- channel blocker)	? (multi-center)	GCS 4-8	404	1998-2001	Completed	Higher mortality, non-significant effect
Cooper et al 2004[25] (Inv. initiated, MRC-Aus: 124330)	Hypertonic saline (Hypovolemia)	12	GCS ≤ 8	229	1998-2002	Completed	No sign. Tx effect
Temkin et al 2007[26] (Inv. initiated, NIH : NS 19643)	Magnesium Sulfate (Multiple mechanisms)	1	GCS ≤ 12	499	1998-2004	Completed	Poorer outcome in treated group

Table 2.1. Overview *(Continued)*

Publication (Funding)	Agent/Intervention (Mechanism)	Centers	Study population	N	Year of study	Status	Result
Maas et al 2006[27] (Pharmos Corp.)	Dexanabinol (Multiple processes)	86	Motor score 2-5 + CT abnormalities	861	2000-2004	Completed	No sign. Tx effect
Edwards et al 2005[28] (MRC UK)	Methylprednisolone (Multiple mechanisms)	239	GCS ≤ 14	10008	1999-2004	Terminated	Higher mortality in Tx group (p=0.001)
Cruz et al 2001[29*] (Investigator initiated)	High dose mannitol (Raised ICP)	1	ASDH	178	1997-2000	Completed	Significant better outcome (p<0.01)
Cruz et al 2002[30*] (Investigator initiated)	High dose mannitol (Raised ICP)	1	Temporal lobe hemorrhage with abnormal pupils	141	1997-2001	Completed	Significant treatment effect
Zhi et al 2003[31] (Investigator initiated)	Mild hypothermia (Various processes)	1	GCS ≤ 8	396	1997-2001	Completed	Reduction of mortality and improved outcome
Lu et al 2003[32] (Investigator initiated)	Decompr. craniectomy (Raised ICP)	1	GCS ≤ 8	230	1998-2001	Completed	Significant reduction of mortality
Jiang et al 2005[33] (Investigator initiated)	Standard trauma craniectomy vs. limited craniectomy (Raised ICP)	5	GCS ≤ 8 + Refractory intracranial hypertension	468	1998-2001	Completed	Better outcome with large craniectomy
Jiang et al 2006[34] (Investigator initiated)	Long term mild hypothermia (Multiple mechanisms)	3	GCS ≤ 8	215	2000-2003	Completed	5 day mild hypothermia is more efficacious than 2 days short term
Xiao et al 2008[35] (Investigator initiated)	Progesterone (Multiple mechanisms)	1	GCS ≤ 8	159	2004-2007	Completed	More fav. outcome in Tx group (p=0.048)

ASDH: acute subdural hematoma; CBF: cerebral blood flow; CT: computed tomography; GCS: Glasgow coma score; HIT: Head Injury Trial; ICP: intracranial pressure; NABIS: National Acute Brain Injury Study; PEGSOD: polyethylene glycol-conjugated bovine superoxide dismutase; Tx: treatment.
*The studies reported by Cruz et al. (36,37,39) have been subjected to severe criticism, and the reliability and validity of the results have been questioned.

This overview permits a number of interesting observations:

First, there is a peak incidence of trial initiations in the mid 1990s, with a sharp decline in the period 2000-2004. This holds both for studies with neuroprotective agents as for studies employing therapeutic strategies, such as for example hypothermia or decompressive craniectomy. We recognize that studies initiated in the latter years may not all have been published yet, and are aware of four Phase III studies on therapeutic strategies (two on decompressive craniectomy, one on hypothermia, and one on the antifibrinolytic agent Tranxamenic Acid (CRASH 2), but only of one trial with a neuroprotective agent (recombinant human erythropoietin, rhEPO) that are currently enrolling or were recently completed (table 2.2). Figure 2.1 summarizes the number of Phase III RCTs initiated per 5-year period in moderate and severe TBI since 1980. The decline in studies on neuroprotective agents does not seem to be caused by a lack of new agents with great potential, but is more likely due to the difficulties experienced in previous trials combined with the high costs.

Second, most studies reporting a significant treatment effect concern those using a therapeutic strategy as investigational modality. Six out of the ten studies on therapeutic strategies demonstrated statistical significance versus only three of the 23 on neuroprotective agents (table 2.3). The three studies on neuroprotective agents showing a significant treatment effect concern the HIT III Nimodipine study (19), the study reported by Xiao et al (35) on progesterone, both showing beneficial effects and the CRASH study (28) which demonstrated higher mortality in the steroid treated patients. These three trials had features which distinguished them from the traditional design of TBI trials: the HIT III study targeted a subgroup of the overall population, only including patients with traumatic subarachnoid hemorrhage; the progesterone study reported by Xiao et al concerned a single center study and the CRASH trials represents the only megatrial conducted in the field of TBI.

Third, we note that six out of the nine single center studies showed a significant treatment effect versus four of the 24 multicenter studies (table 2.4).

Fourth, a certain shift can be noted over the past 5 to 10 years of studies from the Western hemisphere towards the Far East. These primarily concern investigator-initiated studies which we find extremely encouraging, but we also note an increasing interest of pharmaceutical companies for moving trials to include centers in the Far East. This shift is motivated both by higher patient recruitment as also by lower costs. Further collaboration is highly recommended, but at the same time we wish to introduce a word of caution that it is as yet uncertain how results obtained in a different setting may translate into a more global perspective. This is important, as for example various trials conducted in mainland China have shown beneficial effects of decompressive craniectomy, hypothermia and progesterone (31, 32, 33, 34, 35). These studies were all well designed, and of high quality. It is conceivable however, that differences may exist in referral policies, potential for selection bias, access to health care, acute and post-acute treatments and outcome. We consider it a priority to investigate these issues in a comparative, observational study.

Table 2.2. Recently completed, ongoing and expected studies

Funding sponsoring	Study + agent	Mechanism targeted	Type study	Study population	Target N	Start year	Status
Investigator initiated (Zhejiang and Hangzhou Health Department, China)	Decompressive craniectomy	Raised ICP	Phase II	GCS ≤ 8	80	2003	Completed Dec 2008
NIH-NINDS	Cyclosporin A	Mitochondrial dysfunction	Phase II safety	GCS 3-8	50	2003	Completed (Mazzeo et al 2009)[38]
Biotherapeutics	Oxycyte (oxygen carrier)	Increase cerebral oxygenation	Phase II	GCS ≤ 9	8	2005	Completed Jan 2008
Key Neurotek A.G.	KN 38-7271	Cannabinoid receptor agonist	Phase II a	Severe TBI	97	2006	Completed Dec 2008
Investigator initiated (Department of Health, Taipei City Government, Taiwan)	Multiple Cerebral Monitoring	Multiple	Phase II b	GCS 3-8	320	2006	Ongoing
Xytis Pharmaceuticals	XY 2405	Bradykinine (beta 2 receptor) antagonist	Phase II	GCS ≤ 12	400	2007	Stopped early June 2008
Investigator initiated (Cambridge University Hospitals NHS Foundation Trust, UK)	Interleukin-1 receptor antagonist	Inflammatory response	Phase II safety	GCS ≤ 8	26	2008	Ongoing
Biotherapeutics	Oxycyte (oxygen carrier)	Cerebral ischaemia	Phase II safety	Severe TBI	128	July 2009	Ongoing
Solvay Pharmaceuticals	SLV334 (ECE/NEP inhibitor) preventing formation of the vasoconstrictor ET-1	Cerebral vasoconstriction; Atrial Natriuretic Peptide	Phase II	Moderate and severe TBI	72	2009	Ongoing
Neuren Pharmaceuticals	NNZ 2566	Multiple	Phase II dose escalation	?	?	2009?	Expected to start October
Vasopharm	VAS 203	Nitric oxyde synthesis	Phase II	Severe	?	?	Expected

Table 2.2. Recently completed, ongoing and expected studies *(Continued)*

Funding sponsoring	Study + agent	Mechanism targeted	Type study	Study population	Target N	Start year	Status
Investigator initiated (MRC Australia)	DECRA (decompressive craniectomy)	Raised ICP	Phase III	Severe TBI: ICP > 20mmHg for 15 min in 1 hr. refractory to med. measures	165	2004	Ongoing
Investigator intiated (NIH-NINDS)	NABIS: H IIR Hypothermia	Multiple	Phase III	GCS 3-8	240	2005	Stopped
Investigator initiated (MRC UK)	RescueICP (decompressive craniectomy)	Raised ICP	Phase III	ICP > 25 mmHg for 1 hr refractory to med. measures	500	2005	Ongoing
NIH-NINDS	rhEPO (recombinant human Erythropoetin)	Multiple	Phase III	GCS motor ≤ 5	200	2006	Ongoing
Investigator initiated	CRASH-2 (Antifibrinolytic therapy with Tranxamenic acid)	Fibrinolysis	Phase III	All adult TBI patients at risk for developing intracranial hemorrhage	20,000	2007	Ongoing
European Society Intensive Care Medicine (ESICM)	EuroTherm Hypothermia	Multiple	Phase III	Severe	1,800	2009	Expected
Investigator initiated	STITCH (Surgical Trial In Traumatic intraCerebral Haemorrhage)	Contusions with mass effect	Phase III	Traumatic supratentorial ICH or contusion>25ml	840	2010	Expected
BHR Pharma	BHR100 (progesterone)	Multiple	Phase III	GCS 4-8	1,180	2010	Expected
NIH-NINDS	ProTECT (progesterone)	Multiple	Phase III	GCS 4-12	1,140	2010	Expected

ICP: intracranial pressure; NABIS: National Acute Brain Injury Study; TBI: traumatic brain injury.

Two other issues deserve special consideration: the number of unpublished studies and the reasons why some studies were ended prematurely. In total, six studies included in the overview have not been published (Tirilazad-domestic, Eliprodil, Saphir, Cerestat, Parke-Davis and HIT IV). These were all large, pivotal trials which we could include in the overview due to our knowledge of the field. There may however be more 'negative' trials which we are not aware of, particularly single center studies. It is regrettable that these studies have not been published, as we consider it a moral obligation to patients and relatives who consented for participation and to the community in general to publish results, even if a trial does not yield a positive result.

In total, seven trials were terminated prematurely. The NABIS Hypothermia trial was originally initiated with the goal of enrolling 500 patients but was halted prematurely after enrolment of 392 patients, following an unscheduled futility analysis showing that the likelihood for demonstrating benefit on continuation was low. The Tirilazad-Domestic trial was halted prematurely after enrolment of 1,191 of the planned 1,212 patients (98,3%) on recommendation of the data safety monitoring committee having demonstrated a significant difference in mortality between the treatment groups. On more detailed subsequent analysis, it was found that this difference in mortality could be explained by differences in baseline characteristics, but because the target enrolment had nearly been reached the decision was made to terminate the study. The Selfotel trials (US and international) were stopped after enrolment of 693 patients, because of concerns of the safety and monitoring committee about an increased number of deaths and severe brain related adverse events that had occurred in two contemporary trials in stroke. Such adverse events were not noted in the TBI trials, but as analysis indicated a low likelihood of demonstrating benefit on pursuing the trial to completion, the decision was made to definitively stop enrolment.

The Cerestat trial originally aimed to enrol 700 patients, but was halted after enrolment of 532 patients, following a planned interim-analysis conducted on the first 340 patients. This decision was also made against the background of concern about the effects of the agent in patients with stroke. The Parke Davis study on SNX-111 was halted prematurely when the data safety monitoring board observed a 10% increase of mortality in treated patients; the occurrence of hypotension as complication of the use of calcium channel blockers was also noted in the study. The Bradycor trial was designed to enrol 160 evaluable patients, but was stopped after enrolment of 139 patients, following the results of animal toxicology studies conducted during the course of the trial. These results were largely inconclusive after repeat studies, and the decision to terminate the trial is perhaps all the more regrettable as a strong trend towards benefit was noted on the GOS at 6 months (12% improvement) and supported by positive trends seen in ICP, therapy intensity level and neuropsychological tests. The CRASH megatrial was originally aimed to include a total of approximately 20,000 patients, but was stopped in May 2004 following enrolment of 10,008 patients, because of

an excess in early mortality (day 14). These observations demonstrate that there are many in- and external factors influencing the conduct of TBI trials. In three cases, the reasons were safety concerns within the trial, in three cases external factors including experience in simultaneously conducted stroke trials, the results in preclinical studies, and in only one instance (NABIS Hypothermia) resulted from futility analysis. Indirectly however, aspects of futility also probably influenced the decision making in Tirilazad, Selfotel and Cerestat.

Table 2.3. Trial results: differentiated for type of investigational therapy

		Significant effect	No significant effect
Interventional therapy	Neuroprotective agent	3	20
	Therapeutic strategy	6	4

Table 2.4. Trial results: single versus multicenter trials

		Significant effect	No significant effect
Type of study	Single center	7	3
	Multicenter	2	21

The observation that the majority of studies showing efficacy of the interventional therapy concerns single center studies whilst the overwhelming majority of multicenter studies failed to demonstrate effect is of interest and raises the question whether this may simply reflect publication bias, or that center effects may be a factor confounding chances of demonstrating efficacy. Considerable between-center differences in 6-month unfavourable outcome have been reported in the NABIS Hypothermia trial (22). Significant differences in mortality were also found between high and low enrolling centers in one of the more recent trials in TBI, investigating the efficacy of dexanabinol (27). The IMPACT studies have further shown a 3.3 fold difference after correction for random variation in the odds of having an unfavourable outcome between centers ($p<0.001$), which were not explained by patient characteristics (40). These data illustrate some of the complexities involved in multicenter trials. The advantage of multicenter trials over single center studies is enhanced generalizability. Two different approaches may be pursued to minimize variability in choice and sequence of basic therapeutic approaches: one is rigorous standardization of treatment, and the other is to recruit so many patients in a 'megatrial', that variability will be of less concern. An example of the latter approach is the CRASH study. The relative merits of these two contrasting approaches however needs to be further determined.

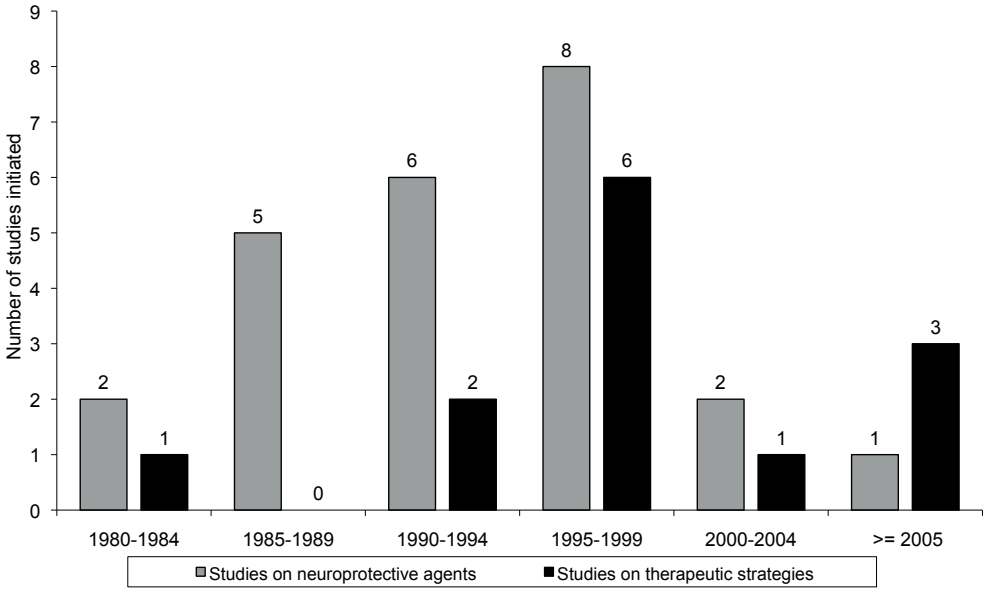


Figure 2.1. Numbers of initiated randomized controlled trials on moderate to severe traumatic brain injury per 5-year time periods. Trials were grouped by studies on neuroprotective agents and studies on therapeutic strategies.

Recently completed, ongoing and expected studies

General information on 'new' and recently completed trials can be found on the websites www.clinicaltrials.gov (NIH trial registry) and www.controlled-trials.com (non-governmental registry initiated by journal editors in collaboration with Springer). Central registration of clinical trials via these registries or the Eudra clinical trials database in the EU has served to promote quality in the design and reporting of clinical trials. Data from the EMEA registry (EudraCT) are not yet publically available, but this is anticipated for 2010. Regulatory agencies consider formal registration of the trial mandatory when considering drug approval and journal editors may refuse publication of trial results if the trial has not been registered prior to initiation. An overview of recently completed, ongoing and expected studies is presented in table 2.2. Much of this knowledge is now in the public domain by virtue of the central trial registries.

This overview indicates that many of the studies concern early stage clinical development. Currently, to our knowledge, there are five Phase III trials ongoing (four on therapeutic strategies and one on the neuroprotective agent rhEPO). A Phase III trial on hypothermia is expected to start soon, supported by the European Society for Intensive Care Medicine. Preparations for a further two trials on progesterone are ongoing, and it is expected that

these studies will be open for enrolment early in 2010.

Compared to the past decade, the number of studies on neuroprotective agents being taken forward into efficacy-oriented studies is dismal. We do not consider this due to lack of potential agents, but rather due to reluctance to embark on a high-risk venture, which is moreover extremely costly. In times of 'trial crisis', a clear need is felt for improved trial methodology in order to decrease the risks and our strong feeling is that if we wish to continue developing the field of neuroprotection in the clinical situation, attempts should be made to reduce costs. In particular, the substantial overhead often imposed in academic institutes is of concern.

Reappraisal of clinical trial methodology in TBI

Overall, the results of the Phase III trials on neuroprotective agents have been disappointing. The disappointing results have stimulated a critical reappraisal of conditional requirements before starting clinical trials with new interventional therapies and of the methodology of trial design and analysis. The reasons that previous trials have not shown convincing benefit reflect a number of factors ranging from preclinical investigations, through problems in translating results from experimental studies to clinical practice to the clinical situation where it is generally uncertain if the pathophysiologic mechanisms that are targeted are indeed active in the patients studied and if so, at what time after injury. Experimental studies should be robust, preferably showing benefit in more than one model, in more than one species with effects on both mechanistic and behavioural endpoints. To what extent time windows determined in the experimental situation may be translated to clinical practice, remains a point of concern. Requirements postulated for successful phase II TBI trials include:

- mechanism demonstrated in animal models
- drug/agent reverses damage in animal models
- mechanism shown to be active in human TBI
- neuroprotective agent passes the brain barrier
- safety/tolerability in humans with TBI
- drug sensitive endpoints.

Phase II trails are aimed at safety, may include dose escalation studies and at best, may fulfil requirements for proof of concept by demonstrating efficacy on a mechanistic endpoint. In general however, clinical efficacy, even indications thereof, may not be expected from Phase II TBI trials.

In most efficacy oriented TBI trials (Phase IIb or Phase III studies), the hypothesis was to increase the absolute proportion of patients with a favourable outcome by at least 10%. Such an effect may not have been proven at this predetermined level, but conversely neither

has inefficacy been proven. In fact, many trials demonstrate some increase in favourable outcome, albeit non-significant in treated patients. For example, the HIT I trial on nimodipine showed an 8% relative improvement (12) and the HIT II trial a 4% absolute improvement in favourable outcome in treated patients (13). In the Triamcinolone trial, a 5% increase in good recovery was observed in patients on active treatment (11). In the PEGSOD trial a 9% improvement in favourable outcome was noted (18) and in the Bradycor trial a 12% improvement (23), and in the Pfizer study (24) a decrease of mortality of 7% and an increase in favourable outcome in the treated group. It would hence appear that at least some trials show an indication of efficacy, and this may be interpreted as further evidence in support of the concept that TBI trials have been underpowered. Other factors, such as confounding effects of heterogeneity of TBI populations, overly optimistic expectations and insensitive methodology have also contributed to the difficulties experienced in demonstrating benefit of investigational treatments. In comparison to other fields of medicine, efficacy-oriented clinical trials in TBI pose complicated methodological challenges. These methodological challenges are in particular related to three important factors: 1) the great heterogeneity of TBI populations, 2) the lack of relevant mechanistic early endpoints and 3) insensitivity of more global outcome measures. The consequences of heterogeneity in TBI populations from the perspective of trial design and analysis and approaches for dealing with this heterogeneity, are the main focus of studies conducted by the IMPACT investigators. Results from these studies have shown that the statistical power in TBI trials may be increased by up to 50% by utilizing more efficient approaches to the analysis. Recommendations and results originating from the IMPACT work are summarized in a dedicated manuscript in this issue (37).

The IMPACT studies were conducted on individual patient data from eight RCTs and three observational studies (38). Relevant parameters from the individual studies were merged into a large dataset, forming a ‘culture medium’ for exploring concepts for improving the design of clinical trials in TBI. Merging the individual patient data into a common format proved an even more formidable task than initially anticipated, and involved over 10 person years of work. TBI datasets demonstrated almost as much heterogeneity as patient populations concerning coding of variables, data storage and most data sets were characterized by general lack of detailed documentation. This experience highlighted the need for standardisation of coding and data storage formats in TBI studies, and helped stimulate an international and interagency initiative to develop recommendations for ‘common data elements’ for TBI studies.

Standardization of data collection: common data elements

The initial steps towards development of standardization of data collection in TBI, were integrated into a much larger interagency initiative towards ‘an integrated approach to

research in Psychological Health and Traumatic Brain Injury'. This initiative, involving the National Institute of Neurological Disorders and Stroke (NINDS), the National Institute on Disability and Rehabilitation Research (NIDRR), the department of Veterans Affairs (VA), the Defense and Veterans Brain Injury Center (DVBIC) and the Defense Centers of Excellence (DCoE), included four working groups in the field of TBI, addressing four domains: demographics and clinical assessment, biomarkers, neuroimaging and outcome. Full results and recommendations of these working groups will be published early 2010. The working group 'Demographics and Clinical Assessment' focused on developing common data elements on all clinical and demographic variables in TBI studies. Specific aspects relating to biomarkers, neuroimaging and outcome assessment, including recommendations for further development required to advance the field, were addressed by the other working groups. The global aim was to develop recommendations on selection and coding of data elements for studies across the broad spectrum of TBI. The elements should be applicable to the entire spectrum on injury, both to milder and more severe injuries, to acute and long term studies and to studies including patients early after injury, but also to those enrolling patients at later time periods; elements should further be applicable to both civilian and military settings. Consistency was sought with the common data elements (CDE) initiative of NIH-NINDS (www.nindscommondataelements.org) initiated by John Marler, at the time vice director of NIH-NINDS with the aim to develop a core set of variables for use across the different fields of clinical neuroresearch.

The process for developing recommendations for common data elements was consensus driven, with multidisciplinary input from experts over a broad range of disciplines, covering the entire trauma chain from emergency medicine to rehabilitation and late outpatient care. Thus, a broad input was ensured. General consensus on selection and coding of data elements was achieved and templates produced summarizing coding formats, motivation of choices and recommendations for procedures. The data elements are contained in modules, which are grouped together in categories. For example, the data elements 'age, gender and race' are contained in the module 'demographics' under the category 'subject characteristics'. It was further realised that the level of detail required in data collection may vary greatly with the design and aim of a specific study. We therefore developed three versions for coding data elements: a basic, an advanced and an extended format with the greatest level of detail in the extended version. The coding of these versions is such that in every case the more detailed coding can be collapsed into the basic version, thus facilitating comparison and meta-analysis of individual patient data between studies. The data elements and modules are intended as 'building blocks' for designing a case report form. They can be used as 'plug in' elements and used multiple times in various sections of the CRF. For example, the module on Glasgow Coma Scale (GCS and pupils) may be recorded only on

admission, or also prehospital as well as daily during the acute care phase. Researchers can select and mix basic, advanced and extended versions of the different data elements, according to the requirement of the study. We note that the data elements proposed are not all inclusive, and the option always remains open to include other elements. The complete package of common data elements, including the templates, may be posted on the website www.commondataelements.org in early 2010. The data elements are currently presented in a 'paper based format'. Work is ongoing to include the elements and modules in a web-based data entry format with pull down menus and automated data checks. We emphasize that the current recommendations represent a beta version; we are in the process of incorporating feedback from a more international forum with the intent to make this a global initiative.

We further note that some of the recommendations include innovative approaches, which require field-testing and validation in clinical practice. Evaluating the applicability of the recommendations across various settings can best be accomplished by an observational study. The initiative to develop standard data elements in the field of TBI will facilitate comparison of research findings across studies, and encourage high quality meta-analysis of individual patient data in the future. This may well constitute one of the most important steps forward in the field of clinical trials in TBI, paving the road for harvesting successful results in the near future.

References

1. Maas AI, Steyerberg EW, Murray GD, et al. **Why have recent trials of neuroprotective agents in head injury failed to show convincing efficacy? A pragmatic analysis and theoretical considerations.** *Neurosurg*. 1999; 44(6): 1286-98.
2. Narayan RK, Michel ME, Ansell B, et al. **Clinical trials in head injury.** *J Neurotrauma*. 2002; 19(5): 503-57.
3. Maas AIR, Marmarou A, Murray GD, Teasdale GM, Steyerberg EW. **Prognosis and clinical trial design in traumatic brain injury: The IMPACT Study.** *J Neurotrauma*. 2007; 24: 232-8.
4. Maas AIR, Stocchetti N, Bullock R. **Moderate and severe traumatic brain injury in adults.** *Lancet Neurol*. 2008; 7(8): 728-741.
5. Finkelstein E, C.P., Miller T and associates, **The Incidence and Economic Burden of Injuries in the United States.** 2006, New York: Oxford University Press.
6. Servadei F, Murray GD, Penny K, et al. **The value of the "worst" computed tomographic scan in clinical studies of moderate and severe head injury. European Brain Injury Consortium.** *Neurosurg*. 2000; 46(1): 70-7.
7. Chang EF, Meeker M, Holland MC. **Acute traumatic intraparenchymal hemorrhage: risk factors for progression in the early post-injury period.** *Neurosurg*. 2006; 58 (4): 647-656.
8. Oertel M, Kelly DF, McArthur D, et al. **Progressive hemorrhage after head trauma: predictors and consequences of the evolving injury.** *J Neurosurg*. 2002; 96(1): 109-16.
9. Braakman R, Schouten HJA, Blaauw-Van Dishoeck M, Minderhoud JM. **Megadose steroids in severe head injury.** *J Neurosurg*.1983; 58: 326-30.
10. Dearden NM, Gibson JS, McDowall DG, Gibson RM, Cameron MM. **Effect of high-dose dexamethasone on outcome from severe head injury.** *J Neurosurg*. 1986; 64: 81-8.
11. Grumme T, Baethmann A, Kolodziejczyk D, et al, **Treatment of patients with severe head injury by triamconolone: a prospective, controlled multicenter clinical trial of 396 cases.** *Res Exp Med (Berl).* 1995; 195: 217-29.
12. Bailey I, Bell A, Gray J, et al. **A trial of the effect of nimodipine on outcome after head injury.** *Acta Neurochir (Wien).* 1991; 110:97-105.
13. European Study Group. **A multicenter trial of the efficacy of nimodipine on outcome after severe head injury. The European Study Group on Nimodipine in Severe Head Injury.** *J Neurosurg.* 1994; 80(5): 797-804.
14. Rockswold GL, Ford SE, Anderson DC, Bergman TA, Sherman RE**. Results of a prospective randomized trial for treatment of severely brain-injured patients with hyperbaric oxygen.** *J Neurosurg* 1992; 76: 929-34.
15. Wolf AL, Levi L, Marmarou A, et al. **Effect of THAM upon outcome in severe head injury: a randomized prospective clinical trial.** *J Neurosurg* 1993; 78: 54-9.
16. Gaab MR, Trost HA, Alcantara A, et al. **"Ultrahigh" dexamethasone in acute brain injury. Results from a prospective randomized double-blind multicenter trial (GUDHIS). German Ultrahigh Dexamethasone Head Injury Study Group.** *Zentralbl Neurochir.* 1994; 55(3): 135-43.
17. Marshall LF, Maas AI, Marshall SB, et al. **A multicenter trial on the efficacy of using tirilazad mesylate in cases of head injury.** *J Neurosurg*. 1998; 89: 519-525.
18. Young B, Runge JW, Waxman KS et al. **Effects of pegorgotein on neurologic outcome of patients with severe head injury. A multicenter, randomized controlled trial.** *JAMA*. 1996; 276: 538-43.
19. Harders A. Kakarieka A, Braakman R. **Traumatic subarachnoid hemorrhage and its treatment with Nimodipine. German tSAH Study Group.** *J Neurosurg*. 1996; 85: 82-9.
20. Robertson CS, Valadka AB, Hannay HJ, et al. **Prevention of secondary ischemic insults after severe head injury.** *Crit Care Med.* 1999; 27(10): 2086-95.
21. Morris GF, Bullock R, Bowers Marshall S, et al. **Failure of the competitive N-methyl-D-aspartate antagonist Selfotel (CGS 19755) in the treatment of severe head injury: results of two Phase III clinical trials.** *J Neurosurg*. 1999; 91: 737-743.
22. Clifton GL, Choi SC, Miller ER, et al. **Intercenter variance in clinical trials of head trauma – experience of the National Acute Brain Injury Study : Hypothermia**. *J Neurosurg* 2001; 95: 7515.
23. Marmarou A, Nichols J, Burgess J et al. **Effects of bradykinin antagonist Bradycor (Deltibant, CP-1027) in severe traumatic brain injury: results of a multi-center, randomized, placebo-controlled trial.** *J Neurotrauma.* 1999;16: 431-44.
24. Yurkewicz L, Weaver J, Bullock MR, Marshall LF. **The effect of the selective NMDA receptor antagonist traxoprodil in the treatment of traumatic brain injury.** *J Neurotrauma.* 2005; 12:1428-43.

25. Cooper DJ, Myles PS, McDermott FT, et al**. Prehospital hypertonic saline resuscitation of patients with hypotension and severe traumatic brain injury.** *JAMA*. 2004; 291: 1350-7.

26. Temkin NR, Anderson GD, Winn HR, et al. **Magnesium sulfate for neuroprotection after traumatic brain injury: a randomized controlled study.** *Lancet Neurol*. 2007; 6: 29-38.

27. Maas AI, Murray G, Henney H 3rd, et al. **Efficacy and safety of dexanabinol in severe traumatic brain injury: results of a phase III randomised, placebo-controlled, clinical trial.** *Lancet Neurol.* 2006; 5(1): 38-45.

28. Edwards P, Arango M, Balica L, et al. **Final results of MRC CRASH, a randomized placebo-controlled trial of intravenous corticosteroid in adults with head injury-outcomes at 6 months.** *Lancet*. 2005; 4-10: 365(9475): 1957-9.

29. Cruz J, Minoja G, Okuchi K. **Improving clinical outcomes from acute subdural hematomas with the emergency preoperative administration of high doses of mannitol: a randomized trial.** *Neurosurg.* 2001; 49: 864-71.

30. Cruz J, Minoja G, Okuchi K. **Major clinical and physiological benefits of early high doses of mannitol for intraparenchymal temporal lobe hermorrhages with abnormal papillary widening: a randomized trial.** *Neurosurg* 2002; 51: 628-38.

31. Zhi D, Zhang S, Lin X. **Study on therapeutic mechanism and clinical effect of mild hypothermia in patients with severe head injury.** *Surg Neurol* 2003; 59: 381-5.

32. Lu LQ, Jiang JY, Yu MK et al. **Standard large trauma craniotomy for severe traumatic brain injury.** *Chin J Traumatol* 2003; 5:302-4.

33. **Jiang J, Xu W, Li W, et al. Efficacy of standard trauma craniectomy for refractory intracranial hypertension with severe traumatic brain injury: a multicenter, prospective, randomized controlled study.** *J Neurotrauma*. 2005; 22: 623-8.

34. Jiang J, Xu W, Li W, et al. **Effect of long-term mild hypothermia or short-term mild hypothermia on outcome of patients with severe traumatic brain injury.** *JCBF&M* 2006; 26: 771-6.

35. Xiao G, Wei J, Yan W, Wang W, Lu Z. **Improved outcomes from the administration of progesterone for patients with acute severe traumatic brain injury: a randomized controlled trial.** *Crit Care.* 2008; 12: R61.

36. Roberts I, Smith R, Evans S. **Doubts over head injury studies.** *BMJ* 2007; 334: 392-4.

37. Maas AIR, Steyerberg EW, Marmarou A et al. I**MPACT recommendations for improving the design and analysis of clinical trials in moderate to severe traumatic brain injury.** *Neurotherapeutics* 2010; 7(1): 127-34.

38. Marmarou A, Lu J, Butcher I, et al. **IMPACT database of traumatic brain injury: design and description.** *J Neurotrauma.* 2007; 24(2): 239-50.

39. Mazzeo AT**, Brophy G, Gilman CB, Alves OL, Robles JR, Hayes RL, Povlishock J, Bullock R. Safety and tolerability of cyclosporin A in severe traumatic brain injury patients: results from a prospective, randomized trial.** *J Neurotrauma*, 2009; 26(12): 2195-206.

40. Lingsma HF, Roozenbeek B, Li B, Lu J, et al. **Large between-center differences in outcome after moderate and severe traumatic brain injury in the international mission on prognosis and clinical trial design in traumatic brain injury (IMPACT) study.** *Neurosurg*. 2011; 68(3): 601-7.

Chapter 3

EARLY PROGNOSIS IN TRAUMATIC BRAIN INJURY: FROM PROPHECIES TO PREDICTIONS

Bob Roozenbeek*
Hester F Lingsma*
Ewout W Steyerberg
Gordon D Murray
Andrew IR Maas

*equal contribution

Lancet Neurol 2010*; 9: 543-54*

Abstract

Traumatic Brain Injury (TBI) constitutes a heterogeneous disease, encompassing a broad spectrum of pathologies. Outcome can be highly variable, particularly in more severely injured patients. Despite the association of many variables with outcome, predictions are notoriously difficult. Multivariable analysis has identified age, clinical severity, Computerized Tomography abnormalities, systemic insults (hypoxia and hypotension) and laboratory parameters as relevant building blocks for combining variables into models to predict outcome in individual patients. A systematic literature search identified 16 studies reporting on prognostic models based upon admission characteristics; many of these showed shortcomings, which may partly explain the limited use of these models in clinical practice.

Advances in statistical modelling and the availability of large datasets has facilitated the development of prognostic models with greater performance and generalizibility. Two prediction models are currently available, that have been developed on large datasets with state of the art methods, offering new opportunities. We see a great potential for use in clinical practice, in research, towards policy making and assessment of the quality of health care delivery. Continued development, refinement and validation is advocated together with assessment of the clinical impact of prediction models. Future directions should include the development of models to predict treatment response.

Introduction

Prognosis is the cornerstone of clinical medicine, because all diagnostic and therapeutic actions aim to improve a patient's prognosis and outcome. Advances in statistical modelling and the availability of large databases have made it possible to consider diagnosis and prognosis in terms of probabilities rather than vague prophecies. Probability estimates can be applied to clinical decision-making, research, and assessment of the quality of health care. Such quantitative estimates are of particular relevance to heterogeneous conditions such as traumatic brain injury (TBI).

TBI poses a major public-health problem, with an estimated annual incidence of up to 500 per 100 000 population and more than 200 hospital admissions per 100000 admissions in Europe each year (1,2). TBI is heterogeneous in terms of cause, pathology, severity, and prognosis, which poses diagnostic challenges. Furthermore, comparison of results between studies is difficult because case mix and treatments may vary substantially.

Various outcomes can be considered in prediction research. A diagnostic perspective is taken in TBI studies, and involves assessment of the probability of structural brain damage or developing an intracranial haematoma, or is used to underpin recommendations for CT scanning (3–5). For example, a recent study used a prediction rule to identify a subset of children who had such low risk for intracranial damage that CT scans were unnecessary (6). These types of diagnostic outcomes are particularly relevant for patients with mild TBI. The ability to predict response to treatment would be highly relevant to patients in the intensive-care setting, in whom intracranial pressure is monitored, but such prognostic rules have not yet been developed.

For patients with moderate and severe TBI, prediction of clinical outcome is also highly relevant. Typically, most studies have defined clinical outcome as mortality or functional outcome assessed with the Glasgow outcome scale (GOS) as their endpoint (7).

In this Review, we focus on the prediction of outcome in terms of mortality and functional outcome in patients with moderate and severe TBI. We aim to describe the basics of prognostic analysis and review the current knowledge about traditional and newly recognised predictors for outcome in TBI. We also discuss prognostic modelling as a novel instrument in medicine, critically review prediction models in TBI, describe the applications for prognostic models in TBI, and provide suggestions for the further development and improvement of prediction research in TBI. We will use the term "outcome" to refer to all endpoints from different studies that use mortality and GOS.

Predictors of outcome

Much research has been done to identify early predictors of mortality and functional outcome, as assessed by the GOS on admission, after moderate or severe TBI. The GOS is usually dichotomised into good recovery and mild disability versus severe disability, vegetative state, and mortality. This is a limitation because we cannot assume that predictors differentiate death from survival as well as they can differentiate good recovery from worse outcomes.

A large body of evidence supports the relation between some predictors and outcome, but for other predictors the relation is less well established. Information obtained during the subsequent clinical course may further contribute to outcome prediction. An overview of the various components of prognostic analysis is shown in figure 3.1, which illustrates the complex relations between potential predictors and highlights gaps in our knowledge (e.g., genomics, biomarkers).

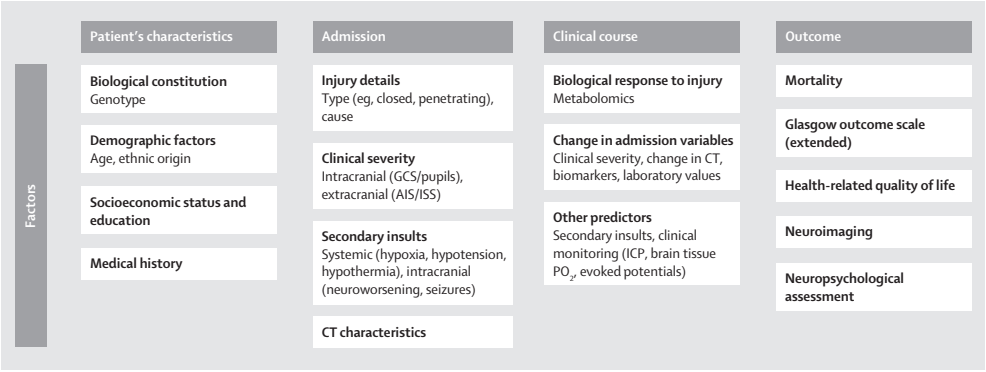


Figure 3.1. Overview of building blocks of prognosis in Traumatic Brain Injury.

Basics of prediction research

Several steps in prediction research have been identified (table 3.1) (8). First, the association between a single predictor and outcome can be studied by univariate analysis, relating a single predictor to the outcome of interest. Such an analysis is of limited value because it does not take into account the role of other confounding factors that may explain (part of) the observed association. Statistical analyses, such as logistic regression, are therefore needed to adjust for confounders in the assessment of relative risks. Odds ratios (ORs) are commonly used to express the strength of prognostic effects. The relation is significant if the 95% CI for the OR does not include the value 1.0. However, the OR does not account for the prevalence of a predictor. A predictor with a high OR but a very low prevalence is thus of

limited predictive value. Predictive value is better reflected in measures such as explained variation (R^2) (9). Other statistical approaches to prognostic analysis include recursive partitioning (prediction trees) and neural network analysis.

Table 3.1. Steps in prognostic analysis in traumatic brain injury

	Univariate analysis	Multivariable analysis	Prediction models
Aim	To estimate the relationship between a single predictor and outcome	To determine the prognostic value of a predictor, adjusting for confounding effects of other predictors.	To combine predictors into a model with the aim to estimate the risk of an outcome for individual patients
Limitations	Does not take the role of confounding factors into account that may explain (part of) the observed association	In individual patients predictors may influence outcome in opposite directions; does not take interactions or differential effects for specific subpopulations into account.	External validation essential to prove generalisibility outside of the development setting.
Performance measures	Sensitivity, specificity Positive predictive value, negative predictive value Odds ratio	Odds ratio Relative Risk R^2	***Discrimination:*** area under the receiver operating characteristic (AUC) ***Calibration:*** graphical representation Hosmer-Lemeshow goodness of fit test
Presentation	Tabular Graphical representation	Tabular Graphical representation	Web-based calculator Score chart

Sensitivity: proportion of patients with the outcome that have the predictor (true positive)
Specificity: proportion of patients without the outcome that do not have the predictor (true negative)
Positive predictive value (PPV): proportion of patients with the predictor that have the outcome
Negative predictive value (NPV): proportion of patients without the predictor that do not have the outcome

Odds ratio (OR): ratio of the odds for better versus poorer outcome in the presence of the parameter, compared to the odds in the absence of the parameter.
Relative risk (RR): risk of outcome in group with the predictor versus group without the predictor
R^2: percentage of variability in outcome that is explained by the predictor. R^2 reflects predictive value better than OR, since also prevalence is taken into account.

Admission characteristics

The prognostic strength of the main predictors of outcome in TBI is summarised in table 3.2. The prognostic value of the different characteristics was quantified in the International Mission for Prognosis and Clinical Trial design in TBI (IMPACT) study data (figure 3.2) (10).

Clinical severity had the highest prognostic value (R^2), followed by CT characteristics. Both measures were significant if included in the model separately and when they were added in the order of availability in clinical practice.

Table 3.2. Strength of the association between predictors and outcome (full ordinal GOS) in TBI in the IMPACT database (n=8686) (10)

Predictor	Reference category	Univariate OR*	Multivariate OR* (adjusted for Age/ Motor score/ Pupils)
Demographics			
- Age	25-75% IQR	2.14 (2.00-2.28)	-
- Gender	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	Localising/		
- Absent	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			
- 1 reacting	Both reacting	2.70 (2.07-3.53)	
- Both non-reacting		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	0.45 (0.35-.067)	0.47 (0.32-0.70)
- CT Class III/IV		2.62 (2.13-3.21)	2.23 (1.83-2.72)
- Mass lesion		2.18 (1.83-2.61)	1.48 (1.27-1.71)
tSAH present	Absent	2.64 (2.42-2.89)	2.01 (1.83-2.21)
Type of intracranial lesion			
- Epidural	No epidural	0.64 (0.56-0.72)	0.63 (0.55-072)
Laboratory parameters			
Glucose	25-75% IQR	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)
Prothrombine time		1.41 (0.99-1.99)	1.63 (1.40-1.89)
Hb		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)

IQR = interquartile range, tSAH = Traumatic Subarachnoid Hemorrhage, * Odds ratios (OR) from proportional odds analysis

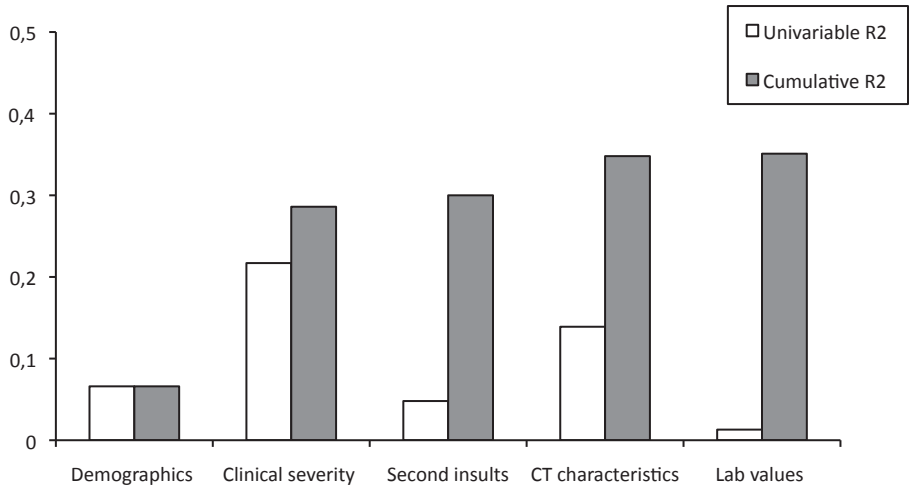


Figure 3.2. Prognostic value of different building blocks of prognosis, expressed in R^2, in the IMPACT dataset (n=8686) (10)

Demographic factors

Age is one of the strongest predictors of mortality and functional outcome in TBI; many studies have shown that older age is associated with poorer outcome (11–21). Most studies analysed the association between age and outcome by use of threshold values, varying from 30 to 60 years of age (11–16). Only a few studies have analysed the association between age and outcome in a continuous way; these studies report both a change around age 30–40 years, above which outcome becomes increasingly poorer, and a fairly continuous relation across all ages, which may be approximated by a linear function (17–21).

Other demographic factors, such as sex and ethnic origin, are also associated with outcome after TBI. Men are more likely to have TBI owing to their higher risk for road traffic accidents and assaults. Although many studies did not find a relation between sex and outcome after adjustment for confounding features (17, 20, 22, 23), a meta- analysis by Farace and Alves (24) found poorer quality of life and worse functional outcomes in women who survived severe TBI compared with men.

The association between ethnic origin and outcome after TBI was controversial until a meta-analysis combining evidence from 5320 patients showed that black patients have a poorer outcome than white or Asian patients. This association has been confirmed by recent studies (20, 25–27). The underlying reasons for this association are speculative, but may include differences in genetic constitution, leading to a different response to injury, and differences in access to acute and post-acute care. Identification of such factors is a priority for further research.

Clinical severity

Clinical severity relates to extracranial and intracranial injuries. The overall severity of extracranial injuries is often assessed with the abbreviated injury score (28) or the injury severity score (28, 29). Most studies on TBI and extracranial injury have included patients with traumatic extracranial injury with or without TBI, and concluded that the coexistence of moderate TBI and extracranial injury is associated with high mortality and morbidity (30–32). By contrast, there is no consensus on the prognostic value of major extracranial injury in patients with TBI. Some studies showed that outcome mainly depends on the severity of the primary cerebral damage and is not worsened by the presence of extracranial injuries (33), but in others, the presence of major extracranial injuries was associated with poorer outcome (34, 35).

We recently did a meta-analysis of individual patient data (129), and found that the conflicting results on the prognostic value of extracranial injuries in previous studies may be explained by an interaction with the severity of brain injury. For patients with more severe brain injuries, the effect of extracranial injury on functional outcome was small, whereas in those with milder brain injuries, extracranial injuries had a more pronounced effect. This suggests that testing for clinically plausible interaction effects is relevant. We also found that extracranial injuries mainly increased the risk of early mortality. Thus, the effect of extracranial injury found in registries that include patients who die early will be larger than the effect found in trials that exclude these patients.

The clinical severity of intracranial injuries is indicated by the level of consciousness, as assessed by the Glasgow coma scale (GCS) (7). Many studies have shown an association between a low score on the GCS and poorer outcome. In patients with more severe injuries, the motor component of the GCS has the greatest predictive value, because eye and verbal response in these patients is often absent (36). The prognostic value of the eye and verbal components of the GCS become more relevant in patients with less severe injuries who can obey commands. Of note, GCS may fluctuate early after injury, and some patients deteriorate whereas others improve (37). From a prognostic perspective, assessment of the GCS should therefore be done at a fixed time period, usually on admission after primary respiratory and haemodynamic stabilisation. However, reliable assessment of the GCS may be obscured in the acute setting by medical sedation, paralysis, or intoxication (36, 38).

Abnormalities in pupillary reactivity indicate brainstem damage or compression and are strongly associated with poorer outcome (39). Pupillary reactivity is a more stable variable in the early phase after injury than is the GCS, because it is less prone to influences of sedation and paralysis.

Secondary insults

An injured brain is more vulnerable to systemic secondary insults (i.e., hypoxia and hypotension) than is a normal, healthy brain. Secondary insults are common after TBI, particularly in the prehospital setting, and can increase the degree of damage (40–42). The association of secondary insults with poorer outcome is well established in the prehospital setting or during acute care, and 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 (42–46).

Most studies have used a cut-off value for early hypotensive and hypoxic events (e.g., any episode with a systolic blood pressure <90 mmHg). However, detailed analysis of the association between the blood pressure measured on admission and outcome showed that this relation is continuous: low and high blood pressure are both associated with poorer outcome (U-shaped relation). After adjustment for age, motor score, and pupillary reactivity, the effects of higher blood pressure on outcome largely disappear, which suggests that higher blood pressure values are merely indicative of more severe injuries and could possibly be caused by raised intracranial pressure (Cushing response) (45).

Structural abnormalities

The prognostic value of CT characteristics has been well documented, including the status of basal cisterns, midline shift, the presence and type of intracranial lesions, and traumatic subarachnoid haemorrhage (47). Obliteration of the basal cisterns and the presence of subarachnoid haemorrhage are the strongest CT predictors of outcome (47). In 1991, Marshall and colleagues (48) introduced a descriptive classification of head injury based on CT characteristics, which focuses on the presence or absence of a mass lesion and differentiates diffuse injuries by signs of increased intracranial pressure (compression of basal cisterns, midline shift). Although the Marshall CT classification has prognostic value, combination of individual CT characteristics in a model, such as in the Rotterdam CT score, provides better discrimination between patients with good versus poor outcomes (tables 3.3 and 3.4) (46, 48, 49).

Prognostic studies have mainly focused on CT abnormalities and used relatively broad categorisations. For example, in traumatic subarachnoid haemorrhage (one of the strongest CT predictors of outcome), most studies have concentrated on the presence or absence of this abnormality without differentiating the location (basal cisterns vs. cortical) or extent.

More detailed analysis and the use of advanced MRI imaging may yield additional prognostic information.

Table 3.3. Marshall CT Classification

Category	Definition
Diffuse injury I (no visible pathology)	No visible intracranial pathology seen on CT scan
Diffuse injury II	Cisterns are present with midline shift of 0–5 mm and/or lesions densities present; no high or mixed density lesion >25 cm^3 may include bone fragments and foreign bodies*
Diffuse injury III (swelling)	Cisterns compressed or absent with midline shift of 0–5mm; no high or mixed density lesion >25 mm
Diffuse injury IV (shift)	Midline shift >5 mm; no high or mixed density lesion >25 cm^3
Evacuated mass lesion	Any lesion surgically evacuated
Non-evacuated mass lesion	High or mixed density lesion >25 cm^3; not surgically evacuated

* As may be the case in depressed skull fractures. Reproduced form Marshall et al. (48).

Table 3.4. 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	+1

Reproduced from Maas et al (49).

Biomarkers and laboratory variables

Interest in the use of biomarkers, including laboratory variables, has been increasing in recent years. Biomarkers may be used to detect and track pathophysiological phenomena as markers of injury severity, and to help assess prognosis. In mild TBI, a biomarker that could

be used to establish the diagnosis or predict the likelihood of secondary damage would have great clinical use. In more severe injuries, use of a biomarker to assess injury severity could help to avoid problems with unreliable GCS assessments in patients who are intoxicated or intubated. Several putative serum, CSF, and microdialysate biomarkers have been evaluated in clinical studies of TBI: S100 protein and neuron-specific enolase are among the most widely investigated (50–55). Although an association between several biomarkers and outcome has been established, the prognostic value of biomarkers is unclear owing to relatively small numbers analysed in univariate analyses (56). Biomarker concentrations may correlate with other clinical indicators such as GCS (57), but they offer limited additional prognostic value over other predictors and need to be assessed in multivariable analysis (58).

The prognostic value of routinely measured laboratory variables has been more widely investigated. High glucose concentrations, low haemoglobin, low platelets, and coagulation disturbances are the strongest predictors of outcome, and are independently related to poorer outcome (59–62). The advantage of laboratory variables is that they are potentially modifiable. The question of causality is relevant when attempts are made to correct abnormal values in the expectation that this will improve outcome. On the basis of the observed association between higher glucose concentration and poorer outcome, two randomised trials were recently done to assess the effect of intensive insulin therapy to reduce glucose concentrations. However, both studies were small (<150 patients) and results conflicted (63,64). The risks of tight glucose control in TBI have been illustrated in microdialysis studies in the brain and show that normalisation of blood glucose could lead to a depletion of glucose in the extracellular fluid of the brain, thus compromising cerebral metabolism (65–67). Although an association between abnormal concentrations of laboratory variables and poorer outcome might exist, this does not mean that correcting these abnormal values will improve outcome. The observed abnormality may simply be an expression or surrogate marker of the severity of injury. Randomised controlled trials are thus required to establish whether the correction of abnormal concentrations is beneficial.

Clinical course

Changes in admission variables

Deterioration in neurological function is a dire prognostic sign that generally indicates progressive brain damage. Early prognosis studies showed that the worst GCS recorded over a given time period is especially predictive of poorer outcome. Deterioration in neurological function has been defined more objectively as neuroworsening (panel 3.1), and is highly predictive for poor outcome (68). In addition to the initial CT scan, follow-up scans also provide prognostic information. A survey among patients with moderate and severe TBI by the European Brain Injury Consortium showed that a substantial proportion of patients with diffuse injury (no mass lesions) on their first CT scan had progressive intracranial damage

on subsequent CT examinations (69). The worst CT scan was more strongly correlated with outcome. Many other studies have confirmed the frequent occurrence of CT progression, but relatively few have addressed the issue of prognostic significance. This is a complex area to study, because CT progression can often lead to therapeutic intervention.

Secondary insults can occur in the clinical setting, despite clinicians' best attempts to avoid them. Patients are particularly at risk for secondary insults during transport within and between hospitals (70). The severity, duration, and number of secondary insults contribute to a poorer outcome (45, 46, 71).

The same laboratory variables that have high prognostic value at admission (i.e., glucose, platelets, and coagulation disturbances) also have value during the clinical course. Persistently high glucose concentrations are associated with poorer outcomes, even after adjustment for other important predictors (59, 72, 73). The lowest platelet count during the first 24 h after admission is an independent predictor of outcome after 6 months (59). A recent meta-analysis showed that the prevalence of coagulopathy after TBI was 33% and that coagulopathy was related both to mortality and an unfavourable outcome (74).

Panel 3.1. Criteria for neuroworsening

- ❑ Spontaneous decrease GCS motor score >=2 points (compared with previous examination)
- ❑ New loss of pupillary reactivity
- ❑ Development of pupillary asymmetry of >=2 mm
- ❑ Other deterioration in neurological status sufficient to warrant immediate medical or surgical intervention

Clinical monitoring

In more severely injured patients, invasive and non-invasive monitoring in the intensive-care unit can provide much information. However, approaches to analysis have remained relatively crude. It is difficult to draw clear conclusions on the predictive value of monitored variables because the time of initiation and duration of monitoring vary greatly between and within studies. Summary measures, such as for intracranial pressure monitoring, include the highest, lowest, and mean values overall or per day, and the number of episodes or percentage of time that values are above a predefined threshold. This variability in analysis and reporting confounds comparisons between studies. Furthermore, predictive information on repeated measurements might be better captured in patterns than in single values. Modern statistical approaches are available to analyse repeated measurements per patient, but have seldom been used in TBI studies and hence pose a challenge for future research.

Many studies have confirmed an association of high intracranial pressure, low cerebral perfusion pressure, and decreased brain oxygen tension with poorer outcome (59, 75–80). These associations, in combination with our understanding of pathophysiological consequences, form the rationale for guideline recommendations to avoid high intracranial pressure and low cerebral perfusion pressure (47). Additionally, outcome might be more dependent on intracranial pressure variability and on response to treatment of raised intracranial pressure than on absolute mean values (79, 81).

Electroencephalography and evoked potentials

Over past decades, there has been interest in electroencephalography (EEG) and evoked potentials or event-related potentials as predictors of outcome (82, 83). However, some have suggested that the predictive ability of EEG is limited because TBI has a greater effect on subcortical axonal fibres than on cortical grey matter, which generates most of the EEG signal (84). In the post-acute phase, the bispectral index, a measure of level of consciousness by algorithmic analysis of the EEG, has a higher correlation with behavioural scales than does the EEG and may help in differentiating between a vegetative and minimally conscious state after TBI (85). Many studies have shown that somatosensory evoked potentials (SSEPs) are useful predictors of outcome after TBI (86–88). Lew and colleagues (88) reported that bilateral absence of cortically recorded median nerve SSEPs within 8 days of severe TBI was strongly predictive of death or persistent vegetative state. A meta-analysis showed that bilaterally negative SSEPs had a 98.5% positive likelihood ratio for an unfavourable outcome (89). However, the false-positive percentage for bilaterally negative SSEPs may be high in patients with focal lesions, subdural effusions, and after recent decompressive craniectomies. Although results are promising, the evidence on the prognostic effects of these clinical neurophysiological modalities is limited, and the added value over other clinical predictors is uncertain.

Prognostic models

Estimation of prognosis is by definition a multivariable challenge. Predictors should be considered jointly rather than on their own, and can be combined in a multivariable prognostic model to quantify the risk for a particular outcome in individual patients.

Combining individual predictors into a model will increase performance and generalisability, and is important because patients could have characteristics that affect the outcome in opposite directions. For example, for a 24-year-old patient with fixed pupils, we would expect a favourable outcome based on age, but an unfavourable outcome based on pupil reactivity.

In our literature search for this Review, we identified 27 prognostic models in 16 studies that met the following criteria: prognostic model for mortality more than 2 weeks after discharge or at 6-month GOS (in English); predictors measured within 24 h after injury; inclusion of more than 200 patients aged older than 14 years; GCS of less than 14 or motor score less than 6; and non- penetrating injury. Many of these models had shortcomings, in particular a high risk of overfitting (e.g., predictive performance is much poorer in new patients than expected from the development phase) in ten studies, and lack of external validation (table 3.5) (11, 17, 90–104). The number of predictors considered was usually higher than the number actually included in the final model, and was often too high in relation to the available sample size. As a guide, the maximum number of candidate predictors can be approximated by dividing the number of events (e.g., number of patients with poor outcome) by 10 (e.g., at most 10 predictors for 100 events) (105). Overfitting is caused by use of statistical techniques for the selection of predictors that are too data-driven, such as backward selection in a small dataset. Overfitting can be assessed by internal validation techniques, such as bootstrap resampling (106). More important is external validation, such as testing the model's performance in another setting that differs in time or place (107, 108). Only five studies reported external validation (11, 17, 102–104). These findings are consistent with reviews published by Perel and Mushkudiani and their colleagues (34, 109).

The models reported by the Corticosteroid Randomisation After Significant Head injury (CRASH) trial investigators and by the IMPACT study group are the most recent, and were developed on the largest numbers of patients (10008 and 8509, respectively) (17, 104).

Different sets of prediction models were developed by use of logistic regression analysis and cross-validated on each other. The models are available in score-chart format and in a web-based application. Both studies showed that the largest amount of prognostic information is contained in a core set of three predictors (age, GCS or motor score, and pupillary reactivity; table 3.6). Development of the CRASH models also included patients with milder injuries, and these models can thus be used to assess such patients. The IMPACT models focused on moderate and severe TBI. Both sets of models can therefore be recommended for prognostic modelling in TBI because they were developed on large numbers of patients and conform to accepted quality criteria for model development, including external validation.

Table 3.5. Overview of prognostic models in moderate and severe TBI published between 1976 and 2009

Reference	Year	N (development)	N predictors included	Severity	Outcome	Risk of overfitting	External validation
Jennett et al. (90)	1976	600	4	In coma for at least 6h	GOS 6 months	High	No
Braakman et al. (91)	1980	305	3	In coma for at least 6h	GOS 6 months	High	No
Choi et al. (92)	1983	264	4	Severe head injury and GMS ≤5	GOS 6 months	Intermediate	No
Lokkenberg and Grimes (93)	1984	254	2	GCS ≤8	GOS 6 months	Low	No
Braakman et al. (94)	1986	306	3	Severe head injury	GOS 6 months	Intermediate	No
Choi et al. (95)	1988	523	3	Severe head injury	GOS 6 months	High	No
Choi et al. (96)	1991	555	4	GCS ≤8	GOS 12 months	High	No
Fearnside et al. (2 models) (97)	1993	315/218	5	GCS ≤8	Mortality/GOS	Intermediate	No
Marmelak et al. (98)	1996	672	3	GCS ≤8	GOS 6 months	Low	No
Quigley et al. (99)	1997	380	2	GCS 3-5	GOS 6 months	Low	No
Lang et al. (100)	1997	799	4	GCS ≤8	Mortality 6 months	High	No
Signorini et al. (11)	1999	372	5	GCS ≤12 and GCS >12 if ISS >15	Mortality 12 months	Intermediate	Yes
Ratanalert et al. (101)	2002	337	3	GCS ≤8	GOS 6 months	Intermediate	No
Hukkelhoven et al. (2 models) (102)	2005	2269	7	GCS ≤12	Mortality/GOS 6 months	Low	Yes
Cremer et al. (103)	2006	304	5	GCS ≤8 and in coma for at least 24h	GOSE 12 months	Low	Yes
Perel et al. (4 models) (17)	2008	10008	4-9	GCS ≤14	Mortality/GOS 14 days/6 months	Low	Yes
Steyerberg et al. (6 models) (104)	2009	8509	3-10	GCS ≤12	Mortality/GOS 6 months	Low	Yes

Table 3.6. Comparison of the CRASH and IMPACT prognostic models

	IMPACT	CRASH
Predicted outcome	6 month mortality 6 month unfavourable outcome	14 day mortality 6 month unfavourable outcome
Core model	Age Motor score Pupil reactivity	Age GCS Pupil reactivity Major extracranial injury
CT model	*Core model +* Hypoxia Hypotension CT Classification tSAH on CT Epidural mass on CT	*Core model +* petechial haemorrhages Obliteration of the third ventricle or basal cisterns Subarachnoid bleeding Midline shift Non-evacuated haematoma
Lab model	*CT model +* Glucose Hb	

Applications of prognostic models in TBI

Clinical practice

Some estimation of prognosis is consciously or subconsciously used by physicians when informing relatives, making treatment decisions, or allocating resources. Estimates derived from large datasets are preferable to the subjective opinion of a physician, whose experience, no matter how vast, can never match the information contained in data from thousands of patients. The Canadian CT rule and the CHIP (CT in head injury patients) prediction rule for CT scanning in mild TBI are examples of how prediction models can provide evidence to better inform clinical decisions (3, 5). Caution is advocated when outcome prediction models are applied in individual patients. Prognostic estimates are only probabilities and cannot provide certainty on an actual outcome.

Research

The main research applications of prognostic models for outcome in TBI include classification and clinical trials. Prognostic risk estimation on hospital admission enables populations to be classified according to their prognostic risk distribution (figure 3.3) (112). We can therefore use such models to gain insight into differences in the case-mix of different studies. In the design and analysis of randomised controlled trials, prognostic models offer opportunities both in the enrolment and analysis phases. Traditionally, clinical trials use relatively strict enrolment criteria. Some of these criteria are motivated by safety and ethical considerations,

but most (e.g., age and disease severity) aim to exclude patients with a very good or a very poor prognosis. Patients at these extremes are not likely to benefit from the treatment under investigation. It is statistically more efficient to combine these criteria in a prognostic model (113–116). The prognostic estimate can first be used to determine eligibility and then be used for the analysis.

In the analysis phase, prognostic models can adjust for baseline characteristics. This substantially increases statistical power, allows the required sample size to be reduced (by more than 25%) (117). Prognostic analysis can also make use of the sliding dichotomy, whereby the point of dichotomy of the GOS is differentiated according to the baseline prognostic risk (118). For a patient with a very severe injury, survival may be more relevant, whereas for patients with less severe injuries, any outcome worse than good recovery might be considered unfavourable. The sliding dichotomy approach has been adopted for the primary analysis of several phase 3 trials in TBI, stroke, and intracerebral haemorrhage (119–121).

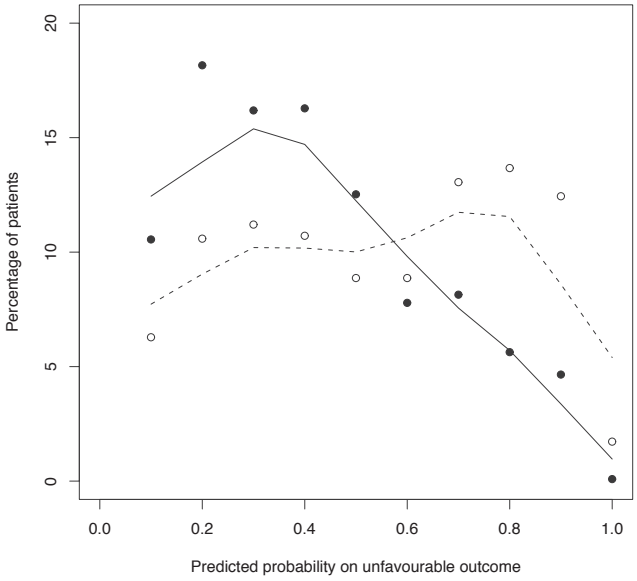


Figure 3.3. Distribution of predicted probabilities on 6-month unfavourable outcome from the IMPACT core model Comparison of an observational study (UK4; open points, dashed line) (110) to a randomised controlled trial (Tirilizad International; solid points, solid line) (111) among patients with TBI. The proportion of patients with a relatively poor prognosis was smaller in the randomised controlled trial than in the observational study. IMPACT=International Mission for Prognosis and Clinical Trial design in TBI. TBI=traumatic brain injury. Adapted from Lingsma et al. (112).

Quality assessment of health-care delivery

Comparison of observed and expected outcomes may give an indication of the quality of care delivered in a specific hospital or in a specific country. An example is the standardised mortality ratio (observed deaths/expected deaths, adjusted for baseline characteristics), which is used as a quality score in intensive-care medicine. Expected mortality is commonly derived from severity-of-disease classification systems used in intensive-care units.

However, these systems were developed for a general intensive-care population, and their applicability to TBI is uncertain. Prognostic models that are specific to TBI are more useful for setting baselines for clinical audits and benchmarking. These models are of great potential relevance for assessing the quality of health-care delivery, because they have been developed not only for mortality, but also for functional outcome, as assessed by the GOS. Of note, however, the cumulative R2 of the IMPACT model was 0.35, which indicates that 65% of the variation is unexplained, so case-mix adjustment is incomplete.

Conclusions and future directions

Prognostic analysis and modelling have great potential in TBI, both for diagnosis and prognosis. Although some of the gaps in our knowledge have been identified, some issues require further investigation. Validated prognostic models have been based mainly on admission characteristics. Although substantial insight has been gained into the prognostic value of variables obtained during the subsequent clinical course, such variables have not yet been widely included in prognostic models. Further research should focus on the quantification of the additional benefit that might be obtained for outcome prediction. The epidemiology of TBI is changing, and approaches to prehospital care, diagnostic capabilities, and intensive-care monitoring and treatment are continually improving. Consequently, prognostic analysis should be seen as a continuous process that needs updating and validation in contemporary series (123).

In the analyses of continuous variables such as age, blood pressure, or laboratory variables, many studies used threshold values, creating a dichotomy or categorisation of continuous predictors (e.g., age ≤50 years vs >50 years). Threshold values are increasingly used in clinical medicine in the practice of goal-directed therapy. However, such values are not natural to biological systems, and the collapsing of continuous variables has many disadvantages (124). We recommend that future prediction studies should analyse continuous predictors in a continuous way, possibly even as non-linear variables (56).

A major gap in our knowledge concerns different responses to similar injuries by different individuals. Such differences could in part be genetically determined, and much research will be needed in the areas of genomics and metabolomics to elucidate variability in response. The relevance of genetics may be shown by the observation that recovery is

poorer in patients with stroke or TBI who have the APOE ε4 allele than in those who do not have this allele (125). Other genes for which evidence exists for an association with poorer outcome are TP53, COMT, DRD2, and CACNA1A (126). Additionally, response to treatment varies between individuals. Research on factors that predict response to treatment in TBI is underway, including various biomarkers and imaging modalities. Predictive factors may lead to targeted therapies, and take into account individual mechanisms of disease (127). Further research is also needed into more sensitive outcome measures, particularly in milder TBI.

Directly relevant to prognostic research in TBI is better standardisation of data collection and coding to facilitate sharing of results and to allow meta-analysis of individual patient data across studies (128). This will provide the opportunity to improve, validate, and update prognostic models on larger numbers of patients.

The challenge for the immediate future is the implementation of prediction models in clinical practice. The tools are now available in the form of reliable and externally validated models. Clinicians and researchers now need to adopt these models in general clinical and research applications, either to improve quality of care, or to improve the prognostic estimate.

Contributors

HFL and BR contributed equally to this Review, in searching and reading literature, and writing. AIRM provided additional important literature sources and developed the outline. GDM and EWS helped to write the Review.

References

1. Maas AI, Marmarou A, Murray GD, Teasdale SG, Steyerberg EW. **Prognosis and clinical trial design in traumatic brain injury: the IMPACT study.** *J Neurotrauma* 2007; 24: 232–38.
2. **Styrke J, Stalnacke BM, Sojka P, Bjornstig U. Traumatic brain injuries in a well-defined population: epidemiological aspects and severity.** *J Neurotrauma* 2007; 24: 1425–36.
3. Smits M, Dippel DW, Steyerberg EW, et al. **Predicting intracranial traumatic findings on computed tomography in patients with minor head injury: the CHIP prediction rule.** *Ann Intern Med* 2007; 146: 397–405.
4. Haydel MJ, Preston CA, Mills TJ, Luber S, Blaudeau E, DeBlieux PM**. Indications for computed tomography in patients with minor head injury.** *N Engl J Med* 2000; 343: 100–05.
5. Stiell IG, Wells GA, Vandemheen K, et al. **The Canadian CT Head Rule for patients with minor head injury.** *Lancet* 2001; 357: 1391–96.
6. Kuppermann N, Holmes JF, Dayan PS, et al. **Identification of children at very low risk of clinically-important brain injuries after head trauma: a prospective cohort study.** *Lancet* 2009; 374: 1160–70.
7. Teasdale G, Jennett B. **Assessment of coma and impaired consciousness. A practical scale.** *Lancet* 1974; 2: 81–84.
8. Steyerberg EW. **Clinical prediction models. A practical approach to development, validating and updating.** New York: Springer, 2009.
9. McHugh GS, Butcher I, Steyerberg EW, et al. **Statistical approaches to the univariate prognostic analysis of the IMPACT database on traumatic brain injury.** *J Neurotrauma* 2007; 24: 251–58.
10. Murray GD, Butcher I, McHugh GS, et al**. Multivariable prognostic analysis in traumatic brain injury: results from the IMPACT study.** *J Neurotrauma* 2007; 24: 329–37.
11. Signorini DF, Andrews PJ, Jones PA, Wardlaw JM, Miller JD. **Predicting survival using simple clinical variables: a case study in traumatic brain injury.** *J Neurol Neurosurg Psychiatry* 1999; 66: 20–25.
12. Ono J, Yamaura A, Kubota M, Okimura Y, Isobe K. **Outcome prediction in severe head injury: analyses of clinical prognostic factors.** *J Clin Neurosci* 2001; 8: 120–23.
13. 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: 326–36.
14. Demetriades D, Murray J, Martin M, et al. **Pedestrians injured by automobiles: relationship of age to injury type and severity.** *J Am Coll Surg* 2004; 199: 382–87.
15. Ratanalert S, Chompikul J, Hirunpat S. **Talked and deteriorated head injury patients: how many poor outcomes can be avoided?** *J Clin Neurosci* 2002; 9: 640–43.
16. 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: 255–61.
17. MRC CRASH Trial Collaborators. **Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients.** *BMJ* 2008; 336: 425–29.
18. Gomez PA, Lobato RD, Boto GR, De la Lama A, Gonzalez PJ, de la Cruz J. **Age and outcome after severe head injury.** *Acta Neurochir (Wien)* 2000; 142: 373–81.
19. 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: 666–73.
20. 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: 259–69.
21. Tokutomi T, Miyagi T, Ogawa T, et al. **Age-associated increases in poor outcomes after traumatic brain injury: a report from the Japan Neurotrauma Data Bank.** *J Neurotrauma* 2008; 25: 1407–14.
22. Colantonio A, Escobar MD, Chipman M, et al. **Predictors of postacute mortality following traumatic brain injury in a seriously injured population.** *J Trauma* 2008; 64: 876–82.
23. Utomo WK, Gabbe BJ, Simpson PM, Cameron PA. **Predictors of in-hospital mortality and 6-month functional outcomes in older adults after moderate to severe traumatic brain injury.** *Injury* 2009; 40: 973–77.
24. Farace E, Alves WM. Do women fare worse? **A metaanalysis of gender differences in outcome after traumatic brain injury.** *Neurosurg Focus* 2000; 8: e6.
25. Sorani MD, Lee M, Kim H, Meeker M, Manley GT. **Race/ethnicity and outcome after traumatic brain injury at a single, diverse center.** *J Trauma* 2009; 67: 75–80.
26. Shafi S, Marquez de la Plata C, Diaz-Arrastia R, et al. **Racial disparities in long-term functional outcome after traumatic brain injury.** *J Trauma* 2007; 63: 1263–70.

27. Arango-Lasprilla JC, Rosenthal M, Deluca J, et al. **Traumatic brain injury and functional outcomes: does minority status matter?** *Brain Inj* 2007; 21: 701–08. Association for the Advancement of Automotive Medicine. The abbreviated injury scale, 1990 revision.
28. Des Plaines, IL: **Association for the Advancement of Automotive Medicine**, 1990: 15–24.
29. Baker SP, O'Neill B, Haddon W Jr, Long WB. **The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care.** *J Trauma* 1974; 14: 187–96.
30. Gennarelli TA, Champion HR, Copes WS, Sacco WJ. **Comparison of mortality, morbidity, and severity of 59,713 head injured patients with 114,447 patients with extracranial injuries.** *J Trauma* 1994; 37: 962–68.
31. Gennarelli TA, Champion HR, Sacco WJ, Copes WS, Alves WM. **Mortality of patients with head injury and extracranial injury treated in trauma centers.** *J Trauma* 1989; 29: 1193–202.
32. McMahon CG, Yates DW, Campbell FM, Hollis S, Woodford M. **Unexpected contribution of moderate traumatic brain injury to death after major trauma.** *J Trauma* 1999; 47: 891–95.
33. Sarrafzadeh AS, Peltonen EE, Kaisers U, Kuchler I, Lanksch WR, Unterberg AW. **Secondary insults in severe head injury—do multiply injured patients do worse?** *Crit Care Med* 2001; 29: 1116–23.
34. Perel P, Edwards P, Wentz R, Roberts I. **Systematic review of prognostic models in traumatic brain injury.** *BMC Med Inform Decis Mak* 2006; 6: 38.
35. Lefering R, Paffrath T, Linker R, Bouillon B, Neugebauer EA; German Society for Trauma Surgery. **Head injury and outcome— what influence do concomitant injuries have?** *J Trauma* 2008; 65: 1036–43.
36. Stocchetti N, Pagan F, Calappi E, et al. **Inaccurate early assessment of neurological severity in head injury.** *J Neurotrauma* 2004; 21: 1131–40.
37. 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: 270–80.
38. Murray GD, Teasdale GM, Braakman R, et al. **The European Brain Injury Consortium survey of head injuries.** *Acta Neurochir (Wien)* 1999; 141: 223–36.
39. 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: 161–62.
40. Stocchetti N, Colombo A, Ortolano F, et al. **Time course of intracranial hypertension after traumatic brain injury.** *J Neurotrauma* 2007; 24: 1339–46.
41. Stocchetti N, Furlan A, Volta F. **Hypoxemia and arterial hypotension at the accident scene in head injury.** *J Trauma* 1996; 40: 764–67.
42. 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: 216–22.
43. Signorini DF, Andrews PJ, Jones PA, Wardlaw JM, Miller JD. **Adding insult to injury: the prognostic value of early secondary insults for survival after traumatic brain injury.** *J Neurol Neurosurg Psychiatry* 1999; 66: 26–31.
44. Walia S, Sutcliffe AJ. **The relationship between blood glucose, mean arterial pressure and outcome after severe head injury: an observational study.** *Injury* 2002; 33: 339–44.
45. 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: 287–93.
46. Chesnut RM. **Secondary brain insults after head injury: clinical perspectives.** *New Horiz* 1995; 3: 366–75.
47. Brain Trauma Foundation**. Guidelines for the management of severe traumatic brain injury (3rd edn).** *J Neurotrauma* 2007; 24 (suppl 1): S1–106.
48. Marshall LF, Bowers S, Klauber MR, et al. **A new classification of head injury based on computerised tomography.** *J Neurosurg* 1991; 75: 1 (suppl): S14–20.
49. Maas AI, Hukkelhoven CW, Marshall LF, Steyerberg EW. **Prediction of outcome in traumatic brain injury with computed tomographic characteristics: a comparison between the computed tomographic classification and combinations of computed tomographic predictors.** *Neurosurgery* 2005; 57: 1173–82.
50. Sawauchi S, Taya K, Murakami S, et al. **Serum S-100B protein and neuron-specific enolase after traumatic brain injury.** *No Shinkei Geka* 2005; 33: 1073–80.
51. Naeimi ZS, Weinhofer A, Sarahrudi K, Heinz T, Vecsei V. **Predictive value of S-100B protein and neuron specific-enolase as markers of traumatic brain damage in clinical use.** *Brain Inj* 2006; 20: 463–68.
52. Nylen K, Ost M, Csajbok LZ, et al. **Serum levels of S100B, S100A1B and S100BB are all related to outcome after severe traumatic brain injury.** *Acta Neurochir (Wien)* 2008; 150: 221–27.

53. Schultke E, Sadanand V, Kelly ME, Griebel RW, Juurlink BH. **Can admission S-100β predict the extent of brain damage in head trauma patients?** *Can J Neurol Sci* 2009; 36: 612–16.

54. Beaudeux JL. **S100B protein: a novel biomarker for the diagnosis of head injury.** *Ann Pharm Fr* 2009; 67: 187–94.

55. Rainey T, Lesko M, Sacho R, Lecky F, Childs C. **Predicting outcome after severe traumatic brain injury using the serum S100B biomarker: results using a single (24h) time-point.** *Resuscitation* 2009; 80: 341–45.

56. Kovesdi E, Luckl J, Bukovics P, et al. **Update on protein biomarkers in traumatic brain injury with emphasis on clinical use in adults and paediatrics.** *Acta Neurochir* 2010; 152: 1–17.

57. Pineda JA, Lewis SB, Valadka AB, et al. **Clinical significance of αII-spectrin breakdown products in cerebrospinal fluid after severe traumatic brain injury.** *J Neurotrauma* 2007; 24: 354–66.

58. Kattan MW. **Judging new markers by their ability to improve predictive accuracy.** *J Natl Cancer Inst* 2003; 95: 634–35.

59. Lannoo E, Van Rietvelde F, Colardyn F, et al. **Early predictors of mortality and morbidity after severe closed head injury.** *J Neurotrauma* 2000; 17: 403–14.

60. 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: 315–28.

61. Saggar V, Mittal RS, Vyas MC. **Hemostatic abnormalities in patients with closed head injuries and their role in predicting early mortality.** *J Neurotrauma* 2009; 26: 1665–68.

62. Rovlias A, Kotsou S. **The blood leukocyte count and its prognostic significance in severe head injury**. *Surg Neurol* 2001; 55: 190–96.

63. Yang M, Guo Q, Zhang X, et al. **Intensive insulin therapy on infection rate, days in NICU, in-hospital mortality and neurological outcome in severe traumatic brain injury patients: a randomized controlled trial.** *Int J Nurs Stud* 2009; 46: 753–58.

64. Bilotta F, Caramia R, Cernak I, et al. **Intensive insulin therapy after severe traumatic brain injury: a randomized clinical trial.** *Neurocrit Care* 2008; 9: 159–66.

65. Bilotta F, Caramia R, Paoloni FP, Delfini R, Rosa G. **Safety and efficacy of intensive insulin therapy in critical neurosurgical patients.** *Anesthesiology* 2009; 110: 611–19.

66. Vespa PM. **Intensive glycemic control in traumatic brain injury: what is the ideal glucose range?** *Crit Care* 2008; 12: 175.

67. Vespa PM. **The implications of cerebral ischemia and metabolic dysfunction for treatment strategies in neurointensive care.** *Curr Opin Crit Care* 2006; 12: 119–23.

68. Morris GF, Juul N, Marshall SB, Benedict B, Marshall LF. **Neurological deterioration as a potential alternative endpoint in human clinical trials of experimental pharmacological agents for treatment of severe traumatic brain injuries.** *Neurosurgery* 1998; 43: 1369–74.

69. Servadei F, Murray GD, Penny K, et al. **The value of the "worst" computed tomographic scan in clinical studies of moderate and severe head injury.** *Neurosurgery* 2000; 46: 70–77.

70. Gentleman D. **Causes and effects of systemic complications among severely head injured patients transferred to a neurosurgical unit.** *Int Surg* 1992**; 77: 297–302.**

71. **Manley G, Knudson MM, Morabito D, Damron S, Erickson V, Pitts L. Hypotension, hypoxia, and head injury: frequency, duration, and consequences.** *Arch Surg* 2001; 136: 1118–23.

72. Rovlias A, Kotsou S. **The influence of hyperglycemia on neurological outcome in patients with severe head injury.** *Neurosurgery* 2000; 46: 335–43.

73. Salim A, Hadjizacharia P, Dubose J, et al. **Persistent hyperglycemia in severe traumatic brain injury: an independent predictor of outcome.** *Am Surg* 2009; 75: 25–29.

74. Harhangi BS, Kompanje EJ, Leebeek FW, Maas AI. **Coagulation disorders after traumatic brain injury.** *Acta Neurochir (Wien)* 2008; 150: 165–75.

75. Huang SJ, Hong WC, Han YY, et al. **Clinical outcome of severe head injury in different protocol-driven therapies.** *J Clin Neurosci* 2007; 14: 449–54.

76. Struchen MA, Hannay HJ, Contant CF, Robertson CS**. The relation between acute physiological variables and outcome on the Glasgow Outcome Scale and Disability Rating Scale following severe traumatic brain injury.** *J Neurotrauma* 2001; 18: 115–25.

77. Juul N, Morris GF, Marshall SB, Marshall LF. **Intracranial hypertension and cerebral perfusion pressure: influence on neurological deterioration and outcome in severe head injury.** *J Neurosurg* 2000; 92: 1–6.

78. Kirkness CJ, Burr RL, Cain KC, Newell DW, Mitchell PH. **Relationship of cerebral perfusion pressure levels to outcome in traumatic brain injury.** *Acta Neurochir Suppl* 2005; 95: 13–16.

79. Kirkness CJ, Burr RL, Mitchell PH. **Intracranial pressure variability and long-term outcome following traumatic brain injury.** *Acta Neurochir Suppl* 2008; 102: 105–08.

80. Vik A, Nag T, Fredriksli OA, et al. **Relationship of "dose" of intracranial hypertension to outcome in severe traumatic brain injury.** *J Neurosurg* 2008; 109: 678–84.

81. Treggiari MM, Schutz N, Yanez ND, Romand JA. **Role of intracranial pressure values and patterns in predicting outcome in traumatic brain injury: a systematic review.** *Neurocrit Care* 2007; 6: 104–12.

82. 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: 751–62.

83. Barelli A, Valente MR, Clemente A, Bozza P, Proietti R, Della Corte F. **Serial multimodality-evoked potentials in severely head-injured patients: diagnostic and prognostic implications.** *Crit Care Med* 1991; 19: 1374–81.

84. Wang JT, Young GB, Connolly JF. **Prognostic value of evoked responses and event-related brain potentials in coma.** *Can J Neurol Sci* 2004; 31: 438–50.

85. Schnakers C, Ledoux D, Majerus S, et al. **Diagnostic and prognostic use of bispectral index in coma, vegetative state and related disorders.** *Brain Inj* 2008; 22: 926–31.

86. Sleigh JW, Havill JH, Frith R, Kersel D, Marsh N, Ulyatt D. **Somatosensory evoked potentials in severe traumatic brain injury: a blinded study.** *J Neurosurg* 1999; 91: 577–80.

87. Mazzini L, Pisano F, Zaccala M, Miscio G, Gareri F, Galante M. **Somatosensory and motor evoked potentials at different stages of recovery from severe traumatic brain injury.** *Arch Phys Med Rehabil* 1999; 80: 33–39.

88. Lew HL, Dikmen S, Slimp J, et al. **Use of somatosensory-evoked potentials and cognitive event-related potentials in predicting outcomes of patients with severe traumatic brain injury.** *Am J Phys Med Rehabil* 2003; 82: 53–64, 80.

89. Carter BG, Butt W. **Review of the use of somatosensory evoked potentials in the prediction of outcome after severe brain injury.** *Crit Care Med* 2001; 29: 178–86.

90. Jennett B, Teasdale G, Braakman R, Minderhoud J, Knill-Jones R. **Predicting outcome in individual patients after severe head injury.** *Lancet* 1976; 1: 1031–34.

91. Braakman R, Gelpke GJ, Habbema JD, Maas AI, Minderhoud JM. **Systematic selection of prognostic features in patients with severe head injury.** *Neurosurgery* 1980; 6: 362–70.

92. Choi SC, Ward JD, Becker DP. **Chart for outcome prediction in severe head injury.** *J Neurosurg* 1983; 59: 294–97.

93. Lokkeberg AR, Grimes RM. **Assessing the influence of non- treatment variables in a study of outcome from severe head injuries.** *J Neurosurg* 1984; 61: 254–62.

94. Braakman R, Habbema JD, Gelpke GJ. **Prognosis and prediction of outcome in comatose head injured patients.** *Acta Neurochir Suppl (Wien)* 1986; 36: 112–17.

95. Choi SC, Narayan RK, Anderson RL, Ward JD. **Enhanced specificity of prognosis in severe head injury.** *J Neurosurg* 1988; 69: 381–85.

96. Choi SC, Muizelaar JP, Barnes TY, Marmarou A, Brooks DM, Young HF. **Prediction tree for severely head-injured patients.** *J Neurosurg* 1991; 75: 251–55.

97. Fearnside MR, Cook RJ, McDougall P, McNeil RJ. **The Westmead Head Injury Project outcome in severe head injury. A comparative analysis of pre-hospital, clinical and CT variables.** *Br J Neurosurg* 1993; 7: 267–79.

98. Marmelak AN, Pitts LH, Damron S. **Predicting survival from head trauma 24 hours after injury: a practical method with therapeutic implications.** *J Trauma* 1996; 41: 91–99.

99. Quigley M, Vidovich D, Cantella D, Wilberger J, Maroon J, Diamond D. **Defining the limits of survivorship after very severe head injury.** *J Trauma* 1997; 42: 7–10.

100. Lang EW, Pitts LH, Damron SL, Rutledge R. **Outcome after severe head injury: an analysis of prediction based upon comparison of neural network versus logistic regression analysis.** *Neurol Res* 1997; 19: 274–80.

101. Ratanalert S, Chompikul J, Hirunpat S, Pheunpathom N. **Prognosis of severe head injury: an experience in Thailand.** *Br J Neurosurg* 2002; 16: 487–93.

102. Hukkelhoven CWPM, Steyerberg EW, Habbema JDF, et al. **Predicting outcome after severe or moderate traumatic brain injury: development and validation of a prognostic score based on admission characteristics.** *J Neurotrauma* 2005; 22: 1025–39.

103. Cremer OL, Moons KG, van Dijk GW, van Balen P, Kalkman CJ. **Prognosis following severe head injury: development and validation of a model for prediction of death, disability, and functional recovery.** *J Trauma* 2006; 61: 1484–91.

104. 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.

105. **Steyerberg EW, Eijkemans MJ, Harrell FE Jr, Habbema JD. Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets.** *Med Decis Making* 2001; 21: 45–56.

106. Steyerberg EW, Harrell FE Jr, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD**. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis.** *J Clin Epidemiol* 2001; 54: 774–81.

107. Bleeker SE, Moll HA, Steyerberg EW, et al**. External validation is necessary in prediction research: a clinical example.** *J Clin Epidemiol* 2003; 56: 826–32.

108. Justice AC, Covinsky KE, Berlin JA. **Assessing the generalizability of prognostic information.** *Ann Intern Med* 1999; 130: 515–24.

109. 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: 331–43.

110. Murray LS, Teasdale GM, Murray GD, Miller DJ, Pickard JD, Shaw MD. **Head injuries in four British neurosurgical centres.** *Br J Neurosurg* 1999; 13: 564–69.

111. Marshall LF, Maas AI, Marshall SB, et al**. A multicenter trial on the efficacy of using tirilazad mesylate in cases of head injury.** *J Neurosurg* 1998; 89: 519–25.

112. Lingsma HF, Maas AIR, Steyerberg EW. **Prognostication of moderate and severe traumatic brain injury [in Dutch].** *Ned Tijdschr Geneeskd* 2010; 154: 107–14.

113. Machado SG, Murray GD, Teasdale GM. **Evaluation of designs for clinical trials of neuroprotective agents in head injury.** *J Neurotrauma* 1999; 16: 1131–38.

114. Roozenbeek B, Maas AI, Lingsma HF, et al. **Baseline characteristics and statistical power in randomized controlled trials: selection, prognostic targeting, or covariate adjustment?** *Crit Care Med* 2009; 37: 2683–90.

115. Kent D, Hayward R. **Subgroup analyses in clinical trials.** *N Engl J Med* 2008; 358: 1199.

116. Hayward RA, Kent DM, Vijan S, Hofer TP. **Multivariable risk prediction can greatly enhance the statistical power of clinical trial subgroup analysis.** *BMC Med Res Methodol* 2006; 6: 18.

117. Hernandez AV, Steyerberg EW, Taylor GS, Marmarou A, Habbema JD, Maas AI. **Subgroup analysis and covariate adjustment in randomized clinical trials of traumatic brain injury: a systematic review.** *Neurosurgery* 2005; 57: 1244–53.

118. Murray GD, Barer D, Choi S, et al. **Design and analysis of phase III trials with ordered outcome scales: the concept of the sliding dichotomy.** *J Neurotrauma* 2005; 22: 511–17.

119. Mendelow AD, Gregson BA, Fernandes HM, et al. **Early surgery versus initial conservative treatment in patients with spontaneous supratentorial intracerebral haematomas in the International Surgical Trial in IntraCerebral Haemorrhage (STICH): a randomised trial.** *Lancet* 2005; 365: 387–97.

120. Maas AI, Murray G, Henney H 3rd, et al**. Efficacy and safety of dexanabinol in severe traumatic brain injury: results of a phase III randomised, placebo-controlled, clinical trial.** *Lancet Neurol* 2006; 5: 38–45.

121. den Hertog HM, van der Worp HB, van Gemert HM, et al. **The Paracetamol (Acetaminophen) In Stroke (PAIS) trial: a multicentre, randomised, placebo-controlled, phase III trial.** *Lancet Neurol* 2009; 8: 434–40.

122. 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.

123. Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD. **Validation and updating of predictive logistic regression models: a study on sample size and shrinkage.** *Stat Med* 2004; 23: 2567–86.

124. Royston P, Altman DG, Sauerbrei W. **Dichotomizing continuous predictors in multiple regression: a bad idea.** *Stat Med* 2006; 25: 127–41.

125. Alexander S, Kerr ME, Kim Y, Kamboh MI, Beers SR, Conley YP. **Apolipoprotein E4 allele presence and functional outcome after severe traumatic brain injury.** *J Neurotrauma* 2007; 24: 790–97.

126. Jordan BD. **Genetic influences on outcome following traumatic brain injury.** *Neurochem Res* 2007; 32: 905–15.

127. Saatman KE, Duhaime AC, Bullock R, et al. **Classification of traumatic brain injury for targeted therapies.** *J Neurotrauma* 2008; 25: 719–38.

128. Maas AI. **Standardisation of data collection in traumatic brain injury: key to the future?** *Crit Care* 2009; 13: 1016.

129. 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.** *Neurosurg* 2012; 70(4): 811-8.

Chapter 4

PREDICTION OF OUTCOME AFTER MODERATE AND SEVERE TRAUMATIC BRAIN INJURY: EXTERNAL VALIDATION OF THE IMPACT AND CRASH PROGNOSTIC MODELS

Bob Roozenbeek
Hester F Lingsma
Fiona E Lecky
Juan Lu
James Weir
Isabella Butcher
Gillian S McHugh
Gordon D Murray
Pablo Perel
Andrew IR Maas
Ewout W Steyerberg

Crit Care Med 2012; 40(5): 1609-17

Abstract

Objective: The International Mission on Prognosis and Analysis of Clinical Trials (IMPACT) and Corticoid Randomisation After Significant Head injury (CRASH) prognostic models predict outcome after traumatic brain injury (TBI) but have not been compared in large datasets. The objective of this is study is to validate externally and compare the IMPACT and CRASH prognostic models for prediction of outcome after moderate or severe TBI.

Design: External validation study.

Patients: We considered 5 new datasets with a total of 9036 patients, comprising three randomized trials and two observational series, containing prospectively collected individual TBI patient data. Measurements: Outcomes were mortality and unfavourable outcome, based on the Glasgow Outcome Score (GOS) at six months after injury. To assess performance, we studied the discrimination of the models (by AUCs), and calibration (by comparison of the mean observed to predicted outcomes and calibration slopes).

Main Results: The highest discrimination was found in the TARN trauma registry (AUCs between 0.83 and 0.87), and the lowest discrimination in the Pharmos trial (AUCs between 0.65 and 0.71). Although differences in predictor effects between development and validation populations were found (calibration slopes varying between 0.58 and 1.53), the differences in discrimination were largely explained by differences in case-mix in the validation studies. Calibration was good, the fraction of observed outcomes generally agreed well with the mean predicted outcome. No meaningful differences were noted in performance between the IMPACT and CRASH models. More complex models discriminated slightly better than simpler variants.

Conclusions: Since both the IMPACT and the CRASH prognostic models show good generalizability to more recent data, they are valid instruments to quantify prognosis in TBI.

Introduction

Traumatic brain injury (TBI) remains the main cause of death and disability in young adults worldwide (1, 2). It is a heterogeneous disease with respect to cause, pathology, severity and prognosis. This causes considerable uncertainty in the expected outcome of individual patients. Prognostic models can be used to combine different characteristics of an individual patient to predict outcome. Reliable outcome predictions can be used to provide realistic information to relatives, for more efficient design and analysis of clinical trials in TBI and can provide a reference for assessing the quality of health care delivery (3).

Many prognostic models for TBI have been developed in the past decades, with varying methodological quality (4, 5). Only two sets of prognostic models were developed on large datasets and according to the latest methodological insights. These are the models developed on the International Mission on Prognosis and Analysis of Clinical trials in Traumatic brain injury database (IMPACT models) and the Corticosteroid Randomisation After Significant Head Injury trial data (CRASH models) (6, 7).

Prognostic models are mathematical models developed from specific populations. External validation is necessary to determine the ability of a model to reliably predict outcome in populations, different from the development setting. Both the IMPACT models and the CRASH models have been externally validated previously, using each other's development datasets, indicating satisfactory generalizability. Nevertheless, a broader validation and comparison of these models has not yet been performed (8).

In this study we aimed to validate externally the IMPACT and CRASH prognostic models in five relatively recent datasets and to compare these different models to predict mortality and unfavourable outcome.

Materials and methods

Models

Details of the development of the prognostic models were reported previously (6, 7). In summary, the IMPACT models were developed on a dataset with 8509 patients from eleven studies (i.e., three surveys and eight randomized controlled trials) on moderate and severe TBI (9). Three different models were developed for the prediction of mortality and unfavourable outcome (Death, Vegetative State, Severe Disability), based on the dichotomized Glasgow Outcome Score (GOS) at six months after injury. These included the Core Model, the Extended Model and the Lab Model (Table 4.1). External validation was performed in patients with moderate to severe TBI enrolled in the CRASH trial dataset (n=6681).

The CRASH models were developed on the CRASH trial dataset (n=10008) with patients who sustained mild, moderate and severe TBI, mostly from low and middle-income countries

(n=7526, 75%) (10, 11). Two different models were developed for both the prediction of mortality at two weeks after injury and unfavourable outcome at six months, using the dichotomized GOS. The models included the Basic Model and the more extensive CT Model (Table 4.1). The IMPACT dataset (n=8509) was used for external validation.

Because the original CRASH models were developed for prediction of outcome in all TBI severity groups (including mild: GCS scores 13-15) and not all variables of the CRASH models were available in the validation datasets (see below), we refitted an adapted version of the CRASH models in a selection of 6681 patients with moderate to severe TBI (GCS < 12), > 14 years of age and with available outcome data. In the refitted models, the total GCS score was replaced by the GCS motor score. Further, the Marshall CT classification and the presence of traumatic subarachnoid hemorrhage were included in the refitted CRASH CT model, instead of the original CT variables (Table 1). New regression coefficients and intercepts were obtained for both the CRASH Basic and the CRASH CT models for prediction of mortality at two weeks and unfavourable outcome at six months (Appendix 4.1 and 4.2).

Table 4.1. Variables included in the IMPACT models and adapted CRASH models

IMPACT Models: predicting unfavourable outcome and mortality at six months		
Core Model	***Extended Model***	***Lab Model***
Age	Core Model variables +	Extended Model variables +
GCS motor score	CT Classification	Glucose
Pupillary reactivity	EDH	Hemoglobin
	tSAH	
	Hypoxia	
	Hypotension	
CRASH Models: predicting unfavourable outcome at six months and mortality at 14 days		
Basic Model	***CT Model***	
Age	Basic Model variables +	
GCS motor score	CT Classification	
Pupillary reactivity	tSAH	
Major extracranial injury		

IMPACT= International Mission on Prognosis and Analysis of Clinical Trials in TBI; GCS= Glasgow Coma Scale; CT= computed tomography; EDH= epidural hematoma; tSAH= traumatic subarachnoid hemorrhage; CRASH= Corticosteroid Randomisation After Significant Head Injury

Datasets for external validation

National Acute Brain Injury Study (NABIS): Hypothermia

The NABIS Hypothermia study (1994-1998) was a multicenter randomized trial performed in the United States investigating the effect of hypothermia in severe TBI (12). The primary outcome was the GOS at six months after injury. The trial was terminated after inclusion of

392 of the planned 500 patients, because an interim analysis showed that the probability of detecting a treatment effect after expansion to 500 included patients was smaller than 1%. For the validation, we excluded patients younger than 14 years of age or with missing outcome (n=3), resulting in a dataset of 385 patients.

Cerestat

Cerestat is an unpublished multicenter randomized controlled trial in a combined group of North American and European centers investigating the effect of Aptiganel HCl, a non-competitive NMDA (N-methyl-D-aspartate) receptor antagonist, in severe TBI (GCS < 8). In 1996 and 1997, 547 patients were enrolled in the trial. An interim analysis performed at the end of 1997 resulted in the decision that continuation of the trial was unjustified (13). No benefit of the study drug was found in TBI patients. As outcome was missing in 30 patients, a total of 517 were available for validation.

Apolipoprotein E (APOE)

This single-center observational cohort study (1996-1999, Glasgow, Scotland) investigated the hypothesis that the possession of the APOE e4 allele is associated with a poor outcome after TBI (14). In total, 1094 patients were enrolled, of whom 513 (54%) had a mild TBI (GCS 13-15). The outcome was the GOS at six months after injury. No overall association between the APOE genotype and outcome was found. To be consistent, we excluded patients with GCS > 12 and age < 14, resulting in a dataset with 404 patients.

Pharmos

The Pharmos Dexanabinol study (2001-2004) was a randomized multicenter, placebo-controlled, phase III trial investigating the safety and efficacy of dexanabinol in patients with severe TBI carried out in Europe, Australia and the United States (15). The study enrolled 861 subjects. The study publication reported dexanabinol to be safe but not efficacious in the treatment of TBI. Because of missing 6-month outcome, five patients were excluded from our analyses, resulting in a dataset of 856 patients.

Trauma Audit and Research Network (TARN)

The Trauma Audit and Registry Network (TARN) is a hospital-based trauma registry in England and Wales including all patients with trauma resulting in immediate admission to hospital for three days or longer or death. The outcome measure is in-hospital mortality. For the validation, we selected patients > 14 years of age enrolled between 2001 and 2009 with moderate or severe TBI (GCS < 12), defined as having an Abbreviated Injury Scale Head of 3 or higher, which was not resulting from scalp laceration, avulsion or penetrating brain injury, with outcome data available, resulting in a dataset of 6874 patients.

Table 4.2. Patient characteristics of the IMPACT Database (used to develop the IMPACT models), the CRASH trial dataset (used to develop the CRASH models) and the NABIS Hypothermia, Cerestat, APOE, Pharmos, and TARN TBI datasets (used for external validation). All datasets were selected for age ≥ 14, GCS ≤ 12 and non-missing outcome data.

Characteristics	Measure or Category Number of patients included in study	IMPACT Database *n* = 8509	CRASH Trial *n* = 6681	NABIS *n* = 385	Cerestat *n* = 517	APOE *n* = 404	Pharmos *n* = 856	TARN TBI *n* = 6874
Origin	Study design	RCT + OBS	RCT	RCT	RCT	OBS	RCT	OBS
Inclusion period		1984-1997	1999-2004	1994-1998	1996-1997	1996-1999	2001-2004	2001-2009
Age, years	Median (25-75 percentile)	30 (21-45)	32 (23-47)	30 (21-40)	28 (20-41)	39 (26-58)	33 (23-46)	39 (24-58)
Motor score of GCS	Total	8509 (100%)	6681 (100%)	385 (100%)	517 (100%)	404 (100%)	856 (100%)	6874 (100%)
	None (1)	1395 (16%)	785 (12%)	82 (21%)	40 (8%)	5 (1%)	43 (5%)	2047 (30%)
	Extension (2)	1042 (12%)	515 (8%)	62 (16%)	88 (17%)	2 (<1%)	91 (11%)	366 (5%)
	Abnormal flexion (3)	1085 (13%)	658 (10%)	55 (14%)	86 (16%)	18 (5%)	136 (16%)	395 (6%)
	Normal flexion (4)	1940 (23%)	1156 (17%)	76 (20%)	180 (35%)	32 (8%)	225 (26%)	668 (10%)
	Localizes/obeys (5/6)	2591 (30%)	3567 (53%)	100 (26%)	123 (24%)	16 (4%)	235 (27%)	2117 (25%)
	Untestable or missing (9)	456 (5%)	0	10 (3%)	0	331 (82%)	126 (15%)	1670 (24%)
Pupillary reactivity	Total	7126 (84%)	6272 (94%)	370 (96%)	426 (82%)	402 (>99%)	822 (96%)	693 (10%)
	Both pupils reactive	4486 (63%)	4956 (74%)	229 (62%)	302 (71%)	299 (74%)	642 (78%)	448 (65%)
	One pupil reactive	886 (12%)	530 (8%)	48 (13%)	72 (17%)	27 (7%)	147 (18%)	63 (9%)
	No pupil reactive	1754 (25%)	786 (12%)	93 (25%)	52 (12%)	76 (19%)	33 (4%)	182 (26%)
Hypoxia	Total	5452 (64%)	0	371 (96%)	478 (93%)	381 (94%)	856 (100%)	4772 (70%)
	Yes or expected	1116 (20%)	NA	124 (33%)	62 (13%)	162 (40%)	212 (25%)	321 (7%)
Hypotension	Total	6420 (75%)	0	371 (96%)	506 (98%)	382 (95%)	855 (>99%)	5700 (83%)
	Yes or expected	1171 (18%)	NA	56 (15%)	85 (17%)	69 (17%)	132 (15%)	419 (6%)

Table 4.2. *(Continued)*

Characteristics	Measure or Category Number of patients included in study	IMPACT Database *n* = 8509	CRASH Trial *n* = 6681	NABIS *n* = 385	Cerestat *n* = 517	APOE *n* = 404	Pharmos *n* = 856	TARN TBI *n* = 6874
CT Classification	Total	5192 (61%)	5654 (85%)	349 (91%)	0	0	849 (99%)	6874 (100%)
	No abnormalities (I)	360 (7%)	954 (17%)	4 (1%)	NA	NA	15 (2%)	482 (7%)
	Diffuse injury (II)	1838 (35%)	1517 (27%)	31 (9%)	NA	NA	412 (49%)	3237 (47%)
	Diffuse injury plus obliteration basal cisterns (III)	863 (17%)	604 (11%)	193 (55%)	NA	NA	203 (24%)	1462 (21%) [a]
	Shift, no mass lesion (IV)	187 (4%)	133 (2%)	5 (2%)	NA	NA	56 (6%)	-
	Any mass lesion (V + VI)	1944 (38%)	2446 (43%)	116 (33%)	NA	NA	163 (19%)	1693 (25%)
Traumatic subarachnoid hemorrhage	Total	7393 (87%)	5653 (85%)	0	0	398 (99%)	856 (100%)	6873 (100%)
	Yes	3313 (45%)	2045 (36%)	NA	NA	141 (35%)	511 (60%)	2468 (36%)
Epidural hematoma	Total	7409 (87%)	0	0	511 (99%)	404 (100%)	856 (100%)	6874 (100%)
	Yes	999 (13%)	NA	NA	60 (12%)	38 (9%)	166 (19%)	1113 (16%)
Glucose (mmol/L)	Total	4830 (57%)	0	217 (56%)	0	0	838 (98%)	0
	Median (25-75 percentile)	8.2 (6.7-10.4)	NA	8.7 (7.4-11.4)	NA	NA	7.4 (6.2-9.1)	NA
Hemoglobin (g/dL)	Total	4376 (51%)	0	236 (61%)	0	0	846 (99%)	0
	Median (25-75 percentile)	12.7 (10.8-14.3)	NA	13.0 (11.6-14.3)	NA	NA	12.7 (11.1- 14.1)	NA
Major extracranial injury	Total	0	6524 (98%)	385 (100%)	517 (100%)	0	856 (100%)	6874 (100%)
	Yes	NA	1735 (27%)	128 (33%)	275 (53%)	NA	440 (51%)	2716 (40%)
Summary measure for heterogeneity	Standard deviation of the LP [b]	1.06	1.09 [c]	0.88	1.04	1.36	0.58	1.75 [c]

[a] CT Classes III and IV were combined and scored as III in TARN TBI.
[b] Standard deviation of the linear predictors for the following logistic regression model: logodds (dead at 6-mo) = $a + b_1 \times age + b_2 \times GCS\ motor + b_3 \times pupillary\ reactivity$
[c] For this dataset 14-d mortality was used instead of 6-mo mortality as outcome for the model
CT, Computed Tomography; GCS, Glasgow Coma Scale; NA, not available.

Data analysis

The IMPACT and CRASH models (table 4.1) were validated in these five validation datasets, all originated from high income countries (Europe, North-America and Australia). Some models could not be applied to all datasets, because not all required predictor or outcome variables were available. Coding of the different predictors in the validation datasets was according to the IMPACT predictor coding as described previously (9). Data on age and GCS motor score, pupillary reactivity, hypoxia and hypotension were available for all validation datasets.

However, in TARN data on pupils were only collected in the latter years of inclusion, resulting in a large number of missing values with only 693 pupils available for 6874 patients, 10%. Missing values occurred for several other predictors, mainly because some predictors were not collected in all studies. Table 4.2 includes information on the total number of patients included in the dataset and on the availability of the parameters of interest in these datasets. In TARN, some CT data were available (EDH and tSAH), but the Marshall CT Classification was not scored directly, and therefore derived with an algorithm using the AIS as a basis (16). Missing values were assumed to be missing at random and were statistically imputed. For the imputation procedure, the *aregImpute* function in *R* software was used (17).

To quantify differences in the distribution of patient characteristics between the different studies, we calculated a summary measure of heterogeneity. This measure is defined as the standard deviation (SD) of the linear predictor (LP) values of a logistic regression model, with mortality as the outcome variable and age, GCS motor score and pupillary reactivity as predictor variables. The LP value is directly related to the predicted probability of mortality.

In a relatively homogeneous population the predicted probabilities - and thus LP values - for all patients will be relatively close together, with a small SD. In a more heterogeneous population, the LPs will be more dispersed and the SD is larger.

Model performance

We assessed the performance of the models in the new datasets in terms of discrimination and calibration. Discrimination describes how well a model distinguishes between patients with and without the outcome of interest. Calibration indicates how closely predicted outcomes match observed outcomes.

To assess discrimination, we calculated the area under the receiver operating characteristic curve (AUC). An AUC of 1 implies perfect discrimination, whereas an AUC of 0.5 implies that a model's discrimination is no better than chance. Discrimination in an external validation setting can be influenced in several ways. The predictor effects in the model can be different, resulting in an invalid model. But differences in the distribution of patient characteristics ("case-mix") between the development and validation data may also influence discriminative power of the model. In a population with a prognostically

homogeneous case-mix, it will be more difficult to distinguish between patients with a good or poor outcome with, than in a heterogeneous population. For example, in an observational study with broad inclusion criteria, discrimination of a model will be better than in a strictly selected trial population.

To take these effects into account, we calculated two recently proposed benchmark values for the AUC: the case-mix-corrected and the refitted AUC (18). The case-mix-corrected AUC indicates the discriminative power of a model, under the assumption that the predictor effects are fully correct for the validation population. It was calculated by simulating new outcome values for all patients in the validation dataset, based on the predicted risks for each patient calculated by the prognostic model. We performed 1000 repetitions to obtain stable estimates. The refitted AUC indicates the discriminative power that can be obtained by refitting the model in the validation dataset, resulting in new, optimal predictor effects. This resulted in an upper bound for the discriminative power, which would be found if the predictor effects in the validation dataset were exactly the same as in the development population.

To assess calibration, we plotted observed versus predicted outcome. We assessed calibration-in-the-large by fitting a logistic regression model with the model predictions as an offset variable. The intercept indicates whether predictions are systematically too low or too high, and should ideally be zero. The calibration slope reflects the average effects of the predictors in the model and was estimated in a logistic regression model with the logit of the model predictions as the only predictor. For a perfect model, the slope is equal to 1.

The current study is part of the IMPACT project that was exempt from institutional review board approval. The analyses were performed using SPSS Statistics (Version 17.0.2, IBM Corporation, Somers, NY, USA) and the *R* software environment (Version 2.7.1, The R Foundation for Statistical Computing, Vienna, Austria) with the *Design* and *Hmisc* packages.

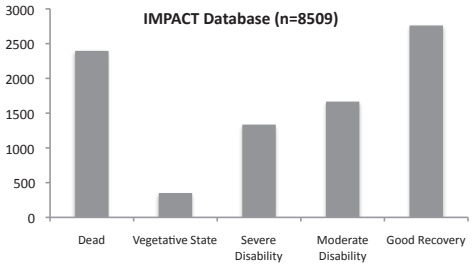
Results

Datasets

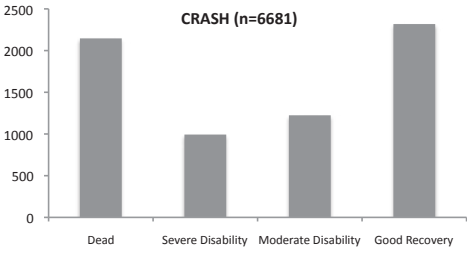
The observational series included a more heterogeneous case-mix than the trial populations (table 4.2). This was illustrated by the summary measure for heterogeneity: the strictly selected trial populations had a lower value (e.g., Pharmos: 0.58) than the observational series (e.g., TARN: 1.75).

The distributions of the Glasgow Outcome Scale score for both the IMPACT and CRASH development datasets were U-shaped (figure 4.1). In NABIS, APOE and Pharmos, the U-shaped outcome distribution had largely disappeared because a higher proportion of patients had severe disability: 26% in both NABIS and APOE and 29% in Pharmos, compared to 16% in IMPACT and 15% in CRASH. This resulted in a greater proportion of patients with unfavourable outcome. In IMPACT and CRASH the proportions of unfavourable outcome were 48% and 47% respectively, while in NABIS this was 56%, in APOE 53% and in Pharmos 51%.

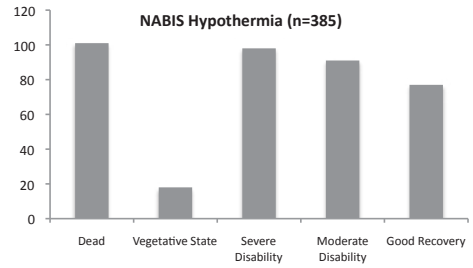
4.1A (1)



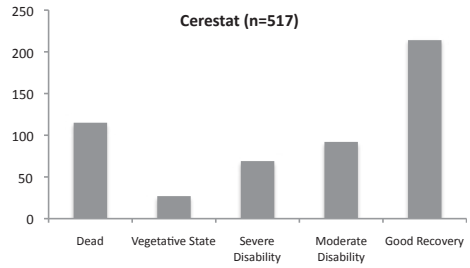
4.1A (2)



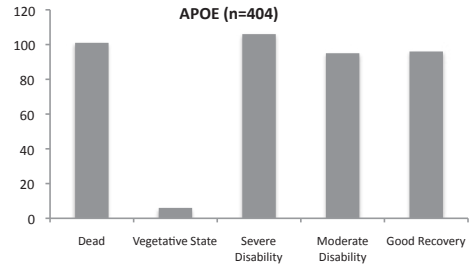
4.1A (3)



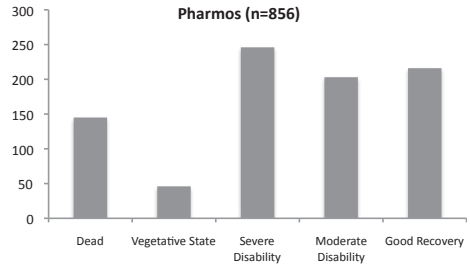
4.1A (4)



4.1A (5)



4.1A (6)



4.1B

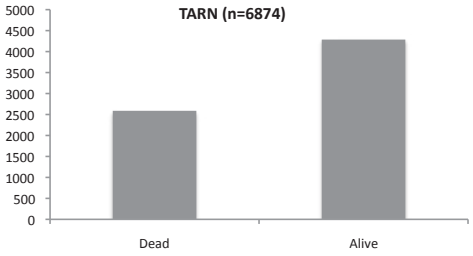


Figure 4.1. A: Six-month Glasgow Outcome Scale in development (IMPACT and CRASH) and validation datasets (NABIS Hypothermia, Cerestat, APOE and Pharmos). **B:** In-hospital mortality in the TARN traumatic brain injury registry.

Model performance

Discrimination

Discrimination of the different models varied between the datasets (table 4.3). For prediction of mortality, AUCs varied between 0.65 and 0.83 for the IMPACT Core model and between 0.66 and 0.85 for the CRASH Basic model. For prediction of unfavourable outcome, AUCs between 0.66 and 0.76 were found for the IMPACT Core model and between 0.68 and 0.76 for the CRASH Basic model. Discrimination improved for both the IMPACT and CRASH models with increasing complexity of the models. Discrimination was better in the observational datasets (TARN and APOE) than in the trials, which was largely explained by a more heterogeneous case-mix, compared to the development population (small difference between development AUC and case-mix-corrected AUC, table 4.3). When the model was refitted in the validation data, the performance was similar to the performance of the externally validated original model (small difference between external validation AUC and refitted AUC).

Calibration

The calibration slope and intercept values of the validation datasets and models are summarized in table 4.3.

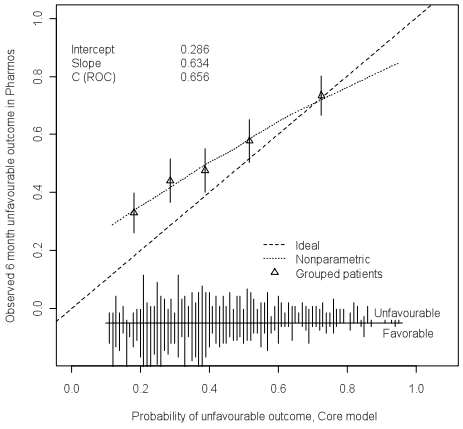
In Pharmos, the calibration plots indicated some degree of miscalibration (figure 4.2A and 4.2B). This was also expressed in the calibration slopes, indicating differences in predictor effects between the development and validation populations. For the IMPACT Core model the slopes were 0.733 for mortality and 0.634 for unfavourable outcome. The CRASH Basic model showed calibration slopes of 0.641 (mortality) and 0.645 (unfavourable outcome). In TARN (figure 4.3), miscalibration was less apparent in the calibration plots. Calibration slopes were 1.318 (IMPACT Core model) and 1.375 (CRASH Basic model).

With regard to calibration-in-the-large, in Pharmos, an intercept of -0.774 was found for mortality prediction with the IMPACT Core model, indicating that the model's predictions were moderately higher than the observed mortality. However, prediction of unfavourable outcome with the IMPACT Core model showed an intercept of 0.286, indicating mild underestimation of the outcome by the model. A similar pattern was observed for the CRASH Basic model. In TARN, the IMPACT Core model gave an intercept of 0.124 and the CRASH Basic model of 0.110 for prediction of mortality.

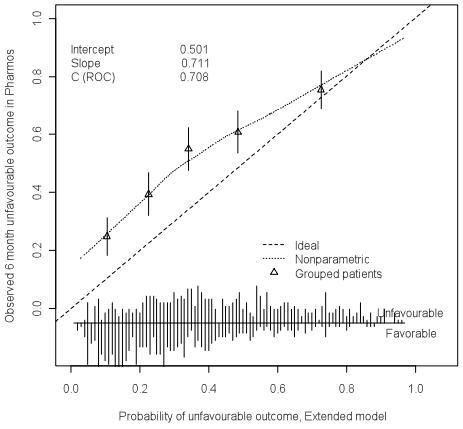
Table 4.3. Performance of the IMPACT Core and CRASH Basic models for prediction of mortality and unfavourable outcome after traumatic brain injury in five new datasets.

		Mortality		Unfavourable outcome	
Studies		IMPACT Core	CRASH Basic	IMPACT Core	CRASH Basic
Development	AUC	0.766	0.776	0.779	0.784
NABIS (n=385)	AUC (external validation)	0.702	0.693	0.728	0.736
	Slope	0.838	0.670	0.890	0.804
	Intercept	-0.403	-0.652	0.424	0.040
	AUC (case-mix-corrected)	0.740	0.771	0.754	0.781
	AUC (refitted)	0.724	0.709	0.751	0.749
Cerestat (n=517)	AUC (external validation)	0.746	0.736	0.712	0.755
	Slope	1.047	0.889	0.815	0.947
	Intercept	-0.222	-0.651	-0.357	-0.725
	AUC (case-mix-corrected)	0.723	0.738	0.751	0.758
	AUC (refitted)	0.776	0.768	0.773	0.769
APOE (n=404)	AUC (external validation)	0.806	NA	0.758	NA
	Slope	1.534	NA	1.154	NA
	Intercept	-0.006	NA	0.144	NA
	AUC (case-mix-corrected)	0.728	NA	0.736	NA
	AUC (refitted)	0.810	NA	0.769	NA
Pharmos (n=856)	AUC (external validation)	0.650	0.655	0.656	0.677
	Slope	0.733	0.641	0.634	0.645
	Intercept	-0.774	-1.005	0.286	-0.014
	AUC (case-mix-corrected)	0.702	0.721	0.728	0.741
	AUC (refitted)	0.658	0.664	0.663	0.682
TARN TBI (n=6874)	AUC (external validation)	0.832	0.848	NA	NA
	Slope	1.318	1.375	NA	NA
	Intercept	0.124	0.110	NA	NA
	AUC (case-mix-corrected)	0.781	0.784	NA	NA
	AUC (refitted)	0.842	0.857	NA	NA

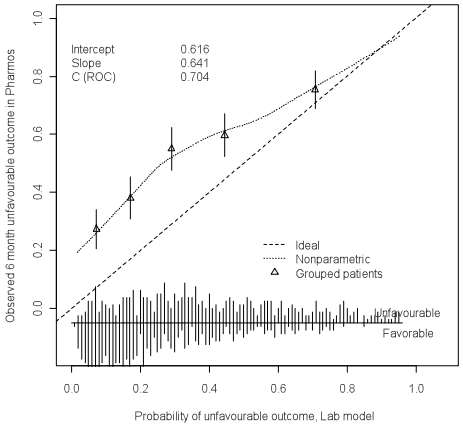
4.2A (1) **IMPACT Core Model in Pharmos (n=856)**



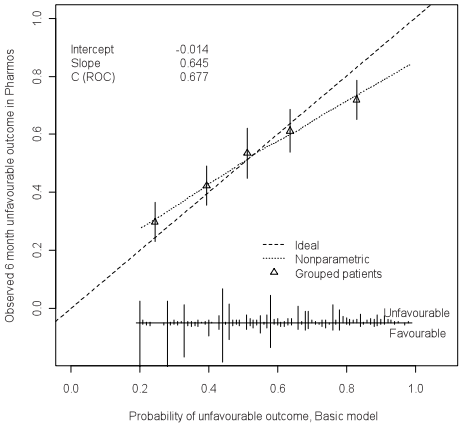
4.2A (2) **IMPACT Extended Model in Pharmos (n=856)**



4.2A (3) **IMPACT Lab Model in Pharmos (n=856)**



4.2A (4) **CRASH Basic Model in Pharmos (n=856)**



4.2A (5) **CRASH CT Model in Pharmos (n=856)**

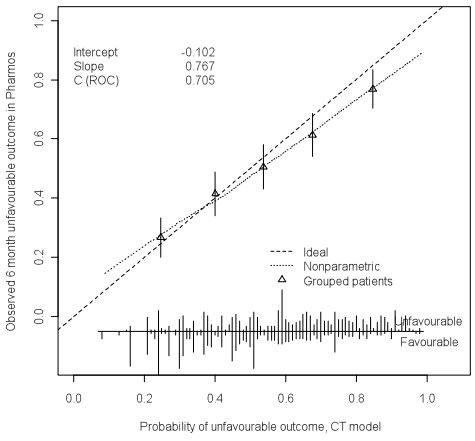
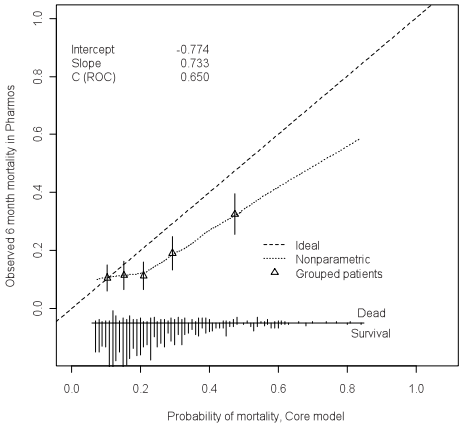
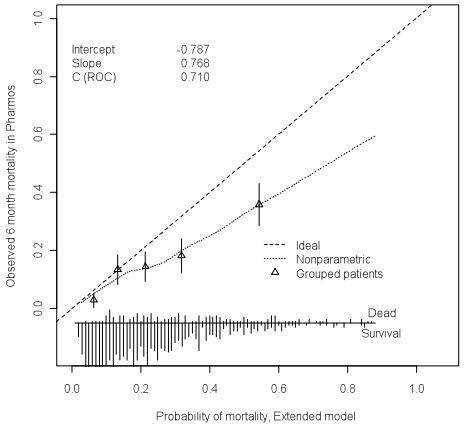


Figure 4.2. A: Calibration plots for external validation of the IMPACT and CRASH models for prediction of 6-month unfavorable outcome in the Pharmos trial data. Predicted probabilities are on the x-axis and observed outcomes on the y-axis. The triangles indicate the observed frequencies by quantiles of predicted probability with 95% confidence intervals (vertical lines). The distribution of the predicted probabilities is shown at the bottom of the graphs, separate for those with and without the outcome of interest.

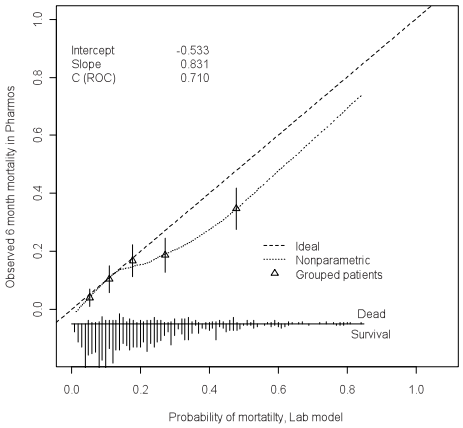
4.2B (1) **IMPACT Core Model in Pharmos (n=856)**



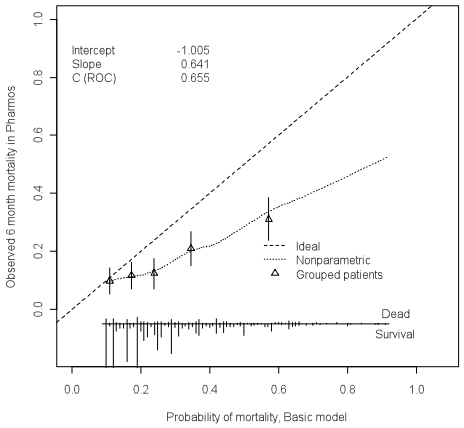
4.2B (2) **IMPACT Extended Model in Pharmos (n=856)**



4.2B (3) **IMPACT Lab Model in Pharmos (n=856)**



4.2B (4) **CRASH Basic Model in Pharmos (n=856)**



4.2B (5) **CRASH CT Model in Pharmos (n=856)**

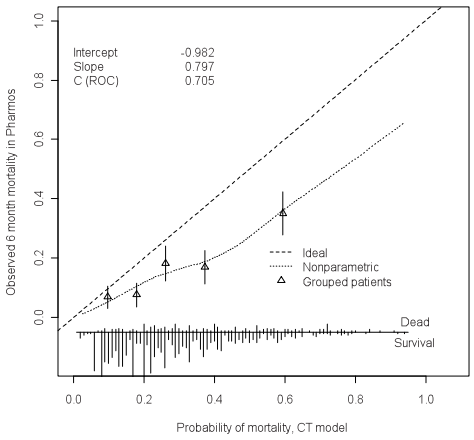


Figure 4.2. B: Calibration plots for external validation of the IMPACT and CRASH models for prediction of mortality in the Pharmos trial data. C= area under the receiver operating characteristic (ROC) curve. CT= computed tomography.

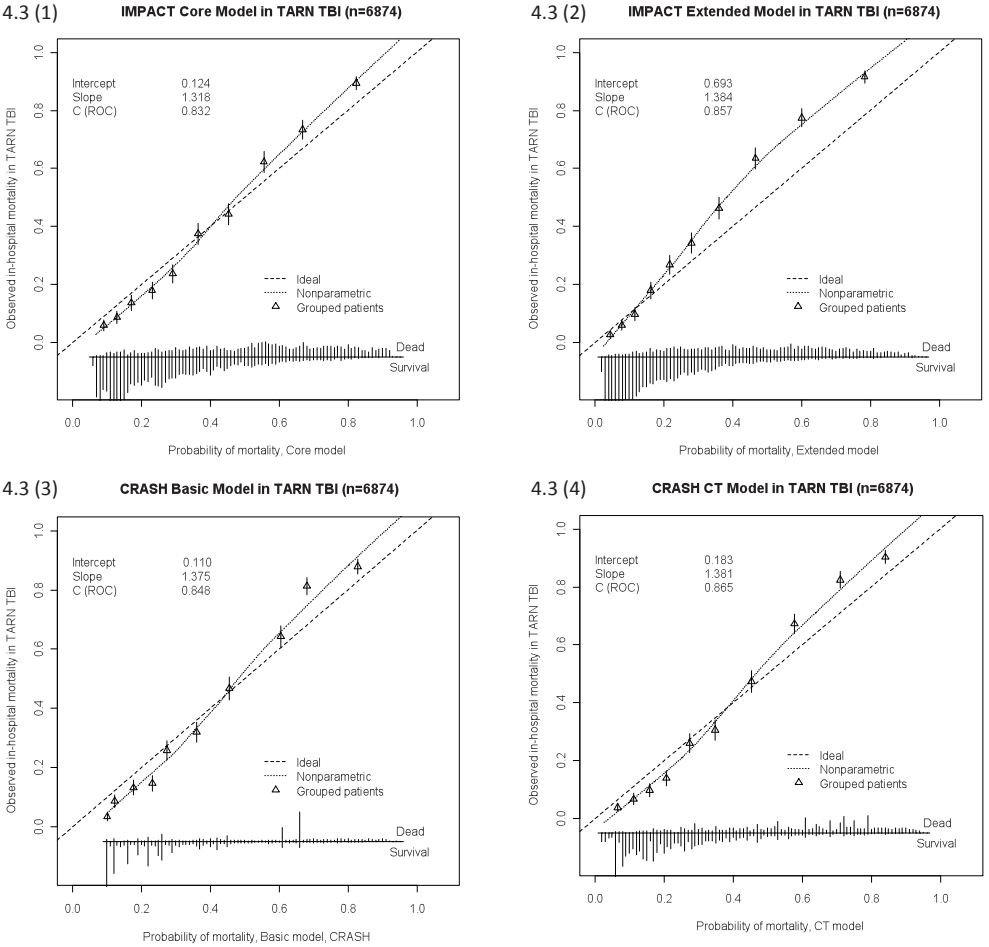


Figure 4.3. Calibration plots for external validation of the IMPACT and CRASH models for prediction of mortality in the TARN traumatic brain injury dataset.
C= area under the receiver operating characteristic (ROC) curve. CT= computed tomography.

Sensitivity analysis

Because of the high proportion of patients with missing data on pupillary reactivity in the TARN data, we also performed the validation in TARN patients with available data on pupillary reactivity (n=693). Compared to the validation with the imputed dataset, the results were very similar for both the IMPACT Models (Core: AUC=0.836, Slope=1.398, Intercept=1081; Extended: AUC=0.864, Slope=1.458, Intercept=0.621) and the CRASH Models (Basic: AUC=0.850, Slope=1.333, Intercept=-0.006; CT: AUC=0.863, Slope=1.330, Intercept=-0.031).

Discussion

We confirmed the external validity of the IMPACT and CRASH prognostic models for prediction of outcome after moderate and severe TBI in five new datasets. Importantly, these datasets included both observational studies and clinical trials, thus permitting conclusions on generalizability. The predictive performance of the IMPACT and CRASH models was largely similar. Discrimination of the models varied between the different validation populations (clinical trials vs. observational studies). Overall, the more complex models performed better than the more simple models.

On average, performance measures did not substantially differ from those obtained in the development populations. Although developed on 6-month mortality or unfavourable outcome, the models also demonstrated good performance when predicting in-hospital mortality in the TARN registry.

These conclusions support the validity of these instruments to quantify prognostic risk in moderate and severe TBI.

Performance and generalizability of models

The CRASH and IMPACT prognostic models have been developed using state of the art approaches on large datasets, and were externally validated reciprocally. They therefore meet methodological quality standards (17). The IMPACT models have however been criticized for having in part been developed on older datasets, and the CRASH models for having been developed exclusively on data from a clinical trial – albeit one with broad enrollment criteria. Establishing their validity for current practice and their generalizability to other settings is therefore important. Panczykowski et al reported good performance of the IMPACT models (AUC: 0.76-0.83) at validation on an unselected series of 587 patients with severe TBI admitted to a single level I trauma center (19). Honeybul et al applied the CRASH model to a selection 147 of 1786 TBI patients who underwent a decompressive craniectomy, to predict outcome at 18 months after injury (20). The authors noted an overestimation of the predicted risk of unfavourable outcome, predominantly when the predicted risk was < 80%. To our knowledge, no other validation studies on the IMPACT and CRASH models have been performed.

Assessing performance – and its interpretation – is a complex procedure, involving multiple levels. Many factors other than those directly related to the model itself may influence the performance of a model. These include changing epidemiology, trauma organization, different treatment policies, but also changes in approaches to outcome assessment. Such external factors could influence the validity of the predictor effects (regression coefficients) and the distribution of the predictor values in the new population (case-mix).

We observed clear effects of the case-mix on discrimination: higher AUCs were found in the observational series (TARN and APOE) compared to the RCTs. These higher AUCs reflect the less restrictive enrollment criteria: the greater the heterogeneity, the better the models can distinguish patients with very low or very high predictions. We therefore strongly discourage the indiscriminate comparisons of reported AUCs between different studies without taking the case-mix of the development and validation population into consideration. Use of the case-mix corrected AUC as proposed by Vergouwe et al would appear more appropriate, as illustrated by the decrease in AUC observed in the observational series versus the increase observed in the RCTs when calculating the case-mix-corrected AUC (table 4.3) (18).

With regard to calibration, we did observe some indication for differences in predictor effects between the development and validation datasets. Particularly noteworthy is the discrepancy in the calibration slopes for the Pharmos and NABIS trials: the observed mortality is lower than predicted, whilst the number of patients with an unfavorable outcome is substantially higher than predicted. This discrepancy could be explained by a greater number of patients with Severe Disability than expected. Although some difference in predictor effects may be present, we consider the different outcome distribution (figure 4.1) a more likely explanation. The main question then becomes what may have caused this change. It may be that approaches to outcome assessment, and in particular to the assignment of outcome categories, have changed. Pharmos and NABIS used the structured interview as a means to better standardize outcome assignment and in the Pharmos trial a central review of outcome assignments was performed (21). Wilson et al reported that central assignments were generally worse, causing a shift towards more severe disability (22). Our findings indicate that changes in the approach to assessment of the GOS, may influence both outcome distribution and predictor effects. Clarification of this issue is considered essential and requires a dedicated prospective study. The need for such a study is becoming more pertinent following the recent proposals by Lu et al for yet another new approach to assessing the extended GOS (23).

IMPACT or CRASH models?

We have often been asked the question which model should be preferred: CRASH or IMPACT? The current validation studies do not indicate any clear preference. Both models demonstrated comparable performance on external validation. A difference between the IMPACT and CRASH models is the inclusion of major extracranial injury (MEI) as a predictor in the CRASH models. We, however, found that the added prognostic effect of MEI in our validation series was negligible (results not shown), which may have resulted from the case-mix of the validation datasets. In a recent meta-analysis of IMPACT, CRASH and TARN data, we found an interaction between MEI and the clinical severity: only small effects of MEI were noted in patients with severe TBI, whilst in mild TBI, the presence of extracranial injury

had a strong prognostic effect (24). Perhaps the most appropriate answer to the question which model should be preferred is to consider the setting and case-mix of the population under study. The IMPACT models were developed on patients with moderate and severe TBI, mostly from high-income countries. The development population for the CRASH models included also patients with mild TBI and a substantial number of patients from low and middle-income countries. Consequently, a preference may be expressed for the CRASH models when the population under study includes patients with mild TBI or patients from low and middle-income countries.

Limitations

Our study has some limitations. First, due to differences in recorded predictor and outcome variables, it was not possible to validate the original CRASH models. As described above, we refitted the CRASH models in a selection of the CRASH trial data (GCS < 12) with a different definition of the CT variables and the GCS motor score instead of the full GCS score. This resulted in an adapted version of the CRASH models, making them more similar to the IMPACT models. Nevertheless, some aspects of the adapted CRASH models were still substantially different from the IMPACT models. These differences included the inclusion of the presence of major extracranial injury as a predictor in the CRASH models and the inclusion of second insults (hypoxia and hypotension) as predictors to the IMPACT Extended and Lab models. Further, the adapted CRASH models were developed to predict early (14-day) mortality and the IMPACT models late (6-month) mortality.

It was not possible to validate all models to all new datasets, since not all required predictor and outcome variables were recorded for each dataset. In TARN, the 6-month GOS was not recorded as an outcome measure and only the models predicting mortality were validated on these data. Although the outcome measure was in-hospital mortality, instead of 14-day or 6-month mortality, both IMPACT and CRASH models performed well.

In TARN, actual CT data were not available, but derived from extensive coding of the Abbreviated Injury Score (AIS) according to an algorithm proposed by Lesko et al. (16). The AIS codes were, however, not only based on data of imaging studies on admission, but also upon information of later imaging studies (CT and MRI), as well as findings during surgery. Servadei et al have previously demonstrated that the “worst” CT may demonstrate more abnormalities than the admission CT (25).

A further limitation was the high proportion of missing values on pupillary reactivity in the TARN dataset (90%). Since it would be a waste of information to exclude all patients with missing pupils (6181 patients with a total of 49496 data points), we decided to include these patients for the validation and to statistically impute the missing values with an imputation model. Both theoretical and empirical support is growing for the use of such imputation methods, instead of traditional complete case analyses (17). As a sensitivity analysis, we also

performed the validation in a complete case dataset of TARN, which showed very similar results, supporting the validity of the approaches taken. The lack of availability of some important predictors in some datasets and missing data in others illustrates the paramount importance of standardization of data collection and coding of variables as proposed in the common data elements initiative (26).

Conclusions

This external validation study confirms the validity of both the IMPACT and CRASH models in more recent datasets, with best performance in more heterogeneous observational series. Model performance improved slightly with increasing complexity of the models, but most information is captured in three key predictors: age, GCS motor score, and pupillary reactivity. No relevant difference in performance was found between the IMPACT and the CRASH models. These findings confirm generalizability of the models in new data and support the use of the models for a wide a range of applications. These include the classification of future TBI patient populations (especially exploiting the discriminative ability of the models) and calculation of the baseline prognostic risk of individual patients (requiring good calibration of the models).

It may be expected that the effects of treatment and health care organization on outcome of individual patients change over time. This underlines the necessity of re-validation of these prognostic models in the future, to re-confirm generalizability or to update the models on more recent patient populations. This should become a continuing process, and highlights the need for a prospective high quality observational study.

Acknowledgements

This work was supported by the U.S. National Institutes of Health (Clinical Trial Design and Analysis in TBI; Project: R01 NS-042691). The authors gratefully acknowledge the permission of the prinicipal investigators of the NABIS-H1 trial (Dr. Guy Clifton), of the Cerestat trial (Dr. Elkan Gamzu), of the APOE study (Sir Graham Teasdale) for granting access to the study data for purposes of analysis by the IMPACT Study Group. The authors also express their gratitude to the CRASH Trial Collaborators and the TARN Executive for making their data available.

Appendices

Appendix 4.1. Logistic regression coefficients and intercepts of the validated IMPACT models.

	6-month Mortality			6-month Unfavourable Outcome		
	Core	***Extended***	***Lab***	***Core***	***Extended***	***Lab***
Intercept	-3.109	-3.787	-3.184	-2.644	-3.023	-2.470
Age	0.0342	0.0321	0.0204	0.0376	0.0332	0.0288
GCS motor=1	1.447	1.205	0.965	1.393	1.218	1.105
GCS motor=2	1.397	1.207	1.109	2.078	1.852	1.801
GCS motor=3	0.797	0.746	0.778	1.266	1.185	1.186
GCS motor=4	0.390	0.313	0.310	0.632	0.575	0.548
GCS motor=5/6	Ref	Ref	Ref	Ref	Ref	Ref
GCS motor=NA	0.522	0.425	0.462	0.885	0.837	0.059
Pupils=1	Ref	Ref	Ref	Ref	Ref	Ref
Pupils=2	0.514	0.334	0.090	0.592	0.442	0.201
Pupils=3	1.239	0.970	0.533	1.216	1.003	0.712
CT Class=1	-	Ref	Ref	-	Ref	Ref
CT Class=2	-	-0.298	-0.150	-	-0.530	-0.567
CT Class=3/4	-	0.774	0.715	-	0.543	0.508
CT Class=5/6	-	0.651	0.807	-	0.497	0.571
tSAH	-	0.606	0.740	-	0.567	0.666
EDH	-	0.379	-0.510	-	-0.572	-0.687
Hypoxia	-	0.237	0.360	-	0.316	0.396
Hypotension	-	0.667	0.366	-	0.614	0.440
Glucose (mmol/L)	-	-	-0.086	-	-	-0.092
Hemoglobin (g/dL)	-	-	0.097	-	-	0.085

Appendix 4.2. Logistic regression coefficients and intercepts of the validated CRASH models. These values resulted from refitting the models in a selection of the original CRASH trial data (GCS ≤ 12, age ≥ 14, non-missing 6-mo outcome, n=6681). Missing values for the variables pupils, MEI, CT class and tSAH were imputed.

	14-day Mortality		6-month Unfavourable Outcome	
	Basic	***CT***	***Basic***	***CT***
Intercept	-2.19	-3.595	-1.356	-2.33
Age*	0.0399	0.0392	0.0614	0.0604
GCS motor=1	0.940	0.937	1.103	1.123
GCS motor=2	1.678	1.472	2.051	1.872
GCS motor=3	1.112	1.028	1.230	1.180
GCS motor=4	0.560	0.500	0.681	0.645
GCS motor=5/6	Ref	Ref	Ref	Ref
Pupils=1	Ref	Ref	Ref	Ref
Pupils=2	0.872	0.695	0.971	0.794
Pupils=3	1.737	1.576	1.717	1.550
MEI	0.210	0.209	0.459	0.467
CT Class=1	-	Ref	-	Ref
CT Class=2	-	0.996	-	0.724
CT Class=3/4	-	1.849	-	1.221
CT Class=5/6	-	1.661	-	1.264
tSAH	-	0.362	-	0.324

GCS=Glasgow Coma Scale; MEI=major extracranial injury; CT Class=Marshall CT Classification; tSAH=traumatic subarachnoid hemorrhage

pupils=1: both pupils reactive; pupils=2: one pupil reactive; pupils=3: no pupil reactive

* The effect of age was modelled in the same way as in the original CRASH models: no association < 40 years, a linear association ≥ 40 years

References

1. Maas AI, Stocchetti N, Bullock R. **Moderate and severe traumatic brain injury in adults.** *Lancet Neurol* 2008;7:728-41
2. Ghajar J. **Traumatic brain injury**. *Lancet* 2000;356:923-9
3. Lingsma HF, Roozenbeek B, Steyerberg EW, et al: **Early prognosis in traumatic brain injury: From prophecies to predictions.** *Lancet Neurol* 2010;9:543-554
4. 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
5. 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;61:331-343
6. 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; discussion e165
7. 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
8. Menon DK, Zahed C: **Prediction of outcome in severe traumatic brain injury.** *Curr Opin Crit Care* 2009;15:437-441
9. Marmarou A, Lu J, Butcher I, et al: **IMPACT database of traumatic brain injury: Design and description.** *J Neurotrauma* 2007;24:239-250
10. 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;364:1321-1328
11. Edwards P, Arango M, Balica L, et al: **Final results of MRC CRASH, a randomised placebo-controlled trial of intravenous corticosteroid in adults with head injury-outcomes at 6 months.** *Lancet* 2005;365:1957-1959
12. Clifton GL, Miller ER, Choi SC, et al: **Lack of effect of induction of hypothermia after acute brain injury.** *N Engl J Med* 2001;344:556-563
13. Miller LP, Hayes R, Newcomb J (Eds): **Head Trauma: Basic, Pre-clinical and Clinical Directions.** Hoboken, NJ, Wiley-Liss Inc., 2001
14. Teasdale GM, Murray GD, Nicoll JA: **The association between APOE epsilon4, age and outcome after head injury: A prospective cohort study.** *Brain* 2005;128:2556-2561
15. Maas AI, Murray G, Henney H,3rd, et al: **Efficacy and safety of dexanabinol in severe traumatic brain injury: Results of a phase III randomised, placebo-controlled, clinical trial.** *Lancet Neurol* 2006;5:38-45
16. Lesko MM, Woodford M, White L, et al: **Using abbreviated injury scale (AIS) codes to classify computed tomography (CT) features in the marshall system.** *BMC Med Res Methodol* 2010;10:72
17. Steyerberg EW: **Clinical prediction models**. Springer Science+Business Media, 2009
18. Vergouwe Y, Moons KG, Steyerberg EW**: External validity of risk models: Use of benchmark values to disentangle a case-mix effect from incorrect coefficients.** *Am J Epidemiol* 2010;172:971-980
19. Panczykowski D, Puccio A, Scruggs BJ, et al: **Prospective Independent Validation of IMPACT Modeling as a Prognostic Tool in Severe Traumatic Brain Injury.** J *Neurotrauma* 2012;29(1):47-52
20. Honeybul S, Ho KM, Lind CR, et al: **Observed versus predicted outcome for decompressive craniectomy: A population-based study.** *J Neurotrauma* 2010;27:1225-1232
21. Wilson JT, Pettigrew LE, Teasdale GM: **Structured interviews for the glasgow outcome scale and the extended glasgow outcome scale: Guidelines for their use.** *J Neurotrauma* 1998;15:573-585
22. Wilson JT, Slieker FJ, Legrand V, et al: **Observer variation in the assessment of outcome in traumatic brain injury: Experience from a multicenter, international randomized clinical trial.** *Neurosurgery* 2007;61:123-8; discussion 128-9
23. Lu J, Marmarou A, Lapane K, et al: **A method for reducing misclassification in the extended glasgow outcome score.** *J Neurotrauma* 2010;27:843-852
24. 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): 811-8
25. Servadei F, Murray GD, Penny K, et al: **The value of the "worst" computed tomographic scan in clinical studies of moderate and severe head injury. European Brain Injury Consortium.** *Neurosurgery* 2000;46:70-5; discussion 75-7
26. Maas AI, Harrison-Felix CL, Menon D, et al: **Standardizing data collection in traumatic brain injury.** *J Neurotrauma* 2011;28:177-187

Chapter 5

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

Bob Roozenbeek
Ya-Lin Chiu
Hester F Lingsma
Linda M Gerber
Ewout W Steyerberg
Jamshid Ghajar
Andrew IR Maas

J Neurotrauma 2012; 29(7): 1306-12

Abstract

Prognostic models for outcome prediction in patients with traumatic brain injury (TBI) are important instruments in both clinical practice and research. To remain current a continuous process of model validation is necessary. We aimed to investigate the performance of the International Mission on Prognosis and Analysis of Clinical Trials in TBI (IMPACT) prognostic models in predicting mortality in a contemporary New York State TBI registry developed and maintained by the Brain Trauma Foundation. The Brain Trauma Foundation (BTF) TBI-trac® database contains data on 3125 patients who sustained severe TBI (GCS ≤8) in New York State between 2000 and 2009. Outcome measure was 14-day mortality. To predict 14-day mortality with admission data, we adapted the IMPACT Core and Extended models.

Performance of the models was assessed by determining calibration (agreement between observed and predicted outcomes) and discrimination (separation of those patients who die from those who survive). Calibration was explored graphically with calibration plots. Discrimination was expressed by the area under the Receiver Operating Characteristic (ROC) curve (AUC). A total of 2513 out of 3125 patients in the BTF database met the inclusion criteria. The 14-day mortality rate was 23%. The models showed excellent calibration. Mean predicted probabilities were 20% for the Core model and 24% for the Extended model.

Both models showed good discrimination with AUCs of 0.79 (Core) and 0.83 (Extended). We conclude that the IMPACT models validly predict 14-day mortality in the BTF database, confirming generalizability of these models for outcome prediction in TBI patients.

Introduction

Traumatic brain injury (TBI) is a heterogeneous disease in terms of injury mechanism, pathology, severity and prognosis (1, 2). Outcome prediction is relevant for both clinical practice and research (3). In clinical practice, evidence-based prediction of individual patients' outcome is important for realistic counseling of patients and relatives. In research, predictions may be used to compare actual outcome to predicted outcome for benchmarking quality of care, for classification, to identify best practices in comparative effectiveness research or as instruments to adjust for heterogeneity in prognostic risk in clinical trials.

In 2000, the Brain Trauma Foundation (BTF) initiated a quality improvement program in New York State to track the treatment of severe TBI patients (Glasgow Coma Scale [GCS] score < 9) which expanded to involve 22 of the 46 state designated level 1, 2 and 3 trauma centers. The program was funded through the New York State Department of Health Bureau of Emergency Medical Services and has a database (TBI-trac®) containing prospective data on currently more than 3000 patients. The project was designed to assess and implement adoption of the evidence-based Guidelines for the Management of Severe TBI (4-6), but the data also offer opportunities for further quality of care research, outcomes research and comparative effectiveness research. Each of these requires valid predictions of outcome for individual patients.

Outcome predictions require a good prognostic model. Methodological requirements to develop a valid prognostic model include sufficiently large patient samples, and internal and external validation (7). In TBI, many prognostic models have been published, but only few fulfill these requirements (8, 9), resulting in limited generalizability beyond the development sample.

The aim of this study was to establish the validity of an existing model to predict 14-day mortality in the TBI-trac® database. The model we used was the International Mission on Prognosis and Analysis of Clinical Trials in TBI (IMPACT) model, which was developed to predict 6-month mortality and unfavorable outcome, and already showed good performance in an external validation set (10). For the present study, the IMPACT model was modified to predict 14-day mortality and applied to the TBI-trac® database.

Materials and methods

Study population

The BTF TBI-trac® database contains prospectively collected data of patients who sustained a severe TBI (GCS < 8). Patients were included between 2000 and 2009 in 22 trauma centers (twenty level 1 and two level 2 designations) enrolled in a New York State quality improvement program. The database includes information on mechanism of injury, demographics (age,

sex, race), injury severity (GCS, pupillary reactivity) and brain CT characteristics. Data were also collected from the patient›s stay in the intensive care unit on physiologic variables, the presence of secondary insults (hypotension and hypoxia and cerebral hypoperfusion) and any therapies to reduce elevated intracranial pressure (ICP). Outcome is assessed at two weeks after injury. For our analyses, we used data recorded after resuscitation on the first day of admission to the hospital. Patients were considered hypoxic when the arterial oxygen pressure on day 1 was less than 60 mmHg or when the lowest oxygen saturation before admission to the hospital was less than 90%. If the patient was sedated during GCS assessment, the variable GCS was recoded into 'untestable'. We excluded patients who died in the emergency department, who had a GCS motor score of 1 or 2 with bilateral fixed and dilated pupils on admission, with GCS > 9 on day 1, GCS motor score of 6 on day 1, age < 14 years, penetrating TBI and missing 14-day outcome.

Models

The IMPACT models were developed in the IMPACT database, which included prospectively collected data of moderate and severe TBI patients from eight randomized controlled trials and three observational series (total n=8509). Details on the development and validation of the IMPACT prognostic models have been described previously (10). The IMPACT Core Model includes the predictors age, GCS motor score and pupillary reactivity. The Extended Model added variables on secondary insults (hypoxia and hypotension) and CT scan characteristics (Marshall CT classification, traumatic subarachnoid hemorrhage and epidural hematoma).

Both models were developed for prediction of mortality and unfavorable outcome at 6 months, according to the dichotomized Glasgow Outcome Scale (Dead, Vegetative State and Severe Disability).

The IMPACT predictors present in the TBI-trac® data were age, GCS motor score, pupillary reactivity, hypoxia, hypotension and traumatic subarachnoid hemorrhage. The Marshall CT classification and epidural hematoma were not recorded in the TBI-trac® data and were replaced by the status of basal cisterns and the presence of midline shift as CT parameters. We therefore refitted the IMPACT models in the IMPACT database to obtain new coefficients and intercepts for prediction of 14-day mortality:

Core model: logodds(14-day mortality)= $\beta 0+\beta 1*age+\beta 2*GCS\ motor\ score + \beta 3*pupillary\ reactivity$

Extended model: logodds(14-day mortality)= $\beta 0+\beta 1*age + \beta 2*GCS\ motor\ score + \beta 3*pupillary\ reactivity + \beta 4*hypoxia + \beta 5*hypotension + \beta 6*midline\ shift + \beta 7*cisterns + \beta 8*subarachnoid\ hemorrhage$

Calibration and discrimination

The external validity of the models was assessed by studying calibration and discrimination. Calibration refers to the agreement between observed and predicted outcomes. The extent of over- or underestimation relative to the observed and predicted rate was explored graphically using validation plots. We assessed calibration-in-the-large by fitting a logistic regression model with the model predictions as an offset variable. The intercept indicates whether predictions are systematically too low or too high, and should ideally be zero.

The calibration slope reflects the average effects of the predictors in the model and was estimated in a logistic regression model with the logit of the model predictions as the only predictor. For a perfect model, the slope is equal to 1.

The Area Under the receiver operating characteristic Curve (AUC) and 95% confidence interval (CI) was used to assess the ability of the model to discriminate between death and survival. The 95% CI of AUC was calculated by a bootstrap resampling method.

Statistical analysis

Missing values in the BTF dataset were statistically imputed using single imputation with the *AregImpute* function in *R* statistical software (R Foundation, Vienna, Austria). In the dataset with eight independent variables and one outcome variable per patient, 832 of the 20,104 data points (4%) were missing and imputed. The highest percentage of missings was in the variables hypoxia and status of basal cisterns. All the analyses were done in both IMPACT and BTF TBI-trac® databases. Patients' baseline characteristics were described as standard summary statistics: median (range) for continuous variables and frequency (percentage) for categorical variables. Univariate and multivariable logistic regressions were performed in both the imputed data and the complete cases and associations were expressed as odds ratios and 95% CIs. Since the means and standard deviations as well as the prognostic effects and the results of the validation were very similar, we report only odds ratios and 95% CI from the imputed data. The calibration plots were created with an adapted version of the *val.prob* function from the *Design* library in the *R* package.

Results

Study population

The BTF TBI-trac® dataset contained data on 3125 severe TBI patients. Patients were excluded if they had a GCS score greater than or equal to nine on day one (152 patients), or a GCS motor score of six on day one (23 patients), since they did not meet the definition of severe TBI. Patients younger than 14 years of age (215 patients), with penetrating or gunshot TBI (132 patients) were excluded, since these were exclusion criteria of the IMPACT

study. Finally, patients who had a GCS motor score of 1 or 2 and bilateral fixed and dilated pupils on admission or died in the emergency department (43 patients), without daily record (7 patients) or 14-day outcome assessment (40 patients) were excluded. After exclusion criteria were applied, a total of 2513 patients were eligible for the validation.

In both the IMPACT database as the TBI-trac® database (over a 10 year period), 23% of the patients had died 14 days after injury. The patients in the TBI-trac® dataset were older (median 35 vs. 30 years), more often had an untestable GCS motor score (15% vs. 5%) and less often had two unreactive pupils (9% vs. 25%). Only 6.7% of the patients in the TBI-trac® data had hypoxia, while this was 20% in the IMPACT data. The percentage of patients with hypotension and the various CT characteristics were similar (table 5.1).

Table 5.1. Baseline characteristics of patients in the IMPACT database and in the BTF TBI-trac® database

Characteristics	Measure or Category	IMPACT Database	BTF Database
		n=8428	***n=2513***
Age, years	Median (25-75 percentile)	30 (21-45)	35 (22-52)
Motor score of GCS	Total	8428 (100%)	2513 (100%)
	None (1)	1381 (16%)	709 (28%)
	Extension (2)	1028 (12%)	147 (6%)
	Abnormal flexion (3)	1078 (13%)	191 (8%)
	Normal flexion (4)	1926 (23%)	537 (21%)
	Localizes/obeys (5/6)	2567 (30%)	540 (21%)
	Untestable / missing (9)	448 (5%)	389 (15%)
Pupillary reactivity	Total	7051 (84%)	2492 (99%)
	Both pupils reactive	4442 (63%)	1789 (72%)
	One pupil reactive	876 (12%)	479 (19%)
	No pupil reactive	1733 (25%)	224 (9%)
		n=6918	***n=2513***
Hypoxia	Total	5375 (78%)	2255 (90%)
	Yes or suspected	1087 (20%)	151 (7%)
Hypotension	Total	6342 (92%)	2513 (100%)
	Yes or suspected	1152 (19%)	505 (20%)
Basal cisterns	Total	3833 (55%)	2266 (90%)
	Partially compressed or absent	1653 (53%)	985 (43%)
Midline shift	Total	4676 (68%)	2365 (94%)
	Yes	1866 (40%)	797 (34%)
Traumatic subarachnoid hemorrhage	Total	5813 (84%)	2355 (94%)
	Yes	2652 (46%)	1096 (47%)

Table 5.2. Associations between predictors and 14-day mortality in the IMPACT database (n=8428) and the BTF TBI-trac® database (n=2513).

		IMPACT			BTF TBI trac		
Characteristics	**Measure or Category**	**Univariate**	**Core Model** n=8428	**Extended Model** n=6918	**Univariate**	**Core Model** n=2513	**Extended Model** n=2513
Age, years	OR 25th-75th percentile	1.8(1.6-1.9)	1.9 (1.8-2.1)	1.7 (1.5-1.9)	1.8 (1.6-2.1)	2.1 (1.8-2.4)	1.9 (1.6-2.3)
Motor score of GCS	None (1)	5.2 (4.4-6.1)	3.1 (2.7-3.8)	2.7 (2.2-3.3)	6.6 (4.7-9.2)	4.1 (2.9-5.9)	3.5 (2.4-5.1)
	Extension (2)	5.2 (4.4-6.3)	3.0 (2.5-3.6)	2.3 (1.8-2.8)	5.7 (3.7-9.0)	4.2 (2.6-6.8)	3.8 (2.3-6.4)
	Abnormal flexion (3)	2.7 (2.2-3.2)	1.9 (1.6-2.3)	1.7 (1.4-2.1)	3.3 (2.1-5.2)	2.5 (1.5-4.0)	2.3 (1.4-3.7)
	Normal flexion (4)	1.5 (1.3-1.8)	1.3 (1.0-1.5)	1.2 (1.0-1.4)	1.7 (1.1-2.4)	1.5 (1.0-2.2)	1.5 (1.0-2.3)
	Localizes/obeys (5/6)	1.0 (Ref)	1.0 (Ref)	1.0 (Ref)	1.0 (Ref)	1.0 (Ref)	1.0 (Ref)
	Untestable / missing (9)	3.1 (2.4-3.9)	2.0 (1.6-2.6)	1.5 (1.2-2.0)	2.7 (1.8-3.9)	2.3 (1.5-3.5)	2.3 (1.5-3.5)
Pupillary reactivity	Both pupils reactive (1)	1.0 (Ref)	1.0 (Ref)	1.0 (Ref)	1.0 (Ref)	1.0 (Ref)	1.0 (Ref)
	One pupil reactive (2)	2.8 (2.4-3.2)	2.2 (1.9-2.6)	1.9 (1.6-2.2)	2.6 (2.1-3.3)	2.3 (1.8-2.9)	1.7 (1.3-2.2)
	No pupil reactive (3)	6.5 (5.8-7.4)	4.8 (4.2-5.5)	4.4 (3.8-5.2)	18.4 (13.2-25.6)	14.2 (10.0-20.1)	8.5 (5.9-12.2)
Hypoxia	Yes or suspected	2.1 (1.8-2.4)		1.3 (1.1-1.6)	1.8 (1.3-2.5)		1.7 (1.1-2.5)
Hypotension	Yes or suspected	2.9 (2.6-3.4)		2.1 (1.8-2.5)	3.1 (2.5-3.8)		2.1 (1.6-2.7)
Basal cisterns	Partially compressed or absent	3.0 (2.6-3.4)		1.9 (1.6-2.2)	5.2 (4.2-6.4)		2.8 (2.1-3.7)
Midline shift	Yes	2.2 (2.0-2.5)		1.3 (1.1-1.5)	3.1 (2.5-3.8)		1.3 (1.0-1.7)
Traumatic subarachnoid hemorrhage	Yes	2.5 (2.2-2.7)		2.1 (1.9-2.4)	2.5 (2.0-3.0)		1.3 (1.0-1.7)

Prognostic effects

Overall, the prognostic effects were very similar in the IMPACT and the TBI-trac® data (table 5.2). The largest difference was in the odds ratios (OR) for no reactive pupils (univariate OR=6.5 (95% CI: 5.8-7.4) in IMPACT and 18.4 (95% CI: 13.2-25.6) in TBI trac®). In the Core Model from both datasets, older age, poor GCS motor score and unreactive pupils were highly predictive for 14-day mortality. In addition, hypoxia, hypotension, compressed or absent basal cisterns, midline shift and traumatic subarachnoid hemorrhage were significant predictors of 14-day mortality from Extended Model in both datasets.

Model performance

The AUCs of the refitted models in the IMPACT data were 0.763 for the Core model and 0.806 for the Extended model (table 5.3). In the validation sample the models discriminated even better between patients who had died within 14 days and those who were alive. AUCs were 0.787 (95% CI: 0.765 - 0.809) for the Core model and 0.827 (95% CI: 0.805 - 0.845) for the Extended model. Also the calibration of the models was good (figure 5.1). The Core Model predicted 20% 14-day mortality and the Extended Model 24%, while observed 14-day mortality was 23%. The calibration slopes were 1.239 for the Core Model and 1.121 for the Extended Model, indicating the prognostic effects were stronger in the validation sample.

The models were also tested in the non-imputed dataset, which showed similar values for the performance measures. The Core Model (n=2492) had an AUC of 0.787 and calibration slope of 1.24. For the Extended Model (n=1976) the AUC was 0.833 and the calibration slope 1.17.

Table 5.3. Discrimination (Area Under the Curve (95% Confidence Interval)) in the development (IMPACT; n=8428 (Core) and n=6918 (Extended)) and validation (BTF TBI-trac®; n=2513) data

	Core model	Extended model
Development	0.763 (0.750-0.775)	0.806 (0.794-0.817)
External validation	0.787 (0.765-0.809)	0.827 (0.805-0.845)

Core model:
logodds (14-day mortality) = -3.342 + 0.027(age) + 1.154*(GCS motor=1) + 1.089*(GCS motor=2) + 0.642*(GCS motor=3) + 0.225*(GCS motor=4) + 0.707*(GCS motor=9) + 0.780*(pupils=2) + 1.567*(pupils=3)*

Extended model:
logodds (14-day mortality) = -3.957 + 0.022(age) + 0.990*(GCS motor=1) + 0.861*(GCS motor=2) + 0.522*(GCS motor=3) + 0.166*(GCS motor=4) + 0.424*(GCS motor=9) + 0.625*(pupils=2) + 1.488*(pupils=3) + 0.282*(hypoxia) + 0.752*(hypotension) + 0.618*(obliterated cisterns) + 0.226*(midline shift) + 0.749*(TSAH)*

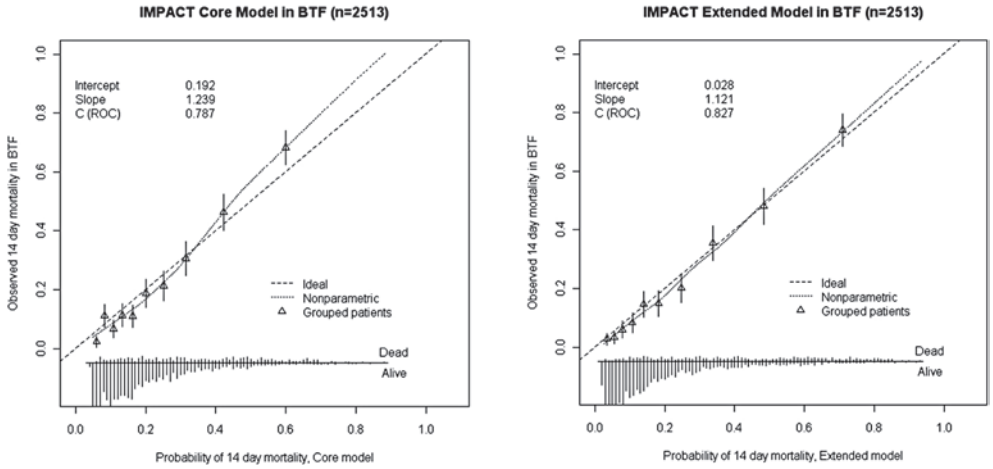


Figure 5.1. Calibration plots of the Core and Extended IMPACT models in BTF TBI-trac® data

Discussion

The relevance of prognostic modelling in TBI is being increasingly recognised (3). This has resulted in the development of various prognostic models. Many of these however have shown methodological shortcomings, in particular overfitting due to incorrect approaches to predictor selection and lack of internal and external validation (8, 9). The IMPACT models were developed on large data sets with state-of-the-art methodology, including external validation (10). The development phase thus meets quality standards for model development (7). The need for validation however does not end with the development phase: in general, prognostic models should be submitted to a continuing process of validation and, where appropriate, updating in order to remain current. In this study, we validated the IMPACT prognostic models in the contemporary data set of the BTF TBI-trac® New York State TBI registry. On this validation the models showed high discrimination and good calibration.

These results confirm generalizability, and further demonstrate the validity of their use for current data.

In this validation study the IMPACT models were refitted, as CT parameters were scored differently in the BTF TBI-trac® database and the outcome in this observational study was 14-day mortality rather than 6-month outcome. In this case, refitting was essential due to the different coding of CT parameters. In general however, the concept of refitting a prognostic model is a highly relevant issue, in particular for the application of prognostic models in the design and analysis of clinical trials. The IMPACT studies have shown that the statistical power of clinical trials can be increased by up to 50% with the combined use of covariate adjustment and ordinal analysis (11, 12). Both covariate adjustment and the

use of the sliding dichotomy as an approach to ordinal analysis require robust prognostic models. Debate exists as to whether these models should be completely pre-specified or whether the weighting of the variables included in the model may be refitted to the data set under study. In the latter case a better discrimination and performance of the model on the new data set may be anticipated. We suggest that such refitting is admissible and perhaps even preferable. A minimum requirement however is the pre-specification of the variables included in the model. We note that this approach to refitting is different from what we did in this study. In the current study the existing IMPACT models were refitted within the development data set rather than on the new data set under study. As a sensitivity analysis we additionally tested the existing IMPACT Core model on the BTF TBI trac® database.

Discrimination was very similar: the AUC for the existing Core model was 0.773, compared to 0.786 for the refitted model. Calibration was less optimal: the intercept was -0.605 for the existing Core model, compared to 0.037 for the refitted variant. The calibration slopes were again quite similar: 1.059 for the existing model versus 1.132 for the refitted.

We found greater discriminative ability of the IMPACT models in the validation data set than previously described in the development data set. This can be explained by the greater heterogeneity of the validation data set (observational series without restrictive enrolment criteria) and illustrates that the AUC as measure for discrimination is also influenced by the case-mix of the validation set (13).

The calibration plots (figure 5.1) for the IMPACT Core model show a small but systematic underestimation of observed mortality. It is tempting to attribute this to improvements in treatment over time. However, differences in coding or distribution of variable between the data sets may offer alternative explanations. We note e.g. a larger number of patients in the TBI-trac® dataset with an absent motor score (most likely including “false absent” due to sedation) and a lower number with unreactive pupils? It would seem possible that this lack of information is compensated for in the extended model by adding details on CT characteristics, as the fit of the extended IMPACT model is excellent. This excellent fit is remarkable as the data sets, underpinning the development of the prognostic models, included series collected between 1984 and 1997 (14). The excellent fit emphasizes the broad generalizability of the IMPACT models both across time and settings.

Several limitations of our study should be acknowledged. First, the BTF TBI-trac® database did not contain all variables in a similar coding format as used in the IMPACT prognostic model. In particular, CT parameters were scored differently. We therefore had to modify and refit the IMPACT extended models to include these parameters; we consider it however unlikely that this will have influenced performance. The discrepancy in coding essential variables emphasizes the necessity for standardization of data collection in TBI studies, as

recommended in the Common Data Elements initiative (15-19).

Second, 14-day mortality was the outcome measure recorded in the BTF TBI-trac® database, whilst the IMPACT models were developed versus 6-month mortality and unfavourable outcome. We therefore refitted the IMPACT models for 14-day mortality in the development population. As a sensitivity analysis we also applied the original IMPACT Core model to the BTF TBI-trac® dataset and found no major differences in performance.

This agreement is possibly explained by the fact that the highest proportion of mortality occurs in the first weeks after the injury. In the CRASH trial investigating the effect of corticosteroid therapy after TBI for example, 85% of deaths occurred in the first two weeks (20, 21). The difference in outcome measures between the IMPACT and BTF TBI-trac® datasets may however be considered more an asset than a limitation, as the good performance of the models on external validation versus 14-day mortality demonstrates the broad generalizability of these models across different settings.

Several other studies have also validated the IMPACT models. Panczykowski et al. reported good performance of the IMPACT models (AUC: 0.76-0.83) at validation on an unselected series of 587 patients with severe TBI admitted to a single level 1 trauma centre (22).

Roozenbeek et al. externally validated the IMPACT models on 5 new datasets (2 observational studies and 3 clinical trials) also finding good performance with regard to both discrimination and calibration (23). Poorer performance was however found in validating the models on one of the most recent phase III clinical trials in TBI (dexanabinol study) (24).

In combination with the results of these other validation studies, the current study has now shown good performance and broad generalizability for the IMPACT models across broad and widely different settings.

Conclusions

The current validation study on the contemporary BTF TBI-trac® database from New York State demonstrates that the IMPACT prognostic models for TBI accurately predict 14-day mortality. In combination with the results of other validation studies, the current study has proven the generalizability of the IMPACT prognostic models across different settings and outcome measures.

Acknowledgements

Development of the IMPACT models was supported by NIH grant 046291.
The BTF TBI-trac® data collection was funded through the New York State Department of Health Bureau of Emergency Medical Services and by the Brain Trauma Foundation.

References

1. Ghajar J. **Traumatic brain injury.** *Lancet* 2000; 356(9233):923-9
2. Maas AI, Stocchetti N, Bullock R. **Moderate and severe traumatic brain injury in adults.** *Lancet Neurol* 2008; 7(8):728-41
3. Lingsma HF, Roozenbeek B, Steyerberg EW, Murray GD, Maas AI. **Early prognosis in traumatic brain injury: from prophecies to predictions.** *Lancet Neurol* 2010; 9(5): 543-54
4. Brain Trauma Foundation. **Guidelines for the Management of Severe Head Injury**; 1995
5. Brain Trauma Foundation. **Management and Prognosis of Severe Traumatic Brain Injury: Part I Guidelines for the Management of Severe Traumatic Brain Injury.** Brain Trauma Foundation, New York, NY; 2000
6. Brain Trauma Foundation. **Guidelines for the Management of Severe Traumatic Brain Injury.** *J Neurotrauma* 2007; 24(supplement 1): S1-S106
7. Steyerberg EW. **Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating.** 2009, **Springer: New York**
8. **Perel P, Edwards P, Wentz R, Roberts I. Systematic review of prognostic models in traumatic brain injury.** *BMC Med Inform Decis Mak* 2006; 6: 38
9. Mushkudiani NA, Hukkelhoven CW, Hernandez AV, Murray GD, Choi SC, Maas AI, Steyerberg EW. **A systematic review finds methodological improvements necessary for prognostic models in determining traumatic brain injury outcomes.** *J Clin Epidemiol* 2008; 61(4): 331-43
10. Steyerberg EW, Mushkudiani N, Perel P, Butcher I, Lu J, McHugh GS, Murray GD, Marmarou A, Roberts I, Habbema JD, Maas AI. **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
11. McHugh GS, Butcher I, Steyerberg EW, Marmarou A, Lu J, Lingsma HF, Weir J, Maas AI, Murray GD. **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(1):44-57
12. Maas AI, Steyerberg EW, Marmarou A, McHugh GS, Lingsma HF, Butcher I, Lu J, Weir J, Roozenbeek B, Murray GD**. IMPACT Recommendations for Improving the Design and Analysis of Clinical Trials in Moderate to Severe Traumatic Brain Injury.** *Neurotherapeutics* 2010; 7(1):127-134
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): 971-80
14. Marmarou A, Lu J, Butcher I, McHugh GS, Mushkudiani NA, Murray GD, Steyerberg EW, Maas AI. **IMPACT database of traumatic brain injury: design and description.** *J Neurotrauma* 2007; 24(2): 239-50
15. Maas AI, Harrison-Felix CL, Menon D, Adelson PD, Balkin T, Bullock R, Engel DC, Gordon W, Orman JL, Lew HL, Robertson C, Temkin N, Valadka A, Verfaellie M, Wainwright M, Wright DW, Schwab K. **Common data elements for traumatic brain injury: recommendations from the interagency working group on demographics and clinical assessment.** *Arch Phys Med Rehabil* 2010; 91(11): 1641-9
16. Maas AI, Harrison-Felix CL, Menon D, Adelson PD, Balkin T, Bullock R, Engel DC, Gordon W, Orman JL, Lew HL, Robertson C, Temkin N, Valadka A, Verfaellie M, Wainwright M, Wright DW, **Schwab K. Standardizing Data Collection in Traumatic Brain Injury.** *J Neurotrauma* 2011; 28:177–187
17. Duhaime AC, Gean AD, Haacke EM, Hicks R, Wintermark M, Mukherjee P, Brody D, Latour L, Riedy G; Common Data Elements Neuroimaging Working Group Members, Pediatric Working Group Members. **Common data elements in radiologic imaging of traumatic brain injury.** *Arch Phys Med Rehabil* 2010; 91(11):1661-6
18. Manley GT, Diaz-Arrastia R, Brophy M, Engel D, Goodman C, Gwinn K, Veenstra TD, Ling G, Ottens AK, Tortella F, Hayes RL. **Common data elements for traumatic brain injury: recommendations from the biospecimens and biomarkers working group.** *Arch Phys Med Rehabil* 2010; 91(11):1667-72
19. Wilde EA, Whiteneck GG, Bogner J, Bushnik T, Cifu DX, Dikmen S, French L, Giacino JT, Hart T, Malec JF, Millis SR, Novack TA, Sherer M, Tulsky DS, Vanderploeg RD, von Steinbuechel N. **Recommendations for the use of common outcome measures in traumatic brain injury research.** *Arch Phys Med Rehabil* 2010; 91(11): 1650-1660.
20. Roberts I, Yates D, Sandercock P, et al on behalf of the CRASH trial collaborators. **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: 1321-28.

21. Edwards P, Arango M, Balica L, et al on behalf of the CRSAH trial collaborators. **Final results of MRC CRASH, a randomised placebo-controlled trial of intravenous corticosteroid in adults with head injury-outcomes at 6 months.** *Lancet*. 2005; 365: 1957-59.

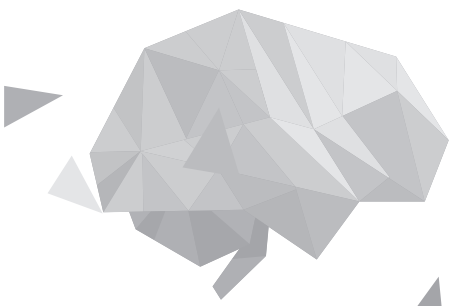
22. Panczykowski D, Puccio A, Scruggs BJ, et al: **Prospective Independent Validation of IMPACT Modeling as a Prognostic Tool in Severe Traumatic Brain Injury.** *J Neurotrauma* 2012; 29(1):47-52

23. Roozenbeek B, Lingsma HF, Lecky FE, Lu J, Weir J, Butcher I, McHugh GS, Murray GD, Perel P, Maas AI, Steyerberg EW: **Prediction of outcome after moderate and severe traumatic brain injury: external validation of the IMPACT and CRASH prognostic models.** *Crit Care Med* 2012; 40(5): 1609-17

24. Maas AI, Murray G, Henney H 3rd, Kassem N, Legrand V, Mangelus M, Muizelaar JP, Stocchetti N, Knoller N; Pharmos TBI investigators. **Efficacy and safety of dexanabinol in severe traumatic brain injury: results of a phase III randomised, placebo-controlled, clinical trial.** *Lancet Neurol* 2006 ; 5(1): 38-45

Chapter 6

THE INFLUENCE OF ENROLLMENT CRITERIA ON RECRUITMENT AND OUTCOME DISTRIBUTION IN TRAUMATIC BRAIN INJURY STUDIES: RESULTS FROM THE IMPACT STUDY



Bob Roozenbeek
Andrew IR Maas
Anthony Marmarou
Isabella Butcher
Hester F Lingsma
Juan Lu
Gillian S McHugh
Gordon D Murray
Ewout W Steyerberg
on behalf of the IMPACT Study Group

J Neurotrauma 2009; 26(7): 1069-75

Abstract

Substantial heterogeneity exists among patients who suffer from traumatic brain injury (TBI). Strict enrollment criteria may diminish heterogeneity in randomized controlled trials (RCTs), but will also decrease recruitment and may affect the outcome distribution. The aim of this study was to investigate the influences of commonly used enrollment criteria for RCTs in TBI on potential recruitment and on outcome distribution. We used individual patient data from the International Mission on Prognosis and Analysis of Clinical Trials in TBI (IMPACT) database, including six therapeutic phase III RCTs (n=5816) and three surveys (n=2217) in TBI.

The primary outcome was the Glasgow Outcome Scale (GOS) at 6 months after injury, which we dichotomized as favorable vs. unfavorable. We investigated the influences of commonly used enrollment criteria on recruitment and outcome distribution: time window between injury and admission to study hospital $\leq$ 8h; age at injury $\leq$ 65 years; $\geq$ 1 reactive pupil; motor score > 1; Glasgow Coma Scale $\leq$ 8. Application of all enrollment criteria resulted in a large reduction of recruitment in both the surveys (up to 65%) and the RCTs (up to 41%).

Among the remaining patients, fewer had an unfavorable outcome in both the surveys (original, 60%; remaining, 44%) and the RCTs (original, 43%; remaining, 38%). Applying these enrollment criteria to patients from the surveys resulted in an outcome distribution that approximated the outcome observed in the RCTs. The use of strict enrollment criteria leads to substantial reductions in the recruitment of RCTs in TBI. The outcome in TBI studies depends strongly on the enrollment criteria.

Introduction

Traumatic brain injury (TBI) represents a serious health problem worldwide, posing high costs to society. Much research is undertaken to improve outcome after TBI. Comparing outcome between different series of TBI patients and demonstrating efficacy of new approaches in randomized clinical trials (RCTs) is notoriously difficult due to the inherent heterogeneity of TBI populations (1). Heterogeneity in terms of type and severity of injury, as well as in terms of prognostic risk, may cause imbalances between treatment groups and dilute treatment effects.

Imbalances in baseline characteristics between treatment groups have been reported in some RCTs (2). Such imbalances may be prevented by block randomization (3), stratified randomization (1, 3, 4), and minimization. We may achieve similar numbers of patients in each arm of the trial by random allocation in blocks (''block randomization''). Stratified randomization may be used to ensure that equal numbers will be allocated to each arm with a characteristic thought to affect prognosis. Furthermore, minimization may be used to ensure balance between trial arms for several patient factors. The first patient is truly randomly allocated; for each subsequent patient, the treatment allocation is identified, which minimizes the imbalance between groups at that time. Approaches for dealing with the heterogeneity in the analysis phase (e.g., covariate adjustment) have also been employed in TBI trials (5).

Heterogeneity amongst the population under study may also be reduced by using strict enrollment criteria. Enrollment criteria aim to select patients that are most likely to benefit from the studied treatment, both from a mechanistic and prognostic perspective (1, 6, 7). Three main goals are aimed for: (a) exclusion of patients with a very good or very bad prognosis, because one may expect that these patients will do well without any treatment or will do poorly no matter what treatment is given; (b) inclusion of patients who are likely to benefit from the treatment, because of the pathophysiologic background of the injury and the studied treatment; (c) reaching a distribution of approximately 50% of patients with a favorable outcome and 50% of patients with an unfavorable outcome in the study population, to increase statistical power for the detection of a true treatment effect.

Machado et al. (6) have demonstrated that targeting a clinical trial to patients with an intermediate prognostic risk would permit a potential reduction in sample size of approximately 30%. The disadvantage of this approach, however, is that it may substantially reduce recruitment.

The objective of this study is to investigate the influences of common enrollment criteria on the potential recruitment and on the outcome distribution of RCTs in TBI.

Methods

Study population

We used individual patient data from the International Mission on Prognosis and Analysis of Clinical Trials in TBI (IMPACT) database. This database contains data of patients with moderate (Glasgow Coma Scale [GCS], 9–12) and severe (GCS ≤ 8) TBI from eight RCTs and three surveys conducted between 1984 and 1996. IMPACT links researchers in Belgium, the Netherlands, the United Kingdom, and the United States in a project addressing methodological problems in design and analysis of RCTs in TBI. Details of the different studies and data management of the IMPACT database have been described previously (8).

The primary outcome measurement was the Glasgow Outcome Scale (GOS) at 6 months after injury. For our analyses, we dichotomized the GOS into unfavorable (GOS 1–3) versus favorable outcome (GOS 4–5). We selected patients over 14 years of age at the time of injury.

The data from two RCTs (HIT-I and SKB) were not used in our analysis, because the numbers of patients in these studies were too small for the purpose of our study. This resulted in a cohort of 8033 adults: 2217 patients from surveys and 5816 patients from RCTs.

Selection by enrollment criteria

We investigated the influences of five commonly used enrollment criteria for TBI studies: time window between injury and admission to study hospital ≤8h; age at injury ≤65 years; ≤1 reactive pupil; motor score>1; and GCS ≤8. We further explored the influence of other cut-off values for time window (≤ 4, ≤ 6, and ≤ 12 h) and for age (≤ 55, ≤ 60, and ≤ 70 years).

The enrollment criteria were applied separately and simultaneously. The influence on the recruitment was studied by comparing the original sample sizes to the new sample sizes with the investigated enrollment criteria applied. The change in outcome distribution was studied by comparing unfavorable outcome fractions at 6 months after the injury.

Statistical analysis

Descriptive statistics included medians and percentages. The software used was SPSS, version 11.0.1 (SPSS Inc., Chicago, IL). Figure 1 was produced with *R* software version 2.6.0 (The *R* Foundation, 2007).

Results

The baseline characteristics of the original study population are summarized in table 6.1.

These reflect the effects of the different initial enrollment criteria per study.

Effects of enrollment criteria on recruitment

Effects of enrollment criteria on recruitment were more pronounced in the surveys than in the RCTs, since these had less stringent initial enrollment criteria (table 6.1). When each of the studied criteria was applied separately in the surveys, ''at least one reactive pupil'' led to the largest reduction in recruitment (32%), followed by ''time window < 8 h'' (20%), ''motor score > 1'' (18%), ''GCS ≤ 8'' (15%), and ''age at injury ≤ 65'' (12%; table 6.2). In the RCTs, the reductions were less (table 6.2).

As expected, using shorter time windows led to a greater reduction of recruitment in both the surveys and the RCTs (table 6.3). In the surveys, the recruitment reduction decreased gradually with the use of higher maximum ages: 22% (≤ 55 years); 17% (≤ 60 years); 12% (≤ 65 years); and 9% (≤ 70 years; table 6.4). The RCTs showed less reduction: 10% (≤ 55 years); 5% (≤ 60 years); 2% (≤ 65 years); and 0% (≤ 70 years; table 6.4).

Application of all five studied enrollment criteria simultaneously led to a mean recruitment reduction of 65% in the surveys (table 6.2, figure 6.1). In the RCTs, the mean reduction of the recruitment was 41%.

Table 6.1. Distribution of patient characteristics across three surveys and six RCTs from the IMPACT study

	Time window *(hours)*	**Age** *(years)*	**Pupillary reactivity**	**Motor score**	**GCS**
	median (interquartile range)	*median (interquartile range)*	*(1) both pupils pos* *(2) one pupil pos* *(3) both pupils neg*	*(1,2) none/extension* *(3) abnormal flexion* *(4) normal flexion* *(5,6) localizes/obeys* *(9) untestable/missing*	median (interquartile range)
TCDB **(n=604)**	1.2 (.8 - 3.1)	26 (21 – 40)	300 (50%) 55 (9%) 249 (41%)	243 (40%) 74 (12%) 122 (20%) 134 (22%) 31 (5%)	6 (4 - 8)
UK4 (n=791)	3.5 (2.0 - 8.9)	36 (22 - 55)	429 (54%) 115 (15%) 247 (31%)	198 (25%) 37 (5%) 141 (18%) 221 (28%) 194 (25%)	6 (4 - 8)
EBIC (n=822)	3.7 (1.5 - 8.0)	38 (24 - 59)	527 (64%) 87 (11%) 208 (25%)	230 (28%) 55 (7%) 113 (14%) 281 (34%) 143 (17%)	7 (4 - 11)
Total surveys (n=2217)	2.8 (1.2 - 6.1)	32 (22 - 53)	1256 (57%) 257 (12%) 704 (32%)	671 (30%) 166 (8%) 376 (17%) 636 (29%) 368 (16%)	7 (4 - 9)

Table 6.1. *(Continued)*

	Time window *(hours)*	**Age** *(years)*	**Pupillary reactivity**	**Motor score**	**GCS**
	median (interquartile range)	*median (interquartile range)*	*(1) both pupils pos* *(2) one pupil pos* *(3) both pupils neg*	*(1,2) none/extension* *(3) abnormal flexion* *(4) normal flexion* *(5,6) localizes/obeys* *(9) untestable/missing*	median (interquartile range)
HIT II (n=819)	2.0 (1.0 - 4.0)	33 (22 - 49)	583 (71%) 101 (12%) 135 (16%)	280 (34%) 92 (11%) 181 (22%) 207 (25%) 59 (7%)	8 (6 - 10)
TIUS (n=1041)	.7 (.5 -1.2)	30 (23 - 41)	709 (68%) 122 (12%) 210 (20%)	152 (15%) 132 (13%) 300 (29%) 457 (44%) 0 (0%)	7 (6 - 10)
TINT (n=1118)	1.2 (.8 - 2.1)	30 (21 - 45)	813 (73%) 170 (15%) 135 (12%)	141 (13%) 237 (21%) 327 (29%) 413 (37%) 0 (0%)	6 (5 - 7)
PEGSOD (n=1510)	1.1 (.6 - 2.2)	27 (20 - 38)	779 (52%) 160 (11%) 571 (38%)	655 (43%) 165 (11%) 334 (22%) 356 (24%) 0 (0%)	5 (3 - 7)
SLIN (n=409)	No data	28 (21 - 43)	316 (78%) 79 (19%) 14 (3%)	55 (13%) 91 (22%) 127 (31%) 136 (33%) 0 (0%)	6 (5 – 7)
SAPHIR (n=919)	2.3 (1.0 - 4.2)	32 (20 - 38)	612 (67%) 307 (33%) 0 (0%)	264 (29%) 143 (16%) 223 (24%) 286 (31%) 3 (0%)	6 (5 - 8)
Total RCTs (n=5816)	1.2 (0.7 - 2.6)	30 (21 - 43)	3812 (66%) 939 (16%) 1065 (18%)	1547 (27%) 860 (15%) 1492 (26%) 1855 (32%) 62 (1%)	7 (5 - 10)

Effects of enrollment criteria on outcome distribution

The application of more stringent enrollment criteria had profound effects on the outcome distribution (table 6.2). Effects were greatest for age, pupils, motor score, and GCS. In contrast to all other selection criteria, the GCS criterion mainly excluded patients with a favorable outcome. Patients excluded for a high age, bilaterally unresponsive pupils or an absent motor score had more unfavorable outcomes. Only modest effects of different time windows were seen on the outcome distribution (table 6.3).

Combining all enrollment criteria had a pronounced effect in the outcome distribution, which was most evident in the surveys, reflecting the greater initial heterogeneity. In the surveys the percentage of unfavorable outcome prior to applying the enrollment criteria was 60%, and decreased to 44% after applying the enrollment criteria, approximating the outcome distribution observed in the RCTs.

Effects of the application of all enrollment criteria on the recruitment and outcome distribution are illustrated in figure 6.1. For example, the first gray bar of figure 6.1 shows that application of all investigated enrollment criteria on the population of the Traumatic Coma Data Bank (TCDB) survey, resulted in selection of 253 patients, from whom 43% had an unfavorable outcome. The first black bar depicts 351 excluded patients from the TCDB population with an unfavorable proportion of 81%.

Table 6.2. Numbers of excluded patients across studies by application of enrollment criteria

	Original study population		Time window ≤ 8 hrs		Age ≤ 65		Unreactive pupils		Motor score > 1		GCS ≤ 8		All criteria applied			Total remaining N if all criteria applied	
	N	% Unfav of N	No. of patients excluded from orig. N	% Unfav of excluded patients	No. of patients excluded from orig. N	% Unfav of excluded patients	No. of patients excluded from orig. N	% Unfav of excluded patients	No. of patients excluded from orig. N	% Unfav of excluded patients	No. of patients excluded from orig. N	% Unfav of excluded patients	No. of patients excluded from orig. N	% of patients excluded form orig. N	% Unfav of excluded patients	N	% Unfav of remaining N
TCDB*	604	65	38	74	41	95	249	88	136	91	40	45	351	58	81	253	44
UK4*	791	66	210	70	90	89	247	90	113	89	80	54	508	64	76	283	47
EBIC*	822	51	198	53	135	79	208	81	150	71	207	29	583	71	55	239	42
Total surveys	2217	60	446	63	266	85	704	87	399	82	327	37	1442	65	67	775	44
HIT II^	819	40	62	42	35	71	135	70	210	47	42	21	391	48	48	428	43
TIUS^	1041	38	0	-	7	100	210	70	9	67	226	16	436	42	43	605	34
TINT^	1118	41	0	-	13	69	135	68	5	60	161	17	310	28	42	808	41
PEGSOD^	1510	51	1	0	32	84	571	69	475	63	68	12	904	60	59	606	40
SLIN^	409	43	0	-	1	100	14	93	0	-	0	-	15	4	93	394	41
SAPHIR^	919	41	32	56	34	82	0	-	141	50	138	25	327	36	43	592	40
Total RCTs	5816	43	95	46	122	80	1065	70	841	57	635	18	2383	41	50	3433	38

* = survey

^ = randomized controlled trial

Table 6.3. Numbers of excluded patients by selection on time window

	Original enrollment criteria	Original study population		Time window ≤ 4 hours		Time window ≤ 6 hours		Time window ≤ 8 hours		Time window ≤ 12 hours	
		N	% Unfav of N	No. of patients excluded	% Unfav of patients excluded	No. of patients excluded	% Unfav of patients excluded	No. of patients excluded	% Unfav of patients excluded	No. of patients excluded	% Unfav of patients excluded
TCDB*	≤ 48 hrs	604	65	106	64	57	70	38	74	22	77
UK4*	≤ 72 hrs	791	66	359	70	252	69	210	70	154	71
EBIC*	≤ 24 hrs	822	51	372	49	252	52	198	53	152	55
Total surveys		2217	60	837	60	561	62	446	63	328	64
HIT II^	≤ 12 hrs no longer obeying command or ≤ 24 hrs after injury	819	40	248	40	121	41	62	42	28	43
TIUS^	≤ 4 hrs	1041	38	-	-	-	-	-	-	-	-
TINT^	≤ 4 hrs	1118	41	-	-	-	-	-	-	-	-
PEGSOD^	≤ 8 hrs	1510	51	107	49	16	38	-	-	-	-
SLIN^	< 8 hrs and within 4 hrs of admission	409	43	0	-	0	-	-	-	-	-
SAPHIR^	≤ 12 hrs	919	41	265	43	103	43	32	56	-	-
Total RCTs		5816	43	620	43	240	42	94	47	28	43

* = survey
^ = randomized controlled trial

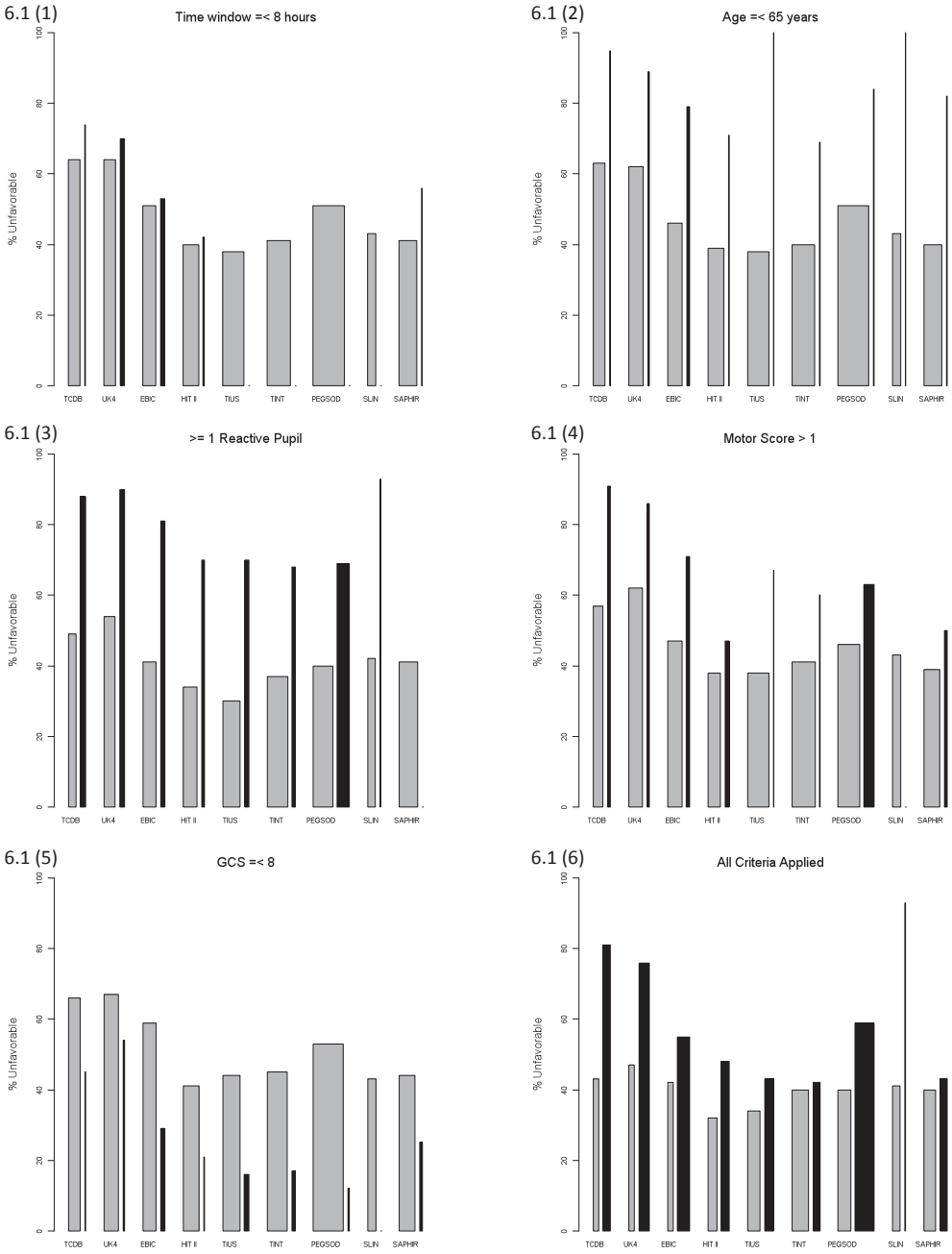


Figure 6.1. The influence of the application of the investigated enrollment criteria (time window ≤ 8 hours; age at injury ≤ 65 years; ≥ 1 reactive pupil; motor score > 1; GCS ≤ 8) on recruitment and outcome distribution in three surveys and six RCTs of the IMPACT database. For each study the selected (grey bars) and the excluded patients (black bars) are depicted. The width of the bars indicates the numbers of selected and excluded patients.

Discussion

Our evaluations show substantial effects of enrollment criteria both on potential recruitment and on outcome. The five studied enrollment criteria led to a joint reduction of recruitment of 65% in the surveys and 41% in the RCTs, and substantially narrowed the differences in outcome between surveys and RCTs that were initially present. Clearly, effects in individual studies varied with the stringency of original criteria.

As a limitation of our study, it must be noticed that we only focused on the influences of five enrollment criteria applied separately and simultaneously. We did not investigate any interactions between the different enrollment criteria. Structural abnormalities as visualized by computed tomography (CT) examination are currently commonly used as selection criterion for enrollment in RCTs, but we did not include these in our analyses, as the available information was not detailed enough in all IMPACT studies.

In TBI, Phase III RCTs traditionally aim to target the population that is most likely to benefit from the therapy under investigation, both from a mechanistic and prognostic perspective. Mechanistic targeting is based upon our—often very poor—potential for identifying patients in whom specific pathophysiologic mechanisms are active. In prognostic targeting, enrollment criteria aim to obtain a homogeneous population with respect to expected outcome, excluding those with either a very poor or very good prognosis. For this reason, most trials have excluded patients over the age of 65, and patients with an absent motor response or unreactive pupils. Indeed, our results show a very high rate of unfavorable outcome in patients excluded for these reasons, indicating that this approach is appropriate.

Nevertheless, interpretation should be with caution: TBI is being seen more frequently in elderly patients and, with the ageing population, the average age of TBI patients is increasing (9). Elderly patients with TBI therefore represent an important subpopulation. The poor outcome in elderly patients observed in our series is primarily caused by severe TBI (GCS $\leq$ 8), and no inference may be made as to patients with moderate TBI.

With the current frequent use of early sedation, intubation and neuromuscular blockade, accurate assessment of the motor scale may be confounded, leading to erroneously recording the motor score as being absent. Likewise, small unreactive pupils may be caused by effects of deep sedation. Consequently, the use of these criteria in isolation may be hazardous. Taken in combination, however, reliability will be increased and for this reason we prefer to classify by prognosis. A limited number of admission characteristics can be used to establish the baseline prognostic risk in individual patients (10, 11).

Machado et al. (6) have shown that targeting a trial towards patients with an intermediate prognosis may reduce the required sample size by 30%. These conclusions were reached from analyses performed on the European Brain Injury Consortium (EBIC) survey. Our results, including also analysis of the EBIC survey, however, show that the reduction in recruitment due to more stringent enrollment criteria may be even higher, thus obviating the potential

benefit. This touches on the debate whether focused, targeted Phase III trials, or large mega trials (with inherent heterogeneity) should be preferred in the field of TBI. To date, only one mega-trial has been reported in TBI (12).

We found that time window is a crucial factor in determining recruitment, but that little effect exists on the outcome distribution. Recruitment reductions by time window selection varied from 38% (≤ 4 h) to 15% (≤ 12 h) in the surveys. In the RCTs, there was a variation of 11% (≤ 4 h) to <1% (≤ 12 h) in reduction of recruitment.

Determination of the appropriate time window for a study should preferably be based on knowledge of the pathophysiologic mechanism targeted, and the time at which this mechanism may be active. Extrapolating results from experimental studies to the clinical situation may, however, be difficult. In practice, time windows are frequently determined based on ''best guess'' and the clinical perception as to the time frame within which a reasonable number of subjects may be recruited. Our studies show a substantial decline in recruitment when time windows are shortened from 6 to 4 hours, but less so on comparing a 6-hour time window to 8 hours. However, these effects of time window on recruitment were based on the available data in the IMPACT database about the time of injury and the time of admission to study hospital. In practice, the actual time windows to study drug administration are longer, due to organizational and logistical reasons such as informed consent procedures (13). Therefore, our observations are likely to be an underestimation of the actual decrease in recruitment by time window selection.

We also probably underestimate the actual recruitment reduction by application of selection criteria because we investigated the effect of selection criteria on pre-selected study populations. In the SAPHIR trial, for example, we have learned from the screening logs that only 19% of the number of patients initially screened was enrolled in the trial; 81% was excluded because of not meeting enrollment criteria, inability to obtain informed consent, or logistic reasons (e.g., unavailability of study personnel or medication) (14).

The profound effect of enrollment criteria on outcome is perhaps not surprising, but has hitherto been insufficiently recognized. Enrollment criteria may be deliberately applied (as in RCTs), but may also inadvertently result from selective hospital admission policies in different institutions, or be caused by effects of local trauma organization, outside the control of investigators. A detailed understanding of such factors is, in our opinion, a prerequisite for comparing outcome results between series, originating from different settings or times.

It is remarkable that the application of common enrollment criteria to surveys resulted in an approximation of outcome distribution observed in RCTs.

Conclusions

Strict selection with common enrollment criteria leads to substantial recruitment reductions in TBI studies. The outcome distribution strongly depends on the enrollment criteria.

References

1. Maas AI, Steyerberg EW, Murray GD, et al. **Why have recent trials of neuroprotective agents in head injury failed to show convincing efficacy? A pragmatic analysis and theoretical considerations.** *Neurosurg* 1999; 44: 1286–1298
2. Marshall LF, Maas AI, Marshall SB, et al. **A multicenter trial on the efficacy of using tirilazad mesylate in cases of head injury.** *J Neurosurg* 1998. 89; 519–525.
3. Bullock, M.R., Merchant, R.E., Choi, S.C., et al. (2002). **Outcome measures for clinical trials in neurotrauma.** *Neurosurg*. Focus. Available at: *www.aans.org/education/journal/neurosurgical/july02/13-1-nsf-toc.asp.* Accessed May 1, 2009
4. Choi, SC, Bullock R. **Design and statistical issues in multicenter trials of severe head injury.** *Neurol Res* 2001; 23: 190–2
5. Hernandez AV, Steyerberg EW, Butcher I, et al. **Adjustment for strong predictors of outcome in traumatic brain injury trials: 25% reduction in sample size requirements in the IMPACT study.** *J Neurotrauma* 2006: 1295–303
6. Machado SG, Murray GD, Teasdale GM. **Evaluation of designs for clinical trials of neuroprotective agents in head injury. European Brain Injury Consortium**. *J Neurotrauma* 1999; 16: 1131–8
7. Narayan RK, Michel ME, Ansell C, et al. **Clinical trials in head injury.** *J Neurotrauma* 2002; 19: 503–557.
8. Marmarou A, Lu J, Butcher I, et al**. IMPACT database of traumatic brain injury: design and description.** *J Neurotrauma* 2007; 24: 239–50
9. Maas AI, Marmarou A, Murray GD, et al. **Prognosis and clinical trial design in traumatic brain injury: the IMPACT study.** *J Neurotrauma* 2007; 24: 232–8
10. Hukkelhoven CW, Steyerberg EW, Habbema JD, et al. **Predicting outcome after traumatic brain injury: development and validation of a prognostic score based on admission characteristics.** *J Neurotrauma* 2005; 22: 1025–39
11. Hukkelhoven CW, Rampen AJ, Maas AI, et al. **Some prognostic models for traumatic brain injury were not valid.** *J Clin Epidemiol* 2006; 59: 132–43
12. Edwards P, Arango M, Balica L, et al. **Final results MRC CRASH, a randomised placebo-controlled trial of intravenous corticosteroid in adults with head injury-outcomes at 6 months.** *Lancet* 2005; 365: 1957–59
13. Kompanje EJ, Maas AI, Hilhorst MT, et al. **Ethical considerations on consent procedures for emergency research in severe and moderate traumatic brain injury.** *Acta Neurochir. (Wien)* 2005; 147: 633–40
14. Slieker FJ, Kompanje EJ, Murray GD, Ohman J, Stocchetti N, Teasdale SG, Maas AI. **SAPHIR and Pharmos TBI investigators.** *Neurosurg* 2008; 62, 1321–28

Chapter 7

BASELINE CHARACTERISTICS AND STATISTICAL POWER IN RANDOMIZED CONTROLLED TRIALS: SELECTION, PROGNOSTIC TARGETING, OR COVARIATE ADJUSTMENT?

Bob Roozenbeek
Andrew IR Maas
Hester F Lingsma
Isabella Butcher
Juan Lu
Anthony Marmarou
Gillian S McHugh
James Weir
Gordon D Murray
Ewout W Steyerberg
on behalf of the IMPACT Study Group

Crit Care Med 2009; 37(10): 2683-90

Abstract

Objective: Heterogeneity of patients is a common problem in randomized controlled trials (RCTs) in various fields of clinical research. We aimed to investigate the potential benefits of different approaches for dealing with heterogeneity in a case study on traumatic brain injury (TBI).

Design and Setting: Statistical modeling studies in three surveys and six randomized controlled trials.

Patients: Individual patient data (n=8033) from the IMPACT database.

Interventions: We investigated the statistical power and efficiency of randomized controlled trials (RCTs) in relation to (1) selection according to baseline characteristics, (2) prognostic targeting (i.e., excluding those with a relatively extreme prognosis), and (3) covariate-adjusted analysis. Statistical power was expressed as the required sample size for obtaining 80% power and efficiency as the relative change in study duration, reflecting both gains in power and adverse effects on recruitment. Uniform and targeted treatment effects were simulated for 6-month unfavorable outcome.

Results: For a uniform treatment effect, selection resulted in a sample size reduction of 33% in the surveys and 5% in the RCTs, but decreased recruitment by 65% and 41%, respectively. Hence, the relative study duration was prolonged (surveys: +95%; RCTs: +60%). Prognostic targeting resulted in sample size reductions of 28% and 17%, and increased relative study duration by +5% in surveys and +11% in the RCTs. Covariate adjustment reduced sample sizes by 30% and 16%, respectively, and did not affect recruitment. For a targeted treatment effect, the sample size reductions by selection (surveys: 47%; RCTs: 20%) and prognostic targeting (surveys: 49%; RCTs: 41%) were larger and adverse effects on recruitment smaller.

Conclusions: The benefits of selection and prognostic targeting in terms of statistical power are reversed by adverse effects on recruitment. Covariate adjusted analysis in a broadly selected group of patients is advisable if a uniform treatment effect is assumed, since there is no decrease in recruitment.

Introduction

Randomized controlled trials (RCTs) have become the principal research tool to investigate medical treatments (1, 2). The way to properly design and analyze RCTs is still under discussion (3). A major methodologic challenge is how we should deal with the inherent heterogeneity of RCT study populations, e.g., patients with sepsis, stroke, or traumatic brain injury (TBI).

TBI populations, for example, are heterogeneous in terms of clinical severity and the pathophysiologic background of the injury. Consequently, a wide variation in baseline prognostic risk may exist among patients. Sample size calculations are generally based on a 50:50 chance of developing a favorable or unfavorable outcome, but statistical power (i.e., the ability to identify a treatment effect when it really exists) will be affected when study populations contain many patients with a more extreme prognosis (e.g., very good or very bad). Previous re- search showed that approximately 40% of the patients enrolled in a Phase III TBI trial are in relatively extreme risk groups (4).

Different solutions for dealing with heterogeneity by using baseline characteristics in both the design and analysis of RCTs in TBI have been described. These include the use of strict enrollment criteria (5), targeting recruitment to patients with an intermediate prognosis (prognostic targeting) (6), and the use of advanced statistical analyses such as covariate adjustment (7, 8).

In practice, clinical trials employ enrollment criteria to select patients most likely to benefit from the treatment under investigation, both from a mechanistic (pathophysiologic mechanism) and prognostic perspective. From a prognostic point of view, prerandomization selection by strict enrollment criteria (e.g., age, sex, disease severity) aims to exclude patients with a more extreme prognosis. This approach is more simple, but similar in nature to prognostic targeting which excludes patients with a relatively extreme prognosis by combining different baseline characteristics in a prognostic model. Prognostic targeting can reduce the required sample size by approximately 30% (6). The effects of selection of patients on the statistical power of TBI trials have, however, not been examined in detail. Furthermore, any expected benefits in terms of statistical power must be viewed in relation to adverse effects in terms of recruitment. The overall efficiencies, reflecting net effects of gains in statistical power and adverse effects on recruitment, have not been established.

An alternative approach of dealing with heterogeneity without adverse effects on recruitment is the use of covariate adjustment. Covariate adjustment provides more individualized effect estimates and improves the statistical power of an RCT. Previous studies on TBI trials demonstrated that covariate adjustment for seven strong predictors yields a reduction in sample size requirements of up to 25% (8). We aimed to investigate the potential benefit of the use of baseline characteristics by selection of patients, by prognostic targeting, and by covariate-adjusted analysis on the statistical power and efficiency of RCTs in TBI.

Methods

Study population

We used individual patient data from the IMPACT I database on TBI. This database contains data of patients with moderate (Glasgow Coma Scale 9–12) and severe (Glasgow Coma Scale ≤ 8) TBI from eight RCTs and three surveys conducted between 1984 and 1996 (table 7.1). IMPACT links researchers in Europe and the United States in a project addressing methodologic problems in the design and analysis of RCTs in TBI (9). Details of the studies and the data management in the IMPACT database have been reported in detail before (10). The primary outcome measurement was the Glasgow Outcome Scale at 6 months after injury. For our analysis, we selected patients more than 14 yrs old at the time of injury. The data from two small RCTs (HIT-I [n=351] and SKB [n=139]) were not used in our analysis. This manuscript was exempt from institutional review board approval.

Table 7.1. Overview of datasets of three surveys and six randomized controlled trials included in IMPACT I

Study Code	Study		Year of study	N*
Surveys:				
TCDB (14)	Traumatic Coma Data Bank		1984-1987	604
UK 4 (15)	UK 4 Center Study		1986-1988	791
EBIC (16)	EBIC Core Data Study		1995	822
Study Code	**Agent**	**Mechanism**	**Year of study**	**N***
Randomized Controlled Trials:				
HIT II (17)	Nimodipine	Calcium Mediated Damage	1989-1991	819
TIUS	Tirilazad Domestic Trial	Lipid peroxidation	1991-1994	1041
TINT (18)	Tirilazad International Trial	Lipid peroxidation	1991-1994	1118
PEGSOD (19)	PEG-Super Oxyde Dismutase International	Glutamate excitotoxicity	1994-1996	409
SLIN	Selfotel	Glutamate excitotoxicity	1995-1997	919
Saphir (20)	D-CPP-ene	Free radical damage	1993-1995	1510

* Numbers may differ from those presented in the original studies due to selection of adult patients only.

Selection

We used the original populations of the IMPACT database, with their study-specific enrollment criteria as a reference, to compare other approaches to prerandomization selection (10). These were labeled “original selections”. We extended the existing enrollment criteria with five additional criteria. These criteria were: time window between injury and admission to study hospital ≤ 8 hrs; age at injury ≤ 65 yrs; ≥ 1 reactive pupil; motor score > 1; Glasgow Coma Scale ≤ 8. We called these “selections by strict enrollment criteria”.

Prognostic targeting

Another approach to selection is the targeting of patients based on their prognostic risk. To define the prognostic risk of the patients in our study population, we used a prognostic model with the predictor's age (continuous variable), the motor component of the Glasgow Coma Scale (five categories, in which the categories "localizing" and "obeying commands" were combined), and pupillary reactivity (three categories: unreactive, one reactive, both reactive) (11). Subsequently, we selected patients based on the calculated risk for unfavorable outcome. We included patients with a risk for unfavorable outcome between 20% and 80%, and thereby excluded patients with an extreme (good or poor) prognosis. These selections were called "selections by prognostic targeting."

Covariate adjustment

Covariate adjustment is an established approach to deal with heterogeneity in clinical Phase III trials. In RCTs, covariate adjustment leads to adjusted estimates of treatment effects, in contrast to unadjusted estimates, for example, adjusting for age results in an estimated treatment effect for a patient of a given age, while unadjusted analysis results in an estimated treatment effect for a patient of average age in the same sample.
Unadjusted analysis may be expressed by the following equation, in which α indicates the intercept and β represents the regression coefficient for the treatment.

$$\text{log odds (unfavorable)} = \alpha + \beta \times \text{treatment}$$

The covariate-adjusted model included three predictors, i.e., age, motor score, and the pupillary reactivity, as well as the treatment variable, as shown in the following equation.

$$\text{log odds (unfavorable)} = \alpha + \beta_1 \times \text{treatment} + \beta_2 \times \text{age} + \beta_3 \times \text{motor} + \beta_4 \times \text{pupil}$$

Simulations

Because none of the studies in the IMPACT database demonstrated a significant treatment effect, we simulated a positive treatment effect as in previous work (8). The regression coefficient for treatment was -0.557, corresponding to an average absolute risk reduction of 10% in unfavorable outcome at 6 months. This was equal to an unadjusted odds ratio of 0.57 (6). The hypothetical treatment was randomly assigned to 50% of the patients. A new outcome variable was generated per study and per simulation, based on the comparison of a random uniform distribution (0 to 1) and the probability of unfavorable outcome.

We simulated two scenarios for the treatment effect. We first simulated a uniform treatment effect in which all included patients had a chance for beneficial effect of the as-

signed treatment. Next, we simulated a targeted treatment effect, based on the prognostic risk per patient for unfavorable outcome. In these simulations, only patients with a prognostic risk between 20% and 80% had a chance for beneficial effect of the hypothetical treatment. One thousand simulations were run using the original sample size of each study for both the uniform and the targeted simulated treatment effect. The software used for the simulations and the sample size calculations was S-Plus 6 (Insightful Inc., Seattle, WA, 2001) and Microsoft Excel 2000 (Microsoft Corporation, Redmond, WA, 1999).

Influence on statistical power and required sample size

To determine the influence on statistical power, we ran the simulations described above using both the unadjusted and the covariate adjusted model. We used three different data- sets for each model, that is, the original selections, the selections by strict enrollment criteria, and the selections by prognostic targeting. The influence on the statistical power was calculated using the mean of the Wald statistic of the treatment effect coefficient derived from the simulations. We expressed the gain in statistical power as the required sample size to obtain 80% power (8). The formula used is shown in the following equation.

required sample size = original sample size x [(mean of Wald statistic for reference model) / (mean of Wald statistic for adjusted model)]^2

Efficiency

The efficiency of different approaches was expressed in the relative study duration, reflecting the net result of benefits in terms of statistical power and adverse effects in terms of lower recruitment rates. The relative study duration was calculated by dividing the ratio of the required vs. original sample size by the relative recruitment rate: ($N_{required}/N_{original}$) / relative recruitment rate. For example, when strict selection of a certain study population (n=800) resulted in a 50% recruitment rate (i.e., 400/800 patients), but the required sample size with strict selection was 600 (25% reduction in required sample size), the relative study duration would be (600/800)/0.5 = 1.50, or +50%.

Results

We analyzed nine studies with 8033 patients: three surveys with 2217 patients, and six RCTs with 5816 patients. The distribution of baseline predictors varied considerably within and between the studies (table 7.2). For example, the median age varied from 26 yrs in the TCDB survey to 38 yrs in the EBIC survey, and the number of patients with unresponsive pupils varied from 32% in the surveys to 18% in the RCTs. We also observed variation with respect to unfavorable outcome at 6 months: 60% in the surveys, compared to 43% in the RCTs.

Table 7.2. Distribution of three baseline predictors and the outcome distribution across three RCTs and six surveys of the IMPACT database

	Age	Pupillary reactivity	Motor score	Unfavorable outcome
	median (25^{th}-75^{th} percentile)	*both responsive one responsive both unresponsive*	*none / extension abnormal flexion normal flexion localizes / obeys untestable / missing*	
TCDB **(n=604)**	26 (21 – 40)	300 (50%) 55 (9%) 249 (41%)	243 (40%) 74 (12%) 122 (20%) 134 (22%) 31 (5%)	393 (65%)
UK4 (n=791)	36 (22 - 55)	429 (54%) 115 (15%) 247 (31%)	198 (25%) 37 (5%) 141 (18%) 221 (28%) 194 (25%)	518 (66%)
EBIC (n=822)	38 (24 - 59)	527 (64%) 87 (11%) 208 (25%)	230 (28%) 55 (7%) 113 (14%) 281 (34%) 143 (17%)	422 (51%)
Total surveys (n=2217)	32 (22 - 53)	1256 (57%) 257 (12%) 704 (32%)	671 (30%) 166 (8%) 376 (17%) 636 (29%) 368 (16%)	1333 (60%)
HIT II (n=819)	33 (22 - 49)	583 (71%) 101 (12%) 135 (16%)	280 (34%) 92 (11%) 181 (22%) 207 (25%) 59 (7%)	328 (40%)
TIUS (n=1041)	30 (23 - 41)	709 (68%) 122 (12%) 210 (20%)	152 (15%) 132 (13%) 300 (29%) 457 (44%) 0 (0%)	395 (43%)
TINT (n=1118)	30 (21 - 45)	813 (73%) 170 (15%) 135 (12%)	141 (13%) 237 (21%) 327 (29%) 413 (37%) 0 (0%)	456 (41%)
PEGSOD (n=1510)	27 (20 - 38)	779 (52%) 160 (11%) 571 (38%)	655 (43%) 165 (11%) 334 (22%) 356 (24%) 0 (0%)	774 (51%)

Table 7.2. *(Continued)*

	Age	Pupillary reactivity	Motor score	Unfavorable outcome
SLIN (n=409)	28 (21 - 43)	316 (78%) 79 (19%) 14 (3%)	55 (13%) 91 (22%) 127 (31%) 136 (33%) 0 (0%)	177 (43%)
Saphir (n=919)	32 (20 - 38)	612 (67%) 307 (33%) 0 (0%)	264 (29%) 143 (16%) 223 (24%) 286 (31%) 3 (0%)	378 (41%)
Total RCTs (n=5816)	30 (21 - 43)	3812 (66%) 939 (16%) 1065 (18%)	1547 (27%) 860 (15%) 1492 (26%) 1855 (32%) 62 (1%)	2508 (43%)

These differences related to the variation in the original enrollment criteria of the investigated studies, with less strict enrollment criteria for the surveys. The maximum time window varied from 4 hrs (SLIN, TIUS) to 72 hrs (UK4). In the RCTs, the minimum age for inclusion was defined as 14 (TINT, TIUS), 15 (PEGSOD, SAPHIR) or 16 (HIT II, SLIN) and maximum age as 65 (SLIN, TINT, TIUS) or 70 (HIT II, PEG- SOD en SAPHIR). Most studies did not use any enrollment criterion on pupillary reactivity, except SLIN ("not both fixed and dilated") and SAPHIR ("at least one reactive pupil"). A minimum motor score of 2 was used in SLIN, TINT, and TIUS. A maximum motor score of 5 was defined in HIT II and SAPHIR. Restrictions to the Glasgow Coma Scale on admission were set as 8 (TCDB, UK4, PEGSOD, and SAPHIR), 4 to 8 (SLIN), 4 to 12 (TINT, TIUS), or $\leq$ 12 (EBIC).

Selection

Following application of strict enrollment criteria, the simulation studies showed a reduction in required sample size of 33% in the surveys and 5% in the RCTs (table 7.3) under the assumption of a uniform treatment effect. Benefits were greater in the more heterogeneous studies. Detailed results are presented in figures 7.1 and 7.2. The application of strict enrollment criteria, however, also induced a substantial reduction in potential recruitment (65% in surveys, 41% in RCTs, table 7.3). Therefore, strict selection would prolong the study duration in both the surveys and the RCTs (95% and 60%, respectively, table 7.3). If the treatment were to be effective only in those with risks between 20% and 80% ("targeted treatment"), the reductions in required sample sizes would be larger (surveys: 47%; RCTs: 20%), and the prolongation of the relative study duration less pro- nounced, but still substantial (surveys: +54%; RCTs: +34%).

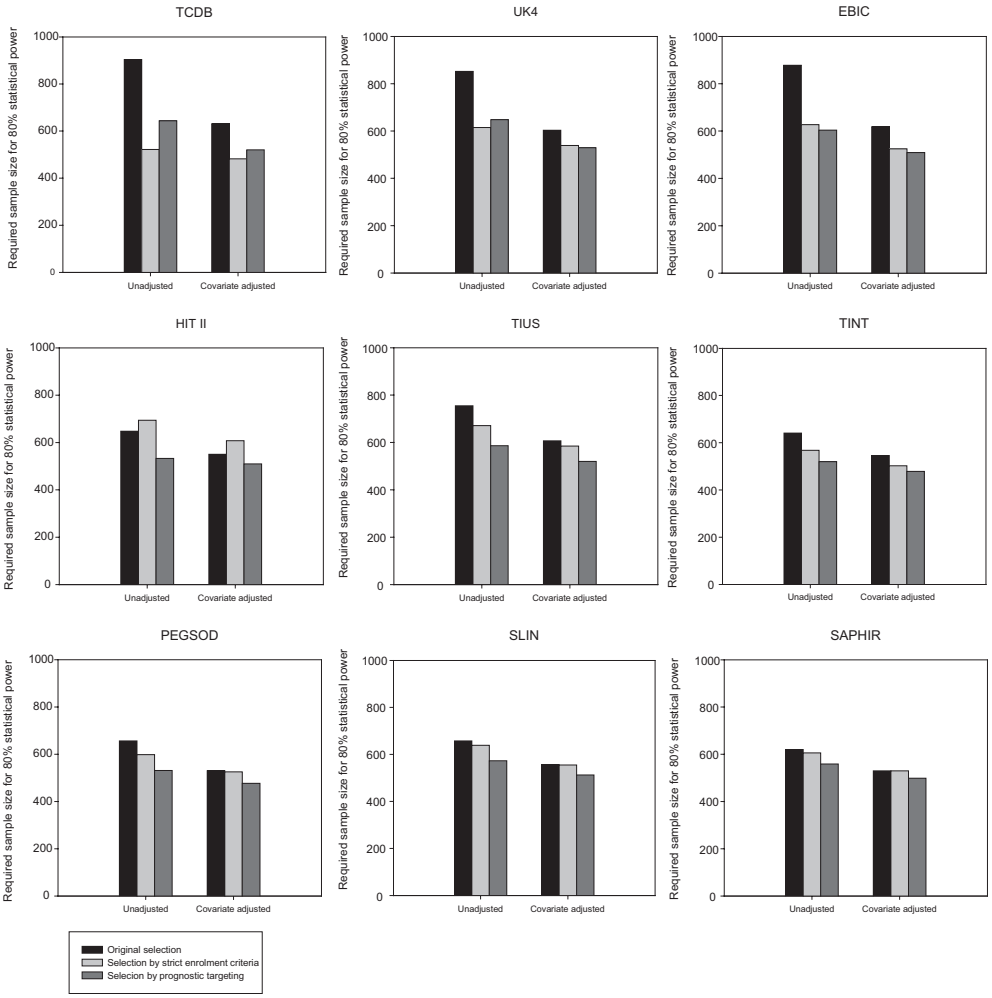


Figure 7.1. Required sample sizes for 80% statistical power in three surveys (TCDB, UK4 and EBIC) and six RCTs (others) of the IMPACT database. Required sample size was calculated using both the unadjusted and the covariate adjusted model. For both models, the required sample size was depicted for the original study population, the strictly selected population (time window ≤ 8 hours; age at injury ≤ 65 years; ≥ 1 reactive pupil; motor score > 1; GCS ≤ 8) and the prognostic targeted population (20%-80% risk for unfavorable outcome). Estimates based on 1000 simulations using a uniform treatment effect.

Prognostic targeting

The exclusion of patients with a relatively extreme prognosis ("prognostic targeting") reduced the required sample sizes by 28% in the surveys and by 17% in the RCTs, with a minor increase in relative study duration (+5% and +11% for surveys and RCTs, respectively, table

7.3; figures 7.1 and 7.2). With a targeted treatment effect, prognostic targeting excludes patients who cause only noise in the analysis, since they do not benefit from treatment. The reductions in required sample sizes were larger than for a uniform treatment effect (49% vs. 28% for surveys; 41% vs. 17% for RCTs). Study duration would be diminished by 26% in the surveys and 22% in the RCTs (table 7.3).

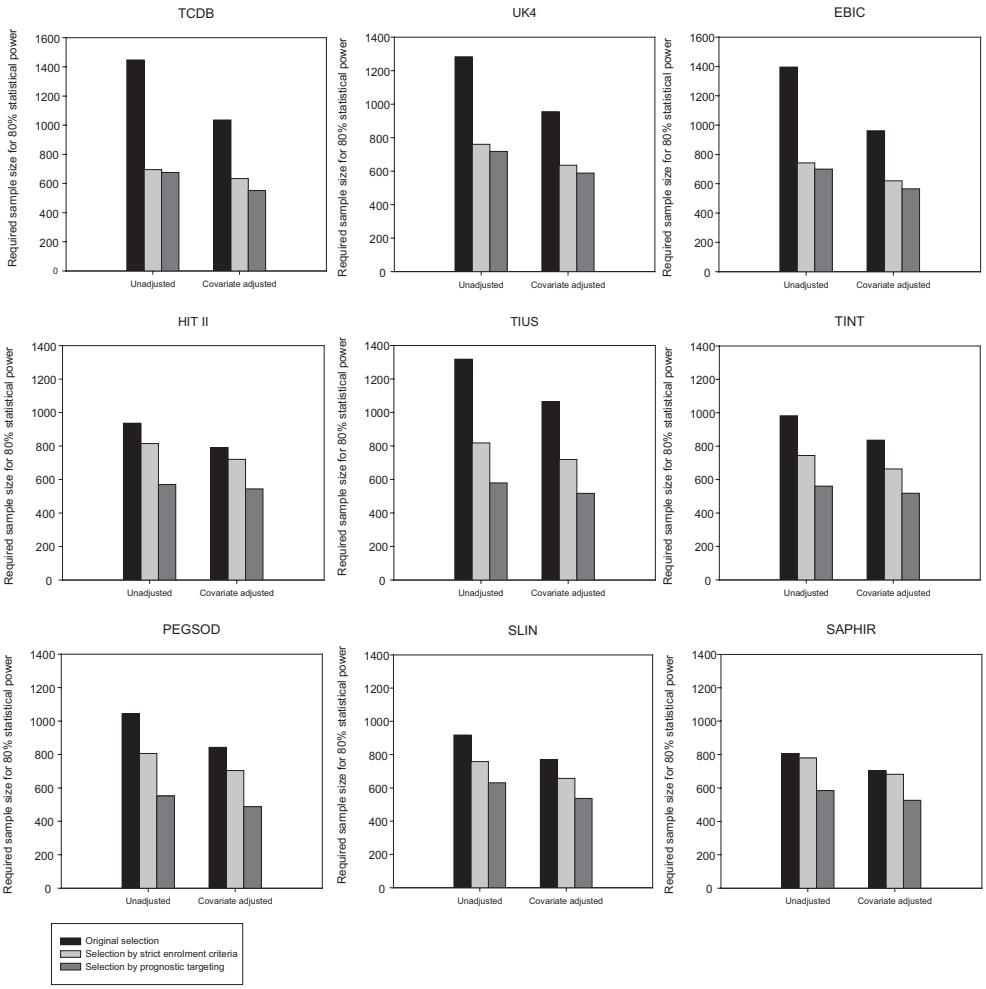


Figure 7.2. Required sample sizes for 80% statistical power in three surveys (TCDB, UK4 and EBIC) and six RCTs (others) of the IMPACT database. Required sample size was calculated using both the unadjusted and the covariate adjusted model. For both models, the required sample size was depicted for the original study population, the strictly selected population (time window ≤ 8 hours; age at injury ≤ 65 years; ≥ 1 reactive pupil; motor score > 1; GCS 8) and the prognostic targeted population (20%-80% risk for unfavorable outcome). Estimates based on 1000 simulations using a targeted treatment effect.

Covariate adjustment

Covariate adjustment was beneficial for the statistical power in all selections (table 7.3). In the original selections, reductions were 30% in the surveys and 16% in the RCTs. Since covariate adjustment had no adverse effects in terms of recruitment, the reduction of required sample size equaled the reduction in the relative study duration. The combination of covariate adjustment and prognostic targeting resulted in a large reduction of required sample size (41% in surveys; 24% in RCTs), but had substantial negative effects on recruitment. In terms of efficiency, this combination of strategies diminished the relative study duration less than when covariate adjustment was applied exclusively, i.e., by 14% in the surveys and by 0% in the RCTs. For a targeted treatment effect, maximum reductions in required sample size were found by combining prognostic targeted selection with covariate adjusted analysis: 59% in the surveys and 47% in the RCTs. The relative study duration would be reduced by 39% in the surveys and by 29% in the RCTs.

Discussion

We compared the statistical power and efficiency of three strategies to deal with heterogeneity in clinical trial populations: strict selection based on enrollment criteria; prognostic targeting (exclusion based on extreme risk); and covariate-adjusted analysis (include all patients, adjust for baseline characteristics in logistic regression analysis). Both use of selection strategies and covariate adjusted analysis led to potential reductions in required sample size or, equivalently, to increases in statistical power. The largest increase in power was found when prognostic targeting was combined with covariate-adjusted analysis, with a reduction of required sample size of up to 41% in surveys and 24% in RCTs. The benefits of prognostic targeting relate to the fact that patients with relatively extreme risks contribute little to the statistical analysis. In terms of efficiency, selection was less efficient than covariate adjustment since study duration would be prolonged.

However, these observations were based on the assumption that the investigated treatment has an effect in every single patient, regardless of the possible risk profile (a uniform treatment effect). It may be questionable if one can expect a treatment effect in patients with extreme risk profiles. Some may anticipate that patients with a very good prognosis will survive well anyway, whereas patients with a very poor prognoses cannot be saved. We therefore also performed simulation studies with a targeted treatment effect, i.e., a beneficial effect in patients with an intermediate prognosis (20% to 80%). These simulations showed that beneficial effects of all three approaches were even larger than with the simulation of a uniform treatment effect. Under this assumption, it is obviously the best strategy to perform prognostic targeting, since patients with an extreme prognosis only contribute noise, combined with covariate adjusted analysis for further power increase.

Table 7.3. Summary of the influences of selection, prognostic targeting and covariate adjustment on required sample size, recruitment rate, and relative study duration in three observational studies and six RCTs of the IMPACT database

	Uniform treatment effect						**Targeted treatment effect**					
	Surveys			***RCTs***			***Surveys***			***RCTs***		
	Req. N	Recruit-ment rate	Relative study duration	Req. N	Recruit-ment rate	Relative study duration	Req. N	Recruitment rate	Relative study duration	Req. N	Recruit-ment rate	Relative study duration
Original selection	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
+ Covariate Adjustment	-30%	0%	-30%	-16%	0%	-16%	-28%	0%	-28%	-16%	0%	-16%
Selection by strict enrolment criteria	-33%	-65%	+95%	-5%	-41%	+60%	-47%	-65%	+54%	-20%	-41%	+34%
+ Covariate Adjustment	-41%	-65%	+70%	-17%	-41%	+40%	-54%	-65%	+32%	-30%	-41%	+18%
Selection by prognostic targeting	-28%	-31%	+5%	-17%	-16%	+11%	-49%	-31%	-26%	-41%	-16%	-22%
+ Covariate Adjustment	-41%	-31%	-14%	-24%	-16%	0%	-59%	-31%	-39%	-47%	-16%	-29%

Estimates based on 1000 simulations, using both a uniform and a targeted treatment effect (effect only if risk 20 – 80%). A decrease in required N reflects gains in statistical power. A decrease in recruitment rate reflects adverse effects on recruitment, e.g. longer time is needed to enrol the same number of patients. The relative study duration expresses the combined effects on power and recruitment in terms of expected study duration.

However, this is the best option only when there is strong suspicion that the studied treatment truly does not work at all in patients with an extreme prognosis. This, however, is perhaps unrealistic from a clinical perspective and decreases generalizability. Similarly, we could exclude patients who are not likely to benefit from treatment based on other criteria, for example, selection on age if we think a treatment will work only in young patients. The simulations with a targeted treatment should hence be seen as proof of principle: if treatment works in a specific selection of patients, the ideal analysis includes only these patients.

The effect of selection in our study population was larger in surveys than in RCTs. This was expected since the RCTs already had more selected study populations. The same pattern was observed when we adjusted for covariates: using the unselected population, the effect in the surveys was larger than in the RCTs. After selection, the effect of covariate adjustment became smaller and more similar for surveys and RCTs. Equivalently, after covariate adjustment, the incremental benefit of selection also became smaller. In our approach to selection, we purposefully mimicked current trends to decrease heterogeneity by selection on time window and baseline predictors. We realize that these are two different aspects, the first being related to mechanistic aspects presuming that treatment may be more effective if given early, and the second related to prognostic aspects and aimed to decrease heterogeneity. The selection criteria we applied are those commonly used in focused Phase III RCTs in TBI.

Strict selection has substantial adverse effects on potential recruitment. This effect is such that beneficial effects in terms of statistical power are counteracted. The additional benefit of combining covariate adjustment with selection is small, while the adverse effects in terms of recruitment remain. In terms of efficiency, it is therefore more effective to perform covariate adjustment instead of using strict selection criteria. This can be motivated both from a scientific and from an economic perspective. Excluding fewer patients will permit enrollment of a broader range of the patient population, thus increasing generalizability of results. Clinical trials are, however, costly, certainly when conducted as Phase III trials with source data verification. Trial costs are determined both by the per-patient costs and the logistic costs of general trial coordination and data management. Thus, in many situations, it is to be expected that the longer a trial lasts, the more expensive it becomes. Furthermore, any delay in obtaining final results would have severe financial consequences if these results were positive, as the process of registration is postponed accordingly. Companies will, hence, wish to reduce the trial duration as much as possible, which supports the argument for a broad rather than a restricted population. An exception would be, however, if targeting a more restrictive population can be motivated from a mechanistic point of view, under the condition that occurrence of this mechanism can indeed be identified or at least considered highly probable.

An important issue with RCTs in general, and in TBI in particular, is that sample sizes are typically too small to detect modest but clinically relevant treatment effects. Sample sizes have commonly been motivated by statistical calculations based on overly optimistic assumptions, such as a 10% absolute improvement in outcome. In the present study, the gain in statistical power resulting from alternative strategies with respect to design and analysis was expressed as the corresponding reduction in sample size. This is intended simply as an easy to understand summary measure or relative efficiency. It is not intended that trialists actually reduce their planned sample size, although the reduction is theoretically well motivated. A further consideration is that the strength of covariate effects would need to be considered in the sample size calculation, which is adding another assumption to the sample size estimation procedure.

In this investigation, we studied only TBI populations. We did not explicitly consider other fields struggling with heterogeneity inpatient characteristics and underlying pathophysiology, but anticipate that results are applicable to other fields characterized by heterogeneity, such as stroke, sepsis, and acute respiratory distress syndrome. In previous work, benefits of covariate adjustment were shown in cardiovascular trials such as GUSTO-I (12). The use of covariate adjustment to increase the statistical power depends strongly on the strength of the association between the covariate and the outcome (3). In TBI research, prior investigations resulted in adequate determination of strong predictors of the outcome (4, 11, 13). These considerations make a strong case for increasing the focus on prognostic analyses.

We conclude that strict selection, prognostic targeting, and covariate adjustment all increase the statistical power in clinical trials in TBI. Covariate-adjusted analysis in a relatively unselected population results in less reduction of required sample size, but is more efficient than strict selection since it has no adverse effects on the study duration. Prognostic targeting, combined with covariate-adjusted analysis, results in a maximal decrease of required sample size, but is inefficient because of substantial reductions of potential recruitment, unless there is truly no treatment effect in the excluded patients. We therefore advise a relatively broad patient selection combined with prespecified covariate-adjusted analysis, unless we anticipate a targeted subgroup specific treatment effect without effect in excluded patients.

References

1. Yusuf S: **Randomised controlled trials in cardiovascular medicine: Past achievements, future challenges.** *BMJ* 1999; 319:564 –568
2. Tunis SR, Stryer DB, Clancy CM: **Practical clinical trials: Increasing the value of clinical research for decision making in clinical and health policy.** *JAMA* 2003; 290:1624 –1632
3. Assmann SF, Pocock SJ, Enos LE, et al: **Subgroup analysis and other (mis)uses of baseline data in clinical trials.** *Lancet* 2000; 355: 1064 –1069
4. Hukkelhoven CW, Steyerberg EW, Habbema JD, et al: **Predicting outcome after traumatic brain injury: Development and validation of a prognostic score based on admission characteristics.** *J Neurotrauma* 2005; 22:1025–1039
5. Saatman KE, Duhaime AC, Bullock R, et al: **Classification of traumatic brain injury for targeted therapies.** *J Neurotrauma* 2008; 25: 719 –738
6. Machado SG, Murray GD, Teasdale GM. Evaluation of designs for clinical trials of **neuroprotective agents in head injury. European Brain Injury Consortium.** *J Neurotrauma* 1999; 16:1131–1138
7. Hernandez AV, Steyerberg EW, Habbema JD. **Covariate adjustment in randomized controlled trials with dichotomous outcomes increases statistical power and reduces sample size requirements.** *J Clin Epidemiol* 2004; 57:454 – 460
8. Hernandez AV, Steyerberg EW, Butcher I, et al: **Adjustment for strong predictors of outcome in traumatic brain injury trials: 25% reduction in sample size requirements in the IMPACT study.** *J Neurotrauma* 2006; 23:1295–1303
9. Maas AI, Marmarou A, Murray GD, et al: **Prognosis and clinical trial design in traumatic brain injury: The IMPACT study.** *J Neurotrauma* 2007; 24:232–238
10. Marmarou A, Lu J, Butcher I, et al: **IMPACT database of traumatic brain injury: design and description.** *J Neurotrauma* 2007; 24:239 –250
11. 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; discussion e165
12. Steyerberg EW, Bossuyt PM, Lee KL**: Clinical trials in acute myocardial infarction: Should we adjust for baseline characteristics?** *Am Heart J* 2000; 139:745–751
13. 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; 336:425–429
14. Foulkes AM, Eisenberg MH, Jane, AJ: **The Traumatic Coma Data Bank: Design, methods, and baseline characteristics.** *J Neurosurg* 1991; 75:s8 –s15
15. Murray LS, Teasdale GM, Murray GD, et al: **Head injuries in four British neurosurgical centres.** *Br J Neurosurg* 1999; 13:564 –569
16. Murray GD, Teasdale GM, Braakman R, et al: **The European Brain Injury Consortium survey of head injuries.** *Acta Neurochir (Wien)* 1999; 141:223–236
17. **A multicenter trial of the efficacy of nimodipine on outcome after severe head injury. The European Study Group on Nimodipine in Severe Head Injury.** *J Neurosurg* 1994; 80:797– 804
18. Marshall LF, Maas AI, Marshall SB, et al**: A multicenter trial on the efficacy of using tirilazad mesylate in cases of head injury.** *J Neurosurg* 1998; 89:519 –525
19. Young B, Runge JW, Waxman KS, et al: **Effects of pegorgotein on neurologic outcome of patients with severe head injury. A multicenter, randomized controlled trial.** *JAMA* 1996; 276:538–543
20. Morris GF, Bullock R, Marshall SB, et al**: Failure of the competitive N-methyl-D-aspartate antagonist Selfotel (CGS 19755) in the treatment of severe head injury: Results of two phase III clinical trials. The Selfotel Investigators.** *J Neurosurg* 1999; 91:737–743

Chapter 8

THE ADDED VALUE OF ORDINAL ANALYSIS IN CLINICAL TRIALS: AN EXAMPLE IN TRAUMATIC BRAIN INJURY

Bob Roozenbeek
Hester F Lingsma
Pablo Perel
Phil Edwards
Ian Roberts
Gordon D Murray
Andrew IR Maas
Ewout W Steyerberg
on behalf of the IMPACT Study Group and the CRASH Trial Collaborators

Crit Care 2011: 15(3); R127

Abstract

Introduction: In clinical trials, ordinal outcome measures are often dichotomized into two categories. In traumatic brain injury (TBI) the 5-point Glasgow outcome scale (GOS) is collapsed into unfavourable versus favourable outcome. Simulation studies have shown that exploiting the ordinal nature of the GOS increases chances of detecting treatment effects. The objective of this study is to quantify the benefits of ordinal analysis in the real-life situation of a large TBI trial.

Methods: We used data from the CRASH trial that investigated the efficacy of corticosteroids in TBI patients (n = 9,554). We applied two techniques for ordinal analysis: proportional odds analysis and the sliding dichotomy approach, where the GOS is dichotomized at different cut-offs according to baseline prognostic risk. These approaches were compared to dichotomous analysis. The information density in each analysis was indicated by a Wald statistic. All analyses were adjusted for baseline characteristics.

Results: Dichotomous analysis of the six-month GOS showed a non-significant treatment effect (OR = 1.09, 95% CI 0.98 to 1.21, P = 0.096). Ordinal analysis with proportional odds regression or sliding dichotomy showed highly statistically significant treatment effects (OR 1.15, 95% CI 1.06 to 1.25, P = 0.0007 and 1.19, 95% CI 1.08 to 1.30, P = 0.0002), with 2.05-fold and 2.56-fold higher information density compared to the dichotomous approach respectively.

Conclusions: Analysis of the CRASH trial data confirmed that ordinal analysis of outcome substantially increases statistical power. We expect these results to hold for other fields of critical care medicine that use ordinal outcome measures and recommend that future trials adopt ordinal analyses. This will permit detection of smaller treatment effects.

Introduction

Traumatic brain injury (TBI) is a major health and socio-economic problem throughout the world. Basic research has elucidated many of the pathophysiological mechanisms underpinning secondary damage and many neuroprotective agents have been developed to counteract these mechanisms. Since the 1980s, at least 33 randomized controlled phase III trials have been performed to investigate the effectiveness of new therapeutic interventions in TBI, but none has convincingly demonstrated benefit in the overall population (1). Heterogeneity of the population and limitations of the conventional statistical analysis of TBI trials contribute to this lack of success (2, 3). We recently published a set of recommendations for improving the design and analysis of future TBI trials [4]. These recommendations were mainly derived from simulation studies and include the use of relatively broad enrolment criteria, covariate adjustment and ordinal rather than dichotomous outcome analysis.

In most phase III TBI trials, the 5-point Glasgow Outcome Scale is used as the primary outcome measure, usually measured at six months after injury, and dichotomized as unfavourable (Dead, Vegetative or Severe Disability) versus favourable outcome (Moderate Disability or Good Recovery) (table 8.1). Similar approaches are often used in the analysis of trials conducted for other indications. For example, in stroke the modified Rankin scale, which is also an ordinal scale, consisting of six categories, is commonly collapsed into a binary scale. This dichotomous outcome is then analysed with a chi-squared test or with binary logistic regression. Simulation studies have demonstrated that ordinal outcome analysis in TBI trials can increase statistical power (5). These results have not yet been validated in empirical data. The aim of this study is to investigate whether the benefits of an ordinal analysis would be upheld on analysis of the largest trial in TBI ever, which did demonstrate a true (but negative) treatment effect.

Table 8.1. The Glasgow Outcome Scale and its traditional dichotomy in favourable versus unfavourable outcome.

Dead Vegetative State Severe Disability	Unfavourable
Moderate Disability Good Recovery	Favourable

Materials and methods

Data

We used the individual patient data of the MRC CRASH trial into which 10,008 patients were enrolled.

The CRASH trial (Corticosteroid Randomisation After Significant Head Injury) was an international, randomised, placebo-controlled trial designed to investigate the effect of early administration of methylprednisolone on the risk of death and disability after head injury. Full results have been reported (6, 7). Enrolment was stopped in May 2004, following demonstration of a higher 14-day mortality rate in the active treatment arm (21.1% versus 17.9% deaths; P = 0.0001). Outcome at six months was available for 9,554 patients. The current study was exempt from institutional review board approval.

Conventional dichotomous analysis

We first estimated the effect of the treatment on the six-month GOS, dichotomized as unfavourable versus favourable, with binary logistic regression. The treatment effect was adjusted for four baseline covariates: age, Glasgow Coma Scale (GCS), pupillary reactivity and presence of major extracranial injury. Age was handled as a continuous variable and GCS as a categorical variable (range 3 to 15). Pupillary reactivity was grouped into three categories: both pupils reactive, one reactive and none reactive to light. The presence of major extracranial injury was included as a binary variable, having a positive value when patients had an extracranial injury that required hospital admission on its own.

Subsequently, we used two approaches exploiting the ordinal nature of the GOS: a proportional odds logistic regression model and the sliding dichotomy approach.

Proportional odds logistic regression

A proportional odds logistic regression model was fitted with the GOS collapsed to a 4-point ordinal scale (Severe Disability and Vegetative State were taken together) as the outcome variable. The proportional odds model has the same structure as the binary logistic regression model, but uses an ordinal outcome variable with more than two possible categories. It estimates a common odds ratio over all possible cut-offs of the outcome scale. The common odds ratio is formally valid if the odds ratios for each cut-off are the same (the proportional odds assumption). We can, however, interpret the common odds ratio as a summary measure of treatment effect, even if the odds ratios differ by cut-off (8). The common odds ratio can also be interpreted as the average shift over the total ordinal outcome scale caused by the treatment under study (5, 9, 10).

Sliding dichotomy

The sliding dichotomy approach dichotomizes the GOS into a binary measure, but the point of dichotomy is tailored to each individual patient's baseline prognosis (11). For example, for a patient with an excellent prognosis only good recovery may be considered as a favourable outcome, whereas for a patient with a very poor prognosis, survival may be regarded as a favourable outcome. First, the baseline prognostic risk of each patient was estimated by calculating the probability of unfavourable outcome with a prediction model including the following variables: age, GCS, pupillary reactivity, and presence of major extracranial injury (12). Subsequently, patients were divided into three prognostic bands of equal size, that is, for the best, intermediate and worst prognosis. For each band a separate cut-off on the GOS was defined and a new outcome variable was generated. For example, in the best prognosis band we only considered Good Recovery as a favourable outcome. The effect of treatment on this newly constructed dichotomous outcome was then estimated with binary logistic regression, with stratification by prognostic band and adjustment for the four covariates mentioned above. The pooled sliding dichotomy odds ratio can be interpreted as the effect of treatment on outcomes being worse than expected (11).

Comparison of the different approaches

We calculated Wald statistics, based on the coefficients of the treatment effect and the corresponding standard error for each analysis. The ratio of the Wald statistics can be interpreted as the gain in information density and is, therefore, a suitable measure for the efficiency of the different approaches.

We adjusted the treatment effect for four baseline covariates in all analyses (age, GCS, pupillary reactivity, major extracranial injury) (12, 13). Missing data occurred for 509 patients on pupillary reactivity and 196 on the presence of extracranial injury. These missing covariates were imputed with a multiple imputation model. Statistical analyses were performed in *R* Statistical Software version 2.7.2 using the Design library (*R* Foundation for Statistical Computation, Vienna, Austria).

Results

The CRASH trial included 10,008 patients. We excluded 454 patients with missing six-month GOS score, leaving 9,554 for the analyses. Median age was 33 years, and 81% of the patients were male (table 8.2). At six months after injury, 2,323 (24%) patients had died and 3,557 (37%) had an unfavourable outcome (figure 8.1). Dichotomous analysis of the six-month GOS showed a non-significant adjusted odds ratio (OR) of 1.09 (95% CI 0.98 to 1.21, P = 0.096).

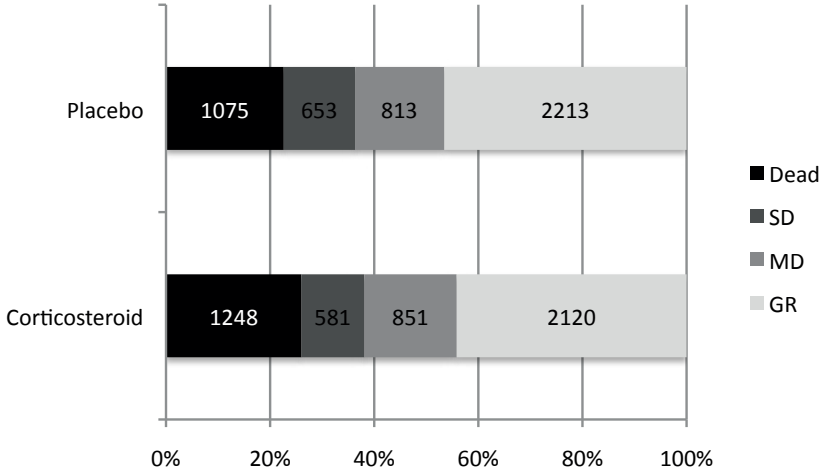


Figure 8.1. Distribution of the Glasgow Outcome Score at six months after injury.
Data from the CRASH trial (n = 9,554). SD, severe disability (including vegetative state); MD, moderate disability; GR, good recovery

Table 8.2. Baseline characteristics of patients enrolled in the CRASH trial with Glasgow Outcome Scale score available.

	Corticosteroid (n = 4,800)	**Placebo (n = 4,754)**
Age (median, IQR)	33, 23 to 47	32, 23 to 48
Gender		
Male	3,892 (81.1%)	3,824 (80.4%)
Glasgow Coma Scale		
Severe (3 to 8)	1,925 (40.1%)	1,890 (39.8%)
Moderate (9 to 12)	1,477 (30.8%)	1,405 (29.6%)
Mild (13 to 14)	1,398 (29.1%)	1,459 (30.7%)
Pupillary reactivity		
Both reactive to light	3,860 (80.4%)	3,822 (80.4%)
One reactive to light	270 (5.6%)	294 (6.2%)
Both not reactive to light	412 (8.6%)	387 (8.1%)
Missing	258 (5.4%)	251 (5.3%)
Major extracranial injury		
Yes	1,106 (23.0%)	1,039 (21.9%)
No	3,600 (75.0%)	3,613 (76.0%)
Missing	94 (2.0%)	102 (2.1%)

IQR: interquartile range.

The use of different splits than the conventional favourable vs. unfavourable outcome resulted in rather different estimates of the treatment effect (table 8.3). Further, the estimated treatment effect was non-significant when the conventional dichotomy was used, while it was significant when the split was taken at less than Good Recovery vs. Good Recovery (OR 1.12, 95% CI 1.01 to 1.23, P = 0.024) and death vs. survival (OR 1.27, 95% CI 1.13 to 1.43, P <0.0001). Application of the proportional odds logistic regression model gave an estimated common odds ratio of 1.15 (95% CI 1.06 to 1.25) with a P-value of 0.0007.

Table 8.3. Analysis of the treatment effect according to different dichotomizations and proportional odds logistic regression.

	Adjusted odds ratio^ (95% CI)	Wald statistic	*P*-value
Dichotomous odds ratios			
Less than good vs. good recovery	1.12 (1.01 to 1.23)	2.26	0.024
Unfavourable vs. favourable outcome	1.09 (0.98 to 1.21)	1.66	0.096
Death vs. survival	1.27 (1.13 to 1.43)	4.16	<0.0001
Common odds ratio (proportional odds model)	1.15 (1.06 to 1.25)	3.41	0.0007

Analyses are based on the six-month Glasgow Outcome Scale. Data from the CRASH trial (n=9,554).
An odds ratio >1 indicates an adverse effect of corticosteroids.
^ Adjustment for age, GCS, pupillary reactivity and major extracranial injury

With the sliding dichotomy approach we divided the study population into three bands of equal numbers, based on the individual prognostic risk for unfavourable outcome of each patient (table 8.4). For each prognostic band a different split for the dichotomization was used (better versus worse than expected). In the 'best prognosis' band the split was taken at Good Recovery versus worse than Good Recovery, in the 'intermediate prognosis' band at Moderate Disability or better versus Severe Disability or worse, and in the 'worst prognosis' band between death and survival. An unadjusted odds ratio was calculated for each prognostic band. These odds ratios varied between 1.06 (95% CI 0.91 to 1.23, P = 0.45) for the 'intermediate prognosis' band and 1.28 (95% CI 1.11 to 1.47, P = 0.0006) for the 'worst prognosis' band. Unadjusted and adjusted pooled odds ratios were similar (1.17, 95% CI 1.07 to 1.27, P = 0.0003 and 1.19, 95% CI 1.08 to 1.30, P = 0.0002).

The logistic regression analysis with dichotomized GOS resulted in a Wald statistic for the treatment effect of 1.66 (P = 0.096). Ordinal analysis with a proportional odds model gave a 2.05-fold higher Wald statistic (3.41, P = 0.0007). The sliding dichotomy approach resulted in an even larger Wald statistic of 3.69 (P = 0.0002), indicating a 2.56-fold increase in information density.

Table 8.4. Analysis of the Glasgow Outcome Scale with the sliding dichotomy approach.

		Dead	SD	MD	GR	Worse than expected	Better than expected	Odds ratio (95% CI)	Wald statistic	*P*-value
Best prognosis	Cortico-steroid	67	86	274	1,162	427	1,162	1.22 (1.03 to 1.43)		0.017
	Placebo	59	84	228	1,227	371	1,227			
Intermediate prognosis	Cortico-steroid	282	215	365	748	497	1,113	1.06 (0.91 to 1.23)		0.45
	Placebo	225	241	357	749	466	1106			
Worst prognosis	Cortico-steroid	899	280	212	210	899	702	1.28 (1.11 to 1.47)		0.0006
	Placebo	791	328	228	237	791	793			
Pooled odds ratio, unadjusted								1.17 (1.07 to 1.27)	3.67	0.0003
Pooled odds ratio, adjusted^								1.19 (1.08 to 1.30)	3.69	0.0002

The prognosis bands were created with model containing the variables age, GCS, pupillary reactivity and major extracranial injury. Odds ratios were given by prognosis band, for the unadjusted treatment effect. Pooled odds ratios were given for the unadjusted and adjusted treatment effect. An odds ratio >1 indicates an adverse effect of corticosteroids. Patients with better outcome than expected are underlined. Data from the CRASH trial (n = 9,554).

^ Adjustment for age; GCS, pupillary reactivity and major extracranial injury.

GCS, Glasgow Coma Scale; SD, Severe Disability (including Vegetative State); MD, Moderate Disability; GR, Good Recovery; OR, odds ratio.

Discussion

Analysis of the MRC CRASH trial data showed that ordinal analysis of the GOS resulted in substantially greater statistical power to detect a treatment effect with equal sample size. Whilst results obtained with the conventional analysis of the dichotomized GOS were non-significant, those obtained with ordinal analysis were highly significant. With ordinal analysis, a 2- to 2.5-fold gain in information density was demonstrated, compared to the dichotomized analysis. Simulation studies had already suggested the potential for ordinal

analysis to increase statistical power in TBI trials, but our current study has proven the value of this approach in the empirical data of a large trial with a true treatment effect.

Earlier research has demonstrated that adjustment for strong predictors of outcome (covariate adjustment) may result in a substantial increase in statistical power and trial efficiency (13-15). In the IMPACT database, we found that the required sample size for a RCT could potentially be reduced by around 25% when covariate adjustment would be applied with seven strong predictors (13). We, therefore, incorporated covariate adjustment in all analyses in the present study.

Why is the use of ordinal outcome analysis beneficial? The common practice of collapsing an ordinal outcome measure to a binary scale results in a loss of information (16). Moreover, dichotomization gives priority to one particular transition in the outcome scale: in the case of the GOS this is the change from severe disability to moderate disability. Patients with a relatively extreme prognosis have little potential to contribute to the detection of a treatment effect on an ordinal functional outcome scale, when this scale is dichotomized for the analysis (17). A patient with a very good prognosis will almost inevitably have a favourable outcome, even without the benefits of a new effective therapy. In contrast, for patients with a very poor prognosis it is extremely unlikely to have a favourable outcome at six months, even with a very beneficial new treatment. This does not mean that these patients may not benefit from the treatment, but simply that the fixed split for dichotomising the outcome measure is not appropriate for these situations. When the outcome is analysed in an ordinal way, all patients can contribute to the detection of a treatment effect.

The idea of exploiting the ordinal nature of ordered outcome scales is far from a new concept in the statistical community (18). Nevertheless, this approach has not been applied to the analysis of clinical trials on a regular basis. The sliding dichotomy approach was recently applied for the primary efficacy in a number of trials: the PAIS trial in stroke (19), the STICH trial in spontaneous intracerebral hemorrhage (20), and the Pharmos trial in TBI (21). The proportional odds model was used in several neurological trials, for example, in the GAIN International trial (22) and the SAINT I trial (23).

Inherent to the proportional odds model is the proportional odds assumption, that is, that the treatment effect is constant across all cut-offs of the outcome scale. This assumption may partly be violated in empirical data. We, therefore, recommend reporting the odds ratios per cut-off if a common odds ratio is reported as the summary measure of the treatment effect. Indeed, we found that the odds ratios were not identical across all cut-offs for the GOS (table 8.2). Also, some variation was seen in the odds ratios across prognostic bands for the sliding dichotomy (table 8.3). The proportional odds assumption was formally

tested with the 'PROC LOGISTIC' test from the SAS software package (SAS Institute Inc., Cary, NC, USA) and was found to be violated. This was confirmed by a graphical test in *R* software (the 'residuals' function from the Design library) to test for parallelism. In a previous study we simulated a non-proportional treatment effect, that is, a treatment that only affected mortality and did not cause a shift for the other categories of the GOS. We found to our surprise that the statistical power of ordinal analyses (proportional odds or sliding dichotomy) remained higher than a dichotomous analysis at the 'correct' cut-off (mortality vs. survival) (11). This robust gain in statistical power is a clear advantage of ordinal analysis, even if one were to object to interpretation of a summary odds ratio when underlying assumptions are violated (8).

The choice between the two ordinal approaches involves primarily a value judgment. The sliding dichotomy approach and its explanation (the effect of treatment on outcomes being worse than expected) may be particularly appealing for clinicians, but it requires a (validated) prognostic model to identify each patient's baseline prognostic risk. The proportional odds method does not necessarily require such a model, but may not have a proper interpretation if effect estimates vary substantially by cut-off (a violation of the proportional odds assumption). A pragmatic approach is to focus more on the underlying concept of 'shift analysis', instead of emphasizing the formal assumptions of this model.

Both approaches to ordinal outcome analysis that were investigated in the present study resulted in substantial power increase. Therefore, we strongly recommend incorporating ordinal methods in the analysis of future trials when an ordered outcome measure is considered. We do not advocate that this power increase should motivate reduced sample sizes for future trials. Since most TBI trials that were published in the past decades have been underpowered (24), the power increase that results from ordinal analysis can thus be used to increase the chance of detecting smaller, but clinically relevant, treatment effects with the same sample size.

The use of ordinal outcome scales is not unique to TBI, but is common to many fields of clinical research. Equally common is the practice of dichotomising ordinal outcome measures. In the field of stroke research, the modified Rankin Scale and the Barthel Index are often used as primary efficacy endpoints – and are also dichotomized (25, 26). The Optimising Analysis of Stroke Trials (OAST) Collaboration has shown the benefit of ordinal analysis in the field of stroke (27). Other examples of ordinal outcome scales can be found in cardiology (for example, NYHA Functional Classification for heart failure), vascular surgery (for example, Rutherford Classification for peripheral artery disease) and pain management (for example, Visual Analogue Scale). The widespread use of ordinal outcome measures and the persisting

practice of collapsing these measures into a binary outcome indicate that our findings in this case study on TBI have much broader implications than for TBI alone. We consider our results directly relevant to clinical trials in other fields of medicine that use ordinal outcome measures, especially if outcomes occur over the full range of the scale.

References

1. Maas AI, Roozenbeek B, Manley GT: **Clinical trials in traumatic brain injury: past experience and current developments.** *Neurotherapeutics* 2010, 7:115-126.
2. Maas AI, Steyerberg EW, Murray GD, Bullock R, Baethmann A, Marshall LF, Teasdale GM: **Why have recent trials of neuroprotective agents in head injury failed to show convincing efficacy? A pragmatic analysis and theoretical considerations.** *Neurosurgery* 1999, 44:1286-1298.
3. Narayan RK, Michel ME, Ansell B, Baethmann A, Biegon A, Bracken MB, Bullock MR, Choi SC, Clifton GL, Contant CF, Coplin WM, Dietrich WD, Ghajar J, Grady SM, Grossman RG, Hall ED, Heetderks W, Hovda DA, Jallo J, Katz RL, Knoller N, Kochanek PM, Maas AI, Majde J, Marion DW, Marmarou A, Marshall LF, McIntosh TK, Miller E, Mohberg N *et al.*: **Clinical trials in head injury.** *J Neurotrauma* 2002, 19:503-557.
4. Maas AI, Steyerberg EW, Marmarou A, McHugh GS, Lingsma HF, Butcher I, Lu J, Weir J, Roozenbeek B, Murray GD: **IMPACT recommendations for improving the design and analysis of clinical trials in moderate to severe traumatic brain injury.** *Neurotherapeutics* 2010, 7:127-134.
5. McHugh GS, Butcher I, Steyerberg EW, Marmarou A, Lu J, Lingsma HF, Weir J, Maas AI, Murray GD: **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.
6. Roberts I, Yates D, Sandercock P, Farrell B, Wasserberg J, Lomas G, Cottingham R, Svoboda P, Brayley N, Mazairac G, Laloe V, Munoz-Sanchez A, Arango M, Hartzenberg B, Khamis H, Yutthakasemsunt S, Komolafe E, Olldashi F, Yadav Y, Murillo-Cabezas F, Shakur H, Edwards P, CRASH trial collaborators: **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, 364:1321-1328.
7. Edwards P, Arango M, Balica L, Cottingham R, El-Sayed H, Farrell B, Fernandes J, Gogichaisvili T, Golden N, Hartzenberg B, Husain M, Ulloa MI, Jerbi Z, Khamis H, Komolafe E, Laloe V, Lomas G, Ludwig S, Mazairac G, Munoz Sanchez Mde L, Nasi L, Olldashi F, Plunkett P, Roberts I, Sandercock P, Shakur H, Soler C, Stocker R, Svoboda P, Trenkler S *et al.*; CRASH trial collaborators: **Final results of MRC CRASH, a randomised placebo-controlled trial of intravenous corticosteroid in adults with head injury-outcomes at 6 months.** *Lancet* 2005, 365:1957-1959.
8. Senn S, Julious S: **Measurement in clinical trials: a neglected issue for statisticians?** *Stat Med* 2009, 28:3189-3209.
9. Saver JL: **Novel end point analytic techniques and interpreting shifts across the entire range of outcome scales in acute stroke trials.** *Stroke* 2007, 38:3055-3062.
10. Valenta Z, Pitha J, Poledne R: **Proportional odds logistic regression--effective means of dealing with limited uncertainty in dichotomizing clinical outcomes.** *Stat Med* 2006, 25:4227-4234.
11. Murray GD, Barer D, Choi S, Fernandes H, Gregson B, Lees KR, Maas AI, Marmarou A, Mendelow AD, Steyerberg EW, Taylor GS, Teasdale GM, Weir CJ: **Design and analysis of phase III trials with ordered outcome scales: the concept of the sliding dichotomy.** *J Neurotrauma* 2005, 22:511-517.
12. MRC CRASH Trial Collaborators, Perel P, Arango M, Clayton T, Edwards P, Komolafe E, Poccock S, Roberts I, Shakur H, Steyerberg E, Yutthakasemsunt S: **Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients.** *BMJ* 2008, 336:425-429.
13. Hernandez AV, Steyerberg EW, Butcher I, Mushkudiani N, Taylor GS, Murray GD, Marmarou A, Choi SC, Lu J, Habbema JD, Maas AI: **Adjustment for strong predictors of outcome in traumatic brain injury trials: 25% reduction in sample size requirements in the IMPACT study.** *J Neurotrauma* 2006, 23:1295-1303.
14. Roozenbeek B, Maas AI, Lingsma HF, Butcher I, Lu J, Marmarou A, McHugh GS, Weir J, Murray GD, Steyerberg EW, IMPACT Study Group: **Baseline characteristics and statistical power in randomized controlled trials: selection, prognostic targeting, or covariate adjustment?** *Crit Care Med* 2009, 37:2683-2690.
15. Steyerberg EW, Bossuyt PM, Lee KL: **Clinical trials in acute myocardial infarction: should we adjust for baseline characteristics?** *Am Heart J* 2000, 139:745-751.
16. Altman DG, Royston P: **The cost of dichotomising continuous variables.** *BMJ* 2006, 332:1080.
17. Machado SG, Murray GD, Teasdale GM: **Evaluation of designs for clinical trials of neuroprotective agents in head injury. European Brain Injury Consortium.** *J Neurotrauma* 1999, 16:1131-1138.
18. Shannon CE: **A mathematical theory of communication.** *Bell Syst Tech J* 1948, 27:379-423, 623-656.

19. den Hertog HM, van der Worp HB, van Gemert HM, Algra A, Kappelle LJ, van Gijn J, Koudstaal PJ, Dippel DW; PAIS Investigators: **The Paracetamol (Acetaminophen) In Stroke (PAIS) trial: a multicentre, randomised, placebo-controlled, phase III trial.** *Lancet Neurol* 2009, 8:434-440.

20. Mendelow AD, Gregson BA, Fernandes HM, Murray GD, Teasdale GM, Hope DT, Karimi A, Shaw MD, Barer DH; STICH investigators: **Early surgery versus initial conservative treatment in patients with spontaneous supratentorial intracerebral haematomas in the International Surgical Trial in Intracerebral Haemorrhage (STICH): a randomised trial.** *Lancet* 2005, 365:387-397.

21. Maas AI, Murray G, Henney H,3rd, Kassem N, Legrand V, Mangelus M, Muizelaar JP, Stocchetti N, Knoller N; Pharmos TBI investigators: **Efficacy and safety of dexanabinol in severe traumatic brain injury: results of a phase III randomised, placebo-controlled, clinical trial.** *Lancet Neurol* 2006, 5:38-45.

22. Lees KR, Asplund K, Carolei A, Davis SM, Diener HC, Kaste M, Orgogozo JM, Whitehead J: **Glycine antagonist (gavestinel) in neuroprotection (GAIN International) in patients with acute stroke: a randomised controlled trial. GAIN International Investigators.** *Lancet* 2000, 355:1949-1954.

23. Lees KR, Zivin JA, Ashwood T, Davalos A, Davis SM, Diener HC, Grotta J, Lyden P, Shuaib A, Hardemark HG, Wasiewski WW; Stroke-Acute Ischemic NXY Treatment (SAINT I) Trial Investigators: **NXY-059 for acute ischemic stroke.** *N Engl J Med* 2006, 354:588-600.

24. Aberegg SK, Richards DR, O'Brien JM: **Delta inflation: a bias in the design of randomized controlled trials in critical care medicine.** *Crit Care* 2010, 14:R77.

25. Rankin J: **Cerebral vascular accidents in patients over the age of 60. II. Prognosis.** *Scott Med J* 1957, 2:200-215.

26. Mahoney FI, Barthel DW: **Functional Evaluation: the Barthel Index.** *Md State Med J* 1965, 14:61-65.

27. Optimising Analysis of Stroke Trials (OAST) Collaboration, Bath PM, Gray LJ, Collier T, Pocock S, Carpenter J: **Can we improve the statistical analysis of stroke trials? Statistical reanalysis of functional outcomes in stroke trials.** *Stroke* 2007, 38:1911-1915.

Chapter 9

NEW CONSIDERATIONS IN THE DESIGN AND ANALYSIS OF CLINICAL TRIALS FOR TRAUMATIC BRAIN INJURY

Bob Roozenbeek
Hester F Lingsma
Andrew IR Maas

Clin Invest (Lond) 2012: 2(2); 153-162

Abstract

Randomized controlled trials (RCTs) in traumatic brain injury (TBI) pose several complicated methodological challenges, related to the heterogeneity of the population. Several strategies to deal with these challenges have been proposed. Recommendations presented by the International Mission on Prognosis and Clinical Trial design in TBI (IMPACT) study group include the use of relatively broad enrollment criteria, combined with covariate adjustment for strong predictors of outcome in the analysis phase, rather than the use of strict enrollment criteria. Further, an ordinal rather than a dichotomized analysis of the Glasgow Outcome Scale - the outcome measure in most TBI trials - will increase statistical power significantly. This review discusses the issue of heterogeneity in TBI trials and summarizes the value of the different innovative methods for the design and statistical analysis of RCTs in TBI. Future directions highlight the opportunities offered by alternative strategies, such as comparative effectiveness research, to investigate the clinical benefits of established and novel therapies in TBI.

Introduction

Traumatic brain injury

Worldwide, many millions of people suffer from traumatic brain injury (TBI) each year (1). Despite all efforts from researchers and clinicians to improve outcome after TBI, it remains a major cause of both death and permanent disability (2).

Rather than a single event, TBI is considered to be a continuous process initiated by the initial impact (primary injury) evolving over subsequent hours and days (secondary injury). Different pathophysiological mechanisms are triggered by the initial injury, which are responsible for further neuronal and glial cell death in the brain following the primary impact of TBI and may result in clinical deterioration of TBI patients. These mechanisms include excitotoxicity, ischemia, edema, oxidative damage, mitochondrial dysfunction, apoptosis and many others (3, 4). These are considered to be possible targets for therapeutic interventions, aiming to limit the disastrous consequences of secondary injury to the brain.

Randomized controlled trials (RCTs) are considered the gold standard for evaluating the efficacy of new treatments. The acute TBI guidelines, covering many aspects of in hospital treatment however, only contain three level 1 recommendations based on the results of clinical trials. Nevertheless, many multicenter Phase III clinical trials have been performed to test different pharmacological agents and therapeutic strategies affecting different pathophysiological mechanisms that are active in TBI. Neuroprotective agents investigated have targeted in particular calcium mediated damage, lipid peroxidation, glutamate excitotoxicity and edema. Mostly, these agents have targeted single mechanisms, but some were thought to target a number of different pathophysiologic mechanisms. Despite very promising results in strictly controlled in vitro and in vivo laboratory experiments, none of these clinical trials have convincingly proven clinical efficacy in the treatment of TBI patients (5, 6). Clearly, a major gap exists in translating experimental findings to the clinical situation (7), for which explanations can be found on both sides of this gap.

Preclinical studies

The classical pathway of the development of a new drug starts in the laboratory by defining molecular and cellular pathways that are active in TBI. The next step is to test efficacy of candidate drugs in the many available animal models of TBI (8). However, none of these models adequately represents the complex picture of TBI seen in patients. In the experimental animal models, type and degree of injury can be standardized, moreover pretreatment is possible, and study endpoints are often mechanistic.

In the clinical situation, wide variability exists in type of pathology and in the severity of injury. Furthermore, pretreatment is impossible and interventions within short therapeutic windows can be challenging. For optimal translation to the clinical situation experimental

studies should preferably be performed in more than one model, in more than one species and have effects on both mechanistic and behavioral endpoints. A workshop organized by the National Institute of Neurological Disorders and Stroke (NINDS) in May 2000 concluded that preclinical studies and early clinical (Phase II) trials have not always been performed with sufficient rigor. As e.g. stated by Dr. Hall in the proceedings of this workshop: "we simply have not done adequate therapeutic window studies in our animal models in most cases. Furthermore, when we do have such data, we tend to ignore it when we go to the clinic".

Differences may exist in onset and duration of pathophysiologic mechanism between lissencephalic animals (rats/mice) and gyrencephalic species like humans. Uncertainty will therefore remain when translating experimental time windows obtained in a rodent model to human pathophysiology. General requirements for initiating Phase II studies include:

- Mechanism demonstrated in animal models
- Drug/agent reverses damage in animal models
- Mechanism shown to be active in human TBIs
- Neuroprotective agent that passes the brain-barrier
- Safety/tolerability in humans with TBI
- Drug sensitive endpoints

Phase III clinical trials

In the past thirty years, more than twenty large multicenter phase III trials have been performed to investigate efficacy of novel neuroprotective agents in TBI (5, 6). However, almost none showed an overall significant treatment effect. Besides deficiencies in the preclinical workup, substantial limitations in study design have been revealed that have contributed to failures in clinical studies (3, 9, 10). The problems include inadequate sample sizes, insensitive outcome measures, inappropriate selection of the study populations, over-optimistic expectations of new therapies, and heterogeneity of TBI patient populations. To investigate possible solutions for these problems, the International Mission on Prognosis and Analysis of Clinical Trials in TBI (IMPACT) study group was initiated in 2003 as a collaborative venture supported by the U.S. National Institutes of Health (11). The study group included clinical, epidemiological and statistical investigators from Belgium, The Netherlands, the U.K. and the U.S. The IMPACT investigators were granted access to individual patient data of initially eight randomized control trials and three observational studies including a total of 9205 patients (12). During the continuation-funding period (2007-2011) the number of studies was expanded providing access to data of over 40,000 patients. Relevant variables from the individual studies were extracted and merged to form a culture medium for exploring concepts to improve the design of clinical trials in TBI. The focus was on methodological

approaches for dealing with the heterogeneity inherent to the TBI population. Identification of robust covariates and the development of prognostic models formed the cornerstone for explorations on how best to deal with the heterogeneity. These explorations involved extensive simulation studies, addressing different approaches to patient selection on enrollment, covariate adjustment and ordinal outcome analysis.

This review discusses the issue of heterogeneity, and summarizes the value of the different innovative methods for the design and statistical analysis of Phase III trials to investigate safety and efficacy of new neuroprotective drugs or other therapeutic interventions in patients with moderate to severe TBI. We will also address alternative strategies for investigating the clinical benefit of established and novel therapies.

Challenges posed by heterogeneity of the population to the design of trials in TBI

TBI is a heterogeneous disease with respect to cause, pathology, severity and prognosis (13). This heterogeneity poses methodological challenges to the design and analysis of clinical trials in TBI. The general perception of the problems related to heterogeneity of populations in clinical trials focuses mainly on the increased risk for imbalances between treatment groups. The problems incurred by heterogeneity, however, are much greater and relate to the methodology for efficacy assessment and sample size calculations.

Imbalances

Imbalances in the distribution of baseline characteristics may influence the results of a study. This is especially the case if the association between a baseline characteristic that is imbalanced and the outcome is strong. It should be noted that even in the absence of (severe) imbalances in individual parameters, their cumulative risk may result in a significant imbalance. It is therefore recommended to report the prognostic risk estimate for treatment and placebo groups separately. This risk can be calculated with existing prognostic models (14, 15).

In TBI subpopulations, with smaller sample sizes, risks for imbalances are even higher. Indeed, various TBI studies have reported imbalances in subgroups. For example in the international Tirilazad study there were imbalances between treatment groups in male patients whose computed tomography (CT) scan showed traumatic subarachnoid hemorrhage (16). These imbalances were related to the CT lesion type, the occurrence of pretreatment hypoxia or hypotension and the presence of subarachnoid hemorrhage on the pre-treatment CT scan favoring outcome in patients treated with placebo. Similar imbalances had also been noted

in the North American Tirilazad study with regard to the Glasgow Coma Scale (GCS), the pattern of brain injury as demonstrated by the pretreatment CT scan and the frequency of bilaterally unreactive pupils.

Heterogeneity and efficacy analysis

Ideally, efficacy analysis should include mechanistic endpoints to allow the detection of drug specific effects. The use of early mechanistic endpoints has advanced the field of e.g. oncology, HIV/aids research and cardiovascular medicine. Such measures are however not yet available for TBI. In the absence of early mechanistic endpoints, TBI investigators and regulatory authorities have both adopted the Glasgow Outcome Scale (GOS) or its extended version (GOSE) as the standard for primary efficacy analysis (17, 18). The GOS is an ordinal scale for functional outcome and consists of five categories. In the extended GOS the categories of Severe Disability, Moderate Disability and Good Recovery are each subdivided into a lower and upper category (table 9.1). In Phase III clinical trials in TBI, the GOS has traditionally been collapsed to a binary outcome of "unfavorable outcome" (Death, Vegetative State, Severe Disability) versus "favorable outcome" (Moderate Disability, Good Recovery), irrespective of the baseline prognostic risk. This prognostic risk can vary greatly, and setting an arbitrary threshold which patients must cross to demonstrate clinical improvement is not reflective of the clinical situation and in fact substantially reduces chances of showing benefit. A patient's prognosis can be so good that no matter how poor the treatment, they may be expected to achieve a favorable outcome, whilst conversely a patient's prognosis may be so poor that it would become very unlikely for even a highly effective intervention to improve that outcome to an extent that it would change from unfavorable to favorable. Such patients will not contribute to the efficacy analysis. Although intuitively perhaps attractive because it is so simple, the traditional approach to dichotomize the GOS is counterproductive and disregards potentially valuable information contained in the ordinal scale (19).

Heterogeneity and sample size calculation

Sample size calculations for TBI trials have commonly aimed to detect an improvement in favorable outcome by an absolute value of 10%, presuming that there will be an approximately 50/50 distribution of outcome in the patient population. These calculations show that approximately 800 patients will be required to detect the postulated treatment effect with a statistical power of 80%. These calculations however are potentially flawed: the assumption of a 50/50 outcome distribution underpinning these calculations does not pertain to the group level but rather to every individual patient. The required sample size is strongly dependent on the outcome distribution, and as this deviates more from the 50/50 level, required sample sizes increase exponentially (figure 9.1). Consequently when

patients are included with a more extreme prognosis sample size calculations need to be adjusted. Even in trials with relatively strict enrolment criteria up to approximately 40% of patients may have an extreme prognosis (low or high risk). The confounding effect of prognostic heterogeneity will inevitably reduce statistical power to an extent, that chances of demonstrating benefit are greatly reduced.

Table 9.1. Glasgow Outcome Scale and its extended version.

	GOS		GOSE
1	**Death**, mortality from any cause	1	**Death**
2	**Vegetative State**, unable to interact with environment, unresponsive	2	**Vegetative State**
3	**Severe Disability** Conscious but dependent	3	**Lower:** dependant on others for activities of daily live
		4	**Upper:** dependant on others for some activities
4	**Moderate Disability** Independent but disabled	5	**Lower:** unable to return to work or participate in social activities
		6	**Upper:** return to work at reduced capacity, reduced participation in social activities
5	**Good Recovery** Return to normal occupation and social activities, may have minor residual deficits	7	**Lower***:* minor social or mental deficits which do not impair normal functioning
		8	**Upper:** full recovery, no residual complaints or deficits

GOS: Glasgow Outcome Scale; GOSE: Glasgow Outcome Scale Extended

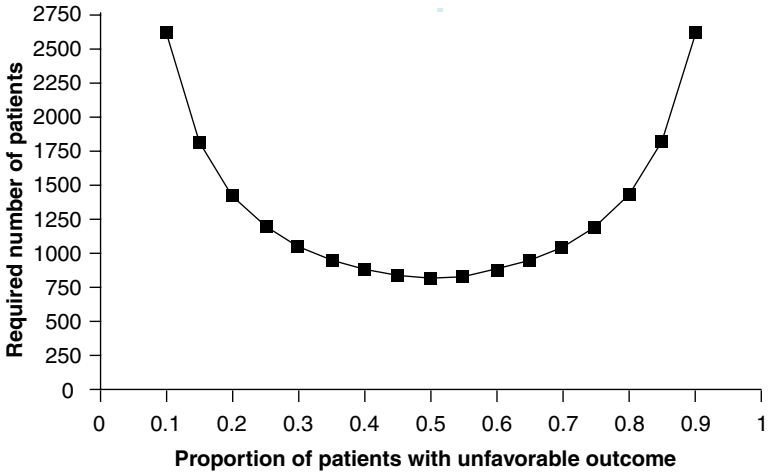


Figure 9.1. Required sample size in relation to outcome distribution. The X-axis displays the (expected) proportion of patients with unfavourable outcome in the study population. The Y-axis displays the required number of patients to detect an absolute difference in unfavourable outcome of 10% between treatment groups with a power of 80%.

Solutions to the challenges posed by heterogeneity of the TBI population

Reducing the risk of imbalances

Three approaches to avoid imbalances can be considered: decrease heterogeneity, increase sample size, or stratified randomization. Strict enrolment criteria have commonly been employed in TBI trials with an aim to decrease heterogeneity. The disadvantage of this approach however, is that the generalizability of findings is decreased and that it is inefficient as it leads to longer study duration (discussed below). Conversely, increasing the sample size and using less restrictive enrolment criteria (accepting the consequent heterogeneity) will increase generalizability and reduce the risk of occurrence of imbalances. In so called "mega trials", these risks are considered very low. The Corticosteroid Randomisation After Significant Head Injury (CRASH) trial was the first "mega trial" in the field of TBI [20; 21], and investigated the efficacy of methylprednisolone in patients with mild to severe TBI (1999-2004). The trial planned to enroll 20,000 patients, but after inclusion of 10,008 it was halted because an interim analysis showed a higher 14-day mortality rate in the active treatment arm (21.1% versus 17.9% deaths, p=0.0001). Liberal enrollment criteria were used to stimulate recruitment, reduce the risk of imbalances and to increase generalizability of results. The large heterogeneity was seen more as an asset than as a problem. Although the approach of selection of the study cohort is substantially different from previous clinical trials in the field of TBI, this trial was one of the first multicentre trials in TBI that demonstrated an overall statistically significant (but negative) treatment effect. Imbalances between treatment groups were not found. Stratifying the randomization on important prognostic factors avoids imbalances in the treatment groups, but has some practical disadvantages, like the need for more advanced (electronic) randomization schemes.

Dealing with prognostic heterogeneity

Conceptually, prognostic heterogeneity can be reduced by the use of stricter enrolment criteria (also reducing the risk of imbalance) or by the application of covariate adjustment in the analysis phase. Both approaches aim to increase statistical power of a trial.

Simulation studies within the IMPACT database have shown that stricter enrollment criteria can indeed increase statistical power by 33% in observational surveys and 5% in RCTs (22). However, a disadvantage of selection is reduction of a trial's recruitment rate (number of patients that can be recruited in a certain period of time) and therefore prolongation of the trial duration. Stricter enrollment criteria lead to a reduction in recruitment of approximately 65% in the observational studies and 41% in the trial populations included in the IMPACT database (23). These studies showed that the benefits of selection in terms of statistical power were outweighed by adverse effects on recruitment rate and trial duration and therefore make this approach inefficient (22) (table 9.2).

Table 9.2. Effects of selection and targeting on statistical power (required sample size) and recruitment.

	Required sample size (statistical power)		Recruitment	
	Surveys	*RCTs*	*Surveys*	*RCTs*
Original selection	ref	ref	ref	ref
Strict enrollment criteria	-33%	-5%	↓ 65%	↓ 41%
Prognostic targeting	-28%	-17%	↓ 45%	↓ 43%

Adapted from Roozenbeek et al., 2009 (22) and Roozenbeek et al., 2009 (23)

Alternatively, one may consider selecting patients not based on separate enrollment criteria, but on their summarized prognostic risk (i.e. targeting patients with an intermediate prognostic risk). This approach is now feasible following the development of validated prognostic models for predicting outcome in TBI (14, 15). The concept of prognostic targeting was previously proposed by Machado [24] and confirmed in the study by Roozenbeek et al (22) (table 9.2). Although the use of a baseline prognostic risk score would seem more efficient than the use of separate inclusion and exclusion criteria, and its benefit in terms of statistical power proven, the exclusion of patients will reduce recruitment and consequently lengthen study duration. The increase in relative study duration as summarized in table 9.3 with the use of both strict enrolment criteria and prognostic targeting render these approaches inefficient. Rather than attempting to decrease heterogeneity on enrollment, one may consider the application of covariate adjustment in the analysis phase. This will not affect recruitment or study duration. Covariate adjustment is a procedure to control for unbalanced prognostic factors in order to obtain less biased estimates of the treatment effect. Stratified randomization is often combined with covariate adjustment, but covariate adjustment does not require stratified randomization.

Simulation studies with TBI trial data showed that logistic regression analysis of the treatment effect, with covariate adjustment for seven strong predictors of outcome resulted in a 25% gain in statistical power, compared to the unadjusted analysis (25). These results were confirmed across the individual patient populations of the IMPACT database, demonstrating an increase in statistical power of up to 30% in more heterogeneous populations of observational surveys and up to 16% in trial populations with stricter criteria (22).

Table 9.3. Net effect of selection, targeting and covariate adjustment on study efficiency.

	Relative study duration	
	Surveys	*RCTs*
Original selection	Ref	Ref
Strict enrollment criteria	+ 95%	+ 60%
Prognostic targeting	+ 05%	+ 11%
Covariate adjustment	- 30%	- 16%

Adapted from Roozenbeek et al., 2009 (22)

Different approaches can be used to decide which covariates should be used for adjustment. One approach is the use of knowledge on predictive factors of outcome from previous studies. The prognostic analysis performed by the IMPACT study group on long-term mortality and functional outcome after TBI have confirmed the predictive value of multiple covariates, including age, the GCS motor score, pupillary reactivity, brain CT scan characteristics and second insults (hypoxia and hypotension) (23). The prognostic models based on these variables have now been extensively externally validated (14, 15, 45). TBI populations of many different settings and therefore can be recommended to be used for covariate adjustment in trials in moderate and severe TBI. Also for mild TBI prognostic models have been published, but not externally validated (26).

Another approach is to use the data of the trial to investigate which covariates may best be used. However, the use of this approach has the risk of subjective selection of variables. Simulation studies have shown that this approach indeed carries the risk for biased estimates of the treatment effect, especially in trials with small sample sizes (27). The use of a prespecified adjustment model, or as a minimum specification of the covariates in the statistical analysis plan is strongly recommended.

Covariate adjustment in the analysis phase does not carry the disadvantage of reducing recruitment. Consequently the IMPACT study group recommends a relatively broad inclusion criteria combined with covariate adjustment in the analysis phase. This approach is expected to yield a reduction of sample size of 20 to 25%, which is much greater than has previously been observed in for example cardiovascular trials (28). An additional advantage of broad inclusion criteria is that it will increase generalizability of the findings in the study to less selected TBI populations.

Ordinal approaches in the primary efficacy analysis

The traditional approach of dichotomizing the GOS has the disadvantages of discarding potentially valuable information [19] and that the point of dichotomization is not related to the initial prognostic risk. Approaches that would either take benefit of the full ordinal nature

of the GOS and/or relate the individuals' outcome to the initial baseline prognostic risk may be expected to be more efficient. In addition, clinically very relevant shifts in outcome, e.g. from moderate disability to good recovery, are taken into account in an ordinal approach. Two possible approaches for this are proportional odds analysis and the application of the sliding dichotomy. Superficially, the two approaches have much in common. They both exploit the ordinal nature of the GOS, but conceptually there are major differences.

In the proportional odds model the pooled odds ratios are based on the treatment effect calculated for each of the possible splits for collapsing an ordinal scale. In this way, the study sample is not subdivided but rather every patient contributes to the estimate of a so-called common odds ratio. Hence an overall estimate of the shift in outcome across the GOS is obtained for better perception in a clinical audience. The use of the proportional odds model has been referred to as a "shift analysis" (28). This approach was used in the efficacy analysis of several stroke trials (30, 31) and in one recently published TBI trial (32).

In the sliding dichotomy approach, the point of dichotomy of the GOS is differentiated according to the baseline prognostic risk. For patients with a poor prognosis, survival may be relevant whilst in those with a good prognosis any outcome worse than good recovery may be considered unfavorable. The concept of the sliding dichotomy approach is intuitively attractive for clinicians and has been adopted for the primary analysis of some phase III trials in TBI, stroke and intracerebral hemorrhage (32-34). In contrast to the traditional dichotomous analysis of the GOS, in which patients are required to cross a pre-specified, fixed and artificially determined boundary, the sliding dichotomy approach takes other transitions, for example that from Moderate Disability to Good Recovery into account. Use of the sliding dichotomy approach (and covariate adjustment) requires identification of robust prognostic models to reliably provide a baseline risk estimate in individual patients, which are available for moderate and severe TBI (14,15).

Extensive simulation studies have been performed in the IMPACT database to explore benefits of ordinal approaches to the outcome analysis (36). Both application of proportional odds analysis and of the sliding dichotomy yielded considerable benefits increasing statistical power by 23 to 30%. Applying covariate adjustment in addition to the ordinal analysis further increased power to a total of up to 40 to 49% (table 9.4). The benefits of this approach were further confirmed in new data sets and when applying both approaches to the data of the MRC CRASH study, the expected benefits were confirmed in the "real life" situation of a clinical trial (37). These findings strongly support adopting an ordinal approach to outcome analysis in TBI trials.

Table 9.4. Increasing trial efficiency by ordinal analysis and/or covariate adjustment.

Statistical Approach	Reduction in sample size (%)		
	Median	*Interquartile range*	*Range*
Conventional dichotomy	Reference	Reference	Reference
Conventional dichotomy + Covariate adjustment	26	20-29	14-29
Proportional odds analysis	23	19-24	18-24
Proportional odds + covariate adjustment	49	45-53	41-57
Sliding dichotomy	30	29-37	16-45
Sliding dichotomy + covariate adjustment	40	34-44	25-51

Adapted from McHugh et al., 2010 (36)

The choice between proportional odds analysis versus the sliding dichotomy involves more a value judgment than that any choice can be scientifically motivated. From a statistical perspective the proportional odds model appears more efficient, but is perhaps less appealing for a clinical audience. Thus, we consider both approaches appropriate for the analysis of TBI trials. Whichever ordinal approach is chosen, evidence strongly indicates that the conventional dichotomous analysis should be discarded from the trialists' toolbox, unless one is exclusively interested in one particular outcome.

Future perspective

The reappraisal of trial methodology in TBI and the recommendations proposed by the IMPACT study group have provided us with tools to conduct trials more efficiently in the field of TBI, providing better chances to detect treatment effects. We see three important directions for further improvements: more comprehensive approaches to outcome assessments, standardization of data collection/coding and alternative approaches to exploring efficacy.

More comprehensive approaches to outcome assessments

More than three decades after Jennett and Bond described the GOS, it is still the most commonly used outcome measure in TBI trials. However, by definition the GOS is as a global assessment and insufficiently recognizes the complex aspects of outcome following TBI. Furthermore, despite the use of the structured interview (38), misclassification of the GOS assessment can occur with adverse effects upon statistical power (39). We see a clear need to further develop a multidimensional approach to outcome assessment and classification (40),

including neuropsychological measure, and to include the patient's perspective on quality of life. We should further consider adding in a time dimension. In the field of health economics the use of quality adjusted life years (QALY) is common practice. A similar approach in which outcome assessments are measured at multiple time points and integrated with mortality may offer a new and more comprehensive way to approach outcome assessments in TBI. This would offer opportunities to additionally capture the speed of recovery.

A major limitation in the conduct of clinical trials in TBI is the striking lack of early mechanistic endpoints. Different surrogate, early mechanistic endpoints for TBI have been suggested in the past, such as intracranial pressure, therapy intensity level, jugular venous oxygen saturation, metabolic measurements and neuro-imaging characteristics [10]. However, none of these measures have proven to be directly related to functional outcome [3]. Consequently, in TBI it is impossible to target existing or novel therapeutic approaches appropriately. Mechanistic targeting - the ideal for clinical trials - would necessitate the identification of occurrence and time course of pathophysiologic mechanisms in individual patients. Advances in neuro-imaging and the emerging field of biomarkers offer hope for the future. With these techniques, disease processes can be tracked and patients could function as "their own controls". Thus, we may determine if the therapy under investigation might reduce incremental tissue injury – the first principle of neuroprotection.

Standardization of data collection

A high quality observational study combined with comparative effectiveness research as well as meta- analysis of individual patient data across existing studies requires standardization of data collection and coding. Initial steps towards development of proposals for standardization have been integrated into an internationally oriented interagency initiative towards: "an integrated approach to research in psychological health and traumatic brain injury". This initiative involving the National Institutes of Neurological Disorders and Stroke (NINDS), the National Institute on Disability and Rehabilitation Research (NIDRR), the Department of Veterans' Affairs (VA), the Defense and Veterans Brain Injury Center (DVBIC) and the Defense Centers of Excellence (DCoE) included four working groups in the field of TBI addressing four domains: 1. Demographics and clinical assessment, 2. Biomarkers, 3. Neuro-imaging and 4. Outcome. General recommendations have been published (34; 35; 36; 37; 41). The diversity and specific characteristics of the topics addressed by the working groups resulted in a different emphasis in the recommendations. In the biomarkers and imaging groups, emphasis was placed on standardization of techniques and procedures, whilst in the outcomes group, the main emphasis was on the selection of instruments. For demographics and clinical assessments the standardization of coding of the variables was most important. The process for developing recommendations for clinical data elements was consensus

driven with multidisciplinary and international input from experts across a broad range of disciplines. A general consensus on selection and coding of data elements was achieved and templates were produced that summarized coding formats, motivation of choices and recommendations for procedures. Some recommendations concerned novel approaches, for example towards assessing the intensity of therapy in severely injured patients. It was recognized that the selection and the required level of detail in data collection would vary greatly with the design and aim of a specific study. Three levels of detail for coding elements were developed: 1. a basic version, 2. an advanced version and 3. an extended format with the greatest level of detail in the extended version. The coding of these versions was such that in every case the more detailed coding could be collapsed into the basic version, thus facilitating comparisons across studies. The proposed data elements and their coding are intended as "building blocks" for designing a case report form and can be used as "plug in" elements and used multiple times in various sections of a case record form. The overall structure of the CDE recommendations for coding and the templates can be viewed on the IMPACT website and on the CDE website.

The process for standardization of data collection in TBI is considered crucial to advancing the care for TBI patients and has been well received in the field. Endorsements have been obtained from the AANS / CNS section on Neurotrauma and Critical care, the International and National Neurotrauma Societies and the European Brain Injury Consortium. It should be recognized however that the process of standardization is and will remain an ongoing process, for which continuous feedback, refinement and updating is required. This could best be overseen by a non-commercial, scientific authority such as the National Institutes of Health in the U.S.

Alternative approaches to exploring efficacy analysis

Randomized controlled trials are considered the gold standard for proving efficacy of new treatments. Classical clinical trials are characterized by a strong reductionist approach and suffer from some limitations. RCTs address efficacy rather than effectiveness. Efficacy reflects the extent to which an intervention provides benefit under carefully controlled conditions chosen to maximize the likelihood of observing an effect, whilst effectiveness relates more to benefits and harms of an intervention in ordinary settings and broader populations. RCTs may therefore not always address the effect of different clinical practices in "the real world". Moreover they are costly and we should recognize that we will never be able to execute adequately powered trials to address all the existing uncertainties in clinical TBI management. Alternative approaches to exploring efficacy should therefore be considered in addition to the gold standard of clinical trials. The existing heterogeneity in TBI populations, management approaches and outcome offers opportunities for exploring

causes for these differences and identifying underlying reasons for a given outcome, or individual patient response to a selected intervention. This represents comparative effectiveness research (CER). The concept of CER in TBI is not new. In 1983 Gelpke et al (42) analyzed differences and outcome between two centers from the Netherlands and in 1989 Colohan et al (43) performed comparative analysis of treatment results between Charlottesville (U.S.) and New Delhi (India). TBI may be considered particularly suited for application of CER for a number of reasons. First, the strong heterogeneity with large differences in both management and outcome between centers and countries (44) provide a major opportunity to compare alternative interventions and managements. Second, the availability of robust and validated risk adjustment models for both mortality and functional outcome permit approaches to adjust for prognostic heterogeneity. Third, sophisticated statistical approaches such as random effect analysis are available to support CER.

A recent workshop "Promoting effective traumatic brain injury research" held during the National Neurotrauma Symposium in 2010, organized by the EU and the NIH stated that improved clinical care of TBI patients will likely depend on a range of research approaches including CER and systems biology. An urgent need for standardization of treatments and validation of effective clinical guidelines in TBI was identified. It was proposed that these goals could best be achieved through a large observational study coupled to comparative effectiveness research. Subsequent to this workshop an "international initiative for TBI research" has been formalized by the EU and NIH/ NINDS.

Acknowledgements

The authors gratefully acknowledge the collaboration within the IMPACT study group from which this manuscript originates. The work was supported by the NIH/NINDS (NS-042691).

References

1. Faul M, Xu L, Wald MM, Coronado VG. **Traumatic brain injury in the United States: emergency department visits, hospitalizations, and deaths.** Atlanta (GA): Centers for Disease Control and Prevention, National Center for Injury Prevention and Control, 2010.
2. Maas AI, Stocchetti N, Bullock R. **Moderate and severe traumatic brain injury in adults.** *Lancet Neurol.* 2008; 7: 728-41.
3. Tolias CM, Bullock MR. **Critical appraisal of neuroprotection trials in head injury: what have we learned.** *NeuroRx* 2004*:* 1(1); 71-9.
4. Werner C, Engelhard K. **Pathophysiology of traumatic brain injury.** *Br J Anaesth* 2007; 99(1): 4-9.
5. Maas AI, Roozenbeek B, Manley GT. **Clinical Trials in Traumatic Brain Injury: Past Experience and Current Developments.** *Neurotherapeutics* 2010; 7: 115-126.
6. Maas AI, Steyerberg EW, Marmarou A, et al**. IMPACT Recommendations for Improving the Design and Analysis of Clinical Trials in Moderate to Severe Traumatic Brain Injury.** *Neurotherapeutics* 2010; 7: 127-134.
7. Janowitz T, Menon DK. **Exploring new routes for neuroprotective drug development in traumatic brain injury.** *Sci Transl Med* 2010; 14: 2(27).
8. Povlishock JT, Hayes RL, Michel ME, McIntosh TK. **Workshop on animal models for TBI.** *J Neurotrauma* 1994: 11(6); 723-32.
9. Maas AI, Steyerberg EW, Murray GD, et al. **Why have recent trials of neuroprotective agents in head injury failed to show convincing efficacy? A pragmatic analysis and theoretical considerations.** *Neurosurgery* 1999; 44(6): 1286-98.
10. Narayan RK, Michel ME, Ansell B, et al. **Clinical trials in head injury.** *J Neurotrauma* 2002; 19(5): 503-57.
11. Maas AI, Marmarou A, Murray GD, Teasdale SG, Steyerberg EW. **Prognosis and clinical trial design in traumatic brain injury: the IMPACT study.** *J Neurotrauma* 2007; 24(2): 232-8.
12. Marmarou, A, Lu J, Butcher I, et al. **IMPACT database of traumatic brain injury: design and description.** *J Neurotrauma* 2007; 24: 239-250.
13. Lingsma HF, Roozenbeek B, Steyerberg EW, Murray GD, Maas AI. **Early prognosis in traumatic brain injury: from prophecies to predictions.** *Lancet Neurol* 2010; 9(5): 543-54 .
14. 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, e16.
15. MRC CRASH Trial Collaborators, Perel P, Arango M, Clayton T, Edwards P, Komolafe E, et al. **Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients.** *BMJ* 2008; 336: 425-9.
16. Maas AI. **Clinical trials in head injury: Europe.** In: *Head Trauma basic, preclinical and clinical directions.* Miller LP, Hayes RL, Newcomb JK (Ed.), Wiley-Liss New York, Chichester, Weinheim, Brisbaine, Singapore, Toronto, 2001.
17. Jennett B, Bond M. **Assessment of outcome after severe brain damage.** *Lancet* 1975; 1(7905), 480-4.
18. Jennett B, Snoek J, Bond MR, Brooks N. **Disability after severe head injury: observations on the use of the Glasgow Outcome Scale.** *J Neurol Neurosurg Psychiatry* 1981; 44(4): 285-93.
19. Altman DG, Royston P. **The cost of dichotomising continuous variables.** *BMJ* 2006; 332(7549): 1080.
20. 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.
21. Edwards P, Arango M, Balica L, et al. **Final results of MRC CRASH, a randomised placebo-controlled trial of intravenous corticosteroid in adults with head injury-outcomes at 6 months.** *Lancet* 2005*;* 365(9475): 1957-1959.
22. Roozenbeek B, Maas AI, Lingsma HF, et al. **Baseline characteristics and statistical power in randomized controlled trials: selection, prognostic targeting, or covariate adjustment?** *Crit Care Med* 2009; 37(10): 2683-90.
23. Roozenbeek B, Maas AI, Marmarou A, et al. **The influence of enrollment criteria on recruitment and outcome distribution in traumatic brain injury studies: results from the impact study.** *J Neurotrauma* 2009; 26(7): 1069-75.
24. Machado SG, Murray GD, Teasdale GM. **Evaluation of designs for clinical trials of neuroprotective agents in head injury. European Brain Injury Consortium.** *J Neurotrauma* 1999*;* 16: 1131-8.

25. Hernandez AV, Steyerberg EW, Butcher I, et al. **Adjustment for strong predictors of outcome in traumatic brain injury trials: 25% reduction in sample size requirements in the IMPACT study.** *J Neurotrauma* 2006; 23(9): 1295-303.

26. Jacobs B, Beems T, Stulemeijer M, et al. **Outcome prediction in mild traumatic brain injury: age and clinical variables are stronger predictors than CT abnormalities.** *J Neurotrauma* 2010; 27(4): 655-68.

27. Beach ML, Meier P. **Choosing covariates in the analysis of clinical trials. Control Clin Trials.** 10(4 Suppl), 161S-175S (1989).

28. Steyerberg EW, Bossuyt PM, Lee KL. **Clinical trials in acute myocardial infarction: should we adjust for baseline characteristics?** *Am Heart J* 2000; 139(5): 745-751.

29. Saver JL. **Novel end point analytical techniques and interpreting shifts across the entire range of outcome scales in acute stroke trials.** *Stroke* 2007; 38: 3055-62.

30. Lees KR, Asplund K, Carolei A, et al. **Glycine antagonist (gavestinel) in neuroprotection (GAIN International) in patients with acute stroke: a randomised controlled trial.** GAIN International Investigators. *Lancet* 2000; 355: 1949-1954.

31. Lees KR, Zivin JA, Ashwood T, et al. **NXY-059 for acute ischemic stroke.** *N Engl J Med* 2006; 354:588-600.

32. Cooper DJ, Rosenfeld JV, Murray L, et al. **Decompressive craniectomy in diffuse traumatic brain injury.** *N Eng J Med* 2011; 364(16): 1943-502.

33. Mendelow AD, Gregson BA, Fernandes HM, et al. **Early surgery versus initial conservative treatment in patients with spontaneous supratentorial intracerebral haematomas in the International Surgical Trial in Intracerebral Haemorrhage (STICH): a randomised trial.** *Lancet* 2005; 365: 387-397.

34. Maas AI, Murray G, Henney H III, et al. **Efficacy and safety of dexanabinol in severe traumatic brain injury: results of a phase III randomised, placebo-controlled, clinical trial.** *Lancet Neurol* 2006; 5: 38-45.

35. den Hertog HM, van der Worp HB, van Gemert HM, et al. **The Paracetamol (Acetaminophen) In Stroke (PAIS) trial: a multicentre, randomised, placebo-controlled, phase III trial.** *Lancet Neurol* 2009; 8(5), 434-40.

36. 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 study.** *Clin Trials* 2010; 7(1): 44-57.

37. 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.

38. Wilson JTL, Pettigrew LEL, Teasdale GM. **Structured interviews for the Glasgow Outcome Scale and the Extended Glasgow Outcome Scale: Guidelines for Their Use.** *J Neurotrauma* 1997; 15(8): 573-85.

39. Lu J, Murray GD, Steyerberg EW, et al. **Effects of Glasgow Outcome Scale misclassification on traumatic brain injury clinical trials.** *J Neurotrauma* 2008; 25(6): 641-51.

40. Bagiella E, Novack TA, Ansel B, et al. **Measuring outcome in traumatic brain injury treatment trials: recommendations from the traumatic brain injury clinical trials network.** *J Head Trauma Rehabil* 2010; 25(5): 375-82.

41. Maas AI, Harrison-Felix CL, Menon DK, et al. **Common Data Elements for Traumatic Brain Injury: Recommendations from the Interagency Working Group on Demographics and Clinical Assessment.** *Archives Phys Rehab Medicine* 2010; 91: 1641-1649.

42. Gelpke GJ, Braakman R, Habbema JD, Hilden J. **Comparison of outcome in two series of patients with severe head injuries.** *J Neurosurg* 1983; 59(5): 745-50.

43. Colohan AR, Alves WM, Gross CR, et al. **Head injury mortality in two centers with different emergency medical services and intensive care.** *J Neurosurg* 1989; 71(2): 202-7.

44. Lingsma HF, Roozenbeek B, Li B, et al. **Large between-center differences in outcome after moderate and severe traumatic brain injury in the IMPACT study.** *Neurosurgery* 2011; 68(3): 601-7.

45. Roozenbeek B, Lingsma HF, Lecky FE, Lu J, Weir J, Butcher I, McHugh GS, Murray GD, Perel P, Maas AI, Steyerberg EW: **Prediction of outcome after moderate and severe traumatic brain injury: external validation of the IMPACT and CRASH prognostic models.** *Crit Care Med* 2012; 40(5): 1609-17.

Chapter 10

GENERAL DISCUSSION AND RECOMMENDATIONS

As described in Chapter 1, **the aim** of this thesis was to investigate which adaptations in design and analysis of randomized controlled TBI trials may be best implemented to deal with the heterogeneity of TBI study populations in order to increase chances for detecting clinically relevant treatment effects in future trials for therapeutic interventions for TBI. In this Chapter, the research questions will be answered, the results presented in this thesis discussed, recommendations for future TBI trials provided, and remarks made about the future of randomized controlled trials for TBI.

Summary answers to the specific questions

Five **specific questions** were posed in Chapter 1, which will be answered here:

I. Which problems related to trial design and analysis have contributed to the failure of most randomized controlled TBI trials in the past?

Answer: Only 3 of the 23 large randomized controlled TBI trials on neuroprotective agents that were performed between 1980 and 2009 showed an overall statistically significant treatment effect. Reasons for the failure are the lack of mechanistic endpoints, the relative insensitivity of outcome measures and the heterogeneity of the patient population.
>> Chapter 2 and Chapter 9

II. What is the most efficient approach to selection of patients for randomized controlled TBI trials?

Answer: The traditional approach of using strict enrollment criteria to reduce heterogeneity is inefficient, since it decreases recruitment. Rather, broad enrollment criteria are preferred, combined with covariate adjustment in the analysis phase, unless only a targeted subgroup specific treatment effect is expected (for mechanistic reasons) without an effect in excluded patients.
>> Chapter 6 and Chapter 7

III. What is the influence of covariate adjustment on statistical power in randomized controlled TBI trials?

Answer: Simulation studies with covariate adjustment for three important predictors across the different studies in the IMPACT database demonstrated an increase in statistical power of up to 30% in more heterogeneous populations of surveys and up to 16% in RCT populations with stricter enrollment criteria. It had no adverse effect on recruitment. Covariate adjustment is an advised approach for every TBI trial.
>> Chapter 7

IV. What is the influence of ordinal rather than dichotomous analysis of the Glasgow Outcome Scale as primary outcome measure in randomized controlled TBI trials?

Answer: Dichotomization of the Glasgow Outcome Scale is not recommended, since it is a statistically inefficient and clinically unattractive approach. Instead, ordinal analysis is advised, since it substantially increases statistical power. In the CRASH trial, application of proportional odds regression and the sliding dichotomy approach increased information density with a factor 2.05 and 2.56 respectively, which permits the detection of smaller treatment effects.

>> Chapter 8

V. Are the currently available prognostic models for prediction of outcome after TBI externally valid and can they be used to make design and analysis of randomized controlled TBI trials more efficient?

Answer: Covariate adjustment, the sliding dichotomy approach and prognostic targeting are techniques that all require validated prognostic models to determine the baseline prognostic risk of individual patients. Both the IMPACT and the CRASH prognostic models were proven to be externally valid in more recent databases. No substantial differences in performance were found between the IMPACT and the CRASH models. The IMPACT models can also predict early (14-day) mortality.

>> Chapter 3, Chapter 4 and Chapter 5

Randomized controlled trials for TBI have failed: the incentive for the IMPACT Study

In the past decades, many multicenter Phase III clinical trials have been performed to test different pharmacological agents and therapeutic strategies, affecting different pathophysiological mechanisms that are active in TBI. However, the majority of these studies failed to show an overall significant treatment effect. Substantial limitations in the design and analysis of clinical TBI trials have contributed to this failure. Many of these limitations are related to the large heterogeneity of TBI patient populations. To address these problems, the IMPACT ("International Mission on Prognosis and Analysis of Clinical Trials in TBI") database was created through merging individual TBI patient data, initially from eight randomized controlled trials and three observational surveys (n=9205). Later, during the continuation funding, the database was extended to a total of more than 40,000 patients. This database formed a culture medium in which innovative approaches to improving trial design and analysis were explored. The U.S. National Institutes of Health funded the project. Investigators with clinical, epidemiological and statistical backgrounds from the Antwerp University Hospital, Belgium; the Erasmus MC, University Medical Center

Rotterdam, the Netherlands; the University of Edinburgh, Scotland and the Medical College of Virginia Commonwealth University in Richmond, Virginia, U.S, collaborated within the project. It was hypothesized that the statistical power of TBI trials could be increased by adjusting for prognostic heterogeneity with covariate adjustment and/or prognostic targeting, by exploiting the ordinal nature of the Glasgow Outcome Scale and by relating the outcome obtained in individual patients to their baseline prognostic risk. Extensive prognostic analyses were required as a first step towards the aim of optimizing the chance of demonstrating benefit of new therapies in future trials. These analyses identified a number of strong predictors of mortality and poor functional outcome at six months post-injury. These predictors include the patient's age, the motor score of the Glasgow Coma Scale, pupillary reactivity, abnormalities on the head CT scan and second insults (hypoxia, hypotension). The prognostic studies were presented in a series of eight papers published in one issue of the *Journal of Neurotrauma* (1-8). These analyses were used as a basis for the development of the IMPACT prognostic models (9). The insight in prognosis that was gained from the prognostic analyses and the development of the models, was used a starting point for our research on design and analysis of RCTs for TBI that is (partly) presented in this thesis.

Although all publications that have resulted from the IMPACT project were product of a team effort by all investigators from all participating centers, the individual centers all had their specific focus. The author of this thesis had the honor to work together with the members of both the "Rotterdam team" and the "Antwerp team". His work within these teams focused on performing simulation studies on various patient selection strategies and covariate adjustment, the application of different forms of ordinal analysis to the data of the CRASH trial and the external validation of the IMPACT and CRASH prognostic models in new datasets. He also contributed to a number of review articles. This work is summarized in the separate Chapters of this thesis.

Broad enrollment criteria, covariate adjustment and ordinal outcome analysis: a paradigm shift in the design and analysis of TBI trials

Selection criteria

The commonly applied and accepted approach to patient selection in randomized controlled trials for TBI is the use of strict enrollment criteria. The aim of this approach is to decrease heterogeneity by excluding patients with an "extreme" (either very good or very bad) baseline prognostic risk. This should increase chances for detecting a treatment effect (statistical power). In Chapter 7 of this thesis, we presented evidence that focusing on the middle part of the prognostic risk spectrum – either by separate strict enrollment criteria or by "prognostic targeting" using a multivariable prognostic model – indeed

increased statistical power of a TBI trial. However, as a result of using strict enrollment criteria a substantial proportion of the potential population is excluded from participation in a randomized controlled trial. In the IMPACT database we found that more than 40% of the patients in the three observational studies (with relatively broad enrollment criteria) were in an extreme risk group (Table 10.1). When stricter enrollment criteria are applied, the recruitment rate (number of recruited patients per time unit) will be reduced and study duration prolonged. We found that the beneficial effect of strict selection is outweighed by the adverse effect on recruitment (Chapters 6 and 7). Therefore, we advise to also include patients with a more extreme prognosis to the trial. The *use of broad enrollment criteria* has the benefit of stimulating recruitment and it increases the generalizability of the trial's results. From our point of view, enrollment criteria should not be limited by "prognostic arguments", but should be based mainly on "mechanistic arguments". Patients should only be excluded from enrollment in a trial, if the mechanistic/biological mechanism that is targeted by the treatment under study is not expected to be active in these specific patients.

The vision that we should use broad instead of strict enrollment criteria is rather revolutionary in the field of clinical TBI research. While almost all trials that were performed in the past had relatively small sample sizes (around 600 subjects) and used strict enrollment criteria, the Corticosteroid Randomisation After Significant Head Injury (CRASH) trial (10, 11) was the first trial in TBI that used very broad enrollment criteria and planned to enroll 20,000 patients. This "mega-trial" was an international, multicenter trial investigating the efficacy of methylprednisolone in patients with mild to severe TBI (1999-2004). After inclusion of 10,008 patients it was halted because an interim analysis showed a higher 14-day mortality rate in the active treatment arm (21.1% versus 17.9% deaths, OR 1.22, 95% confidence interval 1.11-1.35, p=0.0001). Although the approach of selection of the study cohort is substantially different from previous clinical trials in the field of TBI, it was one of the first TBI trials that demonstrated an overall statistical significant (but negative) treatment effect. In other fields of medicine, such as the management of acute myocardial infarction, performing very large trials with liberal enrollment criteria is not new at all (12). However, some authors argue that the "mega-trial" design – i.e. the inclusion of thousands of patients combined with usually (very) limited data collection – is based on flawed methodology. For example, Charlton said that "*a measurement can not be made simultaneously more precise and more valid by reducing the rigour of its protocol in order to allow increased recruitment of patients"* (13). In his opinion the reduction of the confidence interval of the treatment effect by increasing the number of recruitment patients, does not result in a more valid estimate of this treatment effect, and he is against the use of the "mega-trial" design (14).

We think that the optimal approach should combine the best of both worlds: we should use relatively broad enrollment criteria to stimulate recruitment and increase generalizability, and we should ascertain good quality data collection with recording of important covariates in order to obtain a valid estimate of the treatment effect.

Table 10.1. Classification of patients from three surveys and eight randomized controlled trials included in the IMPACT database according to their baseline risk for unfavourable outcome. Reproduced from Maas *et al.*, Neurotherapeutics, 2010 (49).

Models	Number of patients eligible for risk estimation	Number of patients per risk category (%)		
		Low risk 0-20%	*Intermediate risk 41-80%*	*High risk 81-100%*
Surveys				
Core Model [1]	2217	202 (9%)	1269 (57%)	746 (34%)
Extended Model [2]	2217	247 (11%)	1208 (54%)	762 (34%)
RCTs				
Core Model [1]	6292	1040 (17%)	4733 (75%)	519 (8%)
Extended Model [2]	4782	1168 (24%)	3160 (66%)	454 (10%)

Categories were defined according to the risk of unfavourable outcome, as estimated by the prognostic models as designed by the IMPACT investigators.
[1] Core model: age, GCS motor score and pupillary reactivity.
[2] Extended model: Core model plus hypoxia, hypotension and CT characteristics (i.e., Marshall CT Classification, and traumatic subarachnoid hemorrhage).
Percentages may not add up to 100% because of rounding. Numbers of patients may differ from numbers reported elsewhere, because of differences in recorded variables per study and selection of non-missing outcome and age $\geq$ 14.

Covariate adjustment

Heterogeneity of trial populations carries the risk of significant imbalances between the study groups, despite the use of randomization, even when the population is relatively large. Furthermore, even in the absence of significant imbalances both groups cannot be considered comparable by definition, since this also depends on the strength of the association between the covariates and the outcome. The relation between imbalances, prognostic strength of the covariates, and the estimation of the treatment effect has been illustrated previously (15). In order to deal with the problem of heterogeneity, we advise to adjust the treatment effect for important predictors of outcome by a procedure called *covariate adjustment*. Covariate adjustment substantially increases statistical power of a trial. Earlier simulation studies with TBI trial data showed that logistic regression analysis of the treatment effect, with covariate adjustment for seven strong predictors of outcome resulted in a 25% gain in statistical power, compared to the unadjusted analysis (16). These results were confirmed across the individual patient populations of the IMPACT database (presented in Chapter 7 of this thesis), demonstrating an increase in statistical power of up to 30% in more heterogeneous populations of observational surveys and up to 16% in trial populations with stricter criteria.

Recently published explorations of the influence of covariate adjustment on the actual (not simulated) results of the previously mentioned CRASH trial, showed that adjustment

for age, GCS motor score and pupillary reactivity (IMPACT Core model) resulted in a relative sample size of 0.79 to obtain the same statistical power as the unadjusted analysis (17).

Adjustment for the variables age, GCS, pupillary reactivity and the presence of major extracranial injury (CRASH Basic model) led to a relative sample size of 0.73. This implies an increase in power from 80% to 88% and 91% respectively. The authors recommend the use of covariate adjustment in future TBI trials in order to increase statistical power, but warn not to use this method to actually decrease sample sizes in the planning of new trials.

In other fields of medicine that are dealing with heterogeneous patient populations, covariate adjustment is applied as well. Methodological research in the field of stroke trials for example, by the "Optimising Analysis of Stroke Trials" (OAST) Collaboration, provided the insight that covariate adjustment reduced the required sample size by at least 20-30% in acute stroke trials (18). The magnitude of these reductions is comparable to those observed in TBI trials, but is much greater than those previously observed in for example cardiovascular trials (15). The size of the gain in statistical power depends on the strength of the association between the covariates and the outcome.

Although the potential beneficial effect of covariate adjustment on statistical power is becoming more and more accepted, only a minority of randomized controlled trials report an adjusted treatment effect. In a systematic literature review on the use of covariate adjustment in TBI trials Hernández et al. found that only five out of the 18 selected RCTs (28%) applied covariate adjustment (19). A systematic review by Austin and colleagues (20) on the handling of baseline covariates in randomized controlled trials, published in the first half of 2007 in four high-impact general medical journals, found that only approximately one-third of the trials reported an adjusted treatment effect. In this review the disagreement on which measure is most informative for physicians is discussed: the unadjusted or the adjusted treatment effect? The unadjusted treatment effect can be interpreted as the population-average estimate for the study. In contrast, an adjusted treatment effect can be seen as a subject-specific estimate, or the estimate of the treatment effect for a patient with a specific set of covariate values. Hauck et al. pose that the adjusted effect is more relevant from a clinician-patient perspective and should therefore be the preferred measure to present in reports of randomized clinical trials (21). Martens et al. suggest that the population-averaged estimate (unadjusted treatment effect) is better defined and appears to be of greater interest (22). Especially in non-linear regression (as for example logistic regression analysis that is used in most TBI trials) this is a particularly relevant discussion, since adjusted and unadjusted point estimates of the effect differ from each other (23). More specifically, unadjusted treatment effects in logistic regression are closer to the 'null' than adjusted treatment effects. In other words, unadjusted analysis attenuates the estimation of the treatment effect in logistic regression. This is the result of an issue called 'non-collapsibility'. An estimator is collapsible when the population-average effect

is identical to the adjusted effect. This is the case in linear regression. However, non-linear regression techniques, such as logistic regression, result in non-collapsible estimators (e.g., the odds ratio). The phenomenon of non-collapsibility of the odds ratio is nicely illustrated in a paper by Kent et al. published in *Stroke* in 2009, summarized in Table 10.2 (24). It shows a hypothetical stroke trial enrolling patients who belong to 4 different risk/severity strata. The treatment effect is uniform across the 4 strata, i.e. 50% improvement in the odds of having a good outcome (OR=1.5). Surprisingly, if one calculates the OR for the overall results, one would find that treatment increased the odds of a good outcome by only 38%. This represents an underestimation of the within-stratum treatment effect of almost 25%.

This example supports our view that some TBI trials may have failed because of the way these trials were analyzed.

Table 10.2. Unadjusted summary results underestimate stratum-specific odds ratios when the treatment effect is uniform. Reproduced from Kent et al., *Stroke*, 2009 (24).

	Good Functional Outcome		
Risk/Severity Strata	Control Rate	Treatment Rate	OR
Severe stroke	20.0%	27.3%	1.5
	40.0%	50%	1.5
	60.0%	69.2%	1.5
Mild stroke	80.0%	85.7%	1.5
Summary of trial result	50.0%	58.1%	1.38

In line with the previously mentioned argument that the adjusted treatment effect is less well defined, a criticism of the advocacy of covariate adjustment in RCTs is that it only allows adjustment for measured prognostic variables and not for unmeasured variables. For the population-average or the unadjusted model only one 'true' model exists. In contrast, the 'true' patient-specific estimate of the treatment effect can only be estimated with an 'ideal' adjustment model containing all variables that together explain the full variance, which in real life will not be the available. So, the (magnitude of the) adjusted treatment effect depends on the decision about which covariates to include into the adjustment model.

In the systematic review by Austin et al. only 13 of the 39 (33%) trials that reported an adjusted treatment effect clearly described which method was used to select covariates for the adjustment model (20). Six of these 13 trials reported that the variables had been selected 'a priori'. Senn stated that the selection of covariates should be done 'a priori' based on an analysis of which variables are strong predictors of outcome (25). A post hoc or inappropriate variable selection strategy may result in biased estimates of the treatment effect. This underlines the importance of prognostic analysis.

Ordinal analysis

The current practice of dichotomizing the Glasgow Outcome Scale (GOS) in unfavorable (Death, Vegetative State and Severe Disability) versus favorable (Moderate Disability and Good Recovery) is a statistically inefficient and a clinically unattractive approach. Because of this dichotomization, patients with an extreme prognosis have little potential to contribute toward the statistical power of a trial. The prognosis of patients in an extreme risk group can be so good that they will almost inevitably achieve a favorable outcome, even without the benefits of an effective new intervention, or so poor that it is extremely unlikely that even an effective intervention would improve their outcome to such an extent that it would move from being unfavorable to favorable. Moreover, it is not intuitively attractive from a clinical perspective to only focus on one specific split of the outcome distribution and not to take other clinically relevant splits into account. It was shown that approaches that exploit the ordinal nature of the GOS will result in higher statistical power, compared to the traditional dichotomized analysis (26). This will increase the chance of detecting smaller, but clinically relevant treatment effects. We therefore advise to use an approach that exploits the ordinal nature of the GOS. Two approaches that were investigated in this thesis are *proportional odds logistic regression* and the *sliding dichotomy approach* (Figure 10.1).

In the *sliding dichotomy approach*, the point of dichotomy of the GOS is differentiated according to the baseline prognostic risk. For patients with a poor prognosis, survival may be relevant whilst in those with a good prognosis any outcome worse than good recovery may be considered unfavorable. Practically, the concept is applied as follows. First, the trial population in divided in a number of equally large not overlapping prognostic bands (e.g. three bands: good, intermediate and poor prognosis). Patients are divided in these bands based on their individual baseline prognostic risk, calculated by a prognostic model.

Secondly, the point of dichotomization for the GOS is determined for each of the prognostic bands. For example, in the good prognosis band the split is placed between Moderate Disability and Good Recovery, in the intermediate prognosis band between Severe Disability and Moderate Disability, and in the poor prognosis band between Vegetative State/Death and Severe Disability. According to these splits, it is determined whether each individual patient has a favorable or an unfavorable outcome. Thirdly, three odds ratios are obtained by fitting binary logistic regression models. Finally, a pooled odds ratio can be calculated as a summary estimate of the treatment effect. This is the odds ratio for having a better outcome than expected. The pooling over strata is only reasonable if the three odds ratios are similar, i.e. no interaction between prognosis and treatment effect. The concept of the sliding dichotomy approach has been adopted for the primary analysis of several phase III trials in TBI, stroke and intracerebral hemorrhage (27-30). Use of the sliding dichotomy approach – with or without combination with covariate adjustment – requires identification of valid prognostic models to reliably provide a baseline risk estimate in individual patients.

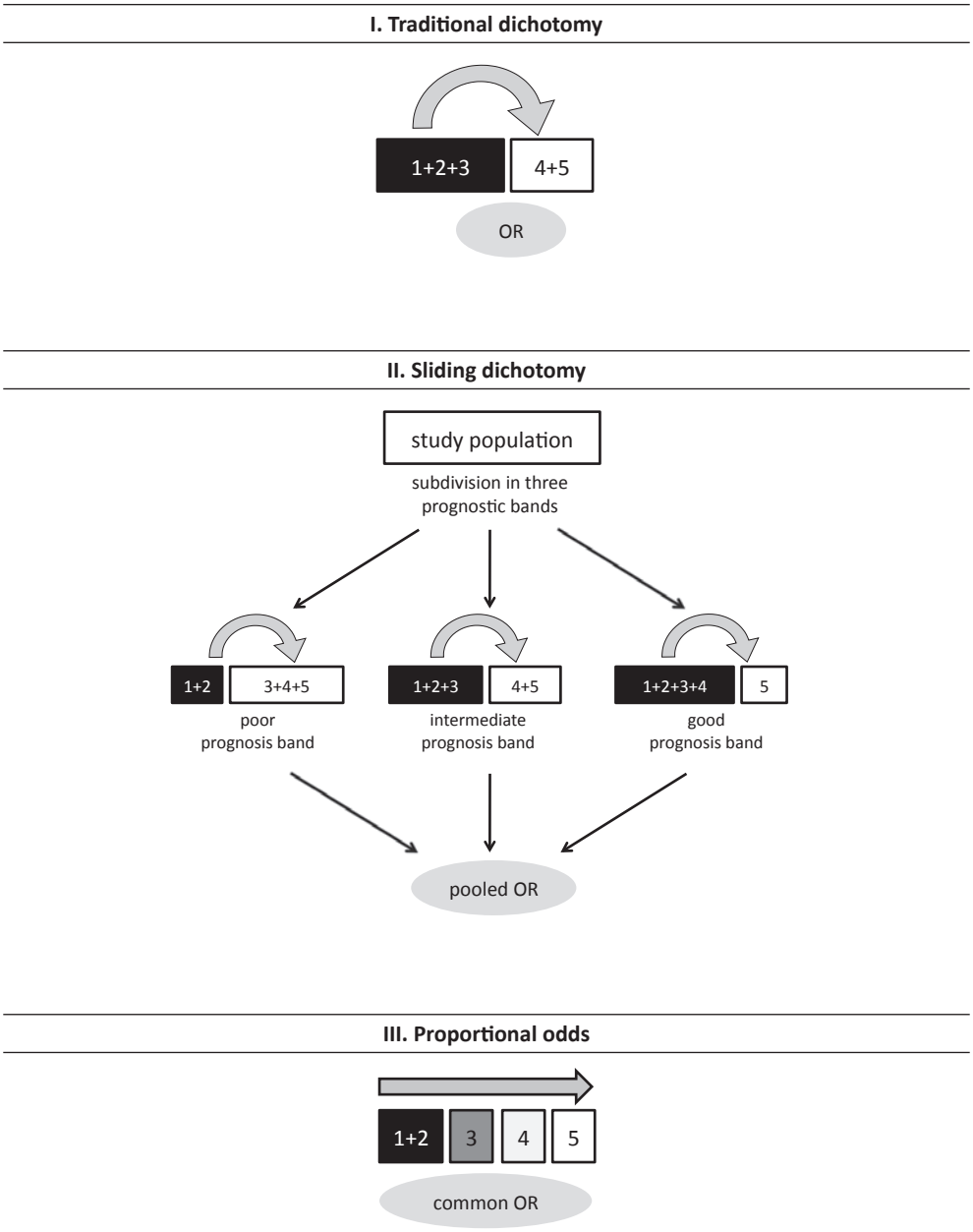


Figure 10.1. Graphical illustration of three different approaches to the analysis of the Glasgow Outcome Scale as outcome measure in randomized controlled TBI trials.

In the *proportional odds approach*, the regression model considers every possible way in which an ordinal scale can be dichotomized, assuming that the odds ratio for a better outcome versus a worse outcome is identical wherever the scale is dichotomized (proportional odds assumption). The proportional odds model combines the odds ratios for all possible splits and estimates a common odds ratio. In this approach, the study population is not subdivided, but every patient contributes to the estimation of this common odds ratio. Hence, we obtain an overall estimate of the shift in outcome across the full GOS. From this perspective it has been proposed to refer to the use of the proportional odds model as a "shift analysis" (31). This approach was used in the efficacy analysis of several stroke trials (32, 33) and in one recently published TBI trial (34).

Earlier simulation studies within the IMPACT database showed considerable benefits of ordinal analysis on statistical power (26; Chapter 9). The application of proportional odds analysis and of the sliding dichotomy increased statistical power by 23 to 30%. Applying covariate adjustment in addition to the ordinal analysis further increased power to a total of up to 40 to 49%. The benefits of this approach were further confirmed in new data sets and when applying both approaches to the data of the MRC CRASH study, the expected benefits were confirmed in the real life situation of a clinical "mega trial" (Chapter 6). This study showed that information density (i.e., the ratio of the Wald statistics of the treatment effect of the different strategies) of the ordinal approaches was 2.05-fold (proportional odds) and 2.56-fold (sliding dichotomy) higher compared to the traditional dichotomized analysis of the GOS. Furthermore, the study showed that both ordinal approaches already showed significant treatment effects after enrolling approximately half the number of patients required with the dichotomized outcome (Figure 10.2). These findings strongly support adopting an ordinal approach to outcome analysis in TBI trials. Importantly, other fields using ordinal outcome measures may benefit from the proposed methodology. For example, the OAST Collaboration found that the use of some form of ordinal analysis rather than dichotomized analysis in acute stroke trials is probably more efficient and more likely to yield reliable results (35).

For both approaches, there is the requirement to examine the degree of heterogeneity amongst the three estimated odds ratios before going on to derive the corresponding pooled estimate (pooled odds ratio or common odds ratio). There are standard methods with both approaches to perform a formal statistical test of heterogeneity (36), but generally, the power to detect heterogeneity will be low. Moreover, even when substantial heterogeneity is observed, it might still be judged appropriate to present the pooled estimate as an overall summary measure of the treatment effect.

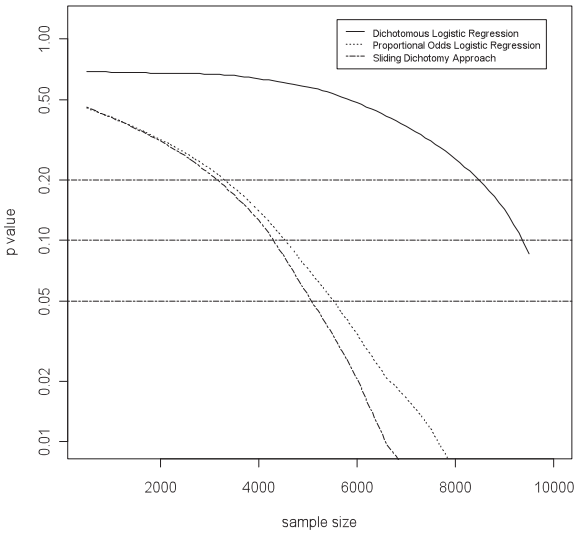


Figure 10.2. Three different approaches to analyze the treatment effect on the 6-month Glasgow Outcome Scale. P values are given for increasing numbers of patients as consecutively enrolled in the CRASH trial (range n=500 to n=9554).

The choice between proportional odds analysis versus the sliding dichotomy is partly a value judgment. From a statistical perspective the proportional odds model is more efficient, but is perhaps less appealing for a clinical audience. Thus, we consider both approaches appropriate for the analysis of TBI trials. Whichever ordinal approach is chosen, evidence strongly indicates that the conventional dichotomous analysis should be discarded from the trialists' toolbox, unless one is exclusively interested in one particular outcome (e.g. mortality).

The use of baseline prognostic information in the analysis of randomized controlled trials requires valid prognostic models

The advanced statistical methods described in this thesis all require validated prognostic models to estimate the baseline prognostic risk for individual patients. Such models are now available for TBI. The CRASH and the IMPACT models have been developed recently, using state of the art methodology on large datasets (9, 37). For both models web calculators are available online (http://www.tbi-impact.org/; http://www.crash2.lshtm.ac.uk/Riskcalculator/).

Both sets of models were externally validated reciprocally. Since the models were never compared head-to-head (38) and external validation is a continuous process (36), we presented two external validation studies in this thesis (Chapter 4 and 5). These

studies resulted in a number of interesting observations. First, the studies confirmed the validity of both sets of models to predict the outcome in new patients (from outside the development datasets). Secondly, comparison of both sets of models (IMPACT vs. CRASH) did not show any substantial differences in performance. Thirdly, we found out that the IMPACT models also reliably predict early (14-day) instead of late (6-month) mortality. And finally, substantial variation in predictive power was seen between the different validation datasets. This was largely explained by the differences in case-mix between the datasets. The best discriminative power was found in the heterogeneous observational studies, while it was less good in the more homogeneous randomized controlled trials. We applied the "case-mix-corrected AUC" as proposed by Vergouwe et al. (39) to adjust for this difference and to make performance measures of the different datasets better comparable.

Model validation is a continuous process. An important issue in the validation of prognostic models is the decision on when updating of an existing model is preferred/necessary and how updating should be performed. Although no specific cut-off values for performance measures are available, it seems reasonable to update models if the incidence of events has changed substantially over time. Different approaches to updating are possible (36). If the mean predicted values no longer agree with the mean observed values ("calibration-in-the-large"), the approach to updating is to re-estimate the intercept of the model. If it appears more likely that the predictive effects of the predictors in the model may have changed or if new predictors are discovered, than a more extensive approach to updating should be followed. Examples of such more extensive approaches are re-estimation of the intercept and the regression coefficients in new data ("re-calibration") or considering new predictors for inclusion in the model ("model revision").

Future perspective

The explorations for improving the design and analysis of clinical trials in TBI illustrate how crude our approaches over the past decades have really been and that we may not have been giving new therapies a fair chance. The recommendations stated above constitute a clear change of direction from current approaches to reduce heterogeneity in randomized controlled TBI trials. It should be recognized however that although these recommendations may partly solve the problem of heterogeneity in terms of injury severity and prognostic risk, they do not address heterogeneity related to mechanism. Pathophysiologic mechanisms are different in patients with for example contusions (large inflammatory component) compared to patients with diffuse axonal injury (less inflammation, more calcium mediated damage). The concept of mechanistic targeting – the ideal for clinical trials – will require reliable identification of occurrence and time course of pathophysiologic mechanisms in individual patients. This may also result in the identification of mechanistic (early) endpoints that can be used in clinical trials. Here we see a great challenge for basic scientists and

clinical researchers (40). Another unsolved problem is the insensitivity of outcome measures that are used in TBI trials. Outcome after TBI is by definition multidimensional, including neurophysical disabilities, disturbances in mental functioning (e.g., cognitive and executive functioning), and consequential problems in social reintegration. Moreover, the patient's perspective of outcome (Health related Quality of Life) should be taken into consideration. We see a clear need to explore the feasibility of developing a multidimensional approach to outcome assessment and classification.

Randomized controlled trials are considered to be the gold standard for investigating efficacy of (new) medical treatments. Nevertheless, not all clinical problems can be investigated with this study design. Randomized controlled trials commonly investigate whether an intervention works compared to placebo in TBI (efficacy), but very few have compared different alternative interventions or approaches to treatment of TBI patients (effectiveness).

An approach to the investigation of especially the effectiveness of treatments is "comparative effectiveness research" (CER). Common characteristics of CER are summarized in Table 10.4. The existing heterogeneity in TBI populations, management approaches and outcome offers opportunity for exploring causes for these differences and identifying underlying reasons for a given outcome, or individual patient response to a selected intervention. The concept of CER in TBI is not new. In 1983 Gelpke and colleagues (41) analyzed differences and outcome between two centers from the Netherlands and in 1989 Colohan et al (42) performed comparative analysis of treatment results between Charlottesville (U.S.) and New Delhi (India). There are a number of specific features of TBI that may make a special argument for CER. First, large differences between centers and between countries exist in both management and outcome. These differences provide a major opportunity to compare alternative interventions and managements. Second, robust and validated risk adjustment models both for mortality and for functional outcome are available. Third, sophisticated statistical approaches, including random effect analysis are available to support CER.

Table 10.4. Common characteristics of Comparative Effectiveness Research.

• The objective of directly informing a specific clinical decision from the patient perspective or a health policy decision from the population perspective.
• Comparison of at least two alternative interventions, each with the potential to be "best practice."
• Description of results at the population and subgroup levels.
• The use of outcomes —both benefits and harms— that are important to patients.
• Employment of methods and data sources appropriate for the decision of interest.
• Conducted in settings that are similar to those in which the intervention will be used in practice.

Recommendations for future TBI trials

Based on the research that was described in this thesis and other IMPACT publications, the IMPACT Study Group has published a set of recommendations for the analysis of TBI trials in the future (49). These recommendations include:

- Details of the major baseline prognostic characteristics should be provided in every report on a TBI study; in trials they should be differentiated per treatment group. We also advocate the reporting of the baseline prognostic risk as determined by validated prognostic models.
- Inclusion criteria should be as broad as is compatible with the current understanding of the mechanisms of action of the intervention being evaluated. This will maximize recruitment rates and enhance the generalizability of the results.
- The statistical analysis should incorporate (prespecified) covariate adjustment to mitigate the effects of the heterogeneity.
- The statistical analysis should use an ordinal approach, based on either sliding dichotomy or proportional odds methodology.

Generalizability of the recommendations

In order to become accepted by researchers, clinicians and regulatory authorities (FDA, EMEA) as a new "standard approach" to clinical trial design in TBI, generalizability of the recommendations needs to be confirmed. A first step to prove generalizability was taken by applying the recommendations in a re-analysis of the MRC CRASH trial data, showing convincing results (as described in Chapter 6). A next step would be to implement the recommendations in new trials. The DECRA trial on decompressive craniectomy in patients with diffuse traumatic brain injury was one of the first large RCTs in TBI that applied most of the recommendations (33). Several of the ongoing TBI trials have also adopted (parts of the) the methodology presented in this thesis. These studies include for example EuroTherm (therapeutic hypothermia for TBI) and SyNAPSe® (progesterone for severe TBI).

As mentioned previously, prognostic heterogeneity of patient populations does not only apply to traumatic brain injury, but also to other (acute) neurological diseases (e.g., ischemic stroke, subarachnoid hemorrhage, intracerebral hemorrhage) and to other fields of medicine (43-45). The methodology and recommendations proposed in this thesis to deal with heterogeneity in randomized controlled trials for TBI might also be of interest for such other fields.

In the field of acute ischemic stroke, most large randomized controlled trials have been neutral as well (46). Stroke populations have in common with those of TBI that there is

substantial prognostic heterogeneity in the population and that an ordinal outcome measure is generally used (modified Rankin Scale), which is often dichotomized for analysis. A recent paper on the methodology of acute ischemic stroke trials advises to tailor the approach to the analysis of the treatment effect to each individual trial, based on how the intervention under study is most likely to modify the distribution of outcomes (35). The paper argues that some form of ordinal analysis is likely to be preferred over the traditional dichotomy and it supports the use of covariate adjustment in stroke trials as well. These findings are very much in line with the recommendations presented in this thesis. As mentioned previously, several acute stroke trials were published that have used different aspects of the methodology described in this thesis (24, 25, 29, 47, 48). Other ongoing acute (ischemic/hemorrhagic) stroke trials, such as STICH-2 (indications for surgery for intracerebral hemorrhage), STASH (statins for subarachnoid hemorrhage), EuroHYP (hypothermia for acute ischemic stroke), MR CLEAN (endovascular treatment for acute ischemic stroke) and others, have planned to adopt (parts) of the recommendations. The application of the recommendations to trials in other fields than TBI, will hopefully contribute to their general acceptance.

Conclusions

Randomized controlled trials in TBI are challenging due to the inherent heterogeneity of the patient population, the lack of early mechanistic end points, and relative insensitivity of outcome measures. Some approaches to deal with the heterogeneity of the patient population were presented in this thesis. The use of strict enrollment criteria is not recommended, as this is inefficient. Rather, broad enrollment criteria may often be preferred combined with covariate adjustment in the analysis phase. Dichotomization of the GOS is not recommended. Ordinal approaches to analysis of treatment effects offer greater statistical power and better sensitivity. To this purpose the use of proportional odds methodology or the sliding dichotomy may be considered. To practically apply these analysis techniques, well-validated prognostic models are currently available. Combining an ordinal approach to the analysis of treatment effects with covariate adjustment can increase statistical power by 40 – 50%. These recommendations are expected to enhance chances for detecting clinically relevant treatment effects for the benefit of future TBI patients.

References

1. Mushkudiani NA, Engel DC, Steyerberg EW, Butcher I, Lu J, Marmarou A, Slieker F, McHugh GS, Murray GD, Maas AI. **Prognostic value of demographic characteristics in traumatic brain injury: results from the IMPACT study. J** Neurotrauma. 2007; 24(2): 259-69.
2. Butcher I, McHugh GS, Lu J, Steyerberg EW, Hernández AV, Mushkudiani N, Maas AI, Marmarou A, Murray GD. **Prognostic value of cause of injury in traumatic brain injury: results from the IMPACT study.** *J Neurotrauma.* 2007; 24(2): 281-6.
3. McHugh GS, Engel DC, Butcher I, Steyerberg EW, Lu J, Mushkudiani N, Hernández AV, Marmarou A, Maas AI, Murray GD. **Prognostic value of secondary insults in traumatic brain injury: results from the IMPACT study.** *J Neurotrauma.* 2007; 24(2): 287-93.
4. Maas AI, Steyerberg EW, Butcher I, Dammers R, Lu J, Marmarou A, Mushkudiani NA, McHugh GS, Murray GD. **Prognostic value of computerized tomography scan characteristics in traumatic brain injury: results from the IMPACT study.** *J Neurotrauma.* 2007; 24(2): 303-14.
5. Van Beek JG, Mushkudiani NA, Steyerberg EW, Butcher I, McHugh GS, Lu J, Marmarou A, Murray GD, Maas AI. **Prognostic value of admission laboratory parameters in traumatic brain injury: results from the IMPACT study.** *J Neurotrauma*. 2007; 24(2): 315-28.
6. Butcher I, Maas AI, Lu J, Marmarou A, Murray GD, Mushkudiani NA, McHugh GS, Steyerberg EW. **Prognostic value of admission blood pressure in traumatic brain injury: results from the IMPACT study.** *J Neurotrauma*. 2007; 24(2): 294-302.
7. Marmarou A, Lu J, Butcher I, McHugh GS, Murray GD, Steyerberg EW, Mushkudiani NA, Choi S, **Maas AI. 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(2): 270-80.
8. Murray GD, Butcher I, McHugh GS, Lu J, Mushkudiani NA, Maas AI, Marmarou A, Steyerberg EW. **Multivariable prognostic analysis in traumatic brain injury: results from the IMPACT study.** *J Neurotrauma.* 2007; 24(2): 329-37.
9. Steyerberg EW, Mushkudiani N, Perel P, Butcher I, Lu J, McHugh GS, Murray GD, Marmarou A, Roberts I, Habbema JD, Maas AI. **Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics.** *PLoS Med.* 2008; 5(8):e165.
10. Roberts I, Yates D, Sandercock P, et al on behalf of the CRASH trial collaborators. **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: 1321-28.
11. Edwards P, Arango M, Balica L, et al on behalf of the CRSAH trial collaborators. **Final results of MRC CRASH, a randomised placebo-controlled trial of intravenous corticosteroid in adults with head injury-outcomes at 6 months.** *Lancet*. 2005; 365: 1957-59.
12. Woods KL. **Mega-trials and management of acute myocardial infarction.** *Lancet.* 1995; 346: 611-14.
13. Charlton BG. **Megatrials are based on a methodological mistake.** *Br J Gen Pract.* 1996; 46(408): 429-31.
14. Charlton BG. **Mega-trials: methodological issues and clinical implications.** *J R Coll Physicians Lond.* 1995; 29: 96-100.
15. Steyerberg EW, Bossuyt PM, Lee KL. **Clinical trials in acute myocardial infarction: should we adjust for baseline characteristics?** *Am Heart J.* 2000; 139(5): 745-51.
16. Hernández AV, Steyerberg EW, Butcher I, Mushkudiani N, Taylor GS, Murray GD, Marmarou A, Choi SC, Lu J, Habbema JD, Maas AI. **Adjustment for strong predictors of outcome in traumatic brain injury trials: 25% reduction in sample size requirements in the IMPACT study.** *J Neurotrauma.* 2006; 23(9): 1295-303.
17. Turner EL, Perel P, Clayton T, Edwards P, Hernández AV, Roberts I, Shakur H, Steyerberg EW; CRASH trial collaborators. **Covariate adjustment increased power in randomized controlled trials: an example in traumatic brain injury.** *J Clin Epidemiol.* 2012; 65(5): 474-81.
18. Optimising the Analysis of Stroke Trials (OAST) Collaboration, Gray LJ, Bath PM, Collier T. **Should stroke trials adjust functional outcome for baseline prognostic factors?** *Stroke*. 2009; 40(3): 888-94.
19. Hernández AV, Steyerberg EW, Taylor GS, Marmarou A, Habbema JD, Maas AI. **Subgroep analysis and covariate adjustment in randomized controlled trials of traumatic brain injury: a systematic review.** *Neurosurg.* 2005; 57(6): 1244-53.
20. Austin PC, Manca A, Zwarenstein M, Juurlink DN, Stanbrook MB. **A substantial and confusing variation exists in handling of baseline covariates in randomized controlled trials: a review of trials published in leading medical journals.** J Clin Epidemiol. 2010; 63: 142–53

21. Hauck WW, Anderson S, Marcus SM. **Should we adjust for covariates in nonlinear regression analyses of randomized trials?** *Control Clin Trials.* 1998; 19:249-56.

22. Martens EP, Pestman WR, Klungel OH. **Conditioning on the propensity score can result in biased estimation of common measures of treatment effect: a Monte Carlo study.** (p n/a) by Austin PC, Grootendorst P, Normand ST, Anderson GM. *Stat Med*. 2007; 26: 3208-10.

23. Gail MH, Wieand S, Piantadosi S. **Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates.** *Biometrika.* 1984; 7: 431-44.

24. Kent DM, Trikalinos TA, Hill MD. **Are unadjusted analysis of clinical trials inappropriately biased toward the null?** *Stroke*. 2009; 40: 672-673.

25. Senn S. **Testing for baseline balance in clinical trials.** *Stat Med.* 1994; 13: 1715-26.

26. McHugh GS, Butcher I, Steyerberg EW, Marmarou A, Lu J, Lingsma HF, Weir J, Maas AI, Murray GD. **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(1): 44-57.

27. Maas AI, Murray G, Henney H, III, Kassem N, Legrand V, Mangelus M et al. **Efficacy and safety of dexanabinol in severe traumatic brain injury: results of a phase III randomised, placebo-controlled, clinical trial.** *Lancet Neurol.* 2006; 5: 38-45.

28. Mendelow AD, Gregson BA, Fernandes HM, Murray GD, Teasdale GM, Hope DT, Kamiri A, Shaw MDM, Barer DH; for the STICH investigators. **Early surgery versus initial conservative treatment in patients with spontaneous supratentorial intracerebral haematomas in the International Surgical Trial in Intracerebral Haemorrhage (STICH): a randomised trial.** *Lancet.* 2005; 365: 387-397.

29. den Hertog HM, van der Worp HB, van Gemert HM, Algra A, Kappelle LJ, van Gijn J, Koudstaal PJ, Dippel DW; PAIS Investigators. **The Paracetamol (Acetaminophen) In Stroke (PAIS) trial: a multicentre, randomised, placebo-controlled, phase III trial.** *Lancet Neurol.* 2009; 8(5): 434-40.

30. Adams HP Jr, Effron MB, Torner J, Dávalos A, Frayne J, Teal P, Leclerc J, Oemar B, Padgett L, Barnathan ES, Hacke W; AbESTT-II Investigators. **Emergency administration of abciximab for treatment of patients with acute ischemic stroke: results of an international phase III trial: Abciximab in Emergency Treatment of Stroke Trial (AbESTT-II).** *Stroke*. 2008; 39(1): 87-99.

31. Saver JL. **Novel end point analytical techniques and interpreting shifts across the entire range of outcome scales in acute stroke trials.** *Stroke.* 2007; 38: 3055-62.

32. Lees KR, Asplund K, Carolei A, et al. **Glycine antagonist (gavestinel) in neuroprotection (GAIN International) in patients with acute stroke: a randomised controlled trial. GAIN International Investigators.** *Lancet*. 2000; 355: 1949-1954.

33. Lees KR, Zivin JA, Ashwood T, et al. **NXY-059 for acute ischemic stroke.** *N Engl J Med*. 2006; 354: 588-600.

34. Cooper DJ, Rosenfeld JV, Murray L, Arabi YM, Davies AR, D'Urso P, Kossmann T, Ponsford J, Seppelt I, Reilly P, Wolfe R; DECRA Trial Investigators; Australian and New Zealand Intensive Care Society Clinical Trials Group. **Decompressive craniectomy in diffuse traumatic brain injury.** *N Engl J Med.* 2011; 364(16): 1493-502.

35. Bath PM, Lees KR, Schellinger PD, Altman H, Bland M, Hogg C, Howard G, Saver JL; on behalf of the European Stroke Organisation Outcomes Working Group. **Statistical Analysis of the Primary Outcome in Acute Stroke Trials.** *Stroke*. 2012; 43(4): 1171-1178.

36. Steyerberg EW. **Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating.** Springer, 2009. ISBN: 978-0-387-77243-1.

37. MRC CRASH Trial Collaborators, Perel P, Arango M, Clayton T, Edwards P, Komolafe E, Poccock S, Roberts I, Shakur H, Steyerberg E, Yutthakasemsunt S. **Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients.** *BMJ.* 2008; 336(7641): 425-9.

38. Menon DK, Zahed C. **Prediction of outcome in severe traumatic brain injury.** *Curr Opin Crit Care.* 2009; 15(5): 437-41.

39. Vergouwe Y, Moons KG, Steyerberg EW. **External validation of risk models: use of benchmark values to disentangle a case-mix effect from incorrect coefficients.** *Am J Epidemiol*. 2010; 172:971–80.

40. Saatman KE, Duhaime AC, Bullock R, Maas AI, Valadka A, Manley GT. **Classification of traumatic brain injury for targeted therapies.** *J Neurotrauma.* 2008; 25: 719-38.

41. Gelpke GJ, Braakman R, Habbema JD, Hilden J. **Comparison of outcome in two series of patients with severe head injuries.** *J Neurosurg.* 1983; 59:745-50.

42. Colohan AR, Alves WM, Gross CR, Torner JC, Mehta VS, Tandon PN, Jane JA. **Head injury mortality in two centers with different emergency medical services and intensive care.** *J Neurosurg.* 1989; 71:202-7.

43. Annane D. **Improving clinical trials in the critically ill: unique challenge—sepsis.** *Crit Care Med.* 2009; 37(1 Suppl): S117-28.

44. Hébert PC, Cook DJ, Wells G, Marshall J. **The design of randomized clinical trials in critically ill patients.** *Chest.* 2002; 121(4): 1290-300.

45. Hernández AV. **Heterogeneity of patients in clinical trials. Subgroup analysis and covariate adjustment in cardiovascular and neurosurgical trials.** 2006. PhD thesis. Erasmus University Rotterdam, The Netherlands.

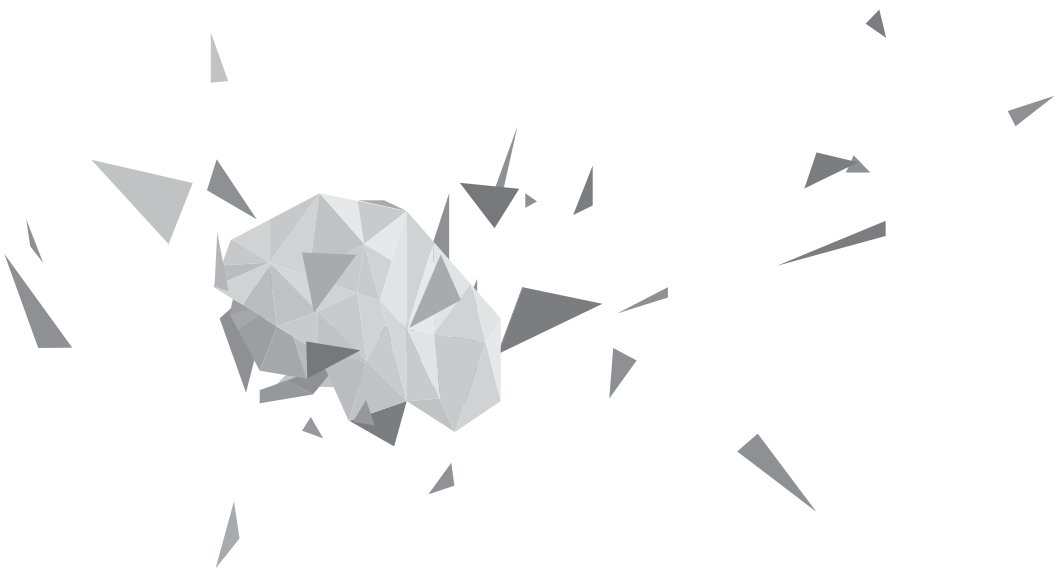
46. Kidwell CS, Liebeskind DS, Starkman S, Saver JL. **Trends in acute ischemic stroke trials through the 20th century.** *Stroke*. 2001; 32: 1349-59.

47. Sandset EC, Bath PM, Boysen G, Jatuzis D, Kõrv J, Lüders S, Murray GD, Richter PS, Roine RO, Terént A, Thijs V, Berge E; SCAST Study Group. **The angiotensin-receptor blocker candesartan for treatment of acute stroke (SCAST): a randomised, placebo-controlled, double-blind trial.** *Lancet.* 2011; 377(9767): 741-50.

48. The IST Collaborative Group. **The benefits and harms of intravenous thrombolysis with recombinant tissue plasminogen activator within 6 h of acute ischaemic stroke (the third international stroke trial [IST-3]): a randomised controlled trial.** *Lancet.* 2012; 379(9834): 2352-63.

49. Maas AI, Steyerberg EW, Marmarou A, McHugh GS, Lingsma HF, Butcher I, Lu J, Weir J, Roozenbeek B, Murray GD. **IMPACT recommendations for improving the design and analysis of clinical trials in moderate to severe traumatic brain injury.** *Neurotherapeutics*. 2010; 7(1): 127-34.

SUMMARY | SAMENVATTING



Summary

Chapter 1, the general introduction, provides the background, aim and research questions that are addressed in this thesis. Randomized controlled trials in traumatic brain injury (TBI) have a poor track record. Most trials that investigated potential interventions to improve outcome in patients with TBI showed a non-significant result. An important factor contributing to this failure is the large heterogeneity that exists among TBI patient populations. The **aim of this thesis** was to investigate which adaptations in design and analysis of randomized controlled TBI trials may be best implemented to deal with the heterogeneity of TBI study populations in order to increase chances for detecting clinically relevant treatment effects in future trials for therapeutic interventions for TBI.

Chapter 2 describes a literature survey on publications of randomized controlled trials (RCTs) that where performed in the last three decades among patients with TBI. 33 RCTs met the search criteria (published between 1980 and 2009). Only thirteen of these trials detected a significant treatment effect. The majority of these trials were 'single-center studies', investigating a 'therapeutic strategy' (e.g. hypothermia, decompressive craniectomy) and not a neuroprotective agent. The vast majority (22 of 33) of the studies were published in the 90s of the last century. Since that time, a strong decline was observed, especially in the number of trials that investigated neuroprotective agents. This decline might be explained by the fact that pharmaceutical companies were less interested in conducting such new (and expensive) trials, since previous neuroprotection trials almost all have failed to detect a significant treatment effect. In contrast, the number of studies investigating 'therapeutic strategies' is rising again.

Chapter 3 describes a literature survey on factors that predict mortality and functional outcome after moderate and severe TBI. Important factors scored on admission that predict the outcome are age, clinical severity of the injury (expressed by the 'motor score' of the Glasgow Coma Scale), structural abnormalities on CT imaging of the brain, hypoxia, hypotension, and laboratory abnormalities such as anemia and dysglycemia. These separate predictors were combined into two prognostic models for the prediction of outcome for individual patients at 2 weeks and/or 6 months after injury (mortality or unfavorable outcome (Glasgow Outcome Scale (GOS) ≤ 3)). Both logistic regression models were developed on datasets with prospectively collected individual patient data. For the first model, data from the CRASH trial was used (Corticosteroid Randomisation After Significant Head injury; n=10008). The other model was developed on the data from the IMPACT Database (International Mission on Prognosis and Clinical Trial Design in Traumatic Brain Injury; n=8509). Both models were externally validated reciprocally and are available online (http://www.crash2.lshtm.ac.uk/RiskCalculator/ and http://www.tbi-impact.org/). The models can be used in clinical practice, for clinical research and for policy making. Important

applications of prognostic models are risk stratification and adjustment of the treatment effect for prognostic factors in randomized controlled trials (covariate adjustment). To ascertain reliable predictions, continued external validation and adaptation of the models is crucial.

Chapter 4 describes the external validation of both the CRASH and the IMPACT models in five different cohorts of TBI patients (total n=9036). Three of these cohorts are from randomized controlled trials (NABIS Hypothermia, Cerestat and Pharmos) and two from observational series (APOE and TARN). We studied both discrimination ("how well does the model differentiate between patients with a favorable and an unfavorable outcome?") and calibration ("do the predicted outcomes agree with the observed outcomes?"). The best discrimination for both models was found in TARN, the poorest in Pharmos. For a large part, this can be explained by the fact that the study population of TARN is much more heterogeneous than the strictly selected homogeneous population of the Pharmos trial. In a heterogeneous population it is easier to predict which patients will do well and which will do poorly than in a homogeneous population. Calibration was reasonably good for both models in all validation cohorts. Comparing performance of the IMPACT models and the CRASH models, no meaningful differences were found. We concluded that both the IMPACT and the CRASH models are valid for use in both clinical research and daily clinical practice.

Chapter 5 describes an external validation study of the IMPACT Core and Extended models in the database of the Brain Trauma Foundation TBI-trac® registry on severe TBI in New York State (US). For this validation, 14-day mortality was used as the outcome measure of interest. Therefore, the models were re-fitted in the original IMPACT data. Excellent calibration and good discrimination were found for the models, proving that the IMPACT models also accurately predict early mortality after TBI.

In **Chapter 6** the influence of strict versus broad inclusion criteria for randomized controlled trials for TBI was described. For this study the data from the IMPACT Database (n=8033) was used, originating from three observational studies (n=2217) with broad enrollment criteria and six randomized controlled trials (n=5816) with relatively strict enrollment criteria. The outcome distribution (favorable versus unfavorable outcome based on the 6-month Glasgow Outcome Scale (GOS)) differed substantially between the studies. In the observational series the mean proportion of patients with an unfavorable outcome was 60%, while in the strictly selected trial population this was 43%. Different 'sets' with stricter inclusion criteria were applied to the original patient selections of these nine studies. This resulted in a change of the outcome distribution: the distribution of the observational series equaled the outcome distribution of the trial populations.

Chapter 7 describes the results of a simulation study on the influence of strict versus broad inclusion criteria with or without covariate adjustment on the chance of detecting a treatment effect when it exists (statistical power) in randomized controlled trials for patients with traumatic brain injury. In none of the eleven studies included in the IMPACT Database a statistically significant treatment effect was found. This is why we simulated a treatment effect (average absolute risk reduction of 10% in unfavorable outcome after 6 months). For the different patient selection strategies (strict versus broad) the required sample size was calculated for a statistical power of 80%. We found that the application of strict inclusion criteria resulted in a decrease in required sample size and thus an increase in statistical power. However, an important adverse effect of strict selection is the prolongation of the study duration, because it will take more time to include enough patients since fewer patients are eligible for inclusion. An alternative strategy is to use more liberal enrollment criteria and to adjust the treatment effect for other relevant predictors of the outcome (covariate adjustment). Our analysis showed that this latter approach is more efficient.

Chapter 8 describes a re-analysis of the CRASH (Corticosteroid Randomisation After Significant Head Injury) trial data with ordinal rather than dichotomous outcome analysis of the GOS. Almost all TBI trials use the GOS as the primary outcome measure, mostly at 6 months after the initial trauma. The 5-point, ordinal scale is dichotomized as unfavorable ('Dead', 'Vegetative State', 'Severe Disability') versus favorable ('Moderate Disability' and 'Good Recovery') outcome. This dichotomization results in a loss of information. Previous simulation studies within the IMPACT Database have shown that the use of ordinal analysis techniques led to a gain in statistical power. Examples of such techniques are 'proportional odds logistic regression' and the 'sliding dichotomy approach'. In the CRASH trial more than 10,000 patients were enrolled to investigate the efficacy of corticosteroids in TBI patients. The trial was stopped early because of a higher proportion of mortality in the active treatment arm of the study. The analyses that are described in Chapter 6 have shown that with the application of ordinal techniques this difference could already have been detected earlier, i.e. after inclusion of approximately half the number of patient that was needed for the unadjusted dichotomous analysis. Moreover, the application of ordinal analysis will make it possible to detect smaller but clinically relevant treatment effects. This observation resulted in the advice to discard dichotomization of the GOS and to adopt ordinal approaches to the analysis of future clinical TBI trials.

Chapter 9 discusses the issue of heterogeneity in TBI trials and summarizes the value of different innovative methods for the design and analysis of randomized controlled trials in TBI, that were all separately discussed in the previous Chapters of this thesis. Further, alternative strategies to investigate clinical benefits of established and novel therapies for TBI are presented and discussed, such as comparative effectiveness research.

In **Chapter 10,** the general discussion, the results that were presented in thesis are discussed and the research questions that were posed in Chapter 1 are answered. A set of recommendations for the design and analysis of future TBI trials is presented. The use of these recommendations in ongoing/planned TBI trials, as well as their applicability in other fields of clinical research are discussed. These recommendations include:

- Details of the major baseline prognostic characteristics should be provided in every report on a TBI study; in trials they should be differentiated per treatment group. We also advocate the reporting of the baseline prognostic risk as determined by validated prognostic models.
- Inclusion criteria should be as broad as is compatible with the current understanding of the mechanisms of action of the intervention being evaluated. This will maximize recruitment rates and enhance the generalizability of the results.
- The statistical analysis should incorporate (prespecified) covariate adjustment to mitigate the effects of the heterogeneity.
- The statistical analysis should use an ordinal approach, based on either sliding dichotomy or proportional odds methodology.

Samenvatting

In **Hoofdstuk 1,** de algemene introductie, wordt de achtergrond van het onderzoek beschreven en worden het doel en de onderzoeksvragen uiteengezet. Gerandomiseerde klinische studies bij patiënten met traumatisch hersenletsel hebben een slechte reputatie. Bijna geen enkele van de trials die de werkzaamheid van potentiële interventies ter verbetering van de uitkomst van patiënten met traumatisch hersenletsel onderzochten, vond een statisch significant behandelingseffect. Een belangrijke verklaring van dit falen zou kunnen liggen in de heterogeniteit van de patiëntenpopulaties met traumatisch hersenletsel. Het was het **doel van dit proefschrift** om te onderzoeken welke aanpassingen in het ontwerp en de analyse van gerandomiseerde klinische studies bij patiënten met traumatisch hersenletsel moeten worden gedaan, om om te gaan met deze heterogeniteit en om zo de kans op het aantonen van klinisch relevante behandelingseffecten in toekomstige studies te vergroten.

In **Hoofdstuk 2** wordt een literatuurstudie beschreven naar publicaties over klinische studies die in de afgelopen decennia werden uitgevoerd bij patiënten met traumatisch hersenletsel. Er werden 33 trials gevonden die voldeden aan de zoekcriteria (gepubliceerd tussen 1980 en 2009). Slechts dertien trials detecteerden een significant behandelingseffect. Het merendeel van de deze trials bleken 'single-center studies' te zijn, die een zogenaamde behandelingsstrategie (hypothermie, decompressieve craniectomie, etc.) en dus niet een neuroprotectief geneesmiddel onderzochten. Het overgrote deel (22 van de 33) van de gevonden studies werd gepubliceerd in de jaren '90 van de vorige eeuw. Sindsdien werd een sterke daling waargenomen, voornamelijk in het aantal gepubliceerde trials dat neuroprotectieve geneesmiddelen onderzocht. Deze daling zou verklaard kunnen worden door het feit dat farmaceutische bedrijven waarschijnlijk minder geïnteresseerd zijn in het opzetten van nieuwe (en dure) neuroprotectie-trials, omdat eerdere studies zulke teleurstellende resultaten hebben opgeleverd. Overigens neemt het aantal studies dat een behandelingsstrategie onderzoekt de laatste jaren weer gestaag toe.

Hoofdstuk 3 beschrijft een literatuurstudie naar factoren die van invloed zijn op de sterfte en functionele uitkomst na middelernstig en ernstig traumatisch hersenletsel.

Belangrijke factoren die worden gescoord bij opname en de uitkomst kunnen voorspellen zijn de leeftijd, de klinische ernst van het hersenletsel (uitgedrukt in de 'motor score' van de Glasgow Coma Score), afwijkende bevindingen op de CT-scan van de hersenen, hypoxie, hypotensie en afwijkende laboratoriumwaarden, zoals anemie en dysglycaemie. Deze 'losse' voorspellende factoren zijn samengevoegd tot twee prognostische modellen ter voorspelling van de uitkomst 2 weken en/of 6 maanden na het trauma (sterfte of ongustige uitkomst (Glasgow Outcome Scale (GOS) ≤ 3)) van individuele patiënten. Beide logistische regressiemodellen werden gebaseerd op individuele patiëntendata, afkomstig uit grote prospectief verzamelde databases. Voor het eerste model werd gebruikgemaakt van de data van de CRASH trial ('Corticosteroid Randomisation After Significant Head injury'; n=10008).

Het andere model werd gebaseerd op patiëntendata uit de IMPACT Database ('International Mission on Prognosis and Clinical Trial Design in Traumatic Brain Injury'; n=8509). Beide modellen werden op elkaars datasets extern gevalideerd en zijn online beschikbaar in de vorm van een webcalculator (http://www.crash2.lshtm.ac.uk/RiskCalculator/ en http://www.tbi-impact.org/). De modellen zijn te gebruiken voor verschillende doeleinden, zowel in de klinische praktijk, voor klinisch wetenschappelijk onderzoek en voor beleidsmakers in de gezondheidszorg. Onder de belangrijkste toepassing van prognostische modellen vallen stratificatie op basis van het prognostisch risico en correctie voor voorspellende factoren in klinische studies ('covariate adjustment').

Hoofdstuk 4 beschrijft de externe validatie van zowel de CRASH als de IMPACT modellen in vijf verschillende cohorten patiënten (n=9036). Drie van deze cohorten zijn afkomstig van gerandomiseerde klinische trials (NABIS Hypothermia, Cerestat en Pharmos), de andere twee van observationale studies (APOE en TARN). Er werd gekeken naar de discriminatie ("hoe goed kan het model differentiëren tussen patiënten met een goede en een slechte uitkomst?") en naar de calibratie ("komen de voorspelde uitkomsten overeen met de geobserveerde uitkomsten?"). De beste discriminatie voor beide modellen werd gevonden in TARN, de slechtste in Pharmos. Dit is goed te verklaren door het feit dat TARN een veel heterogenere studiepopulatie had, terwijl Pharmos een zeer strikt geselecteerde trial betrof.

In een heterogene populatie is het gemakkelijker te voorspellen welke patiënten een goede of slechte uitkomst zullen hebben, dan in een homogene populatie. De calibratie was voor de beide modellen redelijk tot goed in de verschillende validatie-cohorten. Geconcludeerd kan worden dat zowel de IMPACT als de CRASH modellen naar behoren voorspellen in de nieuwe cohorten en dat de modellen betrouwbaar zijn voor gebruik in wetenschappelijk onderzoek en de klinische praktijk.

Hoofdstuk 5 beschrijft een externe validatie studie van de IMPACT 'Core' en 'Extended'modellen in de database van het 'Brain Trauma Foundation TBI-trac® register'voor patiënten met ernstig traumatisch hersenletsel in de staat New York in de Verenigde Staten. We gebruikten de sterfte binnen 14 dagen na het trauma als uitkomstmaat binnen deze validatiestudie. Omdat de originele uitkomstmaat sterfte of ongustige uitkomst op de GOS na 6 maanden was, zijn de modellen voor deze validatiestudie "ge-refit" op basis van een nieuwe uitkomst in de originele data van de IMPACT database. Voor deze nieuwe uitkomst vonden we een excellente calibratie en goede discriminatie voor de modellen. Dit is bewijzend dat de IMPACT modellen ook betrouwbaar in staat zijn om vroege sterfte na ernstig traumatisch hersenletsel te voorspellen.

In **Hoofdstuk 6** wordt beschreven wat de invloed is van ruime of juist strikte inclusiecriteria voor klinische trials bij patiënten met traumatisch hersenletsel. Hiervoor werd gebruikgemaakt van patiëntendata uit de IMPACT Database (n=8033), oorspronkelijk afkomstig uit drie observationele studies (n=2217) met ruime inclusiecriteria en zes klinische

trials (n=5816) met striktere inclusiecriteria. De gemiddelde uitkomstverdeling (ongunstige versus gunstige uitkomst op de GOS na 6 maanden) in de observationele studies en de klinische trials verschilden substantieel van elkaar. Gemiddeld was de proportie patiënten met een ongunstige uitkomst 60% in de observationele studies, terwijl dit binnen de strikt geselecteerde trial 43% was. Verschillende 'sets' met striktere inclusiecriteria werden toegepast op de oorspronkelijke patiëntenselectie van de 9 studies. De uitkomstverdeling van de observationele studies veranderde na toepassing van de strikte inclusiecriteria zodanig, dat deze gelijk werd aan de verdeling binnen de klinische trials.

In **Hoofdstuk 7** worden de resultaten van een simulatiestudie gepresenteerd naar de invloed van strikte of juist ruime inclusiecriteria met of zonder 'covariate adjustment' op de kans op het vinden van een behandelingseffect (power) in klinische trials bij patiënten met traumatisch hersenletsel. In geen van de 11 studies binnen de IMPACT database werd een statisch significant behandelingseffect aangetoond. Daarom werd een behandelingseffect gesimuleerd (gemiddelde absolute risicoreductie van 10% op ongunstige uitkomst na 6 maanden). Voor de verschillende selectiestrategieën (strikt en ruim) werd de benodigde 'sample size' berekend voor het behalen van een power van 80%. Dit leidde tot het inzicht dat strikte selectie leidt tot een afname van de benodigde 'sample size' en dus tot een verhoging van de power. Een belangrijk nadeel van deze strikte selectie is echter dat de duur van de studie hierdoor toeneemt, omdat minder patiënten in aanmerking komen voor inclusie en het dus langer duurt voordat het benodigde aantal patiënten is geïncludeerd.

Een alternatief zou kunnen zijn om een liberaal inclusieregime te gebruiken, waarbij in de analysefase van de studie gecorrigeerd wordt voor patiëntkarakteristieken die voorspellend zijn voor de uitkomst ('covariate adjustment'). Vergelijkende berekeningen resulteerden in het inzicht dat deze laatste strategie efficiënter is vanuit het oogpunt van trial design.

Hoofdstuk 8 beschrijft een heranalyse van de CRASH trial data met ordinale in plaats van dichotome uitkomstanalyse van de GOS. Bijna alle traumatisch-hersenletsel-trials gebruiken de GOS als de primaire uitkomstmaat, meestal 6 maanden na het trauma. Deze 5-punts, ordinale schaal wordt dan gedichotomiseerd als ongunstige (dood; vegetatieve toestand; ernstige invaliditeit (ADL afhankelijk)) versus gunstige (matige invaliditeit (ADL-zelfstandig, maar met beperkende restverschijnselen) of minimale of geen invaliditeit) uitkomst.

Door deze dichotomisatie gaat informatie verloren. Eerdere simulatiestudies binnen de IMPACT database lieten zien dat het gebruik van technieken waarbij de GOS niet wordt gedichotomiseerd, maar ordinaal wordt geanalyseerd leidden tot een winst in power.

Voorbeelden van deze technieken zijn 'proportional odds logistische regressie' en de 'sliding dichotomy' aanpak. In de CRASH trial werden ruim 10.000 patiënten met traumatisch hersenletsel geïncludeerd om te onderzoeken of toediening van corticosteroïden de uitkomst zou verbeteren. De trial werd vroegtijdig gestaakt, in verband met oversterfte in de corticosteroïdgroep. De in Hoofdstuk 6 beschreven analyses laten zien dat bij gebruikmaking

van ordinale analysetechnieken in de CRASH trial dit verschil al na inclusie van een kleinere groep patiënten aangetoond had kunnen worden, namelijk na inclusie van ongeveer de helft van het aantal patiënten dat nodig was voor de ongecorrigeerde dichotome analyse.

Daarnaast wordt het door gebruikmaking van ordinale analysetechnieken mogelijk om kleinere, maar wellicht weldegelijk klinisch relevante behandelingseffecten te detecteren. Er wordt dan ook geadviseerd om ordinale uitkomstschalen niet langer te dichotomiseren, maar in toekomstige trials ordinale analysetechnieken toe te passen.

In **Hoofdstuk 9** wordt het probleem van heterogeniteit binnen de populaties van traumatisch-hersenletsel-trials besproken en wordt de waarde van de eerder beschreven innovatieve methoden voor het ontwerp en de analyse van gerandomiseerde studies samengevat. Daarnaast worden alternatieve studieopzetten besproken waarmee de werkzaamheid en effectiviteit van bestaande en nieuwe behandeling voor traumatisch hersenletsel kunnen worden onderzocht, zoals bijvoorbeeld 'comparative effectiveness research'.

In **Hoofdstuk 10**, de algemene discussie, worden de bevindingen die zijn beschreven in dit proefschrift bediscussieerd en de onderzoeksvragen die staan geformuleerd in Hoofdstuk 1 beantwoord. Er wordt een serie aanbevelingen gedaan voor het ontwerp en de analyse van toekomstige gerandomiseerde klinische studies bij traumatisch hersenletsel. Het gebruik van deze aanbevelingen in lopende en geplande studies voor traumatisch hersenletsel wordt besproken. Daarnaast wordt ook de toepasbaarheid binnen andere vakgebieden in de geneeskunde toegelicht. De aanbevelingen luiden als volgt:

- Iedere publicatie over een klinische studie naar traumatisch hersenletsel dient gedetailleerde informatie over belangrijke voorspellende variabelen bij opname te bevatten; bij RCT's moet deze informatie worden gedifferentieerd naar behandelgroep. Wij pleiten daarnaast voor rapportage van het prognostisch risico van de geïncludeerde patiënten bij opname, geschat met behulp van gevalideerde prognostische modellen.
- Inclusiecriteria dienen zo breed mogelijk te zijn, zo lang als deze verenigbaar zijn met de te verwachte werkzaamheid op mechanistisch/pathofysiologisch niveau. Dit zal de inclusiesnelheid maximaliseren en de generaliseerbaarheid van de resultaten vergroten.
- Binnen de statistische analyse dient correctie voor (vooraf gespecificeerde) voorspellende factoren ("covariate adjustment") te worden toegepast, om de nadelige effecten van een heterogene patiëntenpopulatie te verkleinen.
- Er dient gebruik te worden gemaakt van een ordinale methode voor de analyse van het behandelingseffect, gebaseerd op de "sliding dichotomy" aanpak of op het "proportional odds" model.

Summary | Samenvatting

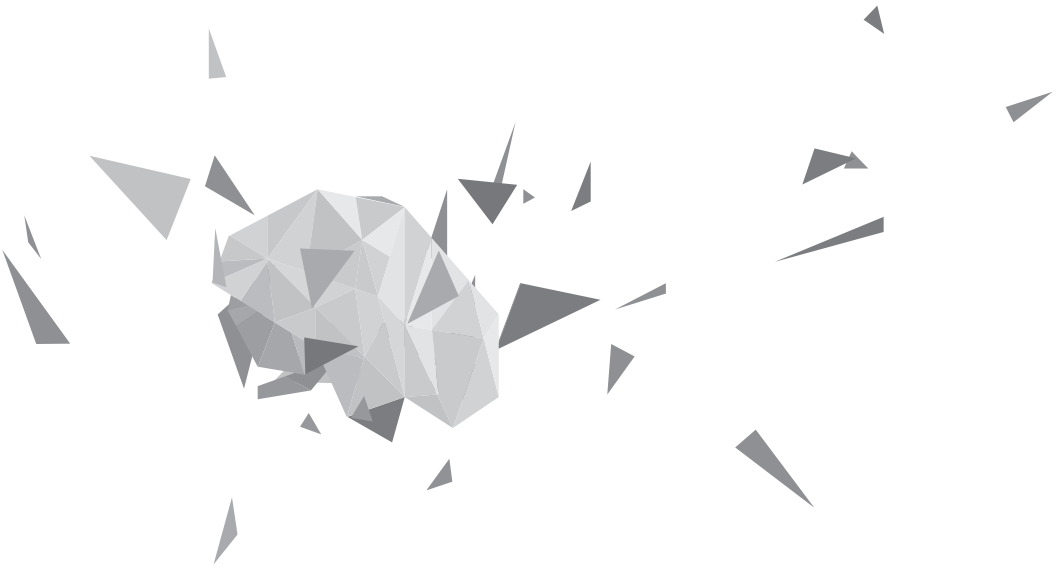
CONTRIBUTING AUTHORS

Dankwoord | Acknowledgements

Curriculum Vitae

List of Publications

Portfolio



Contributing Authors

Isabella Butcher, PhD
Statistician, Research Fellow
Centre for Population Health Sciences, University of Edinburgh, United Kingdom

Ya-Lin Chiu, MSc
Statistician, Research Fellow
Department of Public Health, Division of Biostatistics and Epidemiology, Weill Cornell Medical College, New York, NY, United States

Phil Edwards, BSc, CStat, PhD
Senior Lecturer Statistics
Clinical Trials Unit, London School of Hygiene & Tropical Medicine, London, United Kingdom

Linda M. Gerber, PhD, MA
Professor of Public Health and Epidemiology in Medicine
Department of Public Health, Division of Biostatistics and Epidemiology, Weill Cornell Medical College, New York, NY, United States

Jamshid Ghajar, MD, PhD, FACS
Professor of Neurological Surgery
NewYork-Presbyterian Hospital, NY, United States
Brain Trauma Foundation, New York, NY, United States

Fiona E. Lecky, MD, PhD
Professor of Emergency Medicine, Director of the Trauma Audit and Research Network (TARN)
School of Health and Related Research (ScHARR), University of Sheffield, United Kingdom

Hester F. Lingsma, PhD
Epidemiologist, Postdoctoral Researcher
Department of Public Health, Erasmus MC, Rotterdam, The Netherlands

Juan Lu, MD, MPH, PhD
Assistant Professor of Epidemiology
Department of Epidemiology and Community Health, Virginia Commonwealth University, VA, United States

Andrew I.R. Maas, MD, PhD
Professor and Chairman of Neurosurgery
Department of Neurosurgery, Antwerp University Hospital, Edegem, Belgium

Geoffrey T. Manley, MD, PhD
Professor and Chief of Neurosurgery
San Francisco General Hospital / University of California San Francisco (UCSF), CA, United States

Gillian S. McHugh, PhD
Statistician, Research Fellow
Centre for Population Health Sciences, University of Edinburgh, United Kingdom

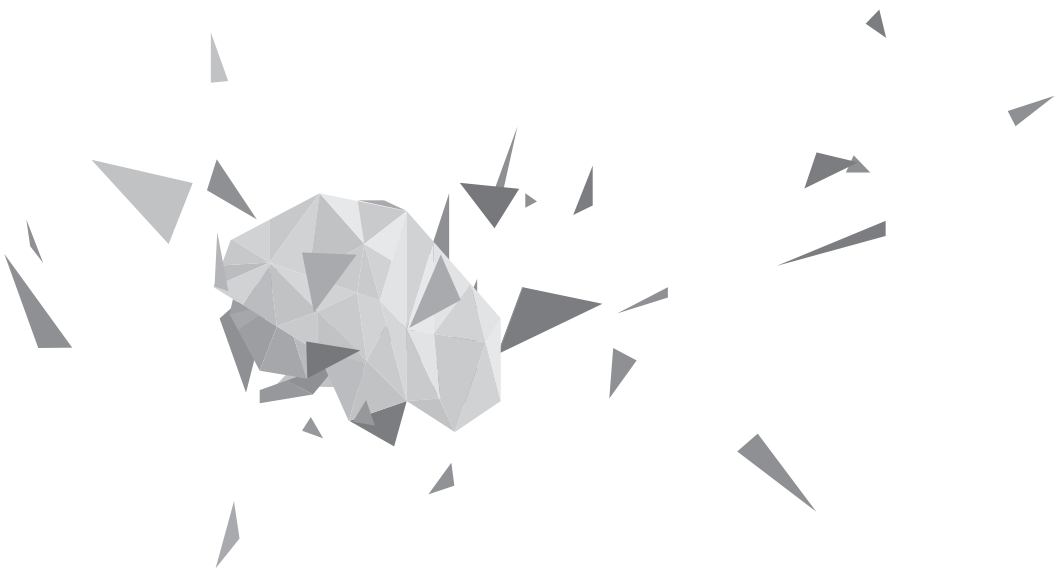
Gordon D. Murray, PhD
Professor of Medical Statistics
Centre for Population Health Sciences, University of Edinburgh, United Kingdom

Pablo Perel, MD, MSc, PhD
Senior Lecturer of Epidemiology
Clinical Trials Unit, London School of Hygiene & Tropical Medicine, London, United Kingdom

Ian Roberts, MB, BCh, FRCP, FPH
Professor of Epidemiology and Public Health, Director Clinical Trials Unit
Clinical Trials Unit, London School of Hygiene & Tropical Medicine, London, United Kingdom

Ewout W. Steyerberg, PhD
Professor of Medical Decision Making
Department of Public Health, Erasmus MC, Rotterdam, The Netherlands

James Weir, PhD
Statistician, Research Fellow
Centre for Population Health Sciences, University of Edinburgh, United Kingdom



Dankwoord |Acknowledgements

Ik beschouw dit proefschrift als het product van een mooi Belgisch-Nederlands avontuur, waaraan met het verschijnen van dit boekje een definitief einde komt. Mijn dank gaat uit naar allen die bij dit avontuur betrokken zijn geweest. Voor hun bijdrage aan de totstandkoming van dit proefschrift wil ik een aantal personen op deze plaats specifiek bedanken.

Mijn Belgische promotor: prof. dr. A.I.R. Maas
Beste Andrew. Wat ben ik blij dat ik als student-assistent een plekje toebedeelt kreeg op de werkkamer van jouw vroegere secretaresse Marja in Rotterdam. Ik zat daar te werken aan een (weinig vruchtbaar gebleken) onderzoekje naar brughoektumoren. Het was in die kamer waar wij elkaar voor het eerst ontmoetten. En jij maakte mij daar al snel duidelijk dat ik maar voor jou moest komen werken. Je gaf mij na mijn artsexamen de mogelijkheid om promotieonderzoek te doen. Daarnaast bood je mij een opleidingsplaats tot neurochirurg in Antwerpen aan. Andrew, ik ben je dankbaar voor alle kansen die jij me hebt geboden. Je schrok even toen ik je nog niet zo lang geleden vertelde dat ik had besloten om geen neurochirurg maar neuroloog te worden. Ondanks dat je nu niet meer mijn “stagemeester” bent, zul je door jouw “skills” als creatieve onderzoeker, lobbyist en dokter met oog voor detail én voor de patiënt als geheel, voor mij een groot voorbeeld blijven.

Mijn Nederlandse promotor: prof. dr. E.W. Steyerberg
Beste Ewout. Andrew bracht ons met elkaar in contact toen ik nog geneeskunde studeerde. Het feit dat ik maar een simpele dokter in de dop was, heeft jou er niet van weerhouden mij te stimuleren om allerlei ingewikkelde R codes te schrijven en statistische analyses te doen. Zeker in de begintijd ging ik na een gesprek met jou met meer vragen de deur uit, dan ik naar binnen was gekomen. Gelukkig heb ik in de afgelopen jaren veel van jou mogen leren. Daarnaast is een goed ontwikkeld gevoel voor humor jou niet vreemd en verliep onze samenwerking steeds in goede harmonie. Hiervoor wil ik je van harte bedanken. De bestaande goede banden tussen jouw onderzoeksgroep en de afdeling Neurologie van het Erasmus MC zullen er hopelijk aan bijdragen dat wij ook in de toekomst weer gezamenlijk aan onderzoeksprojecten zullen werken.

“The IMPACT Family”
At the moment that I joined the IMPACT Study Group in 2008, a lot of work had already been done. Many months - or even years - were invested in the preparation of the IMPACT Database for the analyses that were described in this thesis. I thank the IMPACT Teams in Edinburgh, Richmond and Antwerp for their contributions to this project: Professor Gordon Murray, Dr Izzy Butcher, Dr Gillian McHugh, Dr Jim Weir, Morag Leitch, Dr Juan Lu, Anne-

Claire van Harderwijk and Véronique de Keyser. Professor Anthony Marmarou: with you the neurotraumatology society has lost a great TBI researcher. I thank the members of the IMPACT Advisory Board for their wisdom: Professor Larry Marshall, Dr Elana Farace, Professor Nino Stocchetti, Professor Jim Torner, Professor Alex Valadka, Professor Lindsay Wilson, Professor David Yates and Dr Ramona Hicks.

Contributing authors
Professor Geoff Manley, thanks for your help with Chapter 2. Professor Fiona Lecky, thanks for supplying the TARN-TBI data for Chapter 4. For their contribution to Chapter 5, I would like to thank our colleagues in New York City: Professor Jam Ghajar, Professor Linda Gerber and Dr Ya-Lin Chiu. Professor Ian Roberts, Dr Pablo Perel and Dr Phil Edwards, a.k.a. the CRASH Trial Collaborators: thanks for making the Chapters 4 and 8 possible.

Leden van de Rotterdamse leescommissie
Prof. dr. M.G. Hunink, prof. dr. D.W.J. Dippel en prof. dr. Ph.G. Jorens. Ik dank u allen hartelijk dat u zitting hebt willen nemen in de Rotterdamse leescommissie en dat u het manuscript binnen korte tijd heeft willen beoordelen.

Leden van de Antwerpse individuele doctoraatscommissie
Prof. dr. Ph.G. Jorens, prof. dr. P.M. Parizel en prof. dr. P. Weyler. Dank dat u de voortgang van mijn doctoraat heeft willen bewaken en dat u het proefschrift heeft willen beoordelen. En dank dat u naar Rotterdam bent gekomen om zitting te nemen in de promotiecommissie, of zoals dat in België heet: de doctoraatsjury.

Overige leden van de promotiecommissie
Prof. dr. ir. H. Boersma. Dank dat u bereid bent zitting te nemen in de grote commissie. Prof. dr. P.A. van Doorn, beste Pieter. Bedankt dat je de neuroloog van de grote commissie wilt zijn. Ik kijk uit naar onze verdere samenwerking tijdens mijn opleiding in de kliniek. Prof. dr. J.D.F. Habbema, beste Dik. Dank voor je immer wijze advies als lid van de IMPACT Advisory Board en voor het feit dat je opponent wil zijn. Professor Sir Graham Teasdale. As the co-inventor of the Glasgow Coma Scale and founding father of the current approach to the clinical management of traumatic brain injury, I consider it a great honour that you have accepted our invitation to come over to The Netherlands to be an opponent at my thesis defence. Thanks for the advice that you gave us as chairman of the IMPACT Advisory Board. And thanks for the “data” that you supplied in Noordwijk this year.

Mijn paranimfen: dr. H.F. Lingsma en drs. M.H.T. van der Vaart
Beste Hester. Wat heb ik een geluk gehad dat jij als promovendus werd aangesteld, op het moment dat ik begon als student-assistent bij Ewout. Je hebt me met jouw supersnelle inzicht en daadkracht enorm geholpen. En wat hebben wij enorm veel gelachen samen: in de korte tijd dat ik op MGZ werkte, maar vooral ook tijdens de vele IMPACT-gerelateerde reisjes rond de wereld. Ik vind het een groot genoegen dat jij tijdens de verdediging als paranimf aan mijn zijde wil staan. En misschien mag je ook wel een vraag beantwoorden… ;-)

Beste Thijs. Wat vind ik het leuk dat wij na onze tijd als leden van het 38ste MFVR-bestuur zulke goede vrienden zijn gebleven. Wij hebben veel gemeen. Allebei zijn we 'rete-ambitieus' en samen filosoferen we graag over hoe wij denken dat het in de toekomst allemaal anders (=beter?) kan. Ik vind het geweldig dat je mij als paranimf bij wilt staan. Ik wens je veel succes met de voltooiing van je eigen proefschrift en ik kijk er al naar uit om straks in de kliniek met jou als collega-neuroloog i.o. samen te werken.

Oud-collega-A.S.O.'s neurochirurgie en neurochirurgen in Antwerpen
Beste Niels, Mark, Bart en Süleyman. Veel dank voor de mooie tijd in het Belgische. Dat ik mij mateloos kon storen aan ons "Antwerpse zooitje ongeregeld" is jullie vast niet ontgaan, maar dat ik jullie vanuit de strak georganiseerde neurologie-afdeling in Rotterdam nog wel eens mis zal ik allerminst ontkennen. Ik wens jullie succes met de verdere opleiding en met jullie eigen onderzoek.

Beste Tomas, Dirk, Ricky, Mania en (weer) Niels. Het spreekwoordelijke "diepe" dat ik na mijn artsexamen door jullie werd ingegooid heeft mij een boel ervaring opgeleverd. Ik heb in de jaren die ik in Antwerpen heb doorgebracht veel van jullie mogen leren, waar ik ook als aankomend neuroloog mijn voordeel mee hoop te doen. Mijn hartelijke dank daarvoor.

Collega-A(N)IOS neurologie en neurologen in Rotterdam
In 2011 waren wij al tijdelijk collega's en dit is mij zo goed bevallen, dat ik sinds juni van dit jaar definitief ben teruggekeerd. Dank voor jullie ondersteuning en adviezen in de afrondingsfase van dit proefschrift. Ik kijk uit naar onze verdere samenwerking de komende tijd.

Oud-collega's bij de MGZ en de Heelkunde in het Albert Schweitzer ziekenhuis
Alhoewel ik maar een aantal maanden fysiek op de afdeling Maatschappelijke Gezondheidszorg heb doorgebracht, heb ik hier goede herinneringen aan over gehouden. Alle MGZ-collega's wil ik bedanken voor de prettige samenwerking. Specifiek wil ik Nikki van Leeuwen bedanken voor de samenwerking aan de dankzij jou supersnelle productie van

het “extracranial injury” artikel. En kamergenoot Frans-Willem Goudsmit dank ik voor de gezellige tijd in kamer Ae-103.

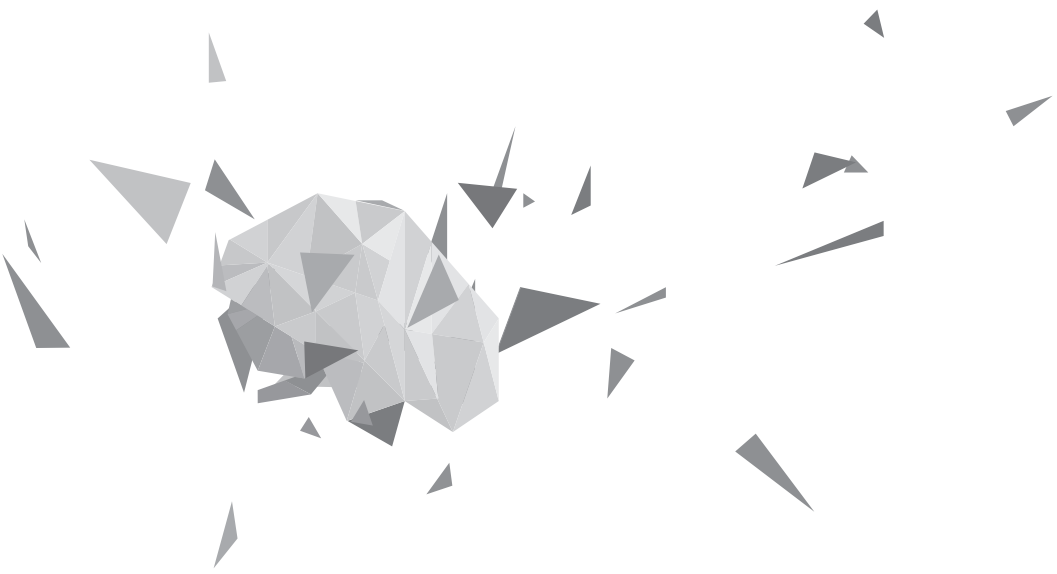
Alhoewel ik nu neuroloog in wording ben, heb ik in 2011 ook nog een halfjaar als AIOS op de Heelkunde stage mogen lopen in het Albert Schweitzer ziekenhuis. Ik wil de Dordtse collega-A(N)IOS en de chirurgen hartelijk danken voor de mooie tijd. De manier waarop jullie voortvarend de geneeskunde bedrijven sprak mij erg aan. Ik zal proberen deze aanpak daar waar mogelijk ook in de neurologie toe te passen.

Zusje en ouders
Lieve Pauline. Ik trots op jou als ‘mijn kleine zusje’ met alle capaciteiten die je bezit. Veel dank dat je naast je drukke baan ook bereid bent om een helpende hand te bieden in de organisatie van de bij een promotie behorende festiviteiten.

Lieve papa en mama. Ik prijs mezelf gelukkig dat jullie mijn ouders zijn. Jullie steunen mij altijd onvoorwaardelijk in alle keuzes die ik maak. Of ik nou als peuter een “wasmachine-tic” had, als kleine jongen op ballet wilde of als student een jaar niet wilde studeren maar besturen: het kon bij jullie allemaal. Dank voor jullie vertrouwen!

Kim
Lieve schat. Met het verschijnen van dit proefschrift komt definitief een einde aan de tijd waarin ik een deel van de week in Antwerpen verbleef. Alhoewel ik wist dat jij het niet erg zou vinden als ik weer in Nederland zou komen werken, heb jij jezelf in de fase dat ik tot dat besluit moest komen op de achtergrond gehouden. Je vond dat het mijn beslissing moest zijn. Dank daarvoor. Wel vroeg je mij onlangs of ik na het afronden van mijn promotie eens zou stoppen om ’s avonds thuis achter de computer te kruipen om weer aan het werk te gaan. Ik zal het proberen… :-). Je hebt me inmiddels volledig laten inzien dat het leven samen veel leuker is dan alleen. Ik hoop dat wij met z’n tweeën nog veel mooie dingen gaan beleven!

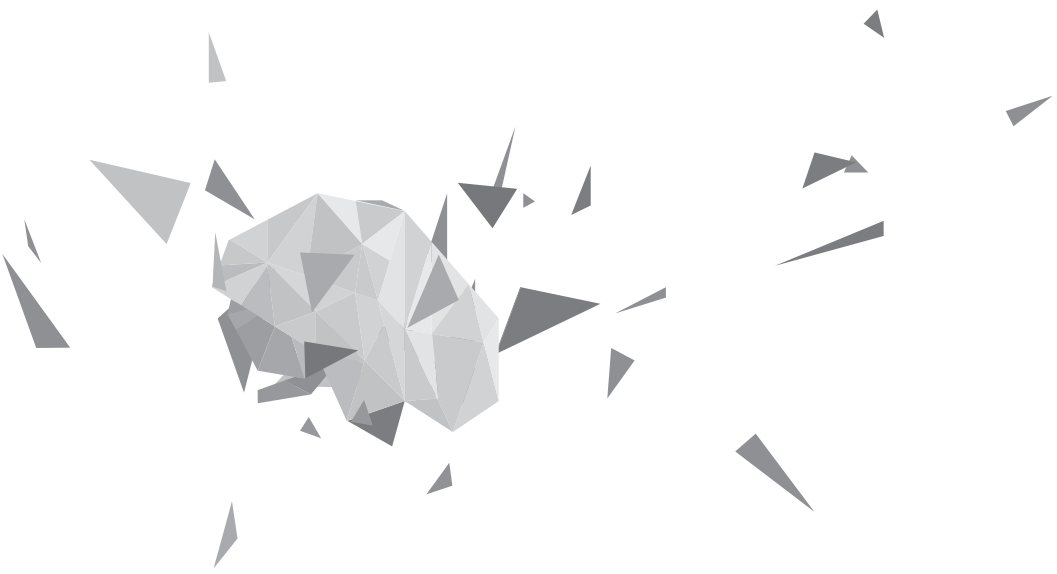
Bob Roozenbeek



Curriculum vitae

Bob Roozenbeek was born on January 14, 1983 in Alkmaar, The Netherlands. After finishing secondary school at the Murmellius Gymnasium Alkmaar, he started Medical School at Erasmus MC Rotterdam in 2002. During his medical training, he worked parttime as a scientific research assistant in the departments of Neurosurgery and Public Health of Erasmus MC. He further was an active member of the Rotterdam Medical Student Association (MFVR); he was President of the board of MFVR in 2004-2005. In 2008, he finished medical school ('doctoraalexamen'). His Master thesis ('keuzeonderzoek') on the influence of strict vs. broad enrollment criteria on the power of randomized controlled trials for traumatic brain injury was awarded with the "Gerrit Jan Mulder Prize" for the best research performed by medical student in 2009. In the same year he graduated 'cum laude' as a medical doctor ('artsexamen').

After his graduation, he worked for seven months as a PhD student at the department of Public Health of Erasmus MC where he worked on the research presented in this thesis (supervisor: prof. dr. E.W. Steyerberg). In September 2009, he started his specialisation in neurosurgery at the department of Neurosurgery of the Antwerp University Hospital in Belgium (supervisor: prof. dr. A.I.R. Maas). A 2-year research program was included as part of the residency program (2009-2011). A collaboration agreement for a 'joint doctorate' was established between Erasmus University Rotterdam and Antwerp University (supervisors: prof. dr. E.W. Steyerberg and prof. dr. A.I.R. Maas), of which this PhD thesis is the final product. In 2011, he fullfilled clinical rotations in general surgery (Albert Schweitzer Hospital, Dordrecht, supervisors: dr. R.J. Oostenbroek and dr. P.W. Plaisier) and neurology (Erasmus MC, Rotterdam, supervisor: prof. dr. P.A.E. Sillevis Smitt). In 2012 he decided to change gears and to switch from neurosurgery to neurology. In June 2012 he started as a neurologist-in-training at the department of Neurology of Erasmus MC (supervisor: prof. dr. P.A.E. Sillevis Smitt). Currently he lives in Rotterdam, together with his girlfriend Kim van Kessel.



List of Publications

Journal Articles

Roozenbeek B, Maas AI, Marmarou A, Butcher I, Lingsma HF, Lu J, McHugh GS, Murray GD, Weir J, Steyerberg EW: **The influence of enrolment criteria on recruitment and outcome distribution in traumatic brain injury studies: results from the IMPACT Study.** *J Neurotrauma* 2009; 26(7): 1069-75

Roozenbeek B, Maas AI, Lingsma HF, Butcher I, Lu J, Marmarou A, McHugh GS, Weir J, Murray GD, Steyerberg EW: **Baseline characteristics and statistical power in randomized controlled trials: selection, prognostic targeting, or covariate adjustment?** *Crit Care Med* 2009; 37(10): 2683-90

Maas AI, Roozenbeek B, Manley GT: **Clinical trials in traumatic brain injury: past experience and current developments.** *Neurotherapeutics* 2010; 7(1): 115-26

Maas AI, Steyerberg EW, Marmarou A, McHugh GS, Lingsma HF, Butcher I, Lu J, Weir J, Roozenbeek B, Murray GM: **IMPACT recommendations for improving the design and analysis of clinical trials in moderate to severe traumatic brain injury.** *Neurotherapeutics* 2010; 7(1): 127-34

Lingsma HF, Roozenbeek B, Steyerberg EW, Murray GM, Maas AI: **Early prognosis in traumatic brain injury: from prophecies to predictions.** *Lancet Neurol* 2010; 9: 543-54

Roozenbeek B, Lingsma HF, Steyerberg EW, Maas AI: **Underpowered trials in critical care medicine: how to deal with them?** *Crit Care* 2010; 14(3): 423

Lingsma HF, Roozenbeek B, Steyerberg EW: **Covariate adjustment increases statistical power in randomized controlled trials.** *J Clin Epidemiol* 2010: 63(12); 1391

Lingsma HF, Roozenbeek B, Li B, Lu J, Weir J, Butcher I, Marmarou A, Murray GD, Maas AI, Steyerberg EW: **Large between-center differences in outcome after moderate and severe Traumatic Brain Injury in the IMPACT Study.** *Neurosurg* 2011: 68(3); 601-7

Roozenbeek B, Lingsma HF, Perel P, Edwards P, Roberts I, Murray GD, Maas AIR, Steyerberg EW: **The added value of ordinal analysis in clinical trials: an example in traumatic brain injury.** *Critical Care* 2011: 15(3); R127

Lingsma HF, Roozenbeek B, Perel P, Roberts I, Maas AI, Steyerberg EW: **Between-centre differences and treatment effects in randomized controlled trials: a case study in traumatic brain injury.** *Trials* 2011: 12; 201

Weir J, Steyerberg EW, Butcher I, Lu J, Lingsma H, McHugh G, Roozenbeek B, Maas A, Murray G: **Does the Extended Glasgow Outcome Scale 'add value' to the conventional Glasgow Outcome Scale?** *J Neurotrauma* 2011: 29(1); 53-8

Hartings JA, Bullock MR, Okonkwo DO, Murray LS, Murray GD, Fabricius M, Maas AI, Woitzik J, Sakowitz O, Mathern B, Roozenbeek B, Lingsma HF, Dreier JP, Puccio AM, Shutter LA, Pahl C, Strong AJ: **Spreading depolarisations and outcome after traumatic brain injury: a prospective observational study.** *Lancet Neurol* 2011: 10(12); 1058-64

Sener S, Roozenbeek B, Maas AI: **Surgical management of traumatic brain injury*:* evidence, controversies and perspectives for the future.** *European Neurological Review* 2011: 6(3); 196–201

Van Leeuwen N, Lingsma HF, Perel P, Lecky F, Roozenbeek B, Lu J, Shakur H, Weir J, Steyerberg EW, Maas AI: **Prognostic value of major extracranial injury in traumatic brain injury: an individual patient data meta-analysis in 39,274 patients.** *Neurosurg* 2012; 70(4): 811-8

Roozenbeek B, Chiu YL, Lingsma HF, Gerber LM, Steyerberg EW, Ghajar J, Maas AI: **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(7): 1306-12

Roozenbeek B, Lingsma HF, Lecky FE, Lu J, Weir J, Butcher I, McHugh GS, Murray GD, Perel P, Maas AI, Steyerberg EW: **Prediction of outcome after moderate and severe traumatic brain injury: external validation of the IMPACT and CRASH prognostic models.** *Crit Care Med* 2012; 40(5): 1609-17

Roozenbeek B, Lingsma HF, Maas AI: **New considerations in the design and analysis of clinical trials in traumatic brain injury.** *Clin Invest (Lond)* 2012: 2(2); 153-62

Book Chapters

Roozenbeek B, Maas AI: **Design and analysis of clinical trials in traumatic brain injury.** In: *Traumatic Brain and Spinal Cord Injury.* Eds: Morganti-Kossmann, Raghupathi, Maas. Cambrigde University Press, New York, 2012

Maas AI, Roozenbeek B, Lingsma HF: **Predicting outcome after traumatic brain injury.** In: *Handbook of Clinical Neurology.* Eds: Grafman and Salazar. In press.

Summary | Samenvatting

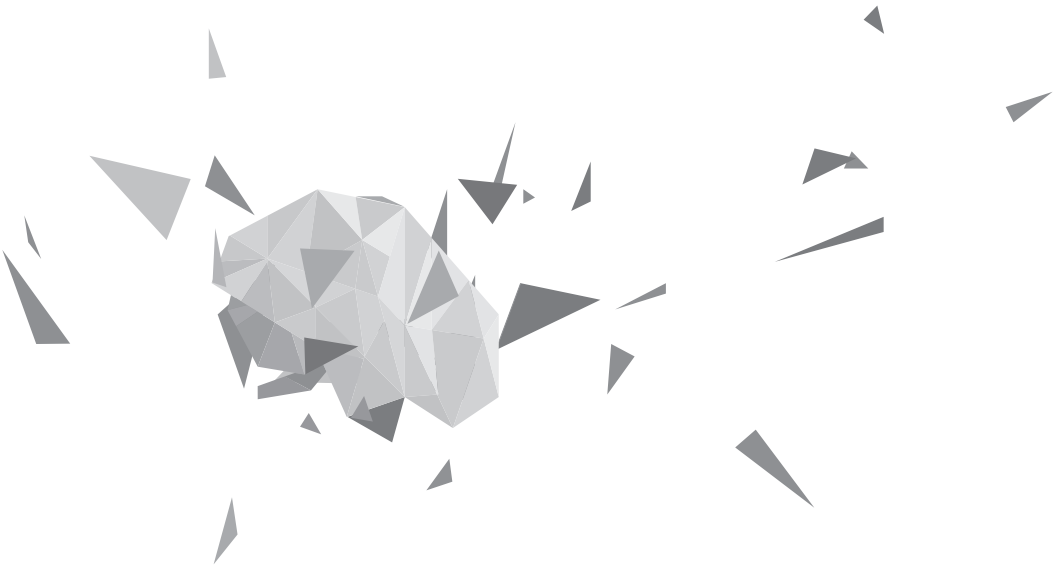
Contributing Authors

Dankwoord | Acknowledgements

Curriculum Vitae

List of Publications

PORTFOLIO



PhD portfolio

Name PhD student: Bob Roozenbeek **Departments:** Dept. Neurosurgery, Antwerp University Hospital (B) Dept. Public Health, Erasmus MC Rotterdam (NL) **Research School:** Nihes	**PhD period:** May 2009 - August 2012 **Promotors:** Prof.dr. A.I.R. Maas Prof.dr. E.W. Steyerberg

	Year	Workload (ECTS)
General Courses		
- "The why and how of readable articles" (Nihes short course)	2009	0.5
- "Introduction to data-analysis" (Nihes course)	2009	1.0
- "Conceptual foundation of epidemiological study design" (Nihes course)	2009	1.0
- "Regression analysis" (Nihes course)	2009	2.0
Specific Courses		
- "Erasmus MRI Course: Central Nervous System II" (Antwerp, Belgium)	2009	1.0
- "Neurocritical Care Course" (Codman®) (Hamburg, Germany)	2009	1.0
- "Advanced Trauma Life Support® Provider Course" (Shock Trauma Center, Baltimore, Maryland, USA)	2011	1.0
- "Practical Neuroanatomy and Neuroradiology" (VUmc, Amsterdam, The Netherlands)	2011	1.0
Seminars and Workshops		
- Weekly seminars Erasmus MC, Department of Public Health	2009	1.0
- Belgian Interuniversity Seminar Neurosurgery on Neurotraumatology	2010	1.0
Presentations		
- World Congress of the International Brain Injury Association (IBIA) 2010, Washington DC, USA >> oral presentation.	2010	1.0
- Annual Meeting of the European Association of Neurosurgical Societies (EANS), Groningen, The Netherlands >> poster.	2010	1.0
- Annual Meeting of the Congress of Neurological Surgeons, San Francisco, California, USA >> poster (awarded as 'best poster presentation in the "Neurotrauma & Neurocritical Care" section').	2010	1.0
- International Neurotrauma Symposium 2011, Sjanghai, China >> oral presentation (awarded with an 'INTS Travel Grant').	2011	1.0
(Inter)national Conferences		
- Joint Symposium of the International and National Neurotrauma Societies, Santa Barbara, California, USA.	2009	1.0
- Joint meeting of the Belgian and Dutch neurosurgical societies, Maastricht, The Netherlands.	2009	1.0
Other		
- Research stay: Brain Trauma Foundation & Dept of Public Health, Weill Cornell Medical College, New York City, New York, USA.	2010	2.0