PROGNOSIS IN TRAUMATIC BRAIN INJURY (TBI)

A thesis submitted to the University of Manchester for the degree of the Doctorate of Philosophy (PhD): Medicine in the Faculty of Medical and Human Sciences

2011

By:
Mehdi Moazzez Lesko

School of Medicine

Volume I of II

Table of contents

Total number of words: 89,977

List of tables

List of figures

List of abbreviations

AIDS	Acquired Immune Deficiency Syndrome
AIS	Abbreviated Injury Score
APACHE	Acute Physiology and Chronic Health Evaluation
AUC	Area Under the ROC Curve
BP	Blood Pressure
CRASH	Corticosteriod Randomisation After Significant Head injury
CSF	Cerebrospinal Fluid
CT	Computed Tomography
ELIZA	Enzyme-Linked Immunosorbent Assay
EPR	Electronic Patient Record
FNA	Fine Needle Aspiration
GCS	Glasgow Coma Score
GOS	Glasgow Outcome Score
GOSE	Glasgow Outcome Score-Extended
ICU	Intensive Care Unit
IMPACT	International Mission for Prognosis and Clinical Trial
ISS	Injury Severity Score
LOC	Loss Of Consciousness
MBP	Myelin Basic Protein
MeSH	Medical Subject Heading
MRI	Magnetic Resonance Imaging
NFS	Not Further Specified
NICE	National Institute for Clinical Excellence
NPV	Negative Predictive Value
NSE	Neuron Specific Enolase
PCPC	Pediartiatic Cerebral Performance Category Score
PCS	Post-Concussion Syndrome
PPV	Positive Predictive Value
ROC	Receiver Operating Characteristic
SPSS	Statistical Package for Social Sciences
TARN	Trauma Audit and Research Network
TBI	Traumatic Brain Injury
WHO	World Health Organisation

The University of Manchester
Mehdi Moazzez Lesko
Doctorate of Philosophy (PhD): Medicine

Prognosis in Traumatic Brain Injury (TBI)

(2010)

Abstract

Introduction: Prognosis in Traumatic Brain Injury (TBI) can be made using prognostic models (the IMPACT and CRASH models) or brain injury biomarkers (S100B). Current prognostic models are derived from historic datasets recruited from heterogeneous countries in terms of trauma care and for the purpose of clinical trials.
Objective: To construct a prognostic model suitable for British trauma care, to compare the prognostic performance of prognostic models with S100B and to assess the combination of prognosticators from the constructed models with S100B.
Methods: A dataset of 802 TBI cases from the Trauma Audit and Research Network (TARN), Manchester, UK was used to construct the prognostic models.. During the modelling, criteria for well-developed models as per the literature review were followed such as the dataset being large, the variables being selected from the literature and missing information being imputed. A further dataset of TBI cases was used to validate these models Moreover, the resulting models were run on a dataset of 100 TBI cases who had their serum S100B recorded at 24 hours to compare their performance with S100B.
Results: Two prognostic models were constructed (models A and B) to predict the discharge survival. Both models share age, admission Glasgow Coma Scale (GCS), admission pupillary reactivity and presence/absence of hypoxia and lowblood pressure (on admission) and brain stem injury. However, model A includes Injury Severity Score (ISS) which is replaced with cause of injury, extracranial injury, brain swelling and interaction of cause of injury and age in model B. Both models have high performance either on the derivation dataset (Area Under the ROC Curve (AUC) of model A: 0.92 and AUC of model B: 0.93) or the external validation set from a later time period in TARN (AUC of model A: 0.92 and AUC of model B: 0.82). Furthermore, in the S100B dataset, it appears that the performance of prognostic models is not significantly different to that of S100B (for example, AUC of model A in this dataset: 0.64 versus 0.69 of the model just including S100B for survival prediction). A combination of S100B and models prognosticators improved performance and S100 improved the performance of models A and B.
Discussion: The proposed prognostic models have very high AUCs and since they have been validated on a different TBI dataset from TARN, they are valid to be used for the purpose of the British trauma care benchmarking. Unfortunately, the results of the analysis on the small S100B dataset are not adequately powerful to be conclusive. However, these findings highlight the importance of future research on this topic in larger datasets.
Conclusion: Two prognostic models have been constructed which can be used for the British TBI patients.

Copyright and the ownership of intellectual property statement

To my parents and sisters

1. Introduction

1.1. Review of the epidemiology of traumatic brain injury and the research question

Worldwide many people die due to Traumatic Brain Injury (TBI) each year or suffer from the resulting neuropsychological disability causing either social or individual problems. As a matter of fact, TBI is a silent epidemic [1-7] which has been disregarded by policy makers compared with other public health challenges such as breast cancer or Acquired Immune Deficiency Syndrome (AIDS). However, the annual incidence of brain injury in the United States surpasses that of breast cancer or AIDS significantly [1]. It is now the leading cause of death and disability among the young generation and anticipated to become one of the major causes of mortality by 2020 worldwide [7].

1.1.1. Definition

Based on Medical Heading Subject (MeSH) database, head trauma or craniocerebral injury is defined as "*a traumatic injury involving the cranium and intracranial structures*" [8] and described as TBI when it involves brain.

Standards for Surveillance of Neurotrauma [9] define a case of TBI as either:

1) An occurrence of injury to the head (arising from blunt or penetrating trauma or from acceleration –deceleration forces) with at least one of the following:

- observed or self reported alteration of consciousness or amnesia due to head trauma
- neurologic or neuropsychological changes (determined from neurologic and neuropsychological examinations) or diagnosis of skull fracture

or intracranial lesions (determined from radiological examination or other neurodiagnostic procedures) that can be attributed to the head trauma.

2) An occurrence of death resulting from trauma with head injury or traumatic brain injury listed on the death certificate, autopsy report, or medical examiner's report in the sequence of conditions that resulted in death.

According to the guideline commissioned by National Institute for Clinical Excellence (NICE) [10] head injury is defined as "*any trauma to the head, other than superficial injuries to the face".*

Recently, a new definition of TBI has been proposed by the Demographic and Clinical Assessment Working Group of the International and Integrity Initiative toward Common Data Elements for Research on Traumatic Brian Injury and Psychological Health {Menon, #303}. The proposed definition is:

'TBI is defined as an alteration in bran function or other evidence of brain pathology caused by an external force.'

The medical term : injury, according to MeSH database, is described as "*damage inflicted on the body as the direct or indirect result of an external force, with or without disruption of structural continuity"* [8]. This description suggests that the term "brain injury" has association with trauma; hence brain injury and TBI can substitute each other in a medical text. In this review, "brain injury" is equated to traumatic brain injury when being referred to. Moreover, brain injury patients are a subgroup of patients who have head trauma with an evidence for brain damage. Even with the current clinical or radiographic diagnostic tools, this distinction is not always possible particularly in mild brain injury.

1.1.2. Incidence and cost

In the United States, each year 1.5 million people suffer from TBI of which 50,000 die, 230,000 are hospitalized and the remaining are discharged home [11]. In Europe, the incidence of hospitalization or death is estimated to be 235 per 100,000 [12]. According to the study of Alan Tennant the incidence rate of traumatic brain injury in England was 229.4 in the year 2001-2 and 229.1 for the year 2002-3 per 100,000 of general population [13].

The cost of TBI is classified as either direct or indirect. The direct cost amounts to hospitals charges after acute care involving prescribed medicines, undertaken procedures or performed rehabilitation strategies and then, further out-patient services. The indirect cost is caused by subsequent patient's lack of productivity, early retirement or sick leave [14]. The total cost related to brain injury in 2000 was estimated to be 6$ billion in the United States [15], while the exact figure from Europe is unknown [12]. However, in Spain during the year 1997, the injuries resulting from traffic crashes imposed 3,397 € to the nation [16] which gives an approximate estimation of brain injury cost in a European country.

1.1.3. High risk groups

Males are at higher risk and age groups of 0 to 4 and 15 to 19 are more prone to sustain TBI than other groups [2]. J David Cassidy and colleagues who undertook a best-evidence synthesis on dimensions of mild TBI epidemiology, including risk factors reported similar results [17]. Recreational and sporting activities [17] and some work-related situations such industrial-type jobs [18]

put the individual at a higher risk level as well. In the United States, being African-American may act as a risk factor [2].

1.1.4. Cause

The causes of TBI are categorized differently in various studies. Fall, motor vehicle –traffic crash, struck by/against an event and assault are conventional categories used by most authors with some variations. In 1995, World Health Organisation (WHO) proposed a guideline for surveillance of nerurotrauma in societies from which the main causes are displayed in table 1[9]. In the United States the most frequent aetiology is fall [2] but in Europe it is said to be motor vehicle-crashes [12].

Vehicular and animal transportation: operator or passenger Vehicular and animal transportation: pedestrian Firearms and other Objects Used as Weapons Violence without Use of Firearm or Other Weapons Sports/Recreation Falls Other (Struck by falling object, other specified or known cause of injury, Unknown)

Table 1 External causes of brain injury (drawn from *Standards for Surveillance of Neurotrauma*, ed. D.J. Thurman, Kraus, J.F., Romer, C.J.. 1995, Geneva: World Health Organisation. [9])

1.1.5. Outcome

Victims of TBI may experience a spectrum of outcome ranging from full recovery to death. Those who survive may continue to suffer from various degrees of physical or mental disabilities affecting their life and some of them, despite being physically fit, may never resume their pre-injury functionality for a long time. Mild physical and mental consequences of TBI may be a number of complaints such as headache, dizziness, fatigue and difficulty concentrating. If a patient demonstrates 3 of these symptoms, he/she would then be diagnosed with Post-Concussion Syndrome (PCS) which commonly occurs following concussion (a mild form of brain injury but severe enough to result in loss of consciousness). The survivors of more severe forms of TBI which result in longer periods of loss of consciousness may suffer from more severe disability.

Mental and physical consequences of TBI do not affect only the individual victim. For example, people with TBI may suffer from relationship difficulties inside the family. Out of 48 participant couples in Wood RL and his colleagues' research [19] twenty three (relatively 47%) were separated or divorced in the end, meaning that just less than 53% continued their partnership. Hawley CA and others [20] reported that the level of stress in parents of TBI children was clinically high with the rate of 40%. One third of

the parents were found to have poor psychological health. Moreover, Montgormery and others [21] observed that 57% of families of children with brain injury witnessed behavioural problems in other siblings. With decreased amount of time worked and subsequent impact on employment status [13, 18] the potential harm of the injury to the family goes beyond psychological matters, ending up as an economic burden as well.

1.1.6. Prognosis

Prognosis means the prediction of the outcome of the disease and as such it would assist the clinicians in having an estimate of the severity. A poor prognosis means the clinician should consider the value of more aggressive intervention.

Knowing the prognosis in TBI is important to assist clinicians dealing with individual TBI patients, to stratify patients based on the severity of their injuries in clinical trials and to perform trauma care benchmarking. Whilst overall, advancement in each of these areas can lead to improvement in patients' outcome, knowing prognosis has a specific advantage in each area. In clinical practice, it would assist clinicians in making timely and appropriate decisions and also in allocating resources. In clinical trials, knowing the prognosis would assist to detect pure effect of the intervention on outcome irrespective of the severity and in the recruitment of patients with a truly intermediate prognosis. And lastly, the quality of trauma care in a given trauma centre can be benchmarked by comparing the observed prognosis of patients cared for in that centre to the expected prognosis at the national level. The expected prognosis is calculated by using a model derived from nationawide data.

Considering the above importance of prognosis in TBI, the research question is:

How can the outcome of a given TBI patient be predicted?

There are currently two common prognostic tools in TBI to predict the outcome: prognostic models and brain injury biomarkers. Prognostic models apply commonly measured patient characteristics and provide the probability of a given outcome at a certain point in time. Brain injury biomarkers are usually serum biomarkers which tend to increase in TBI and also be higher in more severe forms of TBI. In the following sections of the introduction, the common prognostic models in TBI and the most widely acknowledged brain injury biomarker; S100B are discussed.

1.1.7. Common terms in research into TBI prognosis

Glasgow Coma Scale (GCS)

First introduced by Teasdale *et al.* in 1974, this is a scoring system to measure the level of consciousness (Table 2). This system is internationally well-established and ranges from 3 to 15 and is the sum of three subscores assessing three domains of eye, verbal and motor response. Later on, in several studies it was shown that lower GCS is associated with adverse outcome [22-24]

Best verbal response Orientated to time, place and person 5 Appropriate responses but disorientated 4 Inappropriate responses but coherent words 3 Incoherent sounds only 2 No verbal response 1
Best motor response Obeys commands 6 Localise pain 5 Normal flexion to pain 4 Abnormal flexion to pain 3 Extension to pain 2 No motor response 1
Best eye opening response Spontaneous eye opening 4 Opens eyes to speech 3 Opens eyes to pain 2 No eye opening response 1

Table 2 Glasgow Coma Score is the sum of three types of patient's response score to stimulation: verbal, motor and eye opening.

Glasgow Outcome Scale (GOS)

Glasgow Outcome Scale was first introduced by b. Jennet and M. Bond in 1975 [25]. They categorized the outcome into five groups of death, persistent vegetative state, sever disability (conscious but disabled), moderate disability (disabled but independent) and good recovery. In 1981 Jennet *et al.* proposed an extended revised version of the system to enhance its application and reliability [26]. In the extendedextended revision of GOS referred to as GOSE, each three last categories are further divided into upper and lower levels (Table 3). In 1998, Wilson and others [27] proposed a structured pattern of interview to score the outcome of TBI patients based on GOS.

GOSE score	*Performance level*
1	Dead
2	Vegetative state
3	Lower sever disability: completely dependant on others
4	Upper sever disability: dependant on others for some activities
5	Lower moderate disability: unable to return to work or participate in social activities
6	Upper moderate disability: return to work at reduced capacity, reduced participation in social activities
7	Lower good recovery: good recovery with minor social or mental deficits
8	Upper good recovery

Table 3 The extended Glasgow Outcome Scale (GOSE)

1.1.8. Summary

Traumatic brain injury is a major public health problem and has detrimental effects on victims' lives due to physical and psychological morbidity. An understanding of prognosis in TBI assists clinicians with timely and appropriate decisions and allocation of resources. Also, it is important to understand prognosis for benchmarking trauma care quality and clinical trials of new intervention. There are currently two common prognostic tools available in the literature: brain injury biomarkers and prognostic models.

1.2. Prognostic models

Prognostic models employ a number of patients' characteristics such as age, GCS, pupillary reactivity, CT findings etc. to predict a given outcome. In this section, two internationally known models derived from large cohorts of TBI patients plus two models (which have been proposed as well-developed models by a systematic review) are introduced.

Area Under the receiver operator Curve (AUC) is a statistical term to measure the model performance (or its predictive strength) and has been stated in several places of this section. This concept is further discussed in Paper 1 (section 2.7.1).

1.2.1. IMPACT models

International Mission for Prognosis and Clinical Trial (IMPACT) dataset of TBI is the mergence of several previously conducted clinical trials and observational studies which in total contains data on more than 9000 TBI subjects [28]. Included studies were all from developed world (North America and Europe). The initiative started in 2002 targeted at advancing the approach in TBI clinical trials and as part of this initiative, the IMPACT models for TBI prognosis were published in 2008 [24]. The IMPACT dataset contains severe TBI cases as per their admission GCS being 8 or less. Although it contained some very few cases with moderate GCS (i,e, GCS of more than 8 but less than 13),the IMPACT dataset included mainly severe TBI cases.

Overall 6 prognostic models are proposed by the IMPACT collaboration [24]. The *core models* use age, motor GCS score and pupillary reactivity to predict mortality or unfavourable outcome (GOS < 4) at 6 months

after sustaining injury. The *expanded models* use hypoxia, hypotension, the Marshall CT classification, traumatic SAH and epidural haematoma in addition to the covariates in the core models. The *lab models* add glucose and haemoglobin (Hb) to the covariates included in the expanded models. Models are presented in the form of score charts with an online calculator [29].

From the core models to the lab models and as new covariates are added to the models, the performance of the models improves according to AUC. The AUCs are reported in various sub-datasets ranging from 0.66 for the core model on mortality prediction in the Saphir data (phase III trial of the competitive NMDA antagonist DCPP-ene, 1995-1997 [30]) to 0.87 for the expanded model on mortality prediction in EBIC data (European Brain Injury Consortium survey, 1995) (Saphir and EBIC studies were two of many studies pooled into the IMPACT mega dataset [30]).

The important advantage of the IMPACT models is the large dataset from which the models have been derived. Furthermore, the models construction complies with many criteria for a well-developed model proposed by Perel *et al.* [31] and Mushkudiani et al [32]. Briefly and apart from the large derivation set, the variables used by the models such age, GCS etc.. are readily obtainable and the models have been externally validated in another dataset of TBI different to the derivation dataset with no significant change in AUC. In fact, a high external AUC of 0.83 is obtained for the expanded model for mortality prediction when it was run on the CRASH data (section 1.2.2). Furthermore, the missing information had been handled appropriately rather than omitting the cases with missing values. This is because in the modelling procedure, all cases who have a missing value even on one variable are

excluded meaning that they in fact do not exist in the data. Imputing the missing values with an appropriate strategy is preferable to total loss of information [28].

1.2.2. CRASH models

The Medical Research Council (MRC) CRASH (Corticosteroid Randomisation After Significant Head Injury) study was initially a clinical trial to investigate the effect of early administration of methylprednisolone in TBI [33]. The dataset of more than 10 000 cases collected in this trial was subsequently used to derive prognostic models [23]. The participating hospitals were from various countries around the globe with different degrees of contribution to patient recruitment. The trial was commenced in 1999 and was terminated in 2004. The inclusion criteria were head injury sustained at the age of above 16 with admission GCS of 14 or less and clear indication or contraindication for corticosteroid administration.

Overall 6 prognostic models are proposed by the CRASH collaboration. *Basic models* do not contain CT findings and instead use age, GCS, pupillary reactivity and major extra cranial injury. However, there are also separate models by CRASH which include CT findings in addition to the factors in the basic models. Each model (with or without CT) can predict mortality or severe disability within 14 days or six months following the injury. Moreover, separate models are applicable to low-middle income countries versus high-income countries. These models are presented in the tables of odds ratios with an online calculator available [34].

Similar to the IMPACT models, adding more covariates i.e. CT findings to the basic models results in improvement in AUC. The AUC of the

models ranges from 0.81 for death/severe disability prediction to 0.88 for mortality prediction either in high-income or low-income countries.

The upsides of the CRASH models are almost similar to those of the IMPACT models such as easily obtainable variables, external validity and handling of missingness with the most important one being the large derivation dataset. Moreover, the CRASH models reasonably maintain their predictability in the external validation on the IMPACT dataset (a drop from internal AUC of 0.88 to external AUC 0.77).

1.2.3. Other models

Perel *et al.*[31] performed a systematic review of the literature on the prognostic models in TBI and concluded that at the time of the review, there were only two models which fulfilled the methodological requirements with acceptance external validation: models by Signorini *et al.* [35] and Hukkelhoven *et al.* [36]. It is not surprising that since this review was conducted prior to the introduction of the IMPACT and CRASH models, these models are not appraised in Perel's work.

Hukkelhuven *et al.* derived a prognostic model on a TBI series from a clinical trial in Europe and the North America. The trial was designed to assess the effect of Tirilazad Mesylate in TBI and contained over 2,200 cases. Two models were introduced to predict mortality or severe disability at 6 months. The models employed age, GCS (motor score), pupillary reactivity, hypoxia, hypotension, Marshall CT Classification and SAH to make its outcome prediction with AUC of 0.78 for mortality prediction and 0.80 for unfavourable outcome prediction. These AUCs even increased in the external validation set (ranging from 0.83 to 0.89 for various constructed models). The dataset from

which the Hukkelhuven's models were derived, was later pooled in the IMPACT dataset.

Signorini *et al.* analysed a dataset of 124 TBI cases all cared in the intensive care unit and constructed 6 models (one *Baseline model* with 5 *final models*). The outcome measure was mortality at 1 year following injury. The baseline models contained age, GCS, presence/absence of haematoma on CT and pupillary reactivity. The final models contained the covariates in the basic model plus ICP measured at various time points e.g. within 24 hours of injury, between 24 to 48 hours of injury etc. The AUCs of models ranged from 0.84 (for the basic model) to 0.9048 for the basic model plus ICP measured between 24 and 48 hours. The Signorini's models are very old dating back to about 20 years ago (1991).

1.2.4. Summary

Prognostic models are a tool for predicting outcome in TBI. They employ patients' characteristics recorded or known to the clinician and provide the probability of an outcome at a certain time point. Currently, there are two internationally well-developed prognostic models which have been derived from large datasets of TBI: the IMPACT and CRASH prognostic models. The development of these models complies with the methodological requirements of well-developed models. The factors consistently most associated with prognosis are age, GCS, pupillary responses, CT scan findings with measures of injury severity and laboratory variables also having value in some models

1.3. Brain injury biomarkers

These are chemicals which increase in the blood or CSF following TBI and their serum concentration relates to the severity of brain injury and outcome. There are many biomarkers proposed in the literature and Table 4 lists some of them as compiled by Qurreshi in 2002.

Obviously the list of some biomarkers proposed by Qurreshi in 2002 is expected to be longer than that in 2002 especially with the new advances in better understanding of cellular pathways initiated following TBI. For a number of reasons understanding the multifaceted pathogenesis of TBI is important. Firstly, it may assist with diagnosis and identifying the severity of the damage caused by TBI especially that the advanced imaging modalities are not sensitive enough (such as CT to detect diffuse axonal injury) or are not widely available due to their high cost (such as MRI). Secondly, knowing these cellular and molecular can open up the possibility to investigate therapeutic agents to disrupt those pathways which are neurotoxic. It is currently believed that the disruption of Blood Brain Barrier (BBB) following trauma can initiate immunological responses by activating CNS 'resident' cells such as microglia and astrocytes or by causing the infiltration of the peripheral immune cells {Morganti-Kossmann, 2007 #296}. This subsequently results in neuroinflammation. Immune cells react by producing cytokines which can have either neuotoxic or neuroprotecitve effects. Interleukin (IL)-1 and Tumor Necrosis Factor (TNF) are some of the cytokines which may have more neuroprotecrtive effects. Moreover, it has been found that increased patients with severe TBI have increased levels of IL-1, TNF, IL-6, IL-10 in CSF or serum and also some cytokines may be significantly higher in those TBI

patients with unfavorable outcome such as IL-1 {Chiaretti, 2005 #302}. Another expansion in the field of TBI pathogenesis relates to the observation that neuronal death following trauma is not always necrotic and it includes apoptotic cell death as well {Wang, 2000 #294}. The latter refers to a programmed series of events ultimately causing the cellular death. It is currently believed that *Calpain* proteins are mainly involved in necrotic death whereas the apoptotic death is mainly mediated by *Caspase 3 {Wang, 2000 #294}*. Siman *el al.* have shown increased levels of a panel of proteins including a calpain-derived protein (alpha-spectrin) in CSF of patients with TBI {Siman, 2009 #301}. In the same way, Buki *et al.* showed that using a calpain inhibitor can disrupt the axonal injury pointing to the potential pharmacological benefits of understanding the molecules and biomarkers involved in the pathogenesis of nervous system in TBI {Buki, 2003 #298}.

Creatinine kinase, brain type
Glial fibrillary acidic protein
Lactate dehydrogenase isoenzyme 1
Myelin basic protein
Neuron-specific enolase
S100b protein
E-selectin
L-selectin
Soluble intercellular adhesion molecule-1
Cleaved tau protein
Glutamic oxaloacetic transaminase
Glutamic pyruvate transaminase
Malate transaminase
Fructose 1, 6-diphosphate aldolase
α-hydroxybutyric acid dehydrogenase
Tumour necrosis factor-α
Transforming growth factor-β1
Interleukin-1β
Interleukin-6
Interleukin-8
Interleuin-10
Interleukin-12

Neopterin B2-microglobulin Soluble interleukin-2 receptor

Table 4 Brain injury biomarkers which have been proposed for assessment of brain injury severity or prognosis.

Table 5 suggests that a variety of blood markers may have significant clinical role in diagnosis or prognosis of brain injury. Niedeggen A *et al.* observed that all TBI patients who have Creatine Kinase-brain type (CK-BB) > 50 ng/l survive the injury [37]. Mao *et al.* demonstrated that serum Myelin Basic Protein (MBP) is correlated to extra- and intracranial haematoma and the degree of cerebral contusions detected by CT scan [38]. Furthermore, Neuron Specific Enolase (NSE) and MPB have higher serum levels in non-survival as proposed by Yamazaki Y. *et al.* [39]. With regards to cytokines, Gopcevic A. *et al.* found central venous IL-8 concentration has significant prognostic power the same as age and GCS [40]. Similarly, Minambres E. *et at.* showed the correlation of IL-6 arterial and jugular gradient with 6-month GOS [41].

Authors	*Publication year*	*Sample size*	*biomarker*	*outcome*	*Findings*
Nylen *et al.* [42]	2006	59	Serum-GFAP	1-year GOS	- Those with unfavourable outcome had higher peak GFAP - All with GFAP> 15.04 µg/l died
Vose *et al.* [43]	2004	85	Serum S100B, NSE, GFAP	6-month GOS	-S100B was the strongest predictor of death
Pelinka *et al.* [44]	2004	92	Serum S100B, GFAP	3-month GOS	-S100B had higher power for mortality prediction than GFAP.
Niedeggen *et al.* [37]	1989	76	Creatine kinase BB (CK-BB)	Survival	-All with CK-BB> 50 ng/l died -Those with CK-BB < 25 ng/l had minimal neurological deficits.
Mao *et al.* [38]	1995	112	Serum Myelin Basic Protein(MBP)	CT characteristics	- MBP was correlated to the volume of extradural haematoma, intracranial haematoma and the extent of the cerebral contusion.
Yamazaki *et al.* [39]	1995	25	Serum Neuron Specific Enolase (NSE) and MBP	Survival	-The level of NSE and MBP were higher in those who died.
Gopcevic *et al.* [40]	2007	20	Central venous plasma IL-8	Survival	-IL-8 was significantly higher in non-survivors.
Minambres E. *et al.* [41]	2003	62	Transcranial IL-6 (arterial and jugular gradient)	6-month GOS	-The gradient was higher in non-survivals.

Table 5 Some examples of various brain injury biomarkers used in TBI diagnosis or prognosis.

It appears that the most known biomarkers are still S100B, Gilial Fibrillary Acidic Protein (GFAP), and Neuron Specific Enolase (NSE). PubMed was searched for each of these biomarkers utilizing the MeSh (Medical Subject Heading) terms [45] for each marker and brain injury. The number of results for some of these were less than 10 published papers namely CK-BB, E- and L- selectin, Malate transaminase, α-hydroxybuttyric acid dehydrogenase, transforming growth factor-β1, interleukin-1β and 12, Neopterin, β2-microglobulin and soluble interleukin-2 receptor. The search for S100B retrieved 142 papers which are comparable to those for Myelin basic protein (MBP), Neuron-specific Enolase (NSE), and Glial Fibrillary Acidic Protein (GFAP). The results for GFAP remarkably outnumbered that for S100B at more than 299 publications. However, a review of the titles showed that the investigations on GFAP's clinical role in brain injury are much less than that on S100B (< 30). The main focus of general research into GFAP is at molecular level to characterize astrocytic response to injury in vitro and on non-human samples. Overall, it appears the amount of literature for S100B is substantially more than other biomarkers.

In the following section, S100B protein as a brain injury biomarker is discussed.

1.3.1. Summary

There are several biomarkers of brain injury which have been proposed to be associated with outcome in TBI or to be used for outcome prediction. Three of the most known of these biomarkers are GFAP, NSE and S100B. However, so far. S100B has received the most attention in the literature.

1.4. S100B protein

S100 proteins are "*a family of highly acidic calcium-binding proteins found in large concentration in the brain and believed to be glial in origin. They are also found in other organs in the body"* [46]. It was first introduced in 1965 when Moore described it as a characteristic protein of the nervous system [46]. Rather 0.5% of soluble brain proteins are composed of S100 [46]. This family are not exclusively found in nervous system as it was initially thought [47]. It is now clear that it is synthesized in normal as well as pathologic or neoplastic tissues such as melanoma [48].

The S100 protein family comprises at least 25 isomers of which S100B is a member (3) with molecular weight of 21 kilo-Dolton (KD) [49]. Other prominent members are S100A1 to S100A13 [47, 50].

1.4.1. Functions

S100B protein is found in cytoplasmic compartment [46], as a membrane-bound molecule [46] or in the extra cellular compartment [47]. In the extra cellular compartment it is believed to be secreted by astrocytes, pituitary folliculostellate cells and adipocytes in non-pathological situation[47]. For the latter to occur, stimulations by a provocative substance such as nor epinephrine is needed [47]. Amongst the above aforementioned tissues, brain is the main source of knowledge about the extracellular roles of the protein [47]. These roles are primarily divided into physiologic and pathologic functions [49]. The physiologic (trophic) functions exert their effect when the concentration of S100B is at nanomolar levels [47]. Some functions are neurodegeneration [47, 49], energy metabolism [47] and cytoskeleton modification [49]. At micro

molar concentrations, pathologic (toxic) effects [47, 49] of the protein are primarily the induction of iNOS in astrocytes [47] which has a fatal impact on cells. All of the physiologic and pathologic functions are performed through paracrine or autocrine routes [49]. Although in physiologic situations, S100B is thought to be released via secretion but after TBI cell membrane discruption may be the cause of its release into the extracelluar compartment.

1.4.2. Effect of age, gender and race

Gazzolo and others [51] examined these variables in a population of more than 1000 children who were from 1 month to 15 years old with the mean of 8.4 years. Of these, 486 were male and the remaining 522 were female and all of them were in a healthy clinical condition with no apparent neurologic problems. The main observation of this study was higher S100B concentration during the first year and in early adolescence (9 to 10 years old) compared to other age groups. Table 6 presents the obtained reference intervals at various ages during childhood. As seen, the median values in either males or females during first year of life and early adolescence are higher than other time periods. This report also suggests higher levels of S100B in females.

Age at sampling, years (M/F)	*Males (n = 482)*			*Females (n = 522)*			*P value*
	Median	*25th*	*75th*	*Median*	*25th*	*75th*	
0–1 (43/42)	0.81	0.44	1.93	0.95	0.45	2.24	0.27
1–2 (46/40)	0.72	0.32	1.33	0.77	0.49	1.37	0.35
2–3 (52/48)	0.62	0.39	0.97	0.76	0.59	1.61	<0.05
3–4 (52/47)	0.7	0.32	1.27	0.66	0.37	0.89	0.7
4–5 (41/53)	0.61	0.31	1	0.74	0.44	1.1	0.29
5–6 (34/48)	0.68	0.32	1.24	0.56	0.36	1.1	0.59
6–7 (37/49)	0.6	0.43	0.96	0.86	0.65	1.17	<0.05
7–8 (28/39)	0.9	0.65	0.96	0.9	0.39	0.96	0.77
8–9 (35/52)	1.37	1.34	4.56	1.41	1.03	4.6	0.95
9–10 (41/32)	1.44	0.91	1.75	1.67	1.44	2.03	0.45
10–12 (24/18)	1.45	0.81	2.63	1.74	0.97	2.1	0.71
11–12 (16/16)	0.42	0.39	0.45	0.45	0.41	0.48	0.38
12–13 (10/8)	1.23	1.2	2	1.25	0.96	2.21	0.66
13–14 (17/18)	1.13	0.99	1.89	1.35	1.16	2.23	<0.05
14–15 (6/12)	0.66	0.45	0.72	0.91	0.52	0.97	<0.05
Total 1–15 (482/522)	0.8	0.44	1.49	0.95	0.58	1.62	<0.05

Table 6 The effect of gender on S100B concentration at various age groups.

Portela's group [52] recruited 19 healthy neonate and 15 healthy children aged from 4 to 16 and 85 healthy adults from 18 to 70 years old. The impact of age was found significant under the age of 20 and non-significant afterwards. They found the highest concentration of S100B in the neonates. Figure 1 displays the course of serum S100B protein during ageing. As depicted, a healthy individual has the highest level of S100B in serum at more than 0.4 μg/l after birth. The number is set to drop sharply till the age of 20

when its decline rate lessens till the age of 30. After 40, the level tends to increase but not significantly. In fact, despite initial sharp decrease of serum S100B from nearly 0.4 µg/l to just above 0.1 µg/l during the first twenty years of life, the further fluctuations are statistically insignificant. Gazzolo's and Portela's restuls also indicate that S100B in the healthy adult population (i.e. > 20 years old) is expected to be less than 0.1 µg/l.

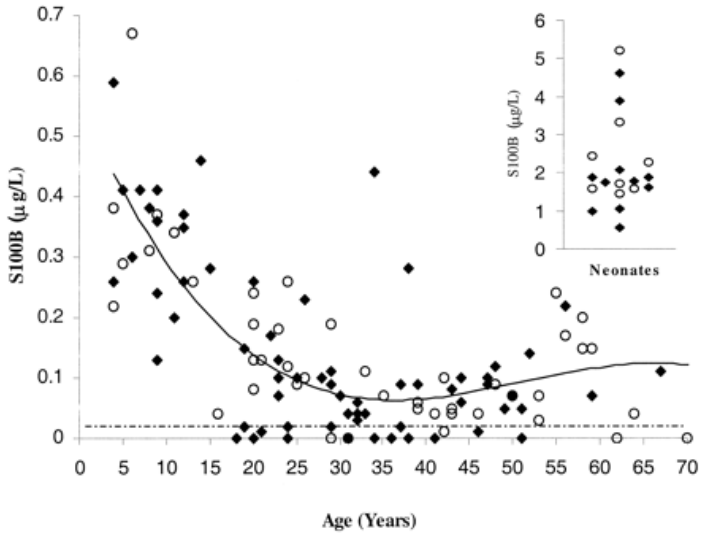


Figure 1 Correlation of serum S100B protein concentrations with age (from: Portela LV, Tort AB, Schaf DV, Ribeiro L, Nora DB, Walz R, Rotta LN, Silva CT, Busnello JV, Kapczinski F, et al: The serum S100B concentration is age dependent. *Clin Chem* 2002, 48:950-952.).

Abdesselam's group [53] compared the S100B concentrations in three groups of "black" (n=40), asian (n=44) and caucasian (N=46) healthy individuals. They did not find any significant difference between asian (0.11 µg/l) and black (0.14 µg/l) groups while in the Caucasian the figure was significantly lower (0.07 µg/l).

1.4.3. Proposed clinical roles

Elevated serum S100B levels have been reported in psychoneurologic diseases such as Alzheimer's disease, Down's syndrome, multiple sclerosis,

amyotrophic lateral sclerosis, schizophrenia, depression, cerebral stroke and TBI [54]. This biomarker also has prognostic value in melanoma and can help to apply an appropriate therapy protocol by detecting metastasis or recurrence after treatment [48, 55].

There are some reports suggesting that serum S100B levels elevate following physical activity compared to values observed before exercise [56-59]. Moreover the pre-post game difference has been claimed to be significantly correlated to the number of headers and trauma events during the game such as a football match [58]. However, despite this elevation, the serum levels remain within normal range [56].

There are some trials on the possible therapeutic roles of S100B after brain injury but on animals so far. In the project of Klenddienst's group, after induced brain injury, the rats received infusion of S100B and consequently experienced improved cognitive performance and enhanced hippocampus neurogenesis [60]. Nearly the same result had been observed by the same scientists in 2004 [61].

In the medico legal field, S100B is assumed to have two applications: firstly; the identification of the cause of death and secondly; attribution of a subsequent disability to brain injury in case of multiple trauma or when such a cause and effect relation should be investigated (in judiciary conditions as it is sometimes unknown whether the brain injury is the actual source of impairment or other elements such as concomitant limb fractures are at play). With regards to cause of death, Li and his colleagues published two papers in 2005 [62, 63]. In both studies 283 autopsy cases were examined within 48 hours post-mortem. Acute death was defined when survival time was shorter

than 6 hours and, in contrast, delayed death happened between 6 hours to 5 months after injury. S100 immunopositivity revealed that positivity in neurones and myelin is more frequent in delayed head injury death and fire fatality while in astrocytes, the rate was lower in acute death due to strangulation and drowning compared to other subcategories of acute death. The authors also reported an inverse relationship of astrocytes S100B positivity with serum S100B in acute death. The initial conclusion was the probable usefulness of combined analysis of S100B positivity in the brain tissue and serum concentrations of the protein for the identification of death resulting from head injury, hypoxic and /or ischemic brain damage. In the second paper [63], the group observed that serum levels are markedly higher in the right heart and subclavian vein after acute death from head injury and asphyxiation due to neck compression while mildly to moderately elevated for other blunt and sharp instrumental injuries. They also found that serum S100B is lower after sub acute than acute death due to head injury. These two studies of Li‘s group propose an important medico legal role for S100B which is now “a hope on the horizon”.

In the following section, the role of S100B in diagnosis or prognosis of TBI has been discussed.

1.4.4. S100B and TBI

In this section, the major proposed roles of S100B in various studies are discussed. This section is divided into 3 parts of mild brain injury, severe brain injury and various severities. This division is based on GCS (mainly on attending or admission to hospital) with GCS < 12 being mild and GCS < 9 being severe. This grouping of studies simplifies understanding of the different

roles of S100B proposed in different studies. At the end, the results of some studies in non-TBI serious which can affect the interpretation of S100B role in TBI, are also presented.

Each part of this section is accompanied by a table (Table 7, Table 8 or Table 9) which summarises the main findings of studies. Additionally, the table contains information on some particular characteristics of each study which is helpful in comparing various studies. These characteristics are age, sampling time, the time point of outcome assessment, the size of the study and whether cases with extracranial injuries were included or excluded (i.e. isolated versus non-isolated brain injury). Age is important because of S100B concentration overall tends to be higher in children than adults and also it decreases as the child ages [51, 52]. Sampling time is important because the elapsed time between injury and blood sampling affects the relationship of S100B and outcome [64]. Furthermore, S100B serum levels tend to decline in a short period of time following their rather immediate surge at the time of injury [65]. Regarding the time point of outcome assessment, one may assume that as the time passes, a higher proportion of TBI victims who have survived recover. Similarly, patients with non-isolated TBI tend to have higher S100B concentrations and perhaps worse outcome than those with isolated TBI [66]. Lastly, the sample size provides an insight about the power of the study.

Mild brain injury

In this severity group of TBI, S100B can be used to predict the outcome (PCS, GOS or GOSE), or the absence/presence of intracranial pathology determined

by CT. The majority of studies tend to report the sensitivity and specificity although AUC is a better tool to compare the results across different studies. PCS is a collection of signs and symptoms such as headache, dizziness, fatigue or difficult concentration which results from head injury severe enough to cause loss of consciousness.

As seen in Table 7, across many papers, sensitivity and specificity do not change in the same direction in that when one increases, the other decreases. This necessitates a compromise between sensitivity and specificity. Thus, it seems investigators prefer more specificity for outcome prediction unlike for CT abnormality which a high sensitivity is preferable (as a screening or ruling out test). A specific test would enable the therapist not to misdiagnose those cases who are prone to develop PCS or GOSE of less than a desirable level for the sake of a timely intervention whilst a sensitive test would enable ruling out of cases from CT scanning when they are unlikely to benefit from it.

Study	*Number of cases*	*Children/adults*	*Sampling time*	*Outcome*	*Outcome assessment time*	*Specificity*	*Sensitivity*	*AUC*	*Isolated*	*Cut-off*
Bazarian et al. [67]	35	10-83	Within 6 hours	PCS	3 months	71.4%	56.3%	0.589	no	
Mussach et. al [68]	139	> 20	On admission	CT		50%	100%	0.864	no	0.21
Savola *et al.* [69]	172	16-49	Within 6 hours	PCS	1 month	93%	27%	0.702	no	0.50
Morochovic *et al.* [70]	102	12-84	Within 6 hours	CT		29.8%	83.3%		nop	0.10
Biberthaler *et al.* [71]	1309	> 18	On admission	CT		30%	99%		yes	0.10
Figueiredo *et al.* [72]	50		On admission	CT		20%	100%	0.82	yes	0.10
Biberthaler *et al.* [73]	52		On admission	CT		40.5%	100%		no	>0.10
Muller *et al.* [74]	226	> 18	In ED	CT		31%	95%	0.73	yes	0.10
Townend *et al.* [75]	112	adults	18-360 hours after injury	GOSE < 5	1 month	83%	90%	0.889	no	0.48

Table 7 Some studies of S100B role in mild TBI (GCS > 12).

The AUC for outcome prediction widely ranges from the minimum of 0.58 [67] to the maximum of 0.88 [75]. This may be explained by differences in the type of outcome (being PCS in Bazarian's study [67] and GOSE < 5 in Townend's study [75]) or the time of outcome assessment (being 3 months in Bazarian's study and 1 month in Towend's study) or other factors affecting the case-mix such as age. Overall, the Savola's study appears to be the most powerful study of this type for outcome prediction in mild TBI as the sample size is 172 [69]. The authors observed an AUC of 0.702 for PCS one month after injury with sensitivity and specificity of respectively 27% and 93%.

Overall, the AUCs for prediction of intracranial pathology determined by CT are higher than those for outcome prediction. The minimum AUC for CT pathology prediction is 0.73 (by Muller *et al.* [74]) with a maximum AUC of 0.86 (by Mussach *et al.* [68]). The most powerful study of this type is the one by Biberthaler *et al.* who managed to recruit 1309 mild TBI cases [71]. Unfortunately, the AUC is not supplied in this paper but a very high sensitivity of 99% with a low specificity of 30% to detect abnormal CT has been obtained. The cut-off serum level for this diagnostic performance is 0.10.

Severe brain injury

In severe TBI, S100B has been commonly proposed for the outcome prediction (Table 8). The outcome measure can be severe disability (measured by GOS or GOSE) or mortality. These studies significantly vary on sampling time and the time point of outcome assessment; hence differing results with regards to cut-offs and AUCs. The sampling time ranges from on admission to some time

later even up to 4 days after injury (Plinka's study [76]). The outcome assessment also ranges from early adverse outcome (i.e. within 12 hours) up to 1 year following injury.

Study	*Number of cases*	*Children/adults*	*Sampling time*	*Outcome*	*Outcome assessment time*	*Specificity*	*Sensitivity*	*AUC*	*Isolated*	*Cut-off*
Olivercrona *et al.* [77]	48	> 15	On admission	GOS < 4	3 month	28%	91.3%	0.585	no	0.32
Olivercrona *et al.* [77]	48	> 15	On admission	Mortality	3 month	95.2%	33.3%	0.687	no	1.67
Olivercrona *et al.* [77]	48	> 15	On admission	GOS < 4	12 month	40%	75%	0.552	no	1
Olivercrona *et al.* [77]	48	> 15	On admission	Mortality	12 month	25	95	0.647	no	1.67
Vos et al. [43]	85	> 15	On admission	GOS < 4	6 months	59%	88%	0.677	no	1.13
Vos et al. [43]	85	> 15	On admission	mortality	6 months	41%	100%		no	1.13
Pelinka *et al.* [76]	23		< 12 hours	Mortality	< 12 hours			0.691	yes	4.42
Pelinka *et al.* [76]	23		13-36 hours	mortality	Within 13-36 hours			0.802	yes	2.24
Pelinka *et al.* [76]	23		37-60 hours	mortality	Within 37-60 hours			0.819	yes	2.71

Table 8 Some studies on S100B role in severe TBI (GCS < 9) *(continued)*

Study	*Number of cases*	*Children/adults*	*Sampling time*	*Outcome*	*Outcome assessment time*	*Specificity*	*Sensitivity*	*AUC*	*Isolated*	*Cut-off*
Pelinka *et al.* [76]	23		> 84 hours	Mortality	> 84 hours			0.971	yes	0.79
Pelinka *et al.* [76]	23		Within 12 hours	mortality	Within 12 hours			0.692	No	7.99
Pelinka *et al.* [76]	23		Within 13-36 hours	mortality	Within 13-36 hours			0.693	No	2.16
Pelinka *et al.* [76]	23		Within 37-60 hours	mortality	Within 37-60 hours			0.747	No	0.70
Pelinka *et al.* [76]	23		> 84 hours	mortality	> 84 hours			0.783	No	1.19
Mussack *et al.* [78]	20		12 hours after admission	GOS < 4	A year after injury	70%	77%	0.90	yes	0.59
Nylen *et al.* [79]	59	> 8	Within 2 days	GOS< 4	A year after injury	80%	50%	0.69	no	0.55

Table 8 Some studies on S100B role in severe TBI (GCS < 9) *(continued).*

The study by Olivercrona *et al.* provides a rough insight about the predictive performance of S100B when sampling time is constant but the type or the time of outcome is different [77]. As presented in Table 9, AUC of S100B tends to be higher for mortality prediction than disability prediction (GOS < 4). For example, the authors observed increase of AUC from 0.58 to 0.68 for mortality versus disability outcome prediction at three months after injury. Furthermore, it appears S100B strength for outcome prediction decreases as the time for outcome assessment prolongs in that, for instance, S100B AUC for disability prediction at three months is 0.585 whereas this is 0.552 for disability prediction at 12 months. However, the sample size in Olivercrona's study is small (only 48 cases).

The study by Plinka *et al.* provides the opportunity to evaluate the effect of sampling time and the presence/absence of extracranial injury on S100B predictive strength [76]. With regards to sampling time, it appears the later the blood taken, the stronger the predictability. For example, S100B measured at less than 12 hours following injury has AUC of 0.691 to predict mortality whereas if the sampling is performed after 84 hours, the AUC jumps to 0.971. Furthermore, excluding cases with extracranial injuries from the analysis led to drop in AUC. This can be observed by comparing AUCs of S100B when sampling time is the same but cases with multiple injuries are ruled in or out.

The AUCs across various studies differ with the minimum of 0.552 to the maximum of 0.971. Not only factors such as the power of the study (sample size), extracranial injuries, sampling time, outcome measure and the time of outcome assessment can explain differing AUCs, this can also occur as a result

of different case-mix as to age or severity of injuries (including extracranial injury). Overall, according to Table 8, the AUC appears to have an average of about 0.80 for outcome prediction in severe TBI irrespective of factors affecting the sampling time, laboratory method of S100B measurement, outcome assessment or case-mix. Moreover, it appears there has to be some degree of compromise between sensitivity and specificity in selection of the cut-off point as, in outcome prediction, sensitivity and specificity tend to hold opposite directions in that if one rises, the other decreases.

According to the sample size, the most powerful study of the table is that by Vos *et al.* having recruited 85 cases [43]. Vos el al. observed an AUC of 0.677 to predict severe disability 6 months after injury. The blood samples were taken on admission and patients with extracranial injury were not excluded. The authors reported the cut-off point of 1.13 having a high sensitivity of 88% (for severe disability) and 100% (for mortality) with a low sensitivity of 59% (for severe disability) and 41% (for mortality).

Various severities

The studies of this type which have included cases with various admission GCSs are useful in that a group of TBI patients who talk and are alert on admission would deteriorate later on. As such, studies with inclusion of various degrees of admission GCS have a higher chance of capturing this group of patients and putting them in the same group as those who present with low GCS to hospital. Like mild brain injury, S100B in these studies is used for detection of intracranial pathology on CT or outcome prediction (Table 9).

Study	*Number of cases*	*Children/adults*	*Sampling time*	*Outcome*	*Outcome assessment time*	*Specificity*	*Sensitivity*	*AUC*	*Isolated*	*Cut-off*
Herrmann *et al.* [80]	39	>16	Median 27 hours after injury	Neuropsychological assessment	2 weeks	90%	69%	0.81	no	140
Herrmann *et al.* [80]	39	>16	Median 27 hours after injury	Neuropsychological assessment	6 months	88.9%	65%	0.77	no	140
Honda *et al.* [81]	34		1 day	CT		100%	27.8%	0.658	no	
Honda *et al.* [81]	34		2 day	CT		100%	33.3%	0.738	no	
Honda *et al.* [81]	34		3 day	CT		100%	33.3%	0.689	no	
Bechtel *et al.* [82]	152	< 18	Within 6 hours	CT		56%	75%	0.67	No	50
Spinella et. at. [83]	27	< 18	Within 12 hours of injury	PCPC ≥ 4	Discharge			0.94	No	

Table 9 Some studies on S100B role in TBI with various severities (i.e. with no limitations on GCS).

Herrrmam *et al.* investigated the effect of the outcome assessment time on S100B prognostic strength and observed AUC of 0.81 and 0.77 for neuropychological deficit respectively at 2 weeks and 6 months after the injury [80]. This may imply that S100B holds less predictive strength for longer term outcome prediction. This is in accordance with what Olivercrona *et al.* observed in severe TBI (GCC < 9) [77]. Honda *et al.* investigated the effect of sampling time on S100B capability to detect abnormal CT and found that the second day sampling has a higher AUC than either first day or third day samples (AUCs of 0.658 versus 0.738 versus 0.689 respectively from the first, second and third day of blood sampling) [81].

Overall the studies including various severities of GCS are small (i.e. < 40 cases). However, that by Betchel *et al.* recruited 152 TBI cases to observe AUC of 0.67 (with sensitivity of 75% and specificity of 56%) for detection of CT abnormality [82].

Extracranial injury

Some studies have examined the effect of extracranial injury on serum S100B concentrations. The importance of this relates to the observation by some authors that the brain is not the only source of S100B release into the blood stream as other body tissues such as fat tissue can contribute to raised S100B serum concentration [66, 84, 85]. There are two types of studies investigating this issue: some studies compared S100B serum levels in subjects of trauma but sparing TBI with subjects in a healthy control group whilst some other

studies compared the serum levels of isolated TBI patients with non-isolated cases.

Unden *et al.* discerned that 29% of patients who had uncomplicated bone fractures with a negligible likelihood of head injury due to the nature of the trauma, had raised S100B (above 0.15 µg/l) [84]. Similarly, Anderson *et al.* compared the S100B concentration in the trauma patients without brain injury (based on physical examination and interview) to healthy individuals and observed that the maximum S100B concentration in the control group was much lower than the minimum S100B contraction in the trauma group (0.13 versus 0.5 µg/l) [85].

Savola *et al.* performed a more detailed analysis of the effect of extracranial injury [66]. They observed the highest S100B concentration was in serum of patients who had moderate to severe brain injury with large concomitant extracranial injury (median 4.01 versus 0.02 µg/l in the normal population). Those patients with isolated moderate to severe brain injury had higher S100B concentration to those without brain injury but with large extracranial trauma (median 0.94 versus 0.35 µg/l). Furthermore, patients with large extracranial injury and without brain injury also had raised S100B but to the lesser degree than to those with both large extracranial injury and brain injury (0.35 versus 4.01 µg/l). These findings were proven to be statistically significant. However, minor extracranial injuries such as soft tissue contusions, wounds, sprains, luxations and small fractures tended not to raise S100B. In fact, serum S100B concentration in the two groups of minor and large extracranial injuries sparing brain injury were significantly different (0.35 versus 0.07 µg/l). Overall, the findings by Savola. *et al.* suggest both brain

injury and larger extracranial injury can contribute to raise S100B serum concentration and cases with non-isolated TBI are expected to have higher S100B concentrations than those sustaining isolated TBI.

1.4.5. Summary

S100B is an isomer of a family of proteins called S100 which are calcium-binding proteins found in the brain with glial origin. They constitute 5% of soluble brain proteins. S100B has positive effects on neuroregeneration, metabolism and cytoskeleton at nano molar concentrations with harmful effects at micro molar concentrations. S100B serum levels appear higher in females than males and also in blacks and Asians than whites. Serum levels are maximal in the first year of life when they decrease till the age of 20 to reach a plateau of less than 0.1 µg/l for the rest of life.

S100B rises in several neurological (such as Alzheimer's disease, Down's syndrome, multiple sclerosis) and non-neurological conditions (such as melanoma). Also physical activity can cause its rise in the blood without trauma. S100B has been proposed in the forensic medicine to assess the cause or nature of the death or disability.

In mild TBI, S100B can be used for prediction of CT abnormality or a certain outcome. For outcome prediction, studies have suggested various AUCs ranging from 0.58 to 0.89. Similarly, in severe TBI, S100B predictive strength varies across studies with AUCs of 0.58 to 0.90. This diversity in AUCs implies differences in sensitivity and specificity too. Overall, it appears that sensitivity and specificity of S100B do not follow the same direction.

1.5. Current issues in TBI prognosis

1.5.1. Problems with the current prognostic models

The current prognostic models in TBI suffer from one or more of the following disadvantages: the sampling population being those included in a clinical trial, the time frame being historic and the location as being fromvarious countries. Clinical trials can not be truly representative of the TBI population for which the prognostic models need to be developed. Firstly, the inclusion criteria in clinical trials are tailored to serve the very purpose of the trial i.e. assessing the efficacy of the intervention. As such, those cases which demonstrate clear indication or contraindication for the intervention are excluded from recruitment according to the uncertainty principle. This poses a selection bias for a prognostic study. Secondly, some patients may be excluded because of the consent issue as to the intervention being administered or not. Further, it is important for a prognostic model to take the temporal advances and regional differences in brain injury trauma care into account. This means if a model has been derived from a historic TBI dataset, it may not necessarily be applicable for the current or more recent TBI populations due to introduction of new improvements in the trauma care policies bound to occur with the passage of time. Similarly, if a model has been developed in a different location, it may not necessarily be applicable in every location with differences in care policies. Murray *et al.* [86] compared some 'interventional' approaches in TBI management across TBI samples from various European countries and observed that in some countries such UK or France, victims of TBI have lower

chance of direct admission to the neurosurgical centre or to undergo intracranial operation than Italy or Spain. Regional differences not only apply to care policies but also to the demographic pattern or to the pattern of injury severities. Murray *et al.* compared TBI sample populations in Europe from this perspective and demonstrated that TBI victims in UK and France tend to be older, less often a vehicle occupant or to sustain major extracranial injury than the victims in Italy or France. All of these countries are, in fact, geographically near each other and all are grouped in the 'high income' or 'developed' countries.

The importance of time and regional differences is even more highlighted when the prognostic models are to be used for the trauma care benchmarking aimed at providing the relative performance of each trauma care centre at a national level. It defines how far (in a positive or a negative direction) the quality of care is from the national standard. This comparison is accomplished through comparing the prognosis in a trauma centre to the standard prognosis at the national level. The prognosis in the national level is obtained through prognostic models which are derived from the pooled datasets of trauma patients across the country. Conspicuously, using a prognostic model in a given country which has been developed in a different country may not be necessarily valid for the purpose of benchmarking of trauma care centres.

The British trauma care is different to that of other countries from either Europe or America. Some of these differences include but are not limited to transfer policies from the scene to hospital (for example via air or land), the skill level of the attending personnel and the initial interventions at scene or during the transfer (such as management of life threatening events, intubation,

drug administration etc.), admission to the neurosurgical unit (primary, secondary or never), adoption of conservative versus interventional approach in many clinical scenarios (such as ventilation, ICP monitoring, drug administration, operation etc.), discharge policies, community care, patients follow-up, rehabilitative strategies etc.

The CRASH dataset was from a clinical trial study and only 3 studies out of 11 studies merged in the IMPACT dataset were observational with the rest being clinical trials. Furthermore, the IMPACT dataset is somewhat historic in terms of the time of data collection since the most recent sub-dataset was Saphir trial which was terminated in 1997. Even, the termination time of some sub-datasets go back to 1980s such as UK4 (the UK four centre study as finished in 1988) or TCDB (Traumatic Coma Data Bank as finished in 1987). With regards to the regional differences in trauma care, this is an issue for both the CRASH and the IMPACT models. The CRASH dataset was collected from 50 countries around the globe and although the originators of the models built individual models for developed and developing countries, this distinction can not address the country's health care expenditure in general or the efficacy of trauma care in particular. This is shown in Murray's study indicating differences in TBI populations as per the demographic pattern, severity pattern or the therapeutic approaches across various developed countries even in the same European geographical region. This could also be the case for the IMPACT dataset which contains cases from various countries mainly in Europe and the North America.

Hukkelhoven's models suffer from the same disadvantages as being derived from clinical trials, being historic (1991-1994) and regionally diverse

(from 15 countries). Signorini's model is also historic as its termination dates back to 1991. Although the design of this study was observational with no regional diversity in the recruitment, the Signorini's model is the oldest amongst all TBI prognostic models.

1.5.2. Problems with S100B

A review of the literature on S100B prognostic role in TBI showed that across studies the performance of this biomarker varies significantly. For example, Townend *et al.* achieved a high AUC of 0.88 to predict disability whereas Nylen *et al.* obtained much lower AUC of 0.69 for the same prediction. The differences in S100B performance can be explained by differences in case-mix, S100B sampling time or laboratory method of measurement and also differences in type or time of outcome measure. Different case-mix implies different patients' characteristics which influence the outcome. For example, if a population sample contains significantly older subjects than the other sample, the older population may have significantly higher proportion of the adverse outcome. However, age is not the only factor which may cause a difference in case-mix. Some other important patients' characteristics are cause of injury, GCS, pupillary reactivity, presence/absence of extracranial injury, various CT abnormalities etc. Depending on these factors, one sample may have more severe TBIs than the other sample and hence different probability of adverse outcome. Furthermore, it appears various laboratory methods of S100B measurement vary in their sensitivity/specificity [87] and in the same way the time of blood sampling with respect to the time of injury affects the prognostic performance of S100B [76]. Lastly, Olivercrona *et al.* observed that S100B may be stronger to predict much shorter- term outcomes (3-month) than

outcomes at one year following injury or it is stronger for mortality prediction than disability prediction [77]. Unfortunately, studies on S100B are not uniform as to the above factors. Therefore, the results differ significantly. As such, it is not possible to draw a conclusion on whether or not S100B can be a good prognostic factor in TBI.

1.5.3. Prognostic models versus brain injury biomarkers

There are currently two available tools to predict the outcome in TBI; brain injury biomarkers and prognostic models.

Brain injury biomarkers can be used by taking a blood sample of the patient and then measuring the concentration of the biomarker and relating that to the outcome. This relationship can be established through knowing the sensitivity and specificity of the biomarker when it is above a certain cut-off. This is a familiar tool to clinicians as they commonly used various blood markers to associate them with a clinical event such as amylase for acute pancreatis, Troponin for myocardial infarction, ALF for liver cancer etc. Furthermore, the concept of sensitivity and specificity is well known to clinicians and they can appreciate the ‘degree of uncertainty’ regarding the predictive strength of these biomarkers. On the other hand, prognostic models involve a different aspect of medical statistics which is currently novel and unfamiliar. The measures to determine the performance of models such as AUC are not widely known by doctors and thus the degree of uncertainty can be hardly appreciated.

Prognostic models, however, do not incur measurement of a factor not commonly collected in trauma management. From this viewpoint, measuring a biomarker can pose extra cost whereas what current prognostic models use for

the outcome prediction are factors which are routinely measured during the physical examination or are collected for the sake of the patient monitoring such as O2 Sat., blood pressure etc.

The above issues are important regarding the potential practicality and popularity of prognostic models versus brain injury biomarkers. However, it may be even more important to understand how performance of prognostic models and brain injury biomarkers differs in that if one prognostic tool significantly outperforms the other, this can influence practicality and popularity issues. To the best knowledge of the investigator, no studies so far have addressed this problem in one single dataset and comparing various studies on the subject of prognostic models versus brain injury biomarkers is not reliably conclusive. This is because the highest AUC which has been achieved by prognostic models is 0.88 (the CRASH model) and 0.84 (the IMPACT model). In the same way, there are studies in the literature which have obtained high AUCs of around 0.90 for S100B outcome prediction [75, 83] beside some other studies which have observed much lower AUC of 0.58 or 0.69 [77, 79]. This means in some studies S100B competes with prognostic models whereas in some other studies prognostic models outperform S100B.

1.5.4. Combination of prognostic models with biomarkers

This combination can have two forms: adding a brain injury biomarker to the existing prognostic models or adjusting the brain injury biomarker with cofounders i.e. other TBI prognosticators. Either of these approaches would address the same topic as to whether or not this combination yields a stronger predictive tool. Adding biomarkers to prognostic models may enhance the

accuracy of models and considering other confounders when using a biomarker may enhance the accuracy of the biomarker predictability.

Adding a biomarker to prognostic models

The maximum AUC of currently well-developed models in TBI are 0.88 (CRASH) and 0.84 (IMPACT) in the derivation dataset which drops to respectively 0.77 and 0.80 in external validation. Although these models appear promising in terms of not making the prediction at random, it is unclear how much accuracy is required for various purposes of prognosis in clinical trials, trauma care benchmarking or clinical practice and whether this degree of accuracy is acceptable or not. When an ideal model has an AUC of 1, it may be argued that these models are far from this ideal. Whilst these models are still the best available models for prognosis, there is still scope to improve their accuracy by examining various ways to 'push' their AUC closer to the ideal model. Conspicuously, there are still patients who do not meet the predictions made by the prognostic models as they may die despite being predicted to survive and vice versa.

One option may be to add other predictors into the models. These models contain the core TBI predictors [88] i.e. age, GCS (or motor subscore) and pupillary reactivity plus some other covariates such as CT and vital signs. However, adding new variables may negatively affect the performance of the model in data other than the derivation dataset. This is commonly referred to as overfitting: when a model performs very well in the derivation set but

significantly drops in performance in different set of data. This may even make the model useless when it can not be applied in other TBI series.

Incorporating a brain injury biomarker to the prognostic models may enhance the accuracy of the models without putting the model at risk of overfitting. This is because some biomarkers such as S100B have been shown to be one of the strongest predictors of outcome in TBI amongst other common TBI prognosticators when considered individually and not in combination. Vos *et al.* compared the R^2 of various TBI predictors and found that S100B was stronger to predict poor outcome (GOS < 4) 6 months after TBI than CT and GCS. This may indicate that adding S100B to models may result in exclusion of other predictors which eventually may lead to a better performing model without holding too many covariates and thus the risk of overfitting.

Adjusting the biomarker with other prognosticators

The obvious conclusion from a literature review on S100B role in TBI presented in Table 7, Table 8 and Table 9 is contradictory results in that in some studies S100B performs very well whereas in other studies the performance is not promising. The different results can be explained by differences in case-mix, sampling time, laboratory method of S100B measurement or the type and time of outcome assessment (being morality, severe disability etc.).

Knowing the reason for this difference in results is important in that it may offer a much better prognostic tool than current prognostic models. In fact,

some authors have reached very high AUCs for S100B as close to 0.90 [75, 76].

1.5.5. Summary

There are three issues regarding the currently proposed prognostic models patient samples are from clinical trials (both IMPACT and CRASH), the dataset being historic (IMPACT) and places for recruitment being diverse (both IMPACT and CRASH). These factors may affect the reliability of models in other TBI populations including British cases of TBI. Regarding S100B, some studies have reached very high AUCs for outcome prediction which can compete with that of prognostic models whereas in some other studies low AUCs were found. The differing AUCs across studies can be due to differences in case mix, sampling time, laboratory methods of S100B measurement, or differences in time or type of outcome measurement. As a result of this inconsistency in the literature, it is not possible to decide between a prognostic model and S100B which one outperforms the other. This is important since, unlike prognostic models, S100B results are easy to interpret although prognosis models use routinely measured patients' characteristics. Trial of the combination of S100B with prognostic models may enable a stronger prognostic tool than either prognostic models or S100B on their own. This is because it is unclear whethe or not the current AUCs of prognostic models are adequately high or why in some studies very high AUCs of close to 0. 90 have been obtained.

1.6. Aim and objectives

1.6.1. Aim

The aim of the PhD is to improve our understanding of brain injury prognosis. For a number of reasons, this area still requires improvement (section 1.5):

1.6.2. Objectives

1. Since, the IMPACT and the CRASH models may not be reliable in a British dataset the first objective of the PhD is:

   To develop a prognostic tool to predict the survival in TBI applicable to the British trauma care.

2. It is important to know which of the two commonly proposed prognostic tools i.e. prognostic models and a brain biomarker are more accurate in their outcome prediction. Regarding this the second objective of the PhD is:

   To ascertain from a multivariate model and a blood test which one is better suited for prognosis in TBI.

3. One way of obtaining more accuracy in TBI prognosis can be combination of prognostic models with brain injury biomarkers. This combination can be either in the form of adding a biomarker to the prognostic model or adjusting biomarker prognosis with confounders. The former may enhance models' accuracy without risk of overfitting. Taking account of other TBI

prognosticators as confounders may enhance the biomarker predictability as some authors have reached AUCs of 0.88 to 0.90 for S100B outcome prediction. Thus the third objective of the PhD is:

To determine whether a combination of multivariate models and a blood test can significantly improve the prognosis in TBI.

1.6.3. Summary

In this section the aims and objectives of the PhD are discussed.

1.7. Approach, design and the importance of the project and the hypothesis formulation

1.7.1. Design

Observational versus experimental

None of the objectives of the PhD involve intervention as to change the course of the disease and as such the design of the study has to be observational.

Cohort versus case-control

It can be assumed that a prognostic tool acts similarly to a diagnostic test to diagnose a medical condition. In fact, both of these predict the occurrence of an event. Therefore, the appropriate study design has to be cohort. Case-control studies only recruit cases with or without the outcome (or disease) and are generally not suitable for assessment of a diagnostic test or prognosis.

Prospective versus retrospective

Although retrospective data can be used for assessing the prognostic accuracy of a model, there are some advantages pertained to prospective research as compared to retrospective research. The prospective studies can be designed and tailored for the specific purpose of the research right from the beginning with regards to many factors such as inclusion/exclusion criteria, data

collection, formatting and sampling size. Unfortunately, retrospective data does not provide much flexibility regarding these factors as the data collection often is part of a different research objective.

Overall, the best study design for the prognostic analysis is observational prospective cohort studies.

1.7.2. Approach and hypothesis

The Trauma Audit and Research Network (TARN) is a trauma registry based at Salford Royal NHS Foundation Trust, Manchester, UK which receives data on trauma patients from participating hospitals across England and Wales [89]. TARN is a part of changes recommended by the College of Emergency Medicine in trauma management through auditing and researching injury and care systems and aims to provide information on the quality of care. The trauma patients' profiles are submitted to TARN if they reach hospital alive after injury and fulfil at least one of the following criteria:

1. stay at hospital for longer than 3 days

or

2. are cared in the Intensive Care (ICU).

or

3. have interhospital transfer

or

4. die at hospital

TBI patients are a subgroup of trauma patients whose characteristics are stored in TARN. Therefore, TARN can provide a set of TBI cases who are from Britain, receive the recent trauma care and are part of an ongoing observational initiative. Such dataset can be used to derive a prognostic model

which is conventionally a better known prognostic tool than a brain injury biomarker. Moreover, the most popular statistical method for model construction is logistic regression. This has been shown by the systematic reviews on prognostic models in TBI performed by Perel *et al.* [31] and Mushkudiani *et al.* [32]. Regarding this the first and second null hypotheses of PhD are:

1. *The probability of survival is not influenced by patients' characteristics in severe TBI.*
2. *The logistic regression model does not explain the pattern of mortality in severe TBI.*

Furthermore, in 2005 a research study was embarked on at Salford Royal NHS Foundation Trust aiming at assessing the strength of S100B to predict death or severe disability 3 months following TBI. The setting was the hospital ICU to enrol severe TBI cases. The blood samples were obtained 24 hours (+/- 2 hours) following the injury and the degree of disability was assessed by GOS. The investigator joined the research group half way through to participate in patient recruitment, data collection and final univariate analysis. The study was initially targeted to recruit 100 subjects and the final dataset provided the opportunity to perform a multivariate analysis on the data. Therefore, it becomes feasible to compare the prognostic strength of a biomarker, S100B, to multivariate models on a single dataset and also to assess whether or not a combination of the two prognostic tools yields a better performance than each tool alone. Therefore, the following null hypotheses are formulated tailored to the second and third objective of the PhD:

3. *Addition of clinicodemographic characteristics of patients does not improve the prognostic performance of S100B in TBI.*
4. *There is no difference between prognostic performance of S100B and multivariate models in severe TBI.*
5. *There is no difference in prognostic performance between multivariate models which do/do not contain S100B as a predictor in severe TBI.*

1.7.3. Importance

Using the TARN data, it is hoped that at the end of PhD, a prognostic model has been developed which can be used for prognosis in British TBI individuals or series. Furthermore, the performance of prognostic models with that of S100B will be compared to provide an insight about which one of these prognostic tools may outperform the other. Finally, by trial of combinations of prognostic models with brain injury biomarkers, it will be investigated whether or not this combination can be an option to improve the predictive strength of prognostic models.

1.7.4. Summary

The PhD includes two parts: construction of a prognostic model for TBI patients using the TARN data and analysing a sample of 100 TBI cases who have their S100B serum levels recorded at 24 hr following the injury. The latter is aimed at comparing S100B prognostic performance to prognostic models or the combination of S100B with other prognosticators. Logistic regression will be used as the statistical method.

1.8. Study protocol

The PhD involves the analysis of two TBI datasets: one from TARN with the other being the S100B dataset. For the rest of this thesis, the TARN study and the S100B study are respectively refer to the analysis, results and the interpretation of results as performed on the TARN and S100B TBI datasets.

AIS codes: various injuries sustained by patients are coded in trauma registry using the AIS dictionary [90, 91]. This dictionary allocates a 6 digit AIS code to each injury which is followed by a post decimal figure representing AIS severity score (briefly severity score or AIS score). AIS scores ranges from 1 to 6. For example, a patient with penetrating abdominal injury with blood loss of more than 20% by volume receives AIS code and severity of 516006.3 (where 516006 is the code and 3 is the severity score).

ISS: it is the sum of the squared of the three highest AIS scores allocated to each patient. It ranges from 1 to 75.

1.8.1. TARN study

1. ***Variables selection:*** This stage addresses what variables are to be investigated in the prognostic analysis. There are many variables proposed for TBI prognosis but in the modeling, selection of every proposed prognosticator may result in overfitting model and it is advised to base this selection on clinical consensus and literature [32] to minimize the number of variables. For this study, the variables were selected from those prognostic studies by the IMPACT and CRASH plus those suggested by Perel *et al.* in their systematic review of prognostic models [31]. All these studies have employed multivariate analysis and the list of chosen variables are: age [23, 31, 35, 36, 92], cause of

injury[92], GCS [23, 31, 35, 36, 92], pupillary reactivity [23, 31, 35, 36, 92], Injury Severity Score (ISS) [31, 35], systolic and mean Blood Pressure (BP) [31, 36, 92], presence/absence of hypoxia [31, 36, 92], CT findings [23, 31, 35, 36, 92] and presence/absence of extracranial injury [23].

2. ***Database selection:*** TARN records the Abbreviated Injury (AIS) code and severity for each injury sustained by patients. Thus a given type of injury of interest can be retrieved from the TARN registry by selecting the cases that have the AIS code related to that very injury. This means that the TBI cases in TARN can be selected through brain injury AIS codes. These codes are those which come under internal organ under the head section of the AIS dictionary [90, 91] plus those AIS codes under the skull which are highly likely to be associated with brain injury. These codes pertain to basal skull fracture and not simple skull fracture and all have AIS severity of 3 and above (the skull AIS codes which are unlikely to be associated with concomitant brain injury all have AIS severities of 1 or 2). Furthermore, only recently TARN commenced recording pupillary reactivity and as such it is necessary that those cases who do not have this variable recorded be excluded from data retrieval. If these cases are not excluded, the number of missing information on pupillary reactivity would be high making the analysis of this variable impossible. This is not desirable since this variable is stated as one of the core TBI prognosticators in TBI along with age and GCS [88]. Overall, the selection criteria of the data from TARN is either head AIS code under internal organ or skull AIS code with AIS severity ≥ 3 AND pupillary reactivity recorded.

3. ***Data preparation:*** This involves determining the best time point (out of at scene and on admission) and also the best sub score of GCS for prognosis

and allocating the appropriate Marshall Class to each case. The latter is because TARN does not record CT images or reports and thus AIS codes have to be used as substitutes to perform this translation.

4. ***Univariate analysis:*** At this stage the linearity of continuous variables with outcome and their association with outcome in a univariate analysis are addressed. Some of the selected variables are obviously categorical by nature namely pupillary reactivity, cause of injury, various CT findings and presence/absence of extracranial injury. However, other variables can be taken either categorical or continuous. Logistic regression makes an assumption that all continuous variables have a linear relationship with logit odds of the outcome of interest (referred to as linearity assumption). This implies that if the continuous variable does not demonstrate this linearity, it has to be transformed to meet this assumption or else be categorized. Furthermore, at this stage it is required to run univariate statistical tests (t test or Mann Whitney U test for continuous variables and Chi Square test for categorical variables) to assess their association with outcome. In the same way, the odds ratio of outcome needs to be derived for each variable through logistic regression without adjustment for other prognosticators.

5. ***Multivariate analysis and model derivation:*** This stage is the actual model construction. It is performed initially with age, GCS, pupillary reactivity, ISS and extracranial injury and proceeds with addition of CT findings and blood pressure and O2 saturation. Age, GCS, pupillary reactivity and extracranial injury are the covariates in the basic CRASH models. Model A of IMPACT also contains these covariates apart from extracranial injury which is not recorded in IMPACT [92]. We added ISS to this list as the extent

of extracranial injury can affect it and thus ISS contains information on extracranial injury too. The next step is adding CT features which is the case in the CRASH models and IMPACT models. Lastly, addition of blood pressure and O2 saturation is based on the order in the IMPACT models.

6. ***Internal and external validation:*** Bootstrapping is used for interval validation and the resulting models are run on the IMPACT dataset and a later TARN dataset as the external validation.

1.8.2. S100B study

1. ***Variables selection:*** this stage is similar to what is done in the literature review for the TARN project. However, it further included those variables which are not recorded in TARN. The list of selected variables are clinico-demographic characteristics namely age [23, 24], GCS [23, 24], pupillary reactivity [23, 24], cause of injury [92], ISS [31, 92, 93], CT characteristics [23, 24, 94], extracranial injury [23], vital signs including mean (BP) [92, 95], systolic BP [92, 95], temperature [92, 96], and laboratory measures including PH [92, 97], Haemoglobin (Hb)[92, 97], Glucose [92, 97], Platelet (Plt) count [92, 97] and prothrombin time (PT) [92, 97] along with O2 saturation (O2 Sat.) [92, 96] and Intracranial Pressure (ICP) [35].

2. ***Retrospective data collection:*** the original S100B project contains data on age, cause of injury, CT classifications and ISS. However, values of other variables are required to be retrospectively recorded with patients identifiers (name or hospital number). Some of these data are recorded in the regional electronic records (the Electronic Patient Record (EPR) at hospital and also ICU for vital signs) and TARN. If the information is not available through these sources then the patients' case notes are reviewed. For GCS and pupillary

reactivity the time point of assessment is on admission whereas for laboratory measures and vital signs the closest time point to 24 hours after injury is selected. This is when the blood sampling was performed in the original S100B study.

3. ***Univariate analysis:*** Chi square test is used for comparing categorical variables across the survivals versus non-survivals or those with favourable outcome versus those with unfavourable outcome. Mann-Whitney U test or t test (depending on the normality of the distribution) is used for continuous variables for the same purpose. Similarly, logistic regression is used to derive the odds ratio for outcome without adjustment for other factors.

4. ***Multivariate analysis and model derivation:*** two types of multivariate analyses are performed. Firstly, the constructed model in the TARN project is run in the S100B dataset and its performance is measured as per AUC, Nagelkerke R^2 (section 2.7.2) and classification accuracy. Then the change in the performance is measured after adding S100B to this model. Secondly, a model is derived with only S100B and its performance is recorded. Subsequently those variables which are either found significant in the univariate analysis or present in the constructed model of TARN project are added to this model to observe the change in performance.

1.8.3. Summary

The protocols of the TARN and S100B studies are presented in this section.

1.9. Thesis structure

This PhD thesis is presented in the alternative format (i.e. by publications) as opposed to the standard format (which contains sections in the following order: introduction/background, methods, results and discussion and future direction). The difference is only the location of the materials for each section of the standard format in that firstly, they are not presented in the standard sequence and secondly, they are interrupted by information from other sections in several places. Thus the coherence of a standard thesis is not maintained although the information on introduction/background, methods, results and discussion is contained in several disparate places.

This PhD thesis contains 7 papers and all across the thesis these papers are referred to base on the number allocated to them (e.g. Paper 1, 2 etc.)

1.9.1. Relevance of each paper to a standard thesis

Paper 1: Utilisation and Assessment of Prognostic Models Derived Through Logistic Regression in Medicine.

This paper covers some background to the concept of predictive models discussing ways as to how to use them or how reliable they are to help us with the purpose they are built for. This paper also provides an insight about the model construction contributing to the methods section.

Paper 2: Predicting Outcome after Severe Traumatic Brain Injury Using the Serum S100B Biomarker: Results Using a Single (24h) Time-point [98].

This paper provides a portion of the evidence/background leading to hypotheses 3 and 4 and also includes partial results with regards to the first objective by assessing S100B as a prognostic tool. However, the prognostic performance of S100B reported in this paper may be far from ideal to be used. The hypothesis 3 addresses the issue that whether accounting for other predictive factors would enhance S100B predicative performance. However, the question still remains as to, among a brain injury biomarker or a multivariate model, which one is the superior prognostic tool. This latter issue is addressed in hypothesis 4. This paper is now published in *Resuscitation [98]*.

Paper 3: Comparing Model Performance for Outcome Prediction Using Total GCS and Its Components in Traumatic Brain Injury

This paper covers an important step taken to make the appropriate decision with regards to methods (step 3 of the TARN study protocol) for the purpose of addressing hypotheses 1 and 2. GCS is recorded by TARN in a way that each patient potentially can hold records of total GCS and each component i.e. motor, eye and verbal at two main time-points of at scene and on admission to emergency department. Considering all these combined records together, there can be 7 various records for GCS (total, motor, eye, verbal, verbal plus motor, motor plus eye and verbal plus eye); each with its own prognostic capability. This paper discusses which GCS component and at what time-point of measurement is expected to have the best predictive value to be contained in a model along with other important predictive factors of outcome in TBI.

Paper 4: Using AIS Codes to Classify CT Findings in the Marshal System [99]

This paper covers an important portion of the methods (step 3 of the TARN study protocol) in PhD to address hypotheses 1 and 2. An algorithm is proposed to assign a Marshall CT class to TBI patients based on recorded head AIS codes in the dataset or trauma registry. The assumptions, upsides and disadvantages of this proposed approach is discussed. This paper is now published in *BMC Medical Research and Methogology* [99].

Paper 5: Prognostic Value of Various Intracranial Pathologies in Traumatic Brain Injury

This paper presents an important step with regards to the methods (step 4 and 5 of the TARN study protocol) employed in PhD and also provides results and discussions to refute hypotheses 1 and 2. Based on literature; various CT findings have significant prognostic value in TBI. However the question relates to the relative influence of each injury such as SAH or contusion on outcome when other important factors such as GCS or pupillary reactivity are taken into account. This paper compares various proposed CT classifications such as the Marshal System or AIS scores and concludes on the strongest intracranial pathologies for outcome prediction. The appropriate intracranial pathologies proposed in this paper are included in the final prognostic models as presented in Paper 6.

Paper 6: Models of Mortality Probability in Traumatic Brain Injury: Results of TARN Modelling

This paper provides the methods, results and the discussion for the PhD with regards to hypotheses 1 and 2 and the first objective of PhD is met. Two prognostic models are presented which use patients' characteristics to make the prediction on survival. Refuting the hypotheses 1 and 2 is mainly based on this paper.

Paper 7: Comparing the Prognostic Performance of S100B with Multivariate Models in Traumatic Brain Injury

This paper includes methods, results and discussion of PhD with respect to hypotheses 3, 4 and 5 and to meet the second and third objectives.

1.9.2. Comparison with a standard thesis

Each paper, apart from Paper 1, contains sections of introduction, methods, results and discussion tailored to the specific objective of that paper. Although the objectives of papers may be different to those of the PhD, they are important in terms of their relevance to the PhD aim (i.e. improvement of our understanding of TBI prognosis). Overall, various sections of a standard thesis are presented in the following manner:

Introduction

This is presented in the introduction section of the PhD thesis and Paper 1.

Methods

This is presented in all papers. Additionally, there are two further sections at the end of Papers 6 and 7 titled as 'Expansion on methods' which provide more details on the methods used to achieve the five PhD objectives. This material was considered overly detailed or irrelevant to standard papers for journal publication.

Results

This is presented in all papers apart from Paper 1. Furthermore, there are some results pertained to the TARN project which follow Paper 6 (titled as 'Further results'). These were considered overly detailed or irrelevant to a standard paper and as such are presented in this way.

Discussion

This is presented in all papers apart from Paper 1 as each paper discusses the limitations, implications of the respective results and their relevance to the literature. However, there is also a separate section at the end of the thesis which provides further discussion on limitations, implications and comparison with the literature in the context of PhD hypotheses, objectives and aim. It is in the section that the discussion on how far the PhD objectives are met is given. In this section, a summary of the discussion from respective papers is provided with respect to limitations, comparison with the literature and implications/interpretation of the results in the first paragraph.

Future direction

Apart form Paper 1, each paper presents information related to this. Furthermore, there is separate section at the end of PhD which suggests other issues to be considered/followed in the future research of TBI prognosis.

1.9.3. Authors contribution

At the beginning of each paper, the names of co-authors of that paper are provided. Apart from Paper 2, the first author of all papers is the investigator (M.M.L.).

Mehdi Moazzez Lesko (the investigator): the investigator joined an ongoing research project on the S100B prognosis to derive its sensitivity and specificity for outcome prediction. Then this project was taken further for PhD along with TARN project. The investigator contributed to prospective data collection half way through. The retrospective data collection (for unrecorded data) was performed by the investigator as well. However access to the hospital EPR was restricted by the ethics committee and therefore this part was performed by F.L., although data retrieval from case notes was done by the investigator himself with the help of Health Record Office at the local hospital. Regarding the vital signs, the data were delivered in password-secured files via email by the IT officer at the local ICU with the patients' names as identifiers. All the statistical analysis of this study (either the original prospective project or the second retrospective part for PhD) was done by the investigator. Regarding the TARN project, the investigator performed all the statistical analysis as far as it was possible in Statistical Package for Social Sciences (SPSS) or Microsoft Excel. However, some parts of the analysis related to fractional polynomial analysis (section 7.5.7) were done by Omar Bouamra in SATA. Moreover, the

investigator drafted all papers apart form Paper 2 in which only results and discussion sections were written by the investigator.

Omar Bouamra (O.B.): As a statistician, O.B. contributed with appropriateness and accuracy of various statistical approaches taken and their results and interpretations for both TARN and S100B projects. Specifically, O.B. performed all the fractional polynomial analysis in STATA. Accordingly, he is the coauthor in all papers except Paper 4.

Timothy Rainey (T.R.): TR started the patients' recruitment of the original S100B project. As such, he is the first author of Paper 1 and the co-author of Paper 7.

Tom Jenks (T.J.): TJ is in charge of data retrieval in TARN and contributed to data retrieval at various parts of PhD. This particularly was for the analysis performed in Papers 3 and 6 and as such he is the co-author of these papers.

Charmiane Childs (C.C.): As the co-supervisor of PhD and the chief investigator of the original ongoing S100B project, C.C. is the main author of Paper 2 and for her supervision and contribution to content accuracy, is the coauthor of Papers 3, 4, 6 and 7.

Sarah O'Brien (S.O.B.): As the co-supervisor of PhD, S.O.B is the co-author of all papers (apart from Paper 2) for her supervision and feedback on the scientific approaches taken and also on the content accuracy and communication efficacy with the broader audience.

Marylyn Woodford (M.W.): M.W. particularly provided advice on the allocation of various brain injury AIS codes to Marshall Classes as presented in Paper 4. Also M.W. set out the objectives for TARN TBI modeling and

provided insight on data retrieval from TARN and thus contributing to Papers 6 and 3.

Laura White (L.W.): L.W. performed an in-depth review of the allocation of AIS codes to Marshall classes and as such she is the co-author of Paper 4.

Hester Lingsma (H.L.): From the IMPACT collaboration, H.L. reviewed the detailed approach taken in the model construction for its appropriateness and provided the relevant comments and suggestions. H.L. is the co-author of Paper 6.

Pablo Perel (P.P.): From the CRASH collaboration, P.P. reviewed the detailed approach taken in the model construction for its appropriateness and provided the relevant comments and feedbacks. P.P. is the coauthor of Paper 6.

Raphael Sacho (R.S.): RS participated in the first part of S100B project in the recruitment and follow-up of a number of patients. Therefore, he is the co-author of Paper 7.

Fiona Lecky (F.L.): As the main supervisor of PhD, F.L. initiated this work and assisted the investigator with the design and the progress of the work throughout the PhD with close supervision and guidance on every part from conduct of the studies and adherence to the protocols, to drafting of each paper and final thesis compilation.

It goes without saying that this work was impossible without invaluable contribution of every co-author to accomplish each piece of PhD.

1.9.4. Summary

The structure of the thesis as the alternative format, the contribution of each paper to the PhD aim and objectives and their relevance to a standard thesis

are discussed in this section. The contribution of each author to various papers is also provided.

2. Paper 1: Utilisation and Assessment of Prognostic Models Derived Through Logistic Regression in Medicine

Authors

- Mehdi Moazzez Lesko
- Omar Boumra
- Sarah O’Brien
- Fiona Lecky

2.1. Abstract

Over recent years quantifying the probability of the existence or occurrence of a medical condition through using various patients' characteristics has become increasingly common. Logistic regression is the commonly used statistical method to make such predictions and its output is called a prognostic model. This article is aimed at clinicians with average knowledge of statistics to introduce the conceptual framework behind the derivation of a prognostic model. Essential statistical concepts to comprehend how a model is used or how the accuracy and reliability of a model is assessed are described. To this aim, various types of model presentation including tables, scoring systems or web-based calculators are also discussed with examples. Similarly, various measures (indices) of model performance such as discrimination and calibration are described. Finally, it is explained what it is meant by model validity or generalisibility. This document can therefore be used to understand the various necessary terminologies used in the application of a prognostic model.

2.2. Introduction

Over recent years quantifying the probability of the *existence* or *occurrence* of a medical condition through using known demographic, clinical, laboratory or other patient's characteristics has become increasingly common. An example of this is to quantify the probability of the *existence* of Pulmonary Embolism (PE) when a number of observations such as presence/absence of tachycardia, haemoptysis or signs and symptoms of Deep Vein Thrombosis (DVT) etc. are known [100] or the probability of the *occurrence* of nausea or vomiting following operation when the patient's age, duration of surgery or anaesthesia are known [101]. To make such predictions, logistic regression is the statistical method which is commonly used and its output is called a *prognostic model*. A prognostic model is not expected to perform perfectly and thus there are several measures to describe its performance like the performance characteristics of a simple clinical or laboratory diagnostic test which is described by its sensitivity, specificity, negative and positive predictive value etc. Understanding the terms which quantify the performance of a prognostic model thus has similar importance to those using them, as understanding these aforementioned terms (describing the performance of a simple diagnostic test) has to clinicians making diagnoses.

With increasing use of prognostic models in medicine, this article aims to provide an overall introduction to the concept of prognostic models constructed through logistic regression, their use and the measures of model performance. A substantial knowledge of statistics is not required to comprehend the concepts since the essentials of medical statistics required to understand logistic regression are first explained in a concise manner herein.

This document can be used to understand various necessary terminologies used while applying a prognostic model.

2.3. Essentials of statistics for prognostic models and logistic regression

The value of statistics in medicine is simple to understand. For instance, we may wish to know whether or not pulmonary outcome is different in patients suffering from Chronic Obstructive Pulmonary Disease (COPD) who take a drug, say Salmetrol, to those patients who do not. Ideally, we have to investigate all patients with COPD in the 'world' by giving a drug (Salmetrol) to some of them and not to others, in a random fashion and to compare the outcome in these two groups. However, this approach is not possible for many reasons and thus the statistics assists us to analyse a sample of the population of patients with COPD and then to infer what goes on in the real world from the analysis of this one sample.

Statistics is the analysis of the variables and variables are objects/characteristics/attributes or 'entities' which vary and are not constant. The parameter is the other term which is alternatively used for the term variable. Alternative terms for the term 'value' are observations, measurements or data. The variables can be categorical or continuous.

Categorical variables vary within certain types/stages of a status or a medical condition (such as gender which varies between being male or female or the type of breast cancer which can vary among being lobular, ductal or mixed).

Continuous variables vary by adopting a measured figure. A continuous variable is in fact any variable which is not categorical (such as age, weight, blood sugar, serum bilirubin etc). However, sometimes continuous variables can be converted to categorical variables by defining a certain 'cut-off points'. For example, for some reason, a statistician may decide to categorise the continuous age as (< 17), (18 to 40), (41-65) and (> 66).

2.4. Position of prognostic models in statistical analysis

Statistics in medical science are used for descriptive or analytical purposes.

In the descriptive form, the average/median and the variability of the data are described. In this manner, the variability is described through presenting graphs (histogram, box-and-whisker plot or bar chart) or through quantifying the frequencies, range, centiles or variance/standard deviation.

In the analytical form of statistics (Figure 2), two or more groups of data are compared or their relation with each other is explored.

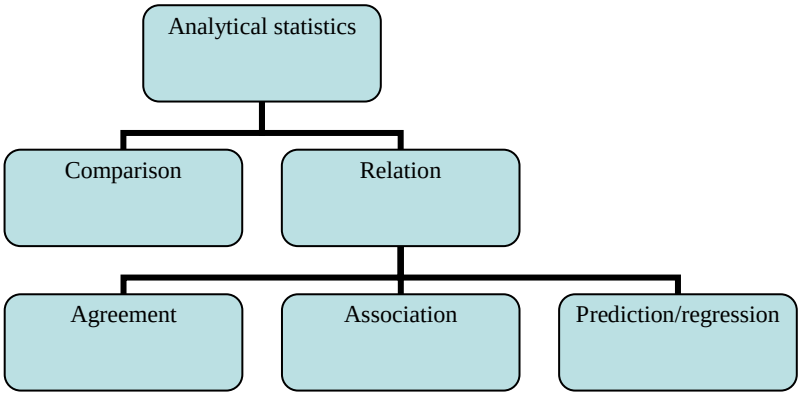


Figure 2 A schematic presentation of various common types of analytical statistics in medicine.

Comparing the data tells us whether or not the values are 'meaningfully' different in two or more groups. For example, whether or not the median of blood cholesterol in a group of patients who had a heart attack is

different to those who did not or whether or not the frequency of anaemia in a group of male population is different to that in a group of female population. Obviously, this ‘meaningful’ difference implies that one variable is expected to be higher or lower than the other in real life as inferred from the difference in the sample. This will be described in more details later (section 2.5).

Exploring the relation among two or more variables involves one of these three aspects: *agreement, association or prediction/regression.*

Agreement refers to similarity of the measurements performed by two people (inter-rater reliability) or at two time points (intra-rater reliability). One example of this is when one wishes to know how Glasgow Coma Scale (GCS) scores are similar (i.e. in agreement) if they are measured by two people or by one person at separate time points.

Association means that, among two variables or more, if one variable is higher, the other is expected to be higher or lower and thus they are not independent. For example, one may be interested to know if the renal clearance of Gentamicin is associated with or independent of the renal creatinine clearance. In such case, increase or decrease in renal clearance may change the serum level of creatinine or these two may be independent.

Prediction or regression is to estimate the value of one variable when the value of one or more other variables are known. This prediction is made through models and models are the mathematical equations used to make this prediction. The term prognostic model is commonly used to refer to these predictive models.

The are several types of models but the simplest one is a linear (univariate) model.

2.4.1. Linear models

The linear model is described by the regression line which is:

$$Y = (a \times X) + b$$

In which, Y is the variable to be predicted (also referred to as dependant, outcome or response variable) and X is the variable used to make the prediction (also referred to as explanatory variable/factor, independent variable/factor, covariate or regressor). a is called the coefficient with b being the constant or intercept.

The above model can be derived through statistical analysis of a sample of the population. It allocates actual figures to "*a*" and "*b*". This then enables a clinician to calculate the unknown value of Y using the known value of X in real life. This formula also shows that X and Y are associated with each other i.e. they are not independent. In fact, one way of evaluating the association of two variables is to perform the regression in that if a is zero, then it can be inferred that there is no association between X and Y.

2.4.2. Multivariate models

Multivariate models incorporate more than one variable to predict the value of the other variable. Such models are called multivariate models and they take the following form:

$$Y = (a_1 \times X_1) + (a_2 \times X_2) \ldots + b$$

An example of such model is when we want to predict the score of a functionality scale called Functional Independence Measure (FIM) total score in stroke patients at discharge. This is a scoring system based on some items related to motor and cognition functionality which ranges from 18 (total dependence or lack of functionality) to 126 (total independence or normal functionality). The following model was derived through analysing the data of 464 stroke patients [102]:

$$\textit{FIM total score at discharge} = -0.32\,(\textit{age}) + 0.80\,(\textit{FIM total score on admission}) - 0.13\,(\textit{onset to admission interval}) + 68.6$$

Using the above formula, then the discharge FIM total score can be calculated if age, admission FIM total score and the elapsed time between onset of symptoms to admission to hospital are known.

However, the information provided by the above formula is more than just a prediction in that:

- Firstly, if other variables are investigated during the construction of a model and they do not appear in the final model, this means those variables are not associated with or have no effect on the dependent variable. For example, during the construction of the above model to predict FIM, the effect of sex and age on FIM was also investigated, but these are not included in the final model. It can therefore be concluded that these factors have no effect on FIM. Obviously, if the effect of a variable is not considered during the model construction, absence of such variable in the model does not necessarily mean no additional effect on the dependant variable.

- Secondly, multivariate models take into account the effect of multiple variables on a dependent factor collectively. This is usually referred to as adjustment with confounders to address the concern that factors may alter the effect of each other. An example of this situation is when one is interested to know whether or not Body Mass Index (BMI) has any effect on developing a cardiac complication following hip fracture in elderly [103]. Here, there are other 'confounders' which may contribute to developing such complication such as older age or male sex. Performing a multivariate analysis by taking BMI, age and gender into account to predict the occurrence of cardiac complication addresses the effect of age and gender on outcome as well. It is then stated that the effect of BMI on risk of developing cardiac complication following hip fracture is adjusted for age and gender (confounders).
- Thirdly, in a model which predicts one variable using one or more other variables, it can be stated that the predicted variable is associated with the predictors or the predictors are risk factors for the predicted variable. For example, when the model to predict the cardiac complication in the elderly with hip fracture contains age, gender and BMI, it can be stated that age, gender and BMI are risk factors for a cardiac event.

2.5. Position of logistic regression in statistical methods

Logistic regression is a 'statistical method' used to derive a prognostic model. Statistical methods are specified mathematical calculations which can serve the purpose of the analysis; being either making a comparison or assessing the

relation between two or more variables. Examples of some commonly used statistical methods are t test, Chi Square test, Mann Whitney U test etc.

The idea is firstly, to describe the population sample based on the purpose of the analysis and secondly, to determine how certain one can be that what is observed in the sample; is going on in the actual population or other samples of this population. The degree of certainty with which one can generalise what is observed in a sample to the whole population or 'the meaningfulness' of what is observed is referred to as the statistical significance.

There are two ways to investigate and demonstrate the statistical significance: estimation and hypothesis testing. Confidence Intervals (CI) are commonly used for estimation and the *p* value is used for hypothesis testing. Most statistical methods in medicine address the issue of clinical significance through hypothesis testing although often knowing the CI is more informative.

2.5.1. Estimation and Confidence Intervals

The Confidence Interval (CI) is presented by two figures between which we can be confident the actual value in the population resides with a 'specified degree of certainty'. This degree of certainty is stated when the CI is reported such as *97. 5%* CI or *95%* CI (but 95% CI is the most commonly used). A 95% CI can be obtained for the mean, frequency, odds ratio and many other statistically descriptive or analytical factors.

It simply means that if we repeat our sampling 100 times (or take 100 samples of the population), the value of our interest (be it mean, frequency etc) is something between the two given figures 95 times (or in 95 samples of population). This information is very useful simply because by analysing one

sample of population, we can estimate what would be observed if we had 100 samples of population instead of only one. For example, a study found that 80% of Intravenous Drug (IV) users in a sample of 499 cases detained for mandatory rehabilitation were positive for Hepatitis C Virus (HCV). The 95% CI from this sample was 76.2-83.6 [104]. This means if the sampling is repeated several times from the population of IV drug users under rehabilitation, the frequency of being HCV (+) is between 76.2 to 83.6, in 95% of the samples.

CIs can also be used to investigate the association between two variables and for this purpose, knowing the CI for the odds ratio is helpful. If this CI does not contain 1 (which demonstrates no association), we can conclude that the association between the two variables are statistically significant. For example, one may be interested to know if diabetes is associated with acute Myocardial Infarction (MI). We can simply obtain the CI for the odds ratio of having MI when the patient is diabetic in a dataset which contains this information. Shaw. *et al.* Observed an odds ratio of 1.41 with 95% CI of 1.26 and 1.57 by analysing a sample of 100,253 patients [6]. This demonstrates a statistically significant association since we can expect the presence of the association described by the confidence limits, on 95% of the occasions if the sampling of the population is performed repeatedly.

2.5.2. Hypothesis testing and the p value

The p value is the probability that what is observed in the sample, actually occurs in the population if the null hypothesis is true. The null hypothesis differs according to the purpose of statistical analysis. For example, if the purpose is to compare a variable among two or more groups of values, then the

null hypothesis is that there is no difference (or the difference is zero). Table 10 presents some null hypotheses related to various statistical analyses and methods. There is a cut-off level for the p value commonly at 5% below which the null hypothesis is declared rejected. *Thus, when the p value of a statistical analysis is less 5%, it is stated that the null hypothesis is refuted.*

Purpose of statistical analysis				*Example of statistical method*	*Null hypothesis*
Comparison		Categorical variable		Chi square test	The difference between frequencies (or proportions) is zero
		Continuous variable		T test, Mann Whitney U test	The difference between the means is zero (t test), the difference between the medians is zero (Mann Whitney U test)
Relation	**Agreement**	-		Kappa (**κ**) test	The ratings are independent
	Association	Categorical variable		Chi Squared Test	The difference between frequencies (or proportions) is zero
		Continuous variable		Simple linear regression	The coefficient is zero
	Prediction /regression	Continuous variable		Multiple linear regression	The coefficient is zero
		Categorical variable	2 categories	Logistic regression	The coefficient is zero
			More than two categories	Multinomial regression	The coefficient is zero

Table 10 Various statistical methods along with their purpose and the hypothesis.

For example, when determining whether the Length Of Stay (LOS) in hospital, in trauma patients with and without underlying medical condition is different or not, the null hypothesis states that there is no difference. If in a sample of trauma patients, we observe a difference between the median LOS in patients with underlying medical condition and those without underlying medical condition with a p value of, say 3%, we can then say that if there is no difference between the LOS in these trauma populations, then there is only 3% chance that we observe such a difference in our sample. In other words, the type I error of our analysis is only 3%.

For regression, the null hypothesis is that the coefficient is zero. In multiple regressions, i.e. there is more than one covariate; each covariate holds a p value related to its coefficient.

2.5.3. Logistic regression

Logistic regression *is a statistical method to construct a model to predict a categorical variable which has only two categories*. Therefore, the outcome variable can be the existence (i.e. absence/presence) or occurrence/non occurrence of a medical condition such as survival, a postoperative complication, having a cancer etc.. The output of the model for a patient with given characteristics is $\log_e$ (odds of what is predicted); for example, $\log_e$ (odds of survival, a postoperative complication, having a cancer, etc.).

Knowing the odds, calculation of the probability is simple since:

$$Odds = probability / (1\text{-}probability)$$

therefore:

$$Probability = odds / (1 + odds)$$

The remaining two sections of this article explain various ways of presenting the output of logistic regression and also the indices which are used to assess how accurate a model is in its predictions. The output of logistic regression can be in two forms: a model (which contains the covariates, coefficients and the constant) or the odds ratio for each covariate. If the output of a logistic regression is in the form of a model, there are several ways to present this.

2.6. Presentation of the results of logistic regression

2.6.1. Table

This is the simplest from of presenting the output of logistic regression analysis. The table contains covariates with their related coefficients, odds ratio (with or without 95% CI) or *p* value. If the table only contains the odds ratio, then it just provides information on the significance of the effect of each covariate on outcome which can be determined by p value or the 95% CI for odds ratio. Such table is not helpful to calculate the probability of the outcome variable.

Knowing the coefficients and the constant from the table, the following equation can then be obtained:

$$Log_e(odds\ of\ outcome) = (coefficient)_1 \times (covariate)_1 + (coefficient)_2 \times (covariate)_2 + (coefficient)_3 \times (covariate)_3 + \ldots + the\ constant$$

If we call Log_e(odds of outcome) as B and the probability of outcome as A, the mathematical calculation of the probability of the outcome is as follows:

$$B = log_e (A/ (1-A))$$

$$e^B = A /(1-A)$$

$$A = e^B / (1+ e^B)$$

Interaction

Two covariates in a model are interacting with each other, when they are affecting the degree to which they each influence the outcome (dependent variable). This is declared in the model by the originators and is simple to mathematically incorporate in the regression equation. In such case, the interaction is presented as an 'extra' covariate which, like other covariates, holds an odds ratio and a coefficient.

Table 11 presents a model which was derived to predict in-hospital morality for patients who undergo a vascular surgery [105]. This model employs the post-operative values of a number of patients' characteristics. In this table, the three interactions are in bold and are: between *out-of-hours surgery* and *chronic statin therapy* and between *last mean daily Heart Rate (HR)* and *last mean daily Systolic Blood Pressure (SBP) < 100 or >179 mmHg'* and between *last mean daily HR* and *withdrawal of chronic beta-blocker*. These interactions mean that the amount of the effect which, for

example, last mean daily HR has on the probability of death would change depending on last mean daily SBP or withdrawal of chronic beta-blockers. This is simple to mathematically describe as the interactions can be treated as individual covariates with their respective coefficients.

Covariate	***Coefficient***
Age (per one-year increase)	0.052
Creatinine > 180 mmol.l-1	1.625
Surgery out of hours × no chronic statin therapy	2.113
Last mean daily HR (per beat per minute increase)	0.017
Last mean daily HR × 'last mean daily SBP < 100 or > 179 mmHg'	0.019
Last mean daily HR × withdrawal of chronic beta-blockade	0.017
Constant	–7.951

Table 11 The model to predict in-hospital mortality following a vascular surgery.

Thus the equation obtained from this table to calculate the $\log_e$ (odds of in-hospital death) is as follows:

$$\log_e(\text{odds of in-hospital death}) = (\text{Age} \times 0.052) + ((\text{Creatinine} > 180\ \text{mmol. l-1}) \times 1.625) + ((\text{Surgery out of hours} \times \text{no chronic statin therapy}) \times 2.113) + (\text{last mean daily HR} \times 0.017) + ((\text{last mean daily HR} \times \text{'last mean daily SBP} < 100 \text{ or } > 179\ \text{mmHg'}) \times 0.019) + ((\text{last mean daily HR} \times \text{withdrawal of chronic beta-blockade}) \times 0.017) - 7.951$$

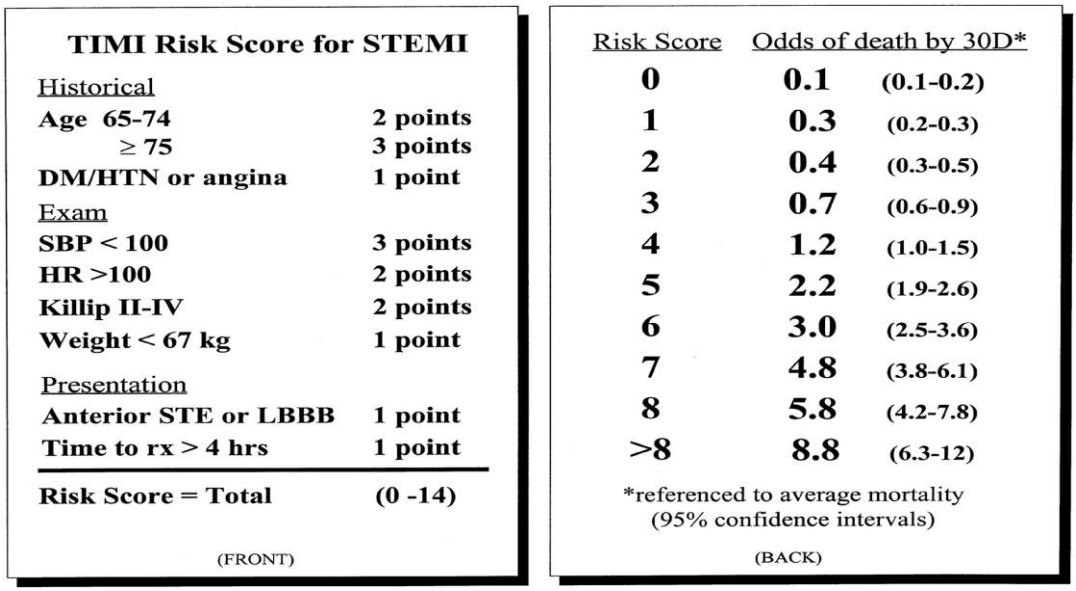
In the above formula, the variables which are of 'yes/no type' such as out-of-hour surgery are 1 when present and are 0 when absent. Therefore, if they are absent, their effect on outcome would be zero.

In the above model, some interactive variables such as *out-of-hour surgery* or *withdrawal of chronic beta-blocker* are not contained as individual covariates in that there is no covariate as out-of-hour surgery or withdrawal of chronic beta-blocker separately. This is not always the case in all models. As can be seen in this model, Last mean daily HR is contained individually and also in interactions with last mean daily SBP and withdrawal of chronic beta-blockade.

2.6.2. Scoring system

In this form of presentation, each covariate receives a score or point and then the sum of all scores (called t*he sum score*) is calculated. Each given sum score holds an equivalent probability or odds ratio of the outcome. There are several ways to obtain the related probability of a sum score. This may be provided in a table, presented in a graph, plotted against a line in the monogram or calculated through a linear equation.

Example 1: predicting 30-day mortality in patients with ST-elevation MI (TIMI risk score): table presentation of the probabilities (Figure 3)



TIMI Risk Score for STEMI

Historical	
Age 65-74	2 points
≥ 75	3 points
DM/HTN or angina	1 point
Exam	
SBP < 100	3 points
HR >100	2 points
Killip II-IV	2 points
Weight < 67 kg	1 point
Presentation	
Anterior STE or LBBB	1 point
Time to rx > 4 hrs	1 point
Risk Score = Total	(0 -14)

(FRONT)

Risk Score	Odds of death by 30D*	
0	0.1	(0.1-0.2)
1	0.3	(0.2-0.3)
2	0.4	(0.3-0.5)
3	0.7	(0.6-0.9)
4	1.2	(1.0-1.5)
5	2.2	(1.9-2.6)
6	3.0	(2.5-3.6)
7	4.8	(3.8-6.1)
8	5.8	(4.2-7.8)
>8	8.8	(6.3-12)

*referenced to average mortality
(95% confidence intervals)

(BACK)

Figure 3 The TIMI risk score to predict mortality within 30 days following ST-elevation MI (from: Morrow DA, Antman EM, Charlesworth A, Cairns R, Murphy SA, de Lemos JA, Giugliano RP, McCabe CH, Braunwald E: TIMI Risk Score for ST-Elevation Myocardial Infarction: A Convenient, Bedside, Clinical Score for Risk Assessment at Presentation : An Intravenous nPA for Treatment of Infarcting Myocardium Early II Trial Substudy. *Circulation* 2000, 102:2031-2037.).

Using this model, a risk score can be calculated which is the sum of points given to each particular characteristic presented in the 'FRONT' table. Subsequently, the equivalent odds ratio of death by 30 days can be spotted in the 'BACK' table knowing the calculated risk score [106].

Example 2: the APCHE III prognostic system: the graph presentation of the probabilities (Figure 4)

This model was developed by analysing a dataset of 17,440 cases to predict the probability of mortality in hospital for patients who are admitted to Intensive Care Unit (ICU) due to various medical conditions (APCHE: Acute Physiology and Chronic Health Evaluation) [107].

Initially, the APACHE III score is calculated which is the sum of points allocated to 17 covariates (the sum score). Then the probability of mortality related to the calculated sum score can be obtained in Figure 4 based on the cause of the admission to the ICU.

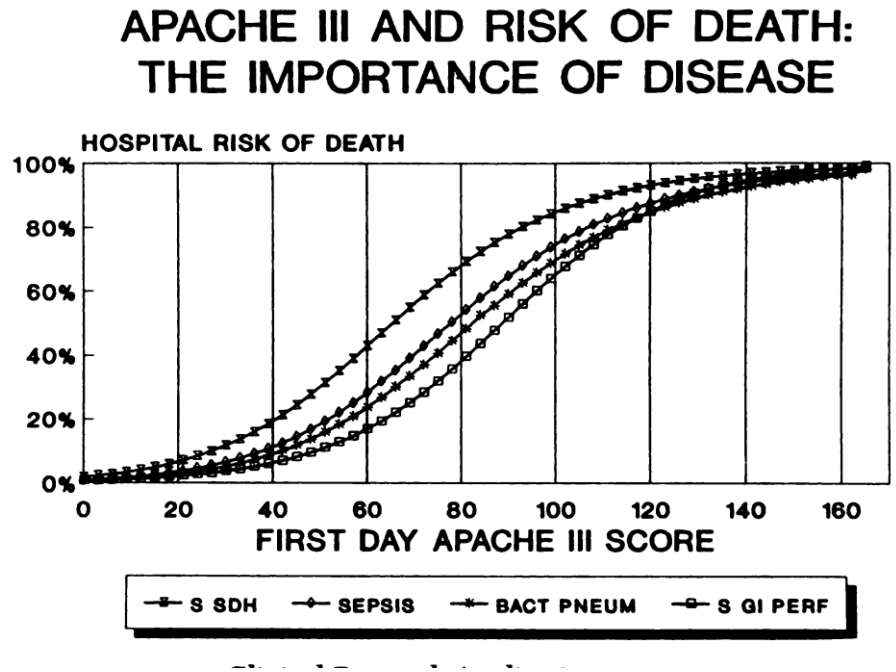


Figure 4 The APCHE III prognostic system to predict mortality in ICU (from: Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, Sirio CA, Murphy DJ, Lotring T, Damiano A, Harrell FE: THE APACHE-III PROGNOSTIC SYSTEM - RISK PREDICTION OF HOSPITAL MORTALITY FOR CRITICALLY ILL HOSPITALIZED ADULTS. *Chest* 1991, 100:1619-1636.)

Example 3: calculation of the probability of distant metastasis in renal cell carcinoma: a nomogram (Figure 5)

A nomogram is a graphical calculator and in multivariate analysis, it contains several *scoring lines*. Each covariate has a scoring line on which its value can be plotted. There is usually three other lines to plot the point given to each value, to spot the total points (the sum score) and finally to spot the final probability (or odds ratio) of the outcome. Figure 4 presents a ‘nomogram’

which can be used to calculate the probability that a patient with renal cell carcinoma has a distant metastasis knowing tumour size and symptom classification (asymptomatic, local or systemic) [108].

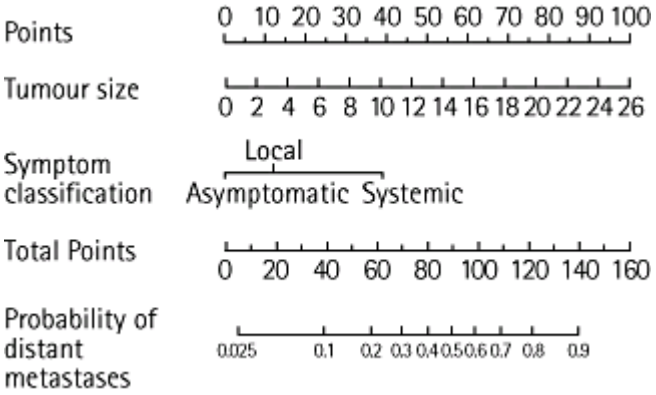


Figure 5 A nomogram to predict distant metastasis in renal cell carcinoma (from: Georg CH, Jean-Jacques P, Claudio J, Paul P, Alexandre de La T, Laurent S, Gregory V, Jacques T, Luca C, Vincenzo F, et al: Patients with distant metastases from renal cell carcinoma can be accurately identified: external validation of a new nomogram. *BJU International* 2008, 101:39-43.)

Example 4: The IMPACT model for the probability of mortality within 6 months following severe traumatic brain injury: the linear equation (Figure 6)

In order to make such prediction, first the sum score is calculated using table in Figure 6 [24].

Characteristics	Value	Score	Sum
Age (years)	≤ 30	0	
	30 – 39	1	
	40 – 49	2	
	50 – 59	3	
	60 – 69	4	
	70 +	5	
Motor score	None/extension	6	
	Abnormal flexion	4	
	Normal flexion	2	
	Localizes/obeys	0	
	Untestable/missing	3	
Pupillary reactivity	Both pupils reacted	0	
	One pupil reacted	2	
	No pupil reacted	4	
	Sum score core model		

Figure 6 The IMPACT model to predict mortality in 6 months following severe traumatic brain injury (from: 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:e165; discussion e165.)

Then, the probability is calculated using the following equations:

$$LP = -2.55 + 0.275 \times sum\ score$$

$$Probability\ of\ mortality\ within\ 6\ months = 1/(1+e^{-LP})$$

2.6.3. Web-based calculator

Oftentimes, an online-calculator is provided which by putting the respective value of a number of variables the probability of outcome is given. An example of such calculators is the one proposed by Chen. *et al.* which predicts the probability of caesarean delivery using a number of mother's characteristics such as age, height, weight, gestational age, pregnancy weight gain etc [109]. This calculator is available at http://www.ise.ufl.edu/rmfe/projects/CSPrediction/CSpredictEnscrpited1.htm.

2.7. Model Accuracy

It is ideal that a model makes an absolutely accurate prediction. However, this does not happen in reality and therefore there are several measures to assess the performance of a model (or how accurate the model's prediction is). This is similar to a diagnostic test for a disease condition which is not always accurate and thus the accuracy of a test is measured through its sensitivity, specificity etc. In the same way, accuracy of a model's prediction can be assessed through several measures. Overall, the performance of a model has two dimensions: discrimination and calibration.

It is first essential that the distinction is made between the *predicted probability* and the *observed probability*. The *predicted probability* is the probability of the outcome as calculated by the model for a patient with given characteristics. The *observed probability* for a patient with given characteristics is the frequency of the outcome among patients who all have those characteristics in the dataset. The idea is that two or more patients with the same characteristics may have different outcomes observed in real life (and thus in the dataset) whereas these two patients have the same probability predicted by the model. The observed probability is the proportion of the outcome observed in the dataset of patients with a given set of characteristics and the predicted probability is the proportion predicted by the model. For example, a model may predict the probability of admission to intensive care for a severe trauma patient at the age of 50 with underlying heart disease as 60%. This means every patient with the above characteristics has 60% chance of ICU admission (predicted probability). However, in the dataset from which the model has been derived, some patients at the age of 50 and with heart problems

may be admitted to ICU and some of them with the same characteristics may be never cared in ICU. The observed probability is the frequency of cases with these given characteristics who were cared in the ICU.

2.7.1. Discrimination

Discrimination is the ability of the model to predict higher probability of outcome for the patients who experience the outcome than those who do not experience the outcome. In a model with a perfect discriminative performance, all patients who experience the outcome at all times have a higher predicted probability than the probability predicted for patients who do not experience the outcome. However, the perfect performance rarely happens in reality.

Area Under the Roc Curve (AUC) or C statistics is the measure to assess the discriminative performance of the model:

Area Under the Roc Curve (AUC) (C statistics)

This is the probability that in a random pair of patients that one has the outcome and the other does not, the patient who experiences the outcome has a higher predicted probability by the model than the one without the outcome. For example, an ideal model for prediction of death in a disease condition is expected to have AUC of 100% (or 1). This means that the predicted probability of death for a patient who dies is always (i.e. at 100% of the times) higher than the predicted probability of death in a patient who survives.

Vergouwe *et al.* stated that in comparing the AUC (or performance) of a prognostic model across various population samples, it is important that the influence by differences in case-mix and also the value of coefficients taken

into account {Vergouwe, #295}. This means that a large difference between performance of a model in the development population compared to the external validation population can partially be explained by the differences in case-mix or the coefficients and a different AUC does not necessarily reflect a different model performance.

However, this ideal model is rarely achieved in reality and thus the AUC is expected to be less than 100%. Nevertheless, if the AUC is more than 50%, it can be said that at least the prediction is not made by chance. (A prediction which is made by chance is like flipping the coin which has 50% chance of being either head or tail).

There is no consensual cut-off for AUC to regard models with AUCs of above that cut-off as good discriminative models. However, an AUC of above 85% may be an indicator of a good discrimination. An example of a prediction with this degree of AUC is to predict the breast cancer through mammography [110].

2.7.2. Nagelkerke R^2

This measure is sometimes given to show a model's performance. It ranges from 0 to 1 (or from 0 to 100%) with 1 indicating the perfect model. Nagelkerke R^2 is sometimes referred to as the equivalent of R^2 in a linear model. Thus understanding R^2 which is used for a linear model can partially help to understand the concept of Nagelkerke R^2 for a logistic regression model. The idea and use of both these measures are the same despite their difference in mathematical calculation.

R^2

By definition, R2 is the amount of variability in the predicted variable (Y) which is explained by the model.

Variability

In the linear regression equation of $Y = (a \times X) + b$, for each value of X, there is only one Y value. However, this is not the case in reality (and thus is the dataset) in that each value of X can potentially hold several Y values.

For example, one may wish to predict the Length Of Stay in hospital (LOS) from the admission age using a linear model. Obviously each value of age can potentially have several different LOS but if a model is derived (where the equation will take this form: $LOS = (a \times age) + b)$ for each value of age there is only one LOS predicted by the model. Thus for each age in the dataset, there can be a number of LOSs which are equal to what the model predicts and a number of LOSs which are different to this prediction.

Variability in a set of data is a mathematical calculation which is performed through *sum of squares* and has three types: *total* sum of squares, *explained* sum of squares and *residual* sum of square.

total sum of squares = $\sum_{i=1}^{n} (y_i - \bar{y})^2$

explained sum of squares (explained variation) = $\sum_{i=1} (\hat{y}_i - \bar{y})^2$

residuals sum of squares (unexplained variation)= $\sum_{i=1} (y_i - \hat{y}_i)^2$

Where y and $\hat{y}$ are the observed and predicted values respectively (Figure 7). According to this graph, the explained value is the difference between the predicted value (the constant line) and the mean of all observed values (the dotted line). This is presented as $\hat{y} - \bar{y}$ in the above formula. The unexplained value is the difference between the observed value (○) and the predicted value (the constant line). This is presented as $\hat{y} - y$ in the above formula.

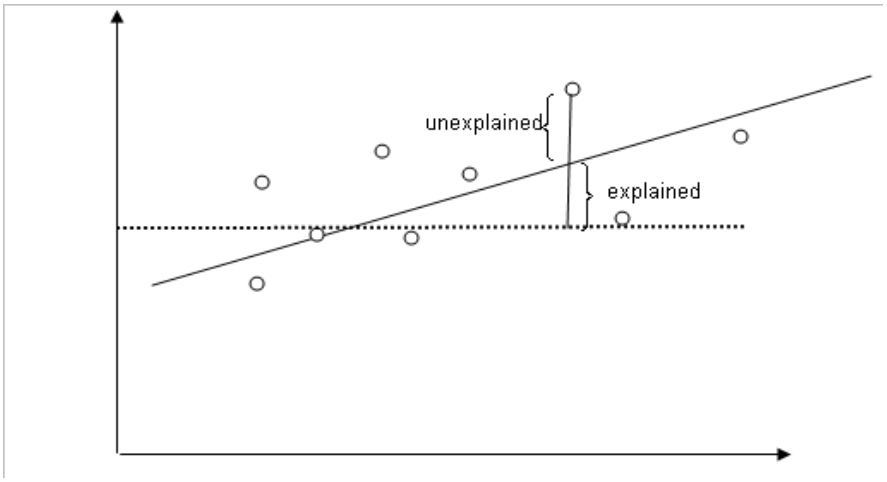


Figure 7 Explained and unexplained values and variation. The dotted vertical line represents the mean of observations.

Calculating R^2

R^2 is calculated using the following formula:

$$R^2 = 1 - (\textit{residuals sum of squares/ total sum of squares})$$

or

$$R^2 = 1 - (\textit{unexplained variation/ total sum of squares})$$

Similar to AUC, there is no cut-off for R^2 above which to regard the performance of the model acceptable. However, R^2 is a good measure to compare the performance of several models if it is known for every model.

2.7.3. Calibration (goodness of fit)

This refers to how close the predicted probability by the model is to the observed probability. . This is expressed as *how the model fits the data*. In an ideal model all predicted probabilities are equal to the observed probabilities. However, similar to discrimination, this perfect performance is rarely achieved.

The following measures are used to mathematically describe the calibration of a model:

Negative Predictive Value (NPV): *This is the proportion of the patients with predicted probability of less than 50% by the model who did not experience the outcome.* For example, if the NPV of a model is 80%, there is 80% chance that a patient with say 30% predicted probability of death (i.e. any value less than 50%) did die in reality.

Positive Predictive Value (PPV): *this is the proportion of the patients with predicted probability of more than 50% by the model who do experience the outcome.* For example, if the PPV of a model is 80%, there is 80% chance for a patient with say 60% predicted probability of death (i.e., any value more than 50%) did die in reality.

Sensitivity: *this is the proportion of the patient with the outcome who hold the predicted probability of more than 50% by the model.* For example, if the sensitivity of a model is 80% this means the model makes the prediction of death of more than 50% for 80% of patients who die.

Specificity: *this the proportion of the patient without the outcome who hold the predicted probability of less than 50% by the model to experience the outcome.* For example, if the specificity of a model is 80% that means the model makes the prediction of no death (i.e. survival) of more than 50% for 80% of patients who do not die.

Accuracy rate (classification accuracy): *this is the proportion of the patients who are correctly classified by the model.* The correct classification implies that if the patient experience the outcome, the predicted probability is more than 50% and if the patient does not experience the outcome, the predicted probability is then less than 50%.

Mathematical calculations of NPV, PPV, sensitivity, specificity and accuracy rate can be presented through a 2 ×2 table (similar to a diagnostic test) as follow:

Outcome *Predicted probability > 50%*	+	-
+	a	b
-	c	d

$$NPV = d / (d + c)$$

$$PPV = a / (a + b)$$

$$Sensitivity = a / (a + c)$$

$$Specificity = d (d + B)$$

$$Accuracy\ rate = (a + d) / (a + b + c + d)$$

The concept of these measures for a model is overall similar to a diagnostic test. For example, sensitivity of a diagnostic test is the proportion or the probability that a patient with the disease has a positive test result. In a model, the (+) test result to experience the outcome is when the predicted probability is more than 50%. This cut-off of 50% is 'by default' and it can also be set at a figure higher than 50% in which case this would be mentioned by the originators of the model.

2.7.4. Hosmer- Lemeshow goodness of fit test (HL statistics)

This is a statistical method referred to as 'goodness of fit' test to assess the calibration of a prognostic model. The test provides a p value and if its value is less than 5%, it is stated that the model does not fit the data. This reflects that the model is not satisfactorily calibrated.

2.7.5. Brier score (the average quadratic)

This is the sum of the squared differences between observed probability and predicted probability divided by the number of patients. Mathematical presentation of this is as follows:

$$Brier\ score = \sum (observed\ probability - predicted\ probability)^2 / n$$

Where n is the number of patients.

A perfect model holds a Brier score of 0 and the worst performing model holds a brier score of 1 (100%).

2.7.6. Calibration plot (Calibration curve)

This is the plot of observed probability against predicted probability. The calibration plot is a valuable tool to assess calibration performance by

providing immediate interpretable information across the whole range of probability which is from 0 to 1. This means in the calibration plot one can observe, for example, the predicted probabilities are closer or more distant to the observed probabilities in the higher ranges of probability of outcome than lower ranges or vice versa.

The X axis in the calibration pot is usually predicted probability and the Y axis is usually the observed probability. There is line of comparison which represents the perfect calibration i.e. observed probability is equal to predicted probability at all times ($X=Y$). In interpretation of a calibration plot, these issues should be considered:

- Overall, how close the predicted probabilities are to observed probabilities. This is just the visual presentation of the calibration because the mathematical answer to this issue can be obtained through HL statistics (section 2.7.4).
- At what ranges of predicted probabilities (i.e. high, moderate or low predicted probability), there is more closeness or more distance from the observed probability.

Figure 8 presents the calibration curves for two models: Admission APACHE II (the line) which uses patients characteristics on Admission and Worst APACHE II (the dotted line) which uses the worst patients characteristics during their stay in ICU to predict mortality [111]. The straight line represents the situation in which the models hold the perfect calibration. This figure depicts following information about Admission APACHE II model:

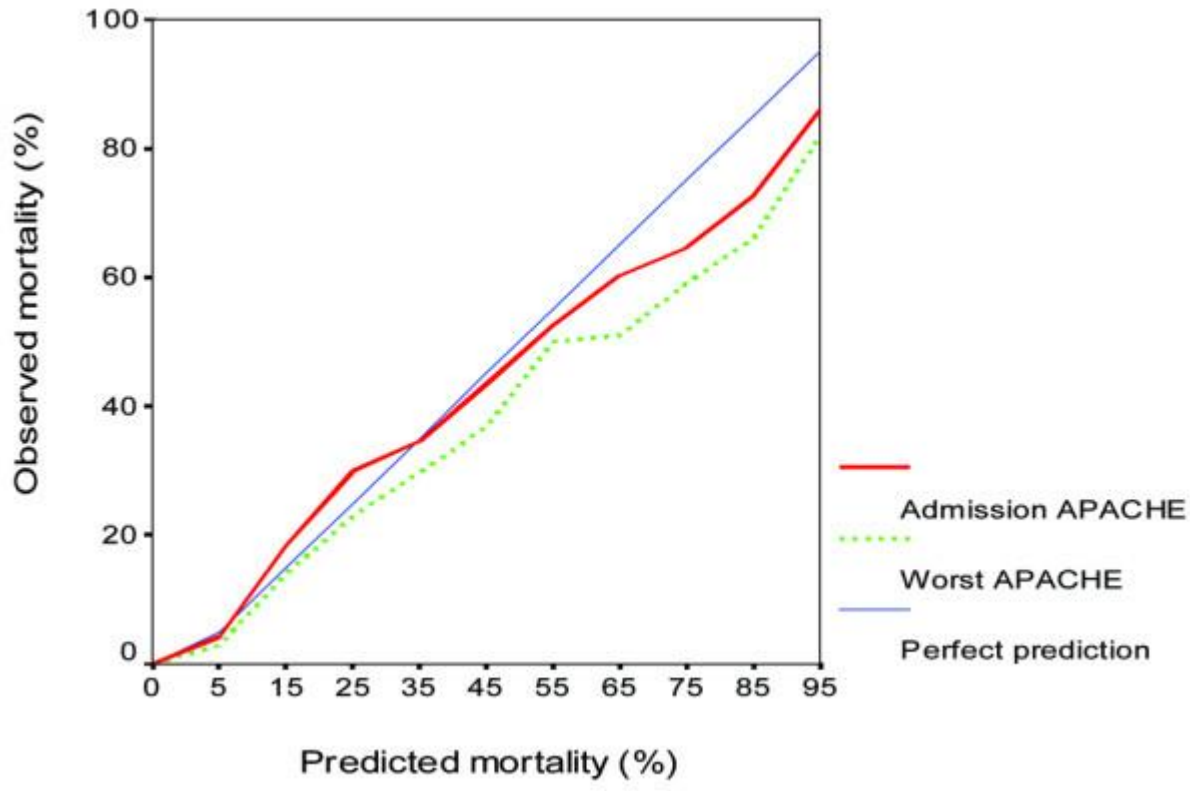


Figure 8 Calibration curves for the Admission APACHE II score and the Worst 24-hour APACHE II score to predict in hospital death (from: Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, Sirio CA, Murphy DJ, Lotring T, Damiano A, Harrell FE: THE APACHE-III PROGNOSTIC SYSTEM - RISK PREDICTION OF HOSPITAL MORTALITY FOR CRITICALLY ILL HOSPITALIZED ADULTS. *Chest* 1991, 100:1619-1636.).

- Overall, the predicted probabilities are close to the observed probabilities (however, this was shown by the authors of this study through HL statistics as well).
- The best calibration is held with very low (i.e. less than 5%) and moderate (i.e. between 35% and 55%) predicted probabilities. This means at the these ranges of predicted probability, the difference between the observed probability and the predicted probability is minimal. The worst calibration is held with low (i.e. between 5% to 35%) and high (i.e. more than 55%) predicted probabilities. This means

at these ranges of predicted probability, the difference between the observed probability and the predicted probability is maximal.

The same information can be obtained for the Worst APACHE II model. For example, as depicted, the higher the predicted probability, the more distant it is from the observed probability.

This figure also shows that overall, the predictions made by Admission APACHE II model is closer to the observed probability than predictions made by the Worst APACHE II model i.e. the admission model has a better calibration performance than the worst model.

2.8. Model's Generalisability (or validity)

Model validity refers to the performance of the model in the dataset from which the model has been derived (internal validly) or in a separate different dataset (external validation). The dataset from which the model is derived is called the derivation dataset. The dataset on which the model is validated is called internal or external validation dataset. If a model is accurate only in the dataset from which it has been derived, then this model is not generalisable i.e. is not reliable to be used in settings other than the derivation set. There are many reasons why an accurate model on the derivation dataset may not hold validity. This may relate to inherent deficiencies of modelling method or the design of the study [112]. However, even when a model has been developed with vigilant study design and methodology, it is still needed to be examined in other settings (or different datasets).

Internal validity (model reproducibility): this is when the validity of the model is examined in a different sample of patients who were not included in the derivation dataset just by random [113]. In this manner, the dataset is, for

instance, divided into two sub datasets (e.g. 2./3 versus 1/3 of the cases) and the derivation is performed in one sample (e.g. on 2./3 of the cases) to obtain the indices of performances. Subsequently, the resulting model is run in the remaining cases and the performance indices are obtained from this second dataset. A model with good internal validation should not demonstrate a huge drop in its measures of performance between these two datasets.

External validation (transportability): this is when the performance of the model is investigated in a different dataset of patients which differ to the derivation sets in many ways [113]. Based on the type of the difference, there are potentially 5 types of external validations: historical, geographic, methodological, spectrum and follow-up interval. Failure of a model in any of these external validations warns the users of such model in settings which have the same type of difference(s) with the derivation set.

Historical validation: this validation is when the validation dataset is recruited from a different calendar time to that from the derivation dataset [113]. The importance of such validation relates to the changes in treatment, management or health care polices bound to occur over time.

Geographic validation: this validation is when the validation dataset is recruited from a different location (s) which refers to the generalisability of the model to a setting in different geographic regions [113]. This type of validation would address regional differences of treatment, management or health care polices.

Methodological Validation: this validation is when the validation dataset is recruited by a different method [113]. This implies to the differences in patients selection or data collection. For example, a dataset of stroke patients

may exclude those cases who have had the symptoms less than 24 hours whereas in the other sample of stroke patients this exclusion criterion does not apply.

Spectrum validity: this validation is when the validation dataset contains cases who, on average, have a different degree of severity or advancement of the disease [113]. For example, a dataset may contain brain injury patients who attended hospital with GCS of less than 8 in contrast to the other dataset who contain brain injury patients with presenting GCS > 12.

Follow-up interval validity: this validation is when the existent of a condition or occurrence of an event pertains to a different period of time in the validation dataset to that in the derivation dataset [113]. An example of such validation is to assess the performance of a model which predicts the degree of disability of stroke patients at discharge from hospital in the other sample of stroke patients to predict the same degree of disability at sometime after discharge.

2.9. Model development

Beside a good performance, a model has to be well-developed. This means construction of a model is much more complicated than performing statistical tests such as a Chi Square test. This is because the modelling has to be performed through several stages such as careful selection of covariates or data preparation for logistic regression. Assessment of the way a presented model has been developed is beyond the scope of this article and is similar to critical appraisal of a research study. For this purpose, the series of publications in the British Medical Journal (BMJ) [114-116] or the quality assessment tool for prognostic models proposed by Perel *et al.* [31] can be referred to [31].

Musshkudiani *et al.* also provided a list of recommendations for constructing and validating models which can be used as quality assessment tool [32].

Briefly, if a model has been derived from a large dataset, is in widespread use, has received international acceptance or has been peer-reviewed prior to its presentation or publication, it may be nearly ensured that the model is well-developed. However, cautions should still be taken as to various types of external validity. For example, a well-developed model may be too old to reflect ongoing changes in patients' management. Similarly, a model which has been constructed in the developed world may not be good for a country which has limited medical facilities or resources.

3. Paper 2: Predicting Outcome After Severe Traumatic Brain Injury Using the Serum S100B Biomarker: Results Using a Single (24h) Time-point [98].

Authors

- Timothy Rainey
- Mehdi Lesko
- Raphael Sacho
- Fiona Lecky
- Charmaine Childs

3.1. Abstract

3.1.1. Background and Objectives

In recent years, biochemical markers have been employed to predict the outcome of patients with traumatic brain injury (TBI). In mild TBI, S100B has shown the most promise as a marker of outcome. The objective of this study in patients with severe TBI was to: show the range of serum S100B levels during the acute phase after trauma: determine if S100B has potential to discriminate favorable from unfavorable outcome in patients with similar brain injury severity scores and to establish an S100B level 'cut-off' predictive for death.

3.1.2. Methods

All patients with severe TBI, admitted to this neurointensive care unit within 24 hours of injury were eligible for inclusion in this study. One serum blood sample was obtained from each patient at the 24 hours post injury time-point. S100B levels were measured using Enzyme-Linked Immunosorbent Assay. Injuries were coded using an internationally recognized injury scoring system. Three month follow up was undertaken with outcome assessed using the Glasgow Outcome Score (GOS).

3.1.3. Results

100 patients were recruited. Serum S100B levels ranged from 0.08µg/l to 12.62µg/l. S100B levels were significantly higher in patients with a GOS of 1 (death) 2 and 3 (unfavorable outcome) compared with those with GOS 4 and 5 (good recovery). In this study a cut-off point of 0.53 µg/l has sensitivity of

>80% and specificity of 60% to predict unfavorable outcome and 49% to predict death.

3.1.4. Conclusion

In 100 patients studied with similar brain injury severity scores, serum S100B measured at the 24 hour time point after injury is significantly associated with outcome but a cut-off 0.53ug/l does not have good prognostic performance.

3.2. Introduction

The severity of brain injury can be assessed in a variety of ways. For the paramedic and clinician the best empirical assessment of injury is the degree of impaired cerebral function. Here assessment of conscious level using the internationally recognised Glasgow Coma Scale (GCS) [117] aids triage, prognosis and family counselling. Subsequent assessments of head and other body injuries can be made using anatomical scoring systems such as the Abbreviated Injury Scale (AIS) [118] and for multiple injuries the Injury Severity Score (ISS) [118]. Physiological scoring systems can also be used using the revised trauma score (RTS) [119] and the four elements composing the TRISS methodology (Trauma Score and Injury Severity Score) [120]. These scoring and survival probability systems are particularly valuable in epidemiological studies for assessment of outcome with respect to severity of injury.

In recent years quantitative biochemical markers have been employed to diagnose a variety of diseases e.g. creatinine for renal failure [121], troponin for myocardial infarction [122] and lipase for pancreatitis [123]. Of the calmodulin/troponin C superfamily of calcium binding proteins [50], S100B has shown most promise as a biochemical marker of outcome after mild head injury [124]. Protein S100B fulfils many of the criteria of an ideal molecular serum biomarker for brain damage in this patient group and has proved more reliable in predicting outcome compared with other markers such as neuron specific enolase [43]. S100B has high specificity for nervous tissue although it is recognised that non-nervous tissue such as fat and muscle also release protein S100B [125]. Increased levels of S100B are associated with a poor

neuropsychological outcome [80, 126]. Consequently, S100B has been proposed as a diagnostic and prognostic tool in mild brain injury. It has also been used to aid decision making about the need for CT scanning in the Emergency Department (ED) [74, 124]. For patients with severe head injury, S100B has proved more reliable in predicting outcome compared with other markers such Neuron Specific Enolase [43, 78]. However, as blood samples have been taken at many different time-points after injury in the various published studies, it remains uncertain as to the time-point at which blood sample should be taken where S100B levels best reflect the severity of brain damage.

In a recent series of investigations from this Centre, the majority of patients with severe TBI admitted to the intensive care unit for medical management of their head injury had the same AIS score for the head (AIS 5) despite considerable differences in outcome when assessed using conventional outcome scores three months after brain damage [17, 18]. To improve our ability to discriminate survivors from non-survivors, the aim of this study was to: show the range of serum S100B levels after severe TBI; determine if S100B has potential to discriminate favourable from unfavourable outcome in TBI patients with the same AIS scores and to establish an S100B level cut off predictive for death.

3.3. Methods

Research ethics approval was obtained before the study commenced. Patients aged ≥16 years with severe head trauma admitted to the intensive care unit (ICU) of this large University teaching hospital within 24 hours of injury were

eligible for recruitment to the study. Late referrals to ICU (admission ≥24 hours after injury) were excluded.

Patients were admitted either as direct referrals from the Emergency Department (ED) or as tertiary referrals from ED of other hospitals within the Greater Manchester region. All the patients were sedated, intubated, and mechanically ventilated and all had an intra or extra-axial lesion on CT, with or without systemic trauma. The patients were treated in accordance with local neurointensive care guidelines to maintain cerebral perfusion pressure (CPP) at 60 mmHg or higher and intracranial pressure (ICP) below 20 mmHg. To manage raised ICP, patients were positioned 30° head up and received sedation, analgesia, neuromuscular blockade and osmotherapy with manitol (0.5g/kg) as required. Advanced therapies such as induction of barbiturate coma, surgical removal of haematoma, were considered as 'second-tier' therapy when ICP was refractory to first-tier treatments [127].

3.3.1. Assessment of Injury Severity

From the patient's case notes details of all injuries sustained at the time of the accident were noted. Using the AIS directory each injury was assigned a code. From the respective codes, a score was given representing the severity of trauma in each of seven separate body regions [118]. Briefly, an AIS code 1 represents minor injury and AIS 5 the most severe of survivable injuries. The AIS for the head region includes trauma to the brain and cranium. For an assessment of the severity of injuries in all the body regions, the internationally recognised Injury Severity Scale (ISS) was used. ISS is calculated by summing the squares of three highest AIS severity of scores allocated. However, if there

is AIS score 6 among scores allocated, then the ISS of 75 will be automatically allocated irerepeticve of other socres.

3.3.2. Blood sampling

Blood samples were obtained from each patient at 24h time-point after the patient had sustained his/her injury to the head. A single 5 ml blood sample was obtained via dwelling arterial cannula. The sample was transported via a pneumatic transport system from the ICU to the laboratory. Samples were centrifuged (2800 rpm) separated and stored at -70^{o}C until batch analysis of serum samples (approx 30/batch) was undertaken. Clinical details were recorded at the time of sampling.

3.3.3. Assay

Stored serum samples were analysed using a one-step immunoassay (enzyme-linked immunosorbent assay, ELISA; Sangtec 100™ Diasorin, Wokingham UK) incorporating S100B antibody coated microtitre plates and tracing antibody conjugated with horse radish peroxidase (HRP) and Tetra Methylbenzidine (TMB) substrate to give a colour reaction proportional to the concentration of S100B in the sample

3.3.4. Outcome

Three months after the primary injury was sustained a follow-up assessment using the GOS [26] was undertaken by one designated member of the research team (RS). Contact was usually made by telephone, speaking either with the patient him/herself (when appropriate) or with a relative, identified at the time of admission, who had agreed to be contacted in the future for the purpose of

outcome assessment. On some occasions patients who remained in the care of NHS rehabilitation services at the time of the three month GOS, were assessed via discussion with the patient's healthcare professional.

3.3.5. Data collection

Details of the anatomical distributions of the patients' injuries, obtained via the medical case notes, were transferred to an Excel™ (Microsoft Corporation) database for AIS and ISS scoring by a member of the research team and checked by the same Trauma Audit Research Network (TARN) officer throughout the study.

3.3.6. Statistics

From previous pilot data of nine patients from whom serial S100B samples were obtained over the course of the first 5 days after severe TBI, the 24-h post-injury time-point was observed to be most promising to detect the point at which injury-induced increase in S100B start to fall (Table 12). A continued elevation (or absence of a fall) in S100B levels might therefore be expected in those patients with the most severe of TBI and is in line with previous publications [128].

Pilot study patient	*Serum S100B (µg/l)*		*outcome*
	Before 24 hr time-point	*After 24hr time-point*	
1	1.12	0.73	NS
2	0.57	0.20	S
3	1.38	0.82	NS
4	1.66	1.57	NS
5	0.50	0.37	S
6	0.29	0.17	S
7	0.84	0.22	NS
8	2.07	3.25	NS
9	1.19	0.45	S

Table 12 Change in S100B serum levels before (range 5-23, median 12.5) the 24h after injury time-point and after (range 25-85; median 35h) 24 h. (NS: Non-Survivors, S: Survivors)

Comparisons of S100B values in the different categories under investigation were made using the Mann-Whitney U-test to compare median values. The Receiver Operating Characteristics (ROC) curve to plot sensitivity versus 1 – specificity was undertaken using the Statistical Package for the Social Sciences (SPSS Inc; Chicago, USA). Each S100B value, in tern, is treated as a cut-off and the sensitivity and specificity was determined for each S100B value., so allowing for the selection of the optimum cut-off value. Where patients have similar AIS (head) scores we predict that a high serum protein S100B level, 24 hours after severe TBI, will enable us to discriminate between patients who have a good versus a poor outcome three months after TBI. A minimum of 96 patients with an AIS for the head of 3 or more are needed to achieve 95% power at the 5% significance level to detect a difference in poor outcome (GOS 3 or less) of 70% *versus* 30% for patients with a high 24 hour protein S100B level.

3.4. Results

One hundred patients (81 male: 19 female) aged 16-86 years (median 31) years with TBI were recruited to the study over a period of 25 months (Table 13). All patients had sustained severe brain trauma. With the exception of two patients only, all injuries to the head scored 4 or 5 (median 5) on the AIS scale corresponding with severe brain damage (Table 1). In 53 patients (53%) brain injury was the only significant trauma. The remaining 47 patients (47%) had additional injuries to the body viscera, bone and soft tissues. In all patients, the ISS (representing the sum of all the injuries to head and body) ranged from 9-50 (median 25), a score corresponding with a classification of severe trauma. Thirty one (31%) patients required emergency neurosurgery (Table 1). Of the 100 patients recruited to the study, 70 patients (70%) survived their injuries and 30 (30%) died. All but one of the deaths occurred during the acute period after injury (first 5 days). Three months after TBI, GOS ranged from 1-5 (median 4) (Table 13)

.

	Number (percentage)
Gender	
Male/female	81/19
Age	
Median	31
Range	16-86
Highest AIS code	
3	2
4	37(37%)
5	61(61%)
ISS	
Median	25
Range	9-57
Multiple injuries	47%
Emergency neurosurgery	31(31%)
Outcome	
GOS 1 (death)	30(30%)
GOS 2	4(4%)
GOS 3	16(16%)
GOS 4	29(29%)
GOS 5	21(21%)
Unfavourable outcome	50%
S100B (µg/l)	
Median	0.74
Range	0.08-12.62

Table 13 Characteristics of the 100 study patients

The concentration of serum S100B obtained 24 hours after injury ranged from 0.08-12.62 (median 0.74) µg/l. In five patients, S100B levels were below the limits of detection for the assay. In the 98 patients with the most severe brain injury (AIS 4, 5) S100B levels were not significantly different between the two AIS categories (Figure 9). However, the p value of 0.06 inidicates that this finding tends to be significant, should the sample size be larger.Table 14 shows S100B levels in patients with and without four key conditions: those who required emergency neurosurgery; those patients with multiple trauma (as opposed to isolated brain injury only); those who had an unfavourable outcome (GOS less than 4) with brain injury severity scored as 4 and 5 respectively and those patients who died. S100B concentrations were not significantly different in patients who underwent emergency neurosurgery compared with those who did not. Similarly, the difference in S100B levels in patients with multiple trauma compared with the levels in patients with isolated brain injury were not significantly different. S100B concentrations (Figure 10) were significantly higher in: those with unfavourable versus favourable outcome (p=0.00) and those who died versus those who survived (p=0.003, Table 14).

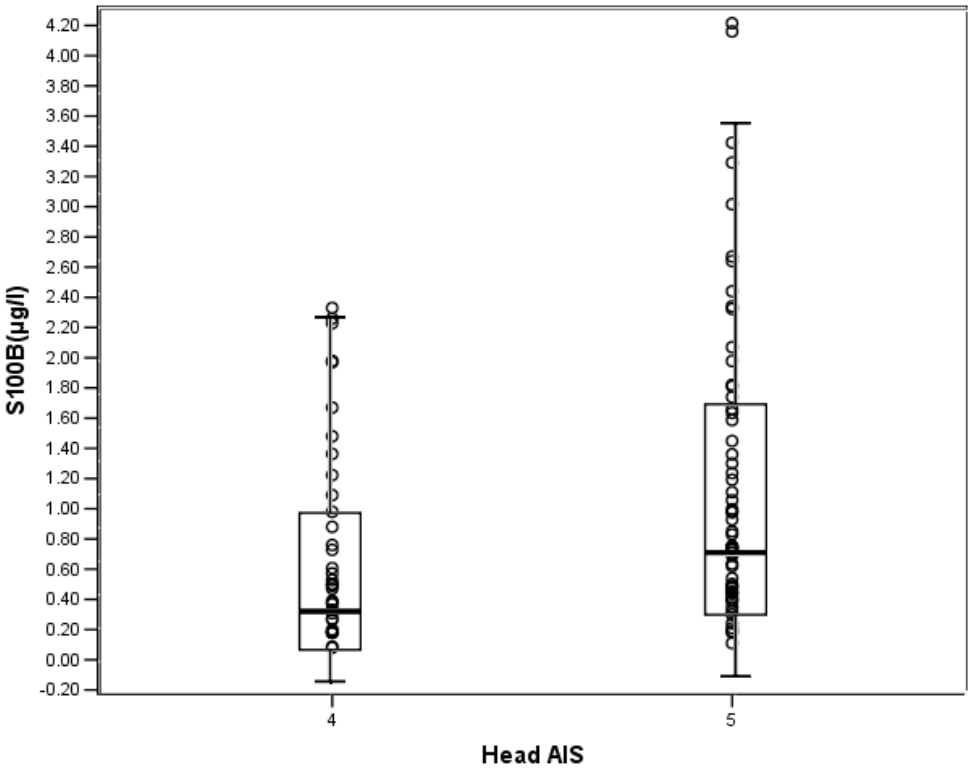


Figure 9 Box and Whisker plot of serum S100B levels (µg/l; open circles) in patients with AIS 4 (n=37) and AIS 5 (n=61). Differences between the two AIS categories are not significant (p=0.06, Mann-Whitney U-test). One extreme outlier (12.62 µg/l) in a patient with AIS 5 has been excluded in the figure but the value has been retained for all calculations including the median and interquartile range.

	Yes		*NO*		*P value*
	N	***Median S100B*** *(25^{th} to 75^{th} centiles)*	***N***	***Median S100B*** *(25^{th} to 75^{th} centiles)*	
Emergency neurosurgery	31	0.83 (0.47-1.66)	69	0.73 (0.39-1.46)	0.78
Multiple injuries	53	0.73 (0.44-1.74)	53	0.75 (0.34-1.42) *	0.53
Unfavourable Outcome (GOS < 4)	47	1.36 (0.60-2.28)	50	0.48 (0.29-0.94)	0
Unfavourable outcome (AIS 4)	14	0.99 (0.55-2.00)	23	0.37 (0.19-0.53)	0
Unfavourable outcome (AIS 5)	35	1.59 (0.64-2.44)	26	0.72 (0.40-0.98)	0
Death	30	1.44 (0.60-2.32)	70	0.59 (0.34-1.20	0.003
Death (AIS 4)	6	0.75 (0.51-1.80)	31	0.47 (0.20-1.09)	0.03
Death (AIS 5)	24	1.61 (0.65-2.41)	37	0.75 (0.45-1.22)	0.03

Table 14 Comparison of serum S100B concentrations (µg/l) in patients who had /did not have the following: emergency neurosurgical management, multiple injuries, unfavourable outcome (AIS 4, 5) or who died/survived severe TBI. *Patients in this category had isolated TBI. Main categories for statistical comparison are emboldened.

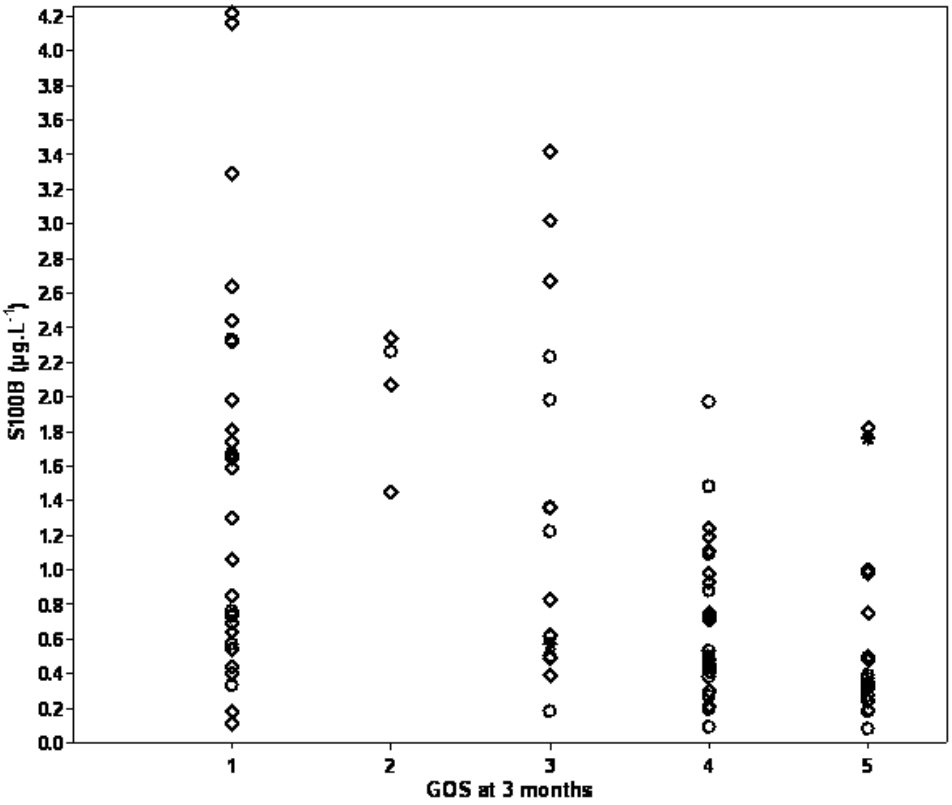


Figure 10 Scatter plot of serum S100B levels for each GOS score. Patients with AIS 3 (*), AIS 4 (○) and AIS 5 (◊). As shown in Figure 9 one outlier, 12.62 µg/l in a patient with GOS score of 1 is excluded.

ROC analysis of these data shows that serum S100B is a significant predictor of outcome with an area under the curve (AUC) of 0.77 (CI: 0.68 - 0.86) for prediction of unfavourable outcome (GOS 1, 2 and 3) at 3 months and AUC 0.69 (CI: 0.57-0.80) for prediction of death (Figure 11)., The best cut-off for S100B (optimizing sensitivity at the cost of lower specificity) is 0.53 µg/l. This cut-off provides a sensitivity of 82% with specificity of 60% to predict GOS of less than 4 at 3 months. To predict death at three months this cut-off has sensitivity of 83% and specificity of 49%

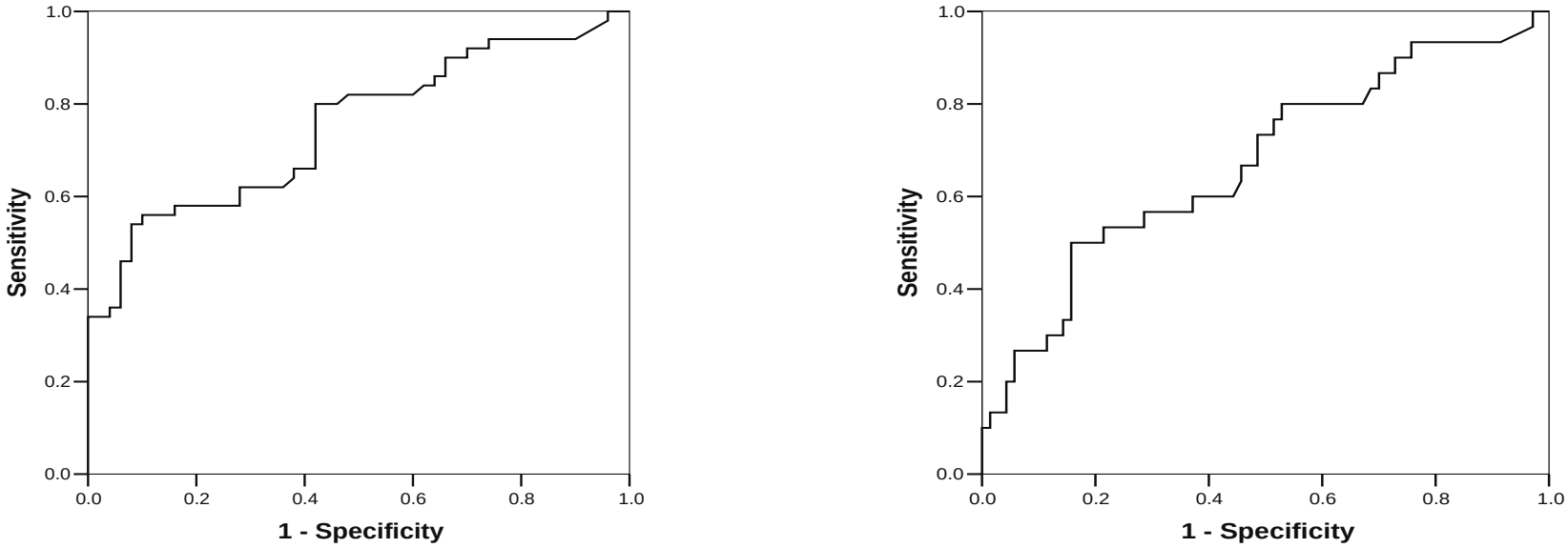


Figure 11 Receiver Operating Characteristic (ROC) curve showing plots of sensitivity versus 1-specificty to predict: A, unfavorable outcome, assessed at 3 months, at various cut-off levels of serum S100B and B, death at various cut-off levels of serum S100B.

3.5. Discussion

In this cohort of 100 patients with severe traumatic brain injury admitted to the intensive care unit of our institution, serum S100B levels measured at the 24 hour time point ranged from 0.08µg/l to 12.62µg/l. Serum levels were significantly higher in those patients who, three months after injury, had an unfavourable outcome compared with those who had made a good recovery. The patients who died also had significantly higher S100B concentrations than the survivors. Although there is no significant difference in S100B values between AIS categories 4 and 5, within each of these categories S100B levels did significantly predict good from a poor outcome. ROC curve analysis showed that S100B can be used to predict outcome at 3 months; be it death or unfavourable outcome (GOS of 3 or less).

We chose the cut-off point of 0.53µg/l to optimise sensitivity without excessive lowering of specificity. This cut-off has sensitivity greater than 80%, meaning that at least 80% of patients who die or have a poor outcome will have serum S100B concentrations of >0.53 µg/l at 24 hours post-injury, however the specificity for unfavourable outcome and death respectively are 60% and 49%. This means that 40% of patients with good outcome and 51% of patients who survive will also have "positive" (>0.53µg/l) S100B levels at this time.

The exact half-life of S100B protein in the blood is still unclear but it is thought that it is close to 97 minutes [78]. Although our target time-point for blood sampling was 24 hours from TBI, we accepted two hours either side; 22 to 26 hours on pragmatic grounds. However, the short half life of S100B presents a limitation to our study because of the potential change in the levels of S100B which might occur between 22 and 24 hours and 24 and 26 hours.

Thus, a two hour time gap before or after the target (24h) sampling time may lead to variability in S100B concentrations simply due to sample time. However, despite the short half-life, S100B serum levels were significantly elevated above baseline for up to 3 to 4 days after injury [22, 23]. The second limitation of our study relates to the cohort of patients recruited. All the patients recruited to the study were admitted to this neurointensive care unit within 24 hours of injury. We cannot be confident, however, that the results we have obtained from our centre are generalisable to all neuro-receiving hospitals i.e. those with and without specialist neurosurgery teams. It has been shown that the provision of early neurosurgical care to TBI patients improves outcome (at discharge or 30 days) by comparison to those patients who are not admitted to a neurosurgical centre [129]. A future multicentre study may, however, provide the results needed to determine if the results we report are generalisable to a wider cohort of severe TBI patients.

Although several previous studies have shown that S100B can predict outcome following TBI, unfortunately the cut-offs and their diagnostic characteristics differ significantly. For example, Vos *et al.* suggested an admission cut-off of 1.13 µg/l for prediction of unfavourable outcome at 6 months [43], which is greater than the cut off identified by our study (0.53 µg/l). However, the diagnostic characteristics are similar (sensitivity 0.88 and specificity 0.43). On the other hand, whilst the S100B cut-off from Mussack's study [78] (0.59µg/l) obtained from samples 12 hours after injury is close to our cut-off, specificity is 100% to predict GOS less than 4 at one year[20] which is significantly higher than our specificity of 60%. Similarly, Nylen and colleagues investigating S100B on admission report a cut-off of 0.55µg/l. This

cut-off provided 100% specificity to predict unfavourable outcome at one year [25] and is close to the cut-off of 0.53µg.L¯which we produced for the 24 hour time-point to predict outcome at three months. The prognostic performance of S100B to predict outcome might be improved in future studies with more consistency in sampling times and follow up intervals. Contrary to our expectations we found neither emergency neurosurgery nor multiple trauma influenced the performance of this serum biomarker but the type of critical care provided (specialist neurosurgical versus general trauma) might, in the future be shown to be of importance in influencing patient outcome [129].

The time point of 24 hour post-injury is probably more reliable than admission serum S100B levels for outcome prediction because the patient has already been resuscitated in the emergency department. This obviates the effects of tissue perfusion on S100B levels [64]. Similarly, and depending on the half-life, the time elapsed from injury to admission to hospital might also be expected to influence the serum concentration [78].

Although extracranial sources of S100B such as fat, muscle and bone marrow [85] may contribute to S100B levels, minor peripheral injuries do not appear to lead to a significant raise S100B levels [66]. This fits with our observations that S100B levels were not significantly different in patients with and without systemic trauma. It is worth noting here that in this group of patients the contribution to the overall ISS due to peripheral trauma was small.

This study shows that although S100B levels tend to be higher in TBI patients with a GOS of less than 4 compared with those patients with a GOS of 4 or 5, an S100B cut-off of 0.53µg/l is not a reliable prognostic indicator. The ROC curves for prediction of death and unfavourable outcome have an AUC of

0.67 and 0.75 respectively which is far from ideal to underpin clinical decision making. Whilst the cut-off of 0.53µg/l has a sensitivity of more than 80% to predict death, the specificity is relatively low (at 60% for prediction of unfavourable outcome and 49% for death prediction). This means that serum S100B measured 24 hours after injury can reliably detect poor outcome/death in more than 80% of cases but 40% of the patients who make a full recovery may also have levels above the 0.53µg/l cut-off.

In conclusion, although serum S100B levels 24 hours after injury are significantly correlated with outcome after severe traumatic brain injury, S100B may not have a good prognostic performance to guide therapy and prognosis. Future research should focus on comparing the prognostic power of S100B to other available well-developed prognostic models in traumatic brain injury.

Acknowledgements

We thank Mrs. Laura White for assistance with AIS coding of injuries and Mr. Omar Bouamra for help with statistics. We would like to thank the nurses and doctors of the intensive care unit for their help and co-operation with the collection of blood samples.

Conflict of interest

None declarable

4. Paper 3: Comparing Model Performance for Outcome Prediction Using Total GCS and Its Components in Traumatic Brain Injury

Authors

- Mehdi Moazzez Lesko
- Tom Jenks
- Sarah J O’Brien
- Charmaine Childs
- Omar Bouamra
- Maralyn Woodford
- Fiona Lecky

4.1. Abstract

4.1.1. Introduction

The Glasgow Coma Scale (GCS) score is used both in clinical practices for patient assessment, communication amongst clinicians and in outcome prediction models such as TRISS. However in clinical practice and during the derivation of prognostic models, it is important to understand which GCS subscore- eye, verbal, motor- contributes most to prognosis.

4.1.2. Objective

To determine which GCS subscore is best correlated with outcome taking time of assessment into account.

4.1.3. Methods

Records of patients with brain injury presenting after 1989 were extracted from the Trauma Audit and Research Network (TARN) database. Using logistic regression, a baseline model was derived with age and Injury Severity Score (ISS) as covariates and discharge outcome (survival) as the dependent variable. Total GCS, its subscores and their combinations at various time points were separately added to the baseline model in order to compare their effect on model performance.

4.1.4. Results

21,657 cases with brain injury were extracted. The total GCS score at scene and its subscore parts had significantly lower predictive power compared with those recorded on arrival at Emergency Department (ED) (scene total GCS:

AUC: 0.89 (95% CI: 0.89-0.90) and Nagelkerke R^2 of 0.54, admission total GCS: AUC of 0.91(95% CI: 0.91-0.91) and Nagelkerke R^2 of 0.58). Eye and verbal subscores had significantly lower performances compared with total GCS, motor subscore and various combinations of GCS subscores. Motor subscore and total GCS appeared to have similar predictive performance (admission total and motor GCS both had AUC of 0.91(95% CI: 0.91-0.91) and Nagelkerke R^2 of 0.58)

4.1.5. Conclusion

GCS on arrival is a significantly better predictor of outcome than that recorded at scene. The reason for this is uncertain. Either GCS is recorded more accurately at hospital or the effect of intoxication has dissipated by the time of arrival at hospital. Motor subscore of GCS is a powerful predictor of outcome and contains most of the predictive power of the total score.

4.2. Introduction

The Glasgow Coma Scale (GCS) was first introduced in 1974 to measure the depth of unconsciousness in patients with acute brain injury [117]. This scale comprises three subscores independently measuring motor response, verbal performance and eye opening. Since its introduction, it has been accepted internationally in clinical practice as the means to estimate the severity of the medical condition or injury affecting the brain. GCS is also used as a predictive subscore in prognostic models. These models are developed to assess prognosis for a given TBI patient (such as the International Mission for Prognosis and Analysis of Clinical Trials (IMPACT) [24] and Corticosteroid Randomisation After Significant Head injury (CRASH) models [23]) or to perform benchmarking of trauma care systems by trauma registries (TRISS methodology) [120].

There are some disadvantages with the measurement of GCS. It is not straightforward to learn [130] and at times it might be impossible to measure quickly, which is an issue in emergency situations. Furthermore, the inter-rater reliability of total GCS is only moderate as Gill *et al.* observed there is a probability of 68% that a pair of two GCS scores measured by two observers at the same time will differ in one subscore or another [131]. It is also well-known that assessing the verbal subscore of GCS is not reliable in sedated or intubated patients. This is also the case for the motor/eye subscore when neuromuscular blockage is used. When using GCS in prognostic models, the problem faced by trauma registries is that it is unclear which time point of GCS assessment holds better prognostic value. GCS is usually measured both at scene (where the injury is incurred) and on admission to the Emergency

Departments (ED) and the trauma registries often hold the record at both time points.

Healey *et al.* reported that the motor subscore contains the most predictive strength of GCS through a careful statistical analysis of a North American dataset of general trauma patients [22]. This analysis was performed without taking other important prognosticators such as age and extracranial injury into account. Ross *et al.* reported that the motor subscore may have similar diagnostic characteristics to total GCS to identify severe structural brain damage [132]. Similarly, Baxt *et al.* suggested a Trauma Triage Rule which employs the pre-hospital motor subscore of GCS along with other factors such as systolic blood pressure or pulse rate to triage major trauma patients [133]. However, the latter two studies did not investigate the prognostic performance of GCS subscores and they were performed on a North American trauma population sample. This is relevant as US trauma patients and care systems differ significantly from Europe, both these aspects influencing patient outcome and also affecting the GCS/outcome inter-relationship.

With regards to reports on the similar association of motor subscore and total GCS with injury severity, one might argue that measurement of motor subscore should suffice. The objectives of this study were twofold: to analyse the prognostic power of various GCS subscores in traumatic brain injury patients under the British trauma care system taking other important prognosticators into account and to investigate which time point of GCS measurement (at scene versus on admission to ED) has more prognostic strength.

4.3. Methods

A subset of TBI patients presenting to Trauma Audit and Research Network (TARN) was studied. TARN is currently the largest trauma registry in Europe and receives information on trauma patients from participating hospitals across England and Wales and increasingly from other countries in Europe. The TARN inclusion criteria are that the injured patient reaches the hospital alive and meets either: (I) more than three days stay in hospital, and/or (II) being cared in the intensive care, and/or (III) inter-hospital transfer and/or (IV) death at any time in hospital. The information is extracted from patients' medical notes or other available electronic sources by the data collector(s) at the participating hospital. Subsequently, TARN staff members code each injury sustained using the Abbreviated Injury Scale (AIS) [90]. The final data are saved in a main server located on the main campus of the University of Manchester. Access to the database is provided by Structured Query Language (SQL) server 2000. The inclusion criteria for this study were all TARN patients sustaining brain injuries of AIS severity score of 3 or above. Patients with head injuries with AIS score 1 and 2 were excluded since these scores refer to cases with mild head injuries such as simple or unspecified skull fractures. Running the appropriate query in SQL, cases fulfilling the above criteria were retrieved from TARN dataset.

For multivariate analysis, logistic regression was employed and to address the linear relationship of continuous variables with log odds of the outcome of interest as a requirement for logistic regression [134], fractional polynomials transformation was used [135]. In this method, the continuous variables are transformed into 1 or more other variables which is referred to as

'functional form' of the original variables. The transformation is a power transformation and the candidates of power are -3, -2, -1, 0, 1, 2, 3 where 0 is $\log_e$ transformation and 1 reflects no transformation (linear). The selection of the best transformation(s) is based on when the power candidate(s) yields a model with a significant improvement (referred to as 'gain') in the goodness of fit of the model which holds no transformations. Table 15 presents the results of fractional polynomials analysis for each covariate. It shows that, for instance, age should be transformed into two variables being $\log_e\left(\frac{age+0.1}{100}\right)+0.93$ and $\left(\frac{age+0.1}{100}\right)^3-0.06$ These two transformed covariates should be supplied to the model instead of the 'true' age.

Variable		*Transformation*
Age		$\log_e\left(\frac{age+0.1}{100}\right)+0.93$
		$\left(\frac{age+0.1}{100}\right)^3-0.06$
ISS s		$\left(\frac{10}{ISS}\right)^2-0.17$
		$\sqrt{\frac{10}{ISS}}-0.64$
Scene Scores	**Total GCS**	$\sqrt{\frac{10}{GCS}}-1.05$
	Motor GCS	$\sqrt{MotorGCS}-1.95$
	Eye GCS	$\frac{1}{EyeGCS}-0.41$
	Verbal GCS	$\frac{1}{\sqrt{VerbalGCS}}-0.6$
	Sum of Motor and Eye GCS	$\log_e(MotorGCS+EyeGCS)-1.82$
	Sum of Motor and Verbal GCS	$\log_e(MotorGCS+VerbalGCS)-1.88$
	Sum of Verbal and Eye GCS	$\left(\frac{1}{VerbalGCS+EyeGCS}\right)-0.03$
		$(VerbalGCS+EyeGCS)^3-137.45$
Admission Scores	**Total GCS**	$\frac{10}{GCS}-0.99$
		$\left(\frac{GCS}{10}\right)^3-1.01$
	Motor GCS	$\left(\frac{1}{MotorGCS}\right)^2-0.53$
		$(MotorGCS)^2-18.66$
	Eye GCS	$\log_e(EyeGCS)-097$
		$(EyeGCS)^3-6.99$
	Verbal GCS	$\left(\frac{1}{VerbalGCS}\right)-0.57$
	Sum of Motor and Eye GCS	$\left(\frac{1}{\sqrt{MotorGCS+EyeGCS}}\right)-0.37$
		$(MotorGCS+EyeGCS)^3-337.74$
	Sum of Motor and Verbal GCS	$\log_e(MotorGCS+VerbalGCS)-2$
	Sum of Verbal and Eye GCS	$\left(\frac{1}{VerbalGCS+EyeGCS}\right)^2-0.3$
		$(VerablGCS+EyeGCS)^3-188.12$

Table 15 Fractional Polynominal transformations of variables included in the modelling

Age and Injury Severity Score (ISS) are covariates with which the GCS prognostic strength is adjusted with. Using logistic regression, a baseline model was derived with age and ISS as regressors and discharge outcome (survival) as the dependent variable. Total GCS, its subscores and their combinations were added separately to the baseline model to assess their effect on model performance. The various combinations of GCS subscores were the sum of motor and eye subscores, motor and verbal subscores and eye and verbal subscores. Overall 15 models were constructed (one baseline model, seven models with admission GCS, subscores or combination of subscores and seven models with scene total GCS, subscores or combination of subscores). Area Under the Roc Cure (AUC), classification accuracy, Nagelkerke R^2 [136] and p value of HL statistics were taken as measures of the performance of each model. Regarding missing information, all missing total GCS scores were imputed with the sum of their subscores in case of lack of availability in the dataset. Similarly, if total GCS was recorded as 15 and one or more subscores were missing, then the missing subscores(s) were imputed with the full score. Apart from this, no other imputation strategies were implemented for missing information. The analysis was performed on the complete dataset with no split sampling.

Whilst the fractional polynomials transformations were performed in Stata, logistic regression was run in Statistical Package for the Social Sciences (SPSS).

4.4. Result

Using the inclusion criteria, a dataset of 21,657 TBI cases were extracted containing all brain injury records in TARN from January 1988 to April 2008. Table 16 presents the clinico-demographic characteristics of the sample population. Median age is 34.4 (interquartile range: 20-57) and 73.3% of the population are male. The median ISS was 24. The median total GCS was 9 at scene with motor, verbal and eye subscores holding respectively medians of 4, 2, and 2. However, the median total GCS on admission is higher than scene score; being 11. Furthermore, the admission motor, verbal and eye subscores hold medians of respectively 5, 4, 3. Sixty nine percent of patients survived their injuries at discharge. The amount of missing information varied across various subscores of GCS and also across the two time points of measurement; at scene or on admission. Overall, there are more missing GCS scores at scene than on admission.

	Median (interquartile range)/frequency (N= 216527)	***Number of missing(percentage)***
Age	34.4(20-57)	0
Male	73.3%	0
ISS	24(16-29)	0
Scene GCS		
○ ***Total***	9(4-14)	9530(44%)
○ ***Motor***	4(1-6)	9876(45.6%)
○ ***Verbal***	2(1-5)	9867(45.5%)
○ ***Eye***	2(1-4)	9852(45.4%)
Admission GCS		
○ ***Total***	11(6-15)	2677(12.3%)
○ ***Motor***	5(3-6)	3726(17.2%)
○ ***Verbal***	4(1-5)	3741(17.2%)
○ ***Eye***	3(1-4)	3721(17.1%)
Survival	69.4	1(0%)

Table 16 Patients characteristics and number of missing values for each parameter.

Adding total GCS, its subscores and various combinations of the subscores resulted in a significant decrease of the deviance of the baseline model at all times. Also in each model the effect of covariates included in the model on outcome was significant i.e. p value < 0.05. Table 17 presents the performance of each constructed model per measures of AUC, Nagelkerke R^2, classification accuracy and HL statistic. The baseline model is the model which contains only age and ISS and other models are named according to the added GCS, subscores or combinations of subscores to the baseline model. Overall the baseline model does not have a good performance as per HL statistics and adding GCS or subscores does not improve this. However, the performance of the baseline model is increased following addition of GCS, subscores or combinations of subscores according to AUC, classification accuracy and Nagelkerke R^2 and the AUC increases are statistically significant. Furthermore, comparing the admission and scene scores, each model containing admission scores outperforms its counterpart model with scene scores in all three measures and the AUC differences are statistically significant. For example, AUC of the admission 'total GCS' model is significantly higher than scene 'total GCS' model (confidence intervals: 0.91-0.92 versus 0.89-0.90 respectively).

		AUC (Confidence Interval)	*Classification accuracy*	*Nagelkerke R^2*	*HL statistics (P value)*
Baseline model		0.84 (0.83-0.84)	78%	0.40	50.11 (0.00)
Admission GCS	**Total GCS (N=12127)**	0.91 (0.91-0.92)	85%	0.58	62.14 (0.00)
	Motor (N=11781)	0.91 (0.91-0.91)	85%	0.58	48.78 (0.00)
	Eye (N=11805)	0.89 (0.89-0.90)	83%	0.53	46.23 (0.00)
	Verbal (N=11890)	0.90 (0.90-0.91)	84%	0.55	53.12 (0.00)
	Motor + Eye (N=11815)	0.91 (0.91-0.91)	85%	0.58	53.55 (0.00)
	Motor + Verbal (N=17993)	0.91 (0.91-0.92)	86%	0.59	46.40 (0.00)
	Verbal + Eye (N=18001)	0.91 (0.90-0.91)	85%	0.57	28.62 (0.00)
Scene GCS	**Total GCS (N=18980)**	0.89 (0.89-0.90)	82%	0.54	40.8 (0.00)
	Motor (N=17931)	0.89 (0.88-0.90)	82%	0.54	38.24 (0.00)
	Eye (N=17936)	0.88 (0.87-0.88)	80%	0.50	48.68 (0.00)
	Verbal (N=17916)	0.88 (0.88-0.90)	81%	0.52	59.63 (0.00)
	Motor + Eye (N=11815)	0.89 (0.89-0.90)	82%	0.55	34.56 (0.00)
	Motor + Verbal (N=11813)	0.89 (0.89-0.90)	82%	0.55	34.51 (0.00)
	Verbal + Eye (N=11821)	0.89 (0.88-0.89)	81%	0.53	45.02 (0.00)

Table 17 Comparison of predicitve models for survival using various GCS subscores and their combinations (N: number of cases included in the modelling).

4.5. Discussion

In this study, we have compared separately the prognostic power of total GCS, its subscores and various combinations of its subscores through multivariate analysis of a large British dataset of TBI cases. A baseline model was constructed with age and ISS and subsequently the improvement in the model performance was investigated following addition of GCS, subscores or

subscores combinations as per four measures of AUC, classification accuracy, Nagelkerke R^2 and HL statistics. Overall adding GCS, subscores or combinations of subscores results in significant improvement of the baseline model performance apart from the goodness of fit assessed by the HL statistics. Furthermore, it appears that eye and verbal subscores on their own, i.e. with no combination with other subscores, hold the least prognostic strength compared with total GCS, motor GCS or combination of motor with eye or verbal GCS. Similarly, the predictive strength of total GCS, motor GCS and combination of eye with verbal subscores or combination of motor with eye subscores appears the same.

4.5.1. Limitations

We acknowledge a number of limitations. Firstly, for this analysis, an existing dataset of TBI patients retrieved from TARN was used. Therefore, the effect of local protocols in GCS collection is unclear as to when the condition of the patient does not permit measurement of one subscore such as intubation or paralysis. In such case, immeasurable subscore might be assigned the lowest score or regarded as missing. Furthermore, the record of GCS on admission to ED does not clarify whether this has been performed prior to or following resuscitation. This depends on each hospital's specified polices. In fact, the acute course of TBI is unstable in that several events occurring over a short time span such as secondary insult or expanding intracranial mass lesion may influence the level of consciousness [137]. Post-resuscitation appears to be a better time point to collect GCS than admission GCS for predictive purposes as in such case the patient is expected to be at least haemodynamically stable. Secondly, the GCS predictability was adjusted only with age and ISS.

However, pupillary reactivity is also one of the important predictors in TBI but was not accounted for in this analysis. This is because recording of pupillary reactivity has only recently commenced for TBI submissions in TARN. Had pupillary reactivity been included in the modelling procedure, the dataset would have then been significantly smaller, yielding less reliable results. Furthermore, we compared the prognostic strength of GCS and its subscores in several models which are not unified in terms of the number of missing information (or the number of TBI cases included). For example, whilst total admission GCS was assessed in a model derived from 18980 cases, the scene score was assessed in much lower number of TBI cases i.e. 12127. This implies that there were a number of cases which were included in the admission model but not in the scene model. This is also the case with comparing various combinations of subscores. It is unclear how this exclusion would change the results.

4.5.2. Comparison with the literature

With regards to the same predictive strength of motor and total GCS, the results of our study are consistent with findings by Healey *et al.* [22]. However, in our study, GCS is adjusted with other TBI predictors i.e. age and ISS. Moreover, unlike Healey *et al.* who performed their analysis on general trauma patients with no exclusion of intoxication or shock potentially affecting level of consciousness, we performed our analysis on a TBI population who all sustained documented brain injury determined by AIS codes. Perhaps this also explains why the population sample in Healey's study consisted of 90% GCS score of 15 whereas our dataset contains more varied GCS scores (e.g. median admission total GCS: 11 with interquartile range of 6 and 15).

Marmarou *et al.* observed that both enrollment (the time point when the patient was enrolled to the study) and pre-enrollment motor GCS have significant prognostic effect on outcome in TBI (Marmarou, Lu et al. 2007). This finding may be the same as our finding in that earlier GCS (scene GCS) is less predictive of survival than later GCS (i.e. admission GCS). The explanation for this finding may relate to the completion of resuscitation as later time points e.g. on admission or on enrolment to the study are more likely to be when the TBI patient has already been resuscitated.

4.5.3. Implications of the study

We cannot explain why admission GCS scores have more predictive strength than scene scores. It might be due to the effects of alcohol or other drugs, which are diminished by the time the patient arrives at hospital. So GCS on admission might be more representative of the true level of consciousness being caused by the injury per se [138]. Also it might highlight inaccurate recording of GCS at scene, which might be due to environmental difficulties or skill level of attending personnel. However, whatever the reason for the difference in scene and admission GCS predictability, it might have pragmatic implications on clinical decisions for therapeutic interventions based on GCS. So we suggest that GCS on admission should be taken into account rather than GCS at scene.

We observed that a model which contains total GCS holds similar prognostic performance to a model which contains only the motor subscore. This suggests that adding eye and verbal subscores to the motor subscore does not hold any prognostic value and thus their measurement may not be necessary. On the one hand omitting eye and verbal scores might result in an

improvement in the overall inter-rater reliability (between personnel with the same level of experience/skill), which is poor for total GCS but better for the motor score [131]. Also measuring only motor subscore would be easier to teach, learn and implement than total GCS since the error rate of eye and verbal GCS is high among unskilled observers compared with skilled ones [130]. On the other hand, although GCS scale is designed to measure the depth of the unconsciousness, which also relates to the outcome, in practise it is not only used for the purpose of prognosis. The results of our study demonstrates a possible similar prognostic strength for total and motor GCS but this cannot be generalised reliably to other applications of GCS such as day-to-day monitoring of patients' alertness (as occurs in the intensive care units) or a clinical decision on intervention. Further, total GCS with respect to its descriptive capability holds more information content compared with motor GCS which does not provide any information on eye and verbal response. Likewise, each GCS score can be the sum of a varied combination of subscores and each combination of subscores might have significantly different mortality rates [22]. Therefore, we assume that for each total GCS with a certain motor subscores, changes in the eye and verbal subscores would then result in different mortality rates. If verbal and eye subscores are not measured, then the influence of eye and verbal response on outcome within the group of patients with the same motor subscore is ignored. It is considered that the added value of the eye and verbal subscores is mainly in trauma patients with more moderate degrees of injuries. In our study, age and ISS were taken into account as confounders, but the analysis was not performed separately on a subgroup of patients with a moderate degree of injury. Overall we believe measurement of motor subscore alone, despite being more simple and perhaps more reliable,

does not outweigh its disadvantages. Had the motor subscore significantly outperformed the total GCS in outcome prediction, omission of eye and verbal subscores might have then been suggested in clinical practice.

The 'motor + verbal' model has similar predictive performance to 'total GCS' model per AUC and even its performance is higher per Nagelkerke R^2. This suggests that measuring the eye subscore might not be necessary although the advantages and disadvantages of this are similar to measurement of the motor score alone, as discussed above. Similarly, the performance of 'motor + eye' model is the same as 'total GCS' model for admission GCS scores. This finding is reassuring in that if measurement of the verbal subscore is not possible (for instance due to intubation), then measurement of only motor and eye subscores should suffice.

4.5.4. Future direction

It is important that the results of our study be validated in a different set of TBI cases from different setting (country) and for a different type and time point of outcome. The quality of trauma care is one factor which affects the outcome and as such has a confounding role. Our dataset was collected from the British hospitals with the specified health care policy. Regarding the outcome, discharge survival is not the only end target of TBI care and management since functionality as close to that prior to the injury is important as well. Furthermore, this normal functionality can be achieved in short-term or long-term.

4.6. Conclusion

In a population of TBI patients whose injuries were managed within England in Wales over the last 20 years, the total GCS and motor subscore may have

similar predictive strength. Furthermore, it appears the eye subscore on its own holds less predictive strength than total GCS or various combinations of subscores. With regard to admission and scene GCS scores, admission scores significantly outperform scene scores for outcome prediction.

5. Paper 4: Using Abbreviated Injury Scale (AIS) Codes to Classify CT Features in the Marshall System [99]

Authors

- Mehdi Moazzez Lesko
- Maralyn Woodford
- Laura White
- Sarah O'Brien
- Charmaine Childs
- Fiona Lecky

5.1. Abstract

5.1.1. Background

The purpose of Abbreviated Injury Scale (AIS) is to code various types of Traumatic Brain Injuries (TBI) based on their anatomical location and severity. The Marshall CT Classification is used to identify those subgroups of brain injured patients at higher risk of mortality or deterioration. The purpose of this study is to determine whether and how AIS coding can be translated to the Marshall Classification

5.1.2. Methods

Initially, a Marshall Class was allocated to each AIS codes through cross-tabulation. This was agreed upon through several discussion meetings with experts from both fields (clinicians and AIS coders). Furthermore, in order to make this translation possible, some necessary assumptions with regards to coding and classification of mass lesions and brain swelling were essential which were all approved and made explicit.

5.1.3. Results

The proposed method involves two stages: firstly to determine all possible Marshall Classes which a given patient can attract based on allocated AIS codes; via cross-tabulation and secondly to assign one Marshall Class to each patient through an algorithm.

5.1.4. Conclusion

This method can be easily programmed in computer softwares and it would enable future important TBI research programs using trauma registry data.

5.2. Background

Trauma registries hold records of patients with Traumatic Brain Injury (TBI) across a designated region mainly for assessment of trauma care centres/systems compared with a national standard e.g. analysing data to predict survival probability (observed – expected survival rates). The demographic and clinical details of trauma patients are submitted to these registries primarily to provide data that will improve clinical outcome for trauma patients but they also form a valuable dataset for epidemiological studies. The Abbreviated Injury Scale (AIS) [90, 91] was proposed by the Association for the Advancement of Automotive Medicine and was designed specifically for coding various types of injury and for scoring them based on the severity. Using a standard dictionary each entry in a trauma registry dataset is assigned a 6-digit AIS code number with a post decimal place representing score of severity. The description for each AIS code is contained in the AIS dictionary. Each post-decimal score of the injury severity ranges from 1 (minimal) to 6 (maximal).

The AIS dictionary is structured by anatomical region of the body such as face, neck, abdomen and pelvic contents etc. One section in this dictionary is allocated to head trauma, which is subdivided into the whole area (massive destruction of cranium and brain, penetrating injury and scalp injury), intracranial vessels, cranial nerves (cranial nerves I to XII), internal organs and skeletal. This part of the AIS dictionary contains information about the anatomical location of the lesion (brain stem, cerebrum and cerebellum), the type of the lesion (e.g. haemorrhage, contusion and brain swelling), various

subtypes of haemorrhage such as Subarachnoid Haemorrhage (SAH), Subdural Haemorrhage (SDH) and the size of the lesion.

Using the AIS dictionary to describe injuries is probably limited to those running trauma registries. It is rarely employed for clinical and therapeutic purposes or in data collection for clinical trials because a trained coder is needed to code the injuries and also because the description and classifications of injuries is more detailed than required for clinical purposes. Alternatively, the Marshall Classification of structural brain damage is based on CT findings of TBI patients [139]. This system was first introduced in 1991 and the main aim was to identify those TBI patients at higher risk of deterioration or mortality although it has been validated as having predictive value for TBI outcome as well [24, 36, 94, 140, 141]. The hierarchy of Marshall Classes represents the increasing risk of developing raised ICP determined by factors relating to this pathology such as mass lesions or brain swelling. This classification challenged the previous perception that patients with compressed or absent cisterns who had a good clinical evaluation could be treated as if their brain CT is normal [139].

Understanding the relationship between AIS coding of brain injury and the Marshall Classification is important for several reasons. First, the AIS and Marshall Classification systems describe slightly different things. The Marshall Classification provides the opportunity to identify a subset of TBI patients at risk of developing intracranial hypertension. It ignores brain stem and cerebellar injuries, which are described in detail in the AIS dictionary. Secondly, the Marshall System is focused on closed head injury and was not designed for penetrating head injuries, for which there are several AIS codes.

Since TBI in trauma registries tends to be coded using the AIS dictionary and in clinical settings using the Marshall Classification, it is impossible to generate a complete picture of TBI incidence, risk factors and outcome without being able to bring these two types of data together.

Therefore we propose a method for allocating a Marshall Class to the AIS codes that are recorded for a given TBI patient. We have assumed that each injury description in the AIS dictionary can be used as an alternative to the CT reports.

5.3. Methods

5.3.1. AIS coding

Coding of brain injuries in the AIS dictionary is based on anatomical location (the brainstem, the cerebellum, the cerebrum and pituitary), the type of injury (Table 18), subtypes of haemorrhage (Table 18) and the degree/extent of the injury. Some types of injuries relate to certain locations of the brain; these being *massive destruction (crush)* which can affect the whole head or can occur in the brain stem, *compression and transection* exclusively occurring in the brain stem and *pneumocephalus* exclusively occurring in the cerebrum. However there are some other types of injuries incurred in more than one anatomical location namely *ischemia, brain swelling or various subtypes of haemorrhage* which may occur in the cerebellum or the cerebrum. Similarly, *penetrating injuries, diffuse axonal injury, contusion, haemorrhage, infarction or laceration* can be potentially sustained in all parts of the brain which include the brain stem, the cerebellum or the cerebrum. The determinants of the degree/extent of each injury include *multiplicity, being uni/bilateral* and

midline shift (for contusions) and *the volume/diameter* (for contusions and various subtypes of haemorrhage). The severity of brain swelling in the cerebrum is determined by *the status of ventricles or the brain stem cisterns* - either or both may be compressed or absent. Where information is not adequately documented, the codes referred to as 'Not Further Specified; NFS' are assigned. Alongside the injuries which fall under the heading

Penetrating injury Diffuse axonal injury Contusion Haemorrhage Brain swelling Infarction Ischemia Pneumocephalus Laceration Compression Massive destruction (crush) Transection ***Subtypes of haemorrhage*** Epidural Intraparenchymal Subdural Subarachnoid Subpial

Table 18 Descriptions of types of injury and subtypes of haemorrhage in the AIS dictionary; update 98

of 'internal organ' in the head section of the dictionary, there are codes which relate to the skeleton and some of them include descriptions of *basal skull fracture* or *not simple vault fractures*, which should, in fact, be considered as traumatic brain injury. Nevertheless, the AIS code 116002, allocated to superficial penetrating injury to the head, should be interpreted as not accompanied by brain injury. It should be noted that TBI cases may be allocated more than one AIS code.

5.4. The Marshall Classification

Table 19 displays the Marshall CT Classification. According to this system, the discriminative features are presence/absence of intracranial pathology, presence/absence of high or mixed density mass lesions, signs of raised intracranial pressure which is status of basal cisterns and midline shift and lastly evacuation of mass lesions. In this classification, a high or mixed density mass lesion implies contusion or hemorrhage. The extent of the lesion is determined by its volume, the cut-off being 25 cc. Moreover, depending on the size and surgical evacuation, a lesion can be one of Mashall Classes II, V or VI. The higher risk of raised ICP is determined by present, absent or compressed basal cisterns and the degree of midline shift – the cut-off point being 5mm. These pathologies fall into classes III or IV based on the severity. Unlike AIS coding, the Marshall System is mutually exclusive in that a TBI case is only allocated to one Marshall Class.

Marshall Class		*Description*
Class I	Diffuse injury I (no visible pathology)	No visible pathology seen on CT scan
Class II	Diffuse injury II	Cisterns are present with midline shift 0-5 mm and/or: lesion densities present no high- or mixed-density lesion > 25 cc may include bone fragments and foreign bodies
Class III	Diffuse injury III (swelling)	Cisterns compressed or absent with midline shift 0-5 mm, no high- or mixed-density lesion > 25 cc
Class IV	Diffuse injury IV (shift)	Midline shift > 5 mm, no high- or mixed- density lesion > 25 cc
Class V	Evacuated mass lesion	Any lesion surgically evacuated
Class VI	Non-evacuated mass lesion	High- or mixed-density lesion > 25 cc, not surgically evacuated

Table 19 The Marshall CT Classification

5.4.1. Cross-tabulation of AIS codes with Marshall Classes

As explained above, the Marshall System and the AIS coding hold two different approaches to brain injury classification and thus reconciliation between the two systems required various assumptions which had to be agreed upon from both the clinical and the coding perspective. A number of meetings were held with participation of two physicians specialising in emergency medicine and neurosurgery and two experts in AIS coding in the UK (from the Trauma Audit and Research Network (TARN)) [89] to discuss the most appropriate Marshall Class allocated to each AIS code performed through cross-tabulation. Table 20 presents the resulting cross-tabulation based on expert consensus where the description for each code can be found in the AIS

document. The mapping was decided to be performed initially on AIS dictionary; update 98 which is still in widespread use despite the new update introduced in 2005. Subsequently, adaptation of this cross-tabulation to suit the AIS dictionary; update 2005 was discussed (Table 21). Likewise, the decision was made to consider only AIS codes which are either apparently brain injuries (such as SAH) or, with a high likelihood, can be regarded to be accompanied with brain injury (such as basal skull fractures). However, codes relating to unconsciousness were excluded from this cross-tabulation since these codes are commonly not used by trauma registries and instead, Glasgow Coma Scale (GCS) with the same value for outcome prediction is used to address the level of consciousness [120, 142].

The rational for mapping various AIS codes of brain injury to the appropriate Marshall Class particularly regarding the assumptions made for brain swelling and mass lesions is provided in the Appnedix.

AIS codes	*Marshall Class*
113000.6	V/VI
116004.5	Penetrating injury
140299.5	Brain stem injury
140202.5	III
140204.5	Brain stem injury
140206.5	Brain stem injury
140208.5	Brain stem injury
140210.5	Brain stem injury
140212.6	Brain stem injury
140214.6	Brain stem injury
140216.6	Penetrating injury
140218.6	Brain stem injury
140499.3	Cerebellar injury
140402.3	Cerebellar injury
140403.3	Cerebellar injury
140404.4	Cerebellar injury
140405.5	Cerebellar injury
140406.5	Cerebellar injury
140410.4	Cerebellar injury
140414.4	Cerebellar injury
140418.4	Cerebellar injury
140422.5	Cerebellar injury
140426.4	Cerebellar injury
140430.4	Cerebellar injury
140434.5	Cerebellar injury
140438.4	Cerebellar injury
140442.4	Cerebellar injury
140446.5	Cerebellar injury
140450.3	Cerebellar injury
140458.3	Cerebellar injury
140462.3	Cerebellar injury
140466.3	Cerebellar injury
140470.3	Cerebellar injury
140474.4	Cerebellar injury
140478.5	Penetrating injury
140699.3	II
140602.3	II
140604.3	II

Table 20 Proposed Marshall Class - AIS code combinations based on the 1998 update of the AIS dictionary

AIS codes	*Marshall Class*
140606.3	II
140608.4	V/VI
140610.5	V/VI
140612.3	II
140614.3	II
140616.4	V/VI
140618.5	V/VI
140611.3	II
140620.3	II
140622.3	II
140624.4	V/VI
140626.5	V/VI
140628.5	II
140629.4	II
140630.4	II
140632.4	II
140634.5	II
140636.5	V/VI
140638.4	II
140640.4	II
140642.4	II
140644.4	II
140646.5	II
140648.5	V/VI
140650.4	II
140652.4	II
140654.5	II
140656.5	V/VI
140660.3	III
140662.3	III
140664.4	III
140666.5	IV
140676.3	II
140678.4	II
140680.3	II
140682.3	II
140684.3	II
140686.3	II

Table 20 Proposed Marshall Class - AIS code combinations based on the 1998 update of the AIS dictionary *(continued)*

AIS codes	*Marshall Class*
140688.4	II
140690.5	Penetrating injury
140799.3	II
150200.3	I
150202.3	I
150204.3	I
150206.4	I
150404.3	I
150406.4	I
150408.4	I

Table 20 Proposed Marshall Class - AIS code combinations based on the 1998 update of the AIS dictionary *(continued)*

Code	*Marshall Class*
140605	II
140613	II
140621	II
140625	II
140627	II
140631	II
140639	II
140643	II
140645	II
140647	II
140649	II
140641	V/VI
140651	II
140655	V/VI
140687	II
140686	II
140691	Penetrating injury
140692	Penetrating injury
140689	II
140701	I
140702	I
140703	I
140675	II
140677	II
140681	II
140683	II
140694	II
140695	II
140697	II
140698	II
150000	I

Table 21 Allocating a Marshall Class to AIS code; update 2005

5.4.2. Selection of one Marshall Class

A TBI patient may receive more than one AIS code whereas each patient should receive only one Marshall Class in the Marshall System. In order to address this, the decision was made to place all AIS codes which fall under the same Marshall Class together as ‘Equivalent to one Marshall Class’. In this manner, Equivalent to Marshall Class I, II, III, IV or V/VI each respectively represents Marshall Classes I, II, III, IV and V/VI. Then an algorithm was

devised to choose one Equivalent to Marshall Class which would be the final single Marshall Class mapped. Using of such algorithms for patients who sustained multiple brain injuries was proposed by Maas *et al.* [94].

5.5. Results

5.5.1. The proposed method to allocate a Marshall Class to a TBI patient

This involves two stages: assignment of Equivalent to Marshall Classes and then selection of the final Marshall Class.

Stage 1: Assignment of Equivalent to Marshall Classes

Table 22 presents various AIS codes which all come under one similar Marshall Class (Equivalent to Marshall Class I, II, and III etc.). According to this table, the unclassified codes relating to brain stem, cerebellar and penetrating injuries were broken down further into penetrating, brain stem/cerebellar codes necessitating addition of two further classes of VII and VIII to represent penetrating and the brain stem/cerebellar injuries respectively. This has been agreed by the authors of previous guides for using the Marshall Classification [140] (personal communication). The other possible options are to further split the brain stem/cerebellar injuries into two distinct individual Marshall Classes or to merge all penetrating, brain stem and cerebellar codes into one class as ‘unclassified’. This depends on the research objective.

	AIS codes
Equivalent to Marshall Class I (no visible pathology)	150200,150202,150204,150206, 150404,150406, 150408
Equivalent to Marshall Class II	140602,140604,140606,140612,140614,140611,140620,140622,140628,140629,140630,140632,140634,140638,140640,140642, 140644,140646,140650,140652,140654,140684,140688, 140686, 140699, 140676, 140678, 140680, 140682, 140799
Equivalent to Marshall Class III (swelling)	140202, 140660, 140662, 140664
Equivalent to Marshall Class IV (shift)	140666
Equivalent to Marshall Class V/VI	140608,140610,140616,140618,140624,140626,140636,140648, 140656, 113000
Cerebellar/ brain stem injuries	140204,140206,140208,140210,140212,140214,140218,140299,140402,140403,140404,140405,140406,140410,140414,140418, 140422,140426,140430,140434,140438,140442,140446,140450,140458,140462,140466,140470,140474,140499,
Penetrating injury	140216, 140478, 140690, 116004

Table 22 Grouping of AIS codes into various 'Equivalent of Marshall Classes'.

Stage 2: Selection of the final Marshall Class

Figure 12 displays an algorithm proposed to select one Equivalent to Marshall Class which can be the mapped final Marshall Class for a given patient. This is based on the fact that the Marshall Classification is ordinal indicating that, in case of multiple injuries, the highest class is the single class allocated to the patient. This is reflected in the algorithm. Initially, all penetrating injuries are contained in Class VIII. This is the point at which the algorithm stops since the Marshall Classification is designed for blunt injuries. At the second step, injuries are screened for Equivalent to Marshall Class V which will result in a class VI designation in case of surgical evacuation or, otherwise, class V. The following steps sequentially take account of Equivalent to Marshall Classes IV, III and II. However, prior to searching for Equivalent to Marshall Class II leading to allocation of class II, Marshall Class VII is mapped in case of the presence of brain stem/cerebellar codes. The algorithm is flexible with the position of this step being implemented prior to screening for Equivalent to Marshall Class I as displayed in Figure 12 or otherwise being placed following exclusion of penetrating codes. In the latter situation, the algorithm begins its detection of the single mapped Marshall Class by exclusion of those who have sustained penetrating, brain stem or cerebellar injuries.

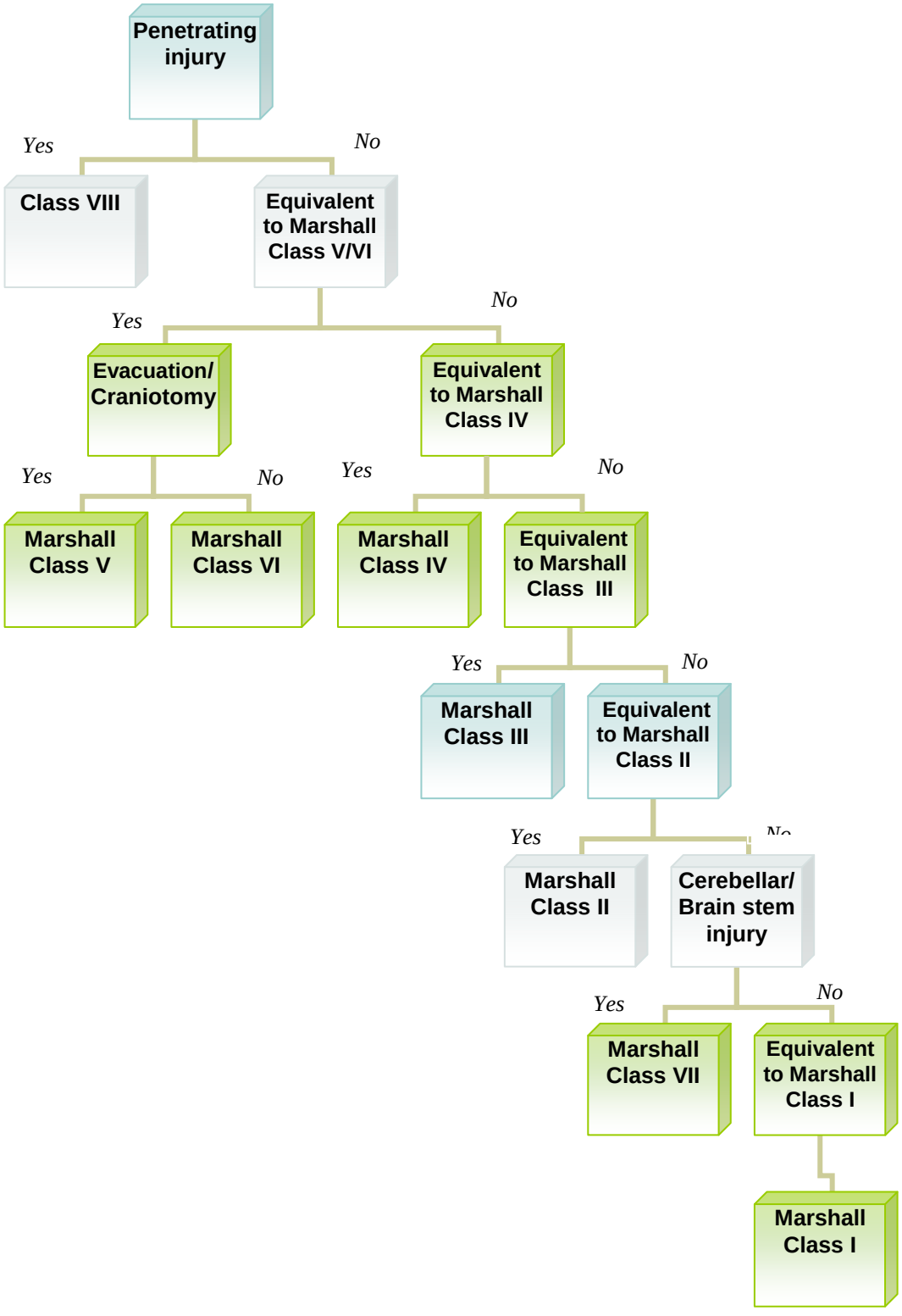


Figure 12 Algorithm to derive the Marshall Class from Equivalent to Marshall Classes.

Programming the procedure to designate a single Marshall Class to a given TBI case for which various AIS codes are recorded is straightforward in

computer softwares such as Statistical Package for the Social Sciences (SPSS) etc. In the first step, all Equivalent to Marshall Classes are computed as nominal variables for each TBI case. Each Equivalent to Marshall Class will then be 'yes' if at least one of the AIS codes allocated to this code (Table 22) is present and otherwise such Equivalent to Marshall Class is 'no'. Second, the computer has to search all computed Equivalent to Marshall Classes step by step in accordance with the algorithm. For example, if a given case has brain stem/cerebellar injuries and Equivalent to Marshall Class V/VI with surgical evacuation, the Marshall Class VI is allocated.

5.6. Discussion

In this study, we have attempted to propose a method to translate the head injury AIS codes into the Marshall CT Classification. This involves two steps; first to cross-tabulate various AIS codes with the Marshall Classes and secondly to select the single Marshall Class allocated to a case of TBI through an algorithm. In order to perform this transformation some assumptions had to be made..

5.6.1. Limitations/assumptions

Although both the Marshall Classification and the AIS dictionary group CT features according to their severity, one important difference between the two systems relates to their purposes. The main aim of the Marshall Classification is to identify those TBI patients who are at higher risk of deterioration or mortality, whereas the AIS scoring system is used to classify injuries based on their anatomy rather than physiological merits. These different approaches to CT classification mean that certain assumptions have to be made when trying

to reconcile the two systems. Ideally the two systems would be completely interchangeable and no assumptions would be required. Since this is not the case an important question is whether or not mapping AIS codes onto the Marshall Classification is worthwhile. We believe adoption of the conceptual approach we have proposed allays some concerns in that, instead of strictly meeting the definition of each Marshall Class, the objective and rational surrounding that class are also employed to spot the appropriate AIS codes. A disadvantage of the Marshall Classification is that it is not a reliable classification to be used in the retrospective research settings in which the access to the real CT obtained during acute phase of therapy is often not possible in case the Marshall Class is not recorded in the existing dataset.

The Marshall Classification should be ideally performed by the expert who views the CT. However, Marshall Class II, unlike other classes, contains a broad range of heterogeneous injury types or severities. Considering the different objective of AIS dictionary which is to anatomically classify injury severities, mapping AIS codes with Marshall Classes is in fact alike ignoring many valuable individual pieces of information by pooling them into one class such as class II. This leads to them all being treated similarly in prognostic analysis despite potentially having varied individual prognostic merit on their own. This defect is substantial when class II contains, for instance, infarction beside laceration which are different in nature and perhaps in prognostic strength.

The method proposed in this study is based primarily on the assumption that the descriptions in the AIS dictionary can be substitutes for CT reports, but this is not always the case. As well as including CT reports, the sources of

information to document injury descriptions also encompass MRI, surgery, x-ray, angiography, post-mortem examinations or clinical diagnosis. Nevertheless, it seems reasonable to assume majority, if not all the information for AIS coding, is obtained from CT scans since CT is the commonest modality for diagnosing structural brain damage for every patient suspected to have sustained severe trauma in the developed world and several other developing nations.

As AIS coding does not rely only on one CT report, the dynamic nature of brain injury as to progressing or regressing over time is reflected in AIS codes unlike the Marshal Classification. This is because the Marshal Classification is collected from CT images/reports at a certain point in time (oftentimes on admission) whilst AIS codes contain information after discharge or death as well. As such using our algorithm to obtain the Marshal Class with AIS codes intermediation would inherit the information on dynamic nature or CT findings evolution as well. This poses a problem since evolution of structural brain damages per se is a prognostic factor indicating higher chance of unfavourable outcome [143]. Servadei *et al.* have shown that the worst CT classification has more prognostic value than less severe CT classification (s) [143]. Consequently, if the Marshal Classification is obtained from AIS codes, there may be some overestimation of its negative prognostic role in TBI as compared to the Marshal Classification using CT images/reports. This may be particularly an issue with patients who sustain more severe brain injuries as they are more subject to various means of investigations such as MRI or operation.

As well as the above fundamental assumption regarding AIS descriptions as substitutes for CT reports, there are two other important assumptions related to the brain swelling and the mass lesion. Unfortunately neither the Marshall Classification nor the AIS dictionary describe precisely the severity of brain swelling. The degree of swelling in AIS dictionary is only determined by cistern/basal cisterns status whereas the degree of midline shift is also an important determining factor. Likewise, although midline shift or cisterns status is important in the Marshall Classification of brain swelling, other causes of midline shift, such as mass lesion, are disregarded. With respect to the size of mass lesions, future research is required to determine the precise size cut-offs for categorising such lesions, in spite of the already-known fact that larger lesions are associated with poorer outcome [29]. Comparing the cut-offs, those for subdural and epidural haemorrhage in the AIS dictionary are larger than those in the Marshall Classification by 25 cc although this difference may be negligible for contusion and intracerebral haemorrhage which is only 5 cc. The evidence base for lesion size in both classifications appears to be limited, despite claims that the cut-offs are backed by substantial experience and are not merely arbitrary [140, 144].

In Table 20, we assumed that codes indicating hypoxic or ischemic brain damage are related to normal CT scan. This may not always be the case as some patients may develop brain swelling as a secondary damage to hypoxia/hypotension. Whilst our assumption of clear CT for hypotension/hypoxia may not be acceptable in our cross-tabulation, we believe the algorithm would address this problem. For example, if a patient develops brain swelling following hypoxia, then two codes of hypoxia and brain

swelling will be allocated. As the injuries are taken through the algorithm, the brain swelling Marshal Class of III or IV (depending on the severity) is allocated prior to the final step of the algorithm as Marshal Class I (normal CT).

The position of various steps of the algorithm is based on the assumption that the Marshal Classification is ordinal in severity. Despite being the case from Class I to IV, this is not true for class IV versus class V as patients in class IV demonstrate lower likelihood of favourable outcome or survival than those in class V [140, 145]. Since the Marshall Classification is mutually exclusive, it is conspicuously necessary to prioritise which type of injuries is more relevant for allocation of the proper Marshal Class in case of multiple brain injuries. Whilst, according to the adverse outcome frequency, the brain swelling with compressed cisterns may have to be placed prior to mass lesion in the algorithm, we believe the current position of each step is more reflective of what occurs in real life of Marshal Classification through observing the actual CT. In fact, the current position of various steps of the algorithm are according to what has been suggested by Maas *et al.* [140]. In Maas's algorithm, the Marshal Classification was taken as ordinal but still class V represented lesser degrees of brain injury than class IV in their subsequent prognostic analysis.

5.6.2. Implication

The Marshall Classification has prognostic value to make predictions on the outcome of TBI patient [24, 36, 94, 140]. AIS coding is also important from prognostic viewpoint [146] but the severity scores (ranging from 3 to 6 in TBI) encompass a wide variety of different injuries that the relationship of the score

and CT findings can not be easily made particularly in clinical settings where the AIS dictionary is not a familiar tool. Hence, it is important for trauma registries to ` avoid exclusive reliance on AIS coding for the sake of better communication with the clinical audience. As the Marshall Classification holds comparable prognostic value to age, GCS, pupillary reactivity, SAH etc. [24, 36] and trauma registries commonly do not have record of this classification, the translation of AIS codes to the Marshall system opens up the possibility for multivariate prognostic analysis of large series of TBI subjects saved in trauma registries. In fact, the internationally known IMPACT prognostic models [24] in TBI employ the Marshall Classification for outcome prediction and using our proposed translation not only permits running the IMPACT models in trauma registries, derivation of new prognostic models including the Marshall Classification becomes feasible. Furthermore, as other TBI series accrued in clinical studies (observational or clinical trials) often do not have AIS coding, our proposed translation facilitates mergence of datasets from trauma registries and clinical studies to conduct more powerful studies or performance of comparative analysis across datasets when data recording is not uniform.

5.6.3. Future direction

The design of the algorithm is such that at the end of the allocation, there must be no cases left with no Marshall Class assigned. We ran our algorithm in a dataset of 802 TBI cases from the Trauma Audit Research and Network (TARN) and noticed that there were no cases left with no Marshall Class allocated (unpublished data). However, we acknowledge that our proposed allocation still requires three possible forms of validation in the future. First, it is yet to be determined how accurate our method is when the Marshall

Classification is performed using the AIS codes. In this manner, AIS codes are applied as substitutes for CT reports and in case all the assumptions are followed, 100% accuracy should be met. The second form of validation is when the allocation is performed with actual CT images at hand. In this manner, the allocations are compared across two groups. In one group, the classification is done through observing the CT and in the other group the Marshall Class is obtained following assignments of AIS coding and subsequently using our proposed cross-tabulation and algorithm. This form of validation is not expected to yield 100% accuracy and it examines how strong the assumptions are. The third form of validation is to compare the Marshall Classification at certain time point with that collected from AIS codes obtained from any available source including CT, MRI, operation notes etc. This form of validation would examine the influence of multiple sources of information or the temporal effect of events on the cross-tabulation and algorithm.

5.7. Conclusion

Using robust assumptions, we have proposed a method to allocate a single Marshall Class to a patient whose AIS codes are available, such as in trauma registries. This would enable future important TBI research programs.

Conflict of interest

The authors declare that they have no competing interests.

Authors' contribution

M.M.L. and F.L. jointly developed the idea and designed the stages of the proposed method. M.M.L. drafted the cross-tabulation and the manuscript. M.W. and L.W. provided input for the cross-tabulation from the AIS coding

perspective and approved the final revision of the cross-tabulation. S.O. and C.C. reviewed the proposed method and edited the final draft of the paper to ensure content accuracy. F.L. provided input from the clinical perspective, approved the final revision of the cross-tabulation and supervised the project. All authors reviewed and approved the final revision of the manuscript.

5.8. Appendix: description of AIS codes to the Marshall Classes cross-tabulation

AIS codes outside cerebrum or unrelated to raised ICP

The Marshall Classification enables categorising a subset of TBI patients at risk of developing intracranial hypertention. Therefore injuries sustained in the brain stem and cerebellum is ignored. However, there are many codes describing the injuries in these two anatomical locations in the AIS dictionary. Almost all these codes do not have a Marshall Class equivalent. Bearing this in mind, we differentiated other non cerebral injuries relating to the brain stem or cerebellum by grouping them as ‘brain stem injury’ or ‘cerebellar injury’ without allocating a Marshall Class. The exception is AIS code 140202, which identifies the brain stem compression and thus corresponds to the Marshall Class III which involves compressed or absent cisterns. Moreover, the Marshall System is not designed for penetrating injuries [139] for which there are several AIS codes. Therefore, we grouped all such AIS codes as “penetrating injury” with no Marshall Class allocated. However, a penetrating injury AIS code related to the brain stem or cerebellum should still be grouped as

penetrating injury rather than a cerebellar or brain stem injury. This is because penetrating and blunt brain injuries pathophysilogically differ.

Similarly, a number of AIS codes representing cerebral injuries are not directly related to raised intracranial pressure. These include massive destruction of both cranium and brain (crush), infarction, intraventricular haemorrhage, ischemia, pneumocephalus, laceration and pituitary injury. Such injuries are best mapped to the Marshall Class II since they do not indicate a normal CT (i.e. Marshall Class I) nor do they indicate brain swelling or mass lesions (i.e. Marshall Classes III and above). However, crush injury should be mapped to the most severe Marshall Class i.e. Class VI because of the very severe nature of this injury.

Not Further Specified (NF)S AIS codes

In allocating the appropriate Marshall Class to the cerebral AIS codes, we assumed that NFS injuries are minimally severe injuries of their type as is always the case in the dictionary. For example, the code 140999 which represents cerebral NFS was allocated to Marshall Class II, which represents the least severe brain injury in the Marshall Classification.

Brain swelling

Although, in the Marshall Classification, only class III is declared as ‘brain swelling’ by Marshall *et al.* , class IV also contains this pathology. This is because midline shift, which denotes class IV, can be caused by brain swelling

as well. Thus, there are two Marshall Classes of III and IV indicating brain swelling, which are distinguished by compressed/absent cisterns for class III and midline shift of more than 5mm for class IV. However, in AIS coding the degree of the brain swelling is determined by the status of ventricles/cisterns being normal, compressed or absent. Therefore, the highest degree of brain swelling in AIS dictionary, i.e. absent cisterns, actually falls in the Marshall Class that indicates the lowest degree of brain swelling (class III) with no equivalent AIS code for Marshall Class IV. This inability in the AIS dictionary to distinguish between Marshall Classes III and IV poses a problem. The decision is whether or not to pool all AIS codes of brain swelling into Marshall Class III and to leave Class IV blank or to allocate AIS codes of mild and moderate brain swelling to Marshall Class III and AIS codes for severe swelling to Class IV. We selected the second option assuming that Marshall Classes III and IV represent mild and severe brain swelling respectively, irrespective of the criteria of the severity.

Mass lesions

There are several separate AIS codes for two kinds of mass lesions; contusion and haemorrhage. There are also several severity groups (small, moderate and large, massive or extensive) into which these lesions can fall depending on the size as ascertained by AIS severity scores. Furthermore, the cut-offs for this classification based on size are different in the AIS dictionary and the Marshall Classification. Whilst the Marshall Classification uses the simple cut-off of 25cc regardless of type and location, those used in the AIS dictionary vary by

the type, anatomical location and, at times, by age of the patient. For instance, a single contusion in the cerebrum is small when < 30 cc, large when between 30 cc and 50 cc and is extensive when > 50 cc (The cut-offs for the size-wise grouping of intracerebral haemorrhage, epidural or subdural hematoma are receptively 30cc, 50cc and 50cc.)

Regarding the size of high density mass lesions, a problem exists on the cut-off or criteria to distinguish small from large lesions being different in the AIS dictionary and the Marshall Classification. Therefore the assumption was made that small haemorrhage and contusion (unilateral or bilateral), SAH and Subpial haemorrhage correspond to the Marshall Class II with all other large, massive or extensive mass lesions coming under class VI.

Skull fractures

Codes indicating several skeletal fractures (*basal skull fracture* or *not simple vault fractures)* were all placed in Marshall Class I, which is described as no intracranial pathology.

5.8.1. AIS 2005

Adapting our proposed mapping for the 2005 update is simple since we know that the update to the head section involves changes in a number of AIS scores and the addition of some new codes. None of the old AIS codes, which have undergone changes in their severity score, are affected in terms of their mapped Marshall Class. However, for new AIS codes, represents the most appropriate Marshall Class mapped. Some of the new AIS codes have arisen because some of the old AIS codes have been further sub-divided to specify the injuries in

more detail. Overall these criteria do not affect the mapping proposed in for each particular injury. For example, in the 2005 AIS dictionary, the severity of Diffuse Axonal Injury (DAI) is further qualified by whether or not it is confined to white matter/basal ganglia or involves the corpus callusom. No matter which is the case, the equivalent Marshall Class II, as allocated in , still holds. Nevertheless, there are 3 new codes (140701, 140702 and 140703) that describe the hypoxic or ischemic brain damage which occurs due to systemic hypoxia, hypotension or shock. Since these causes of brain damage are not directly related to head trauma, we can infer that the head CT of such patients should be clear which indicates Marshall Class I (no visible pathology).

Acknowledgment

We would like to acknowledge the help of Prof. Andrew Maas from the IMPACT collaboration in discussion of the methods used to arrive at our final algorithm. This work is supported by the funding from the Trauma Audit and Research Network (TARN) and Overseas Research Students (ORS) Award Scheme, University of Manchester.

6. Paper 5: Prognostic Value of Various Intracranial Pathologies in Traumatic Brain Injury

Authors

- Mehdi Moazzez Lesko
- Omar Bouamra
- Sarah O’Brien
- Fiona Lecky

6.1. Abstract

6.1.1. Introduction

Various diagnosed intracranial pathologies in Traumatic Brain Injury (TBI) can help to predict patients' outcome. These pathologies can be categorised using the Marshal Classification or the Abbreviated Injury Score (AIS) dictionary or can be described through traditional descriptive terms referring to the type of injury such as Subarachnoid Haemorrhage (SAH), Subdural Haemorrhage (SDH), Epidural Haemorrhage (EDH) etc.

6.1.2. Objective

To assess the prognostic value of AIS scores, the Marshall Classification and various intracranial pathologies in TBI.

6.1.3. Method

A dataset of 802 TBI patients in the Trauma Audit and Research Network (TARN) database was analysed using logistic regression. Initially reference models were constructed with age, Glasgow Coma Scale (GCS), pupillary reactivity, Injury Severity Score (ISS), cause of injury and presence/absence of extracranial injury as predictors and survival at discharge as outcome. Subsequently, AIS score, the Marshall Classification and various intracranial pathologies such as haemorrahge, SAH or brain swelling were added to assess the relative predictive strength of each variable and also to assess the improvement in the predictive performance of the models.

6.1.4. Results

Various AIS scores or Marshal Classes did not appear to significantly affect the outcome. Among traditional descriptive terms, only brain stem injury and brain swelling significantly influenced outcome (odds ratios for survival: 0.17 (95% CI: 0.08-0.40) and 0.48 (95% CI: 0.29-0.80) respectively). Neither haemorrhage nor its subtypes such as SAH, SDH, and EDH were significantly associated with outcome. Adding AIS scores, the Marshall Classification and various intracranial pathologies to the prognostic models resulted in almost equal increase in the predictive performance of baseline models.

6.1.5. Conclusion

In this relatively recent dataset, the significant effect of brain swelling and brain stem injury on outcome in comparison to injuries such as SAH suggests the need to improve therapeutic approaches to patients who have sustained these injuries.

6.2. Introduction

Traumatic Brain Injury (TBI) can cause various types of intracranial pathologies. It may lead to contusion, haemorrhage or diffuse axonal injury (including brain swelling). With regards to predicting outcome, the presence of structural damage indicates poorer outcome compared with a normal CT [147]. There are several traditional terms such as intracranial haemorrhage, Subarachnoid Haemorrhage (SAH), Epidural Haemorrhage (EDH), Subdural Haemorrhage (SDH), brain swelling to describe the structural brain damage. The association of these pathologies such as intracranial haemorrhage [93], SAH [24, 36, 92, 140, 148-152], SDH [92, 94, 153], EDH [24, 92, 94, 140, 148] and brain swelling [94, 143, 149] with outcome has been shown in the TBI literature.

Abbreviated Injury Scale (AIS) dictionary is a document which codes various intracranial injuries sustained due to TBI. The injuries are coded based on anatomical location of the lesion (brain stem, cerebrum and cerebellum), the type of the lesion (e.g. haemorrhage, contusion and brain swelling), various subtypes of haemorrhage such as SAH, SDH and the size/degree of the pathology. Each AIS code is equivalent to a particular injury description and is followed by a figure as one post-decimal place ranging from 1 (the minimal severity) to 6 (the maximal severity). The post-decimal point is referred to as AIS severity score. For example, an SDH less than 50cc in an adult receives AIS code of 140652 which has severity score of 4. This is presented as 140652.4

Furthermore, the Marshal Classification of CT findings in TBI is first introduced in 1991 and is used to have a more accurate predictive assessment

of TBI patients who sustained intracranial hypertension by considering those injuries which are causes of, or somehow related to, raised ICP [139]. Since this classification is based on the degree of brain swelling or the extent of mass lesion, it disregards physiological characteristics and anatomical distribution of injuries in that, for example, contusion, SDH or EDH are all considered as mass lesions. Some studies have shown this classification can also be applied to predict the outcome in TBI [24, 36, 92, 94, 140, 143].

The Trauma Audit and Research Network (TARN) [89], based in the UK, is a trauma registry which receives detailed information on severe trauma patients mainly from trauma receiving hospitals in England and Wales. TARN has embarked on a project to construct prognostic models applicable to a subset of trauma patients who have sustained brain injury using a dataset which includes cases submitted after September 2005. This provides an opportunity to assess the predictive performance of various intracranial pathologies due to TBI. The studies on this issue so far have been conducted on older datasets which do not take account of temporal advances and improvements in trauma care systems [24, 36, 92-94, 140, 143, 148-153]. The objective of this study is twofold: to determine the relative prognostic strength of various AIS severity scores of brain injury and the Marshall Classification and to determine which intracranial pathologies are more important for outcome prediction in TBI using a more recent dataset.

6.3. Methods

6.3.1. Patients included

TARN holds the anonymous record of each TBI patient with various AIS codes and scores. The criteria for submission to TARN are that the patient arrives at hospital alive and fulfils one of the following criteria: (I) more than three days stay in hospital, (II) being nursed in the intensive care unit, (III) inter-hospital transfer or (IV) death at any time point in hospital. If a given patient meets the criteria, trained coders then code the sustained injuries using the AIS dictionary. The criteria to retrieve TBI cases from TARN general trauma registry were AIS severity score of brain injury 3 or more (AIS codes under the 'Internal Organ' in the head section of the dictionary, update 98 [90]) OR AIS cods related to basal and compound/depressed/open skull fracture AND availability of pupillary reactivity at any time point in TARN dataset. The reason for the availability of pupillary reactivity as a criterion was that many well-conducted studies have shown the importance of this variable for outcome prediction in TBI [23, 24, 36, 92, 93, 137] and it has only been recorded in TARN recently i.e. from September 2005 onward. The outcome measure for the analysis was survival at discharge and where applicable, the time point of measurement of covariates was on admission.

6.3.2. Differentiation of various intracranial pathologies

To investigate the predictive importance of AIS scores, the decision was made to choose the highest AIS score in case the patient had more than one AIS code assigned (34.4% of patient had more than 2 AIS codes/severity score assigned). Furthermore, AIS codes were classified according to what is presented in table

23. This classification exclusively describes type, subtypes and location and disregards the extent/degree of injuries as determined by size, volume or other parameters in the AIS dictionary. Injury types include contusion, haemorrhage and brain swelling with EDH, SAH and intracranial haemorrhage as subtypes of haemorrhage only if they are sustained in the cerebrum (i.e. not in the brain stem or cerebellum). These are intracranial pathologies which have been found by the International Mission for Prognosis And Clinical Trial Design (IMPACT) [28, 92, 94] to significantly affect the outcome in TBI. Moreover, there are other injury types sustained in the cerebrum which are not contained in table 23 in order not to make the classification overly detailed and thus complicated. Some examples of these codes are infarction, ischemia or diffuse axonal injury. Furthermore, should the haemorrhage or contusion be sustained in the brain stem or cerebellum, in table 23 , this is classified under the brain stem and cerebellar injuries and not the cerebral injury.

Cerebral contusion
Cerebral brain swelling
Cerebral haemorrhage
Cerebral EDH
Cerebral SDH
Cerebral SAH
Brain stem injury
Cerebellar injury

Table 23 Traditional terms to describe intracranial pathologies in TBI

In the same way, AIS codes were differentiated according to the Marshall Classification [139]. An algorithm was applied which enabled a Marshall Class to be allocated to each patient depending on the brain injuries described by AIS dictionary. This algorithm adds two additional classes to the original Marshall Classification to represent penetrating and brain

stem/cerebellar injuries as classes VII and VIII respectively. However, since in TARN, the distinction between classes V and VI can not be reliably made, these two classes were merged together in this analysis (class V represents evacuated mass lesion versus class VI representing non-evacuated mass lesion).

6.3.3. Examination of prognostic value of AIS severity scores and various intracranial pathologies

Initially the significance of the association of the brain injury AIS severity scores and various intracranial pathologies with survival at discharge was investigated using Chi Square test. Then two models were constructed. One model included age, GCS, pupillary reactivity, Injury Severity Score (ISS) (model A) and the other model included the same covariates except that ISS was replaced with presence/absence of major extracranial injury (i.e. extracranial AIS severity score > 3) and the cause of injury (model B). The reason for this replacement was that ISS and extracranial injury could not be contained in the same model due to multicolinearity effect. Following replacement of ISS with extracranial injury, the cause of injury became significant. The admission values were selected for GCS and pupillary reactivity and all missing information on these covariates were replaced firstly with observations en-route or, secondly, at scene. These models were named as "the baseline models A and B".

Subsequently, brain injury AIS severity scores and various intracranial pathologies were added separately to each model A and B firstly to observe the significance of effect of each variable on outcome in the model and secondly to observe the changes in the model performance according to the significance of

the decrease in the deviance or increase in Area Under the Roc Curve (AUC) and Nagelkerke R^2 [136]. Due to insufficient cases with AIS score 6 (only 6 cases in the whole dataset), this score was merged with AIS score 5. Various intracranial pathologies as presented in table 23 were supplied to the models altogether (i.e. not individually) only if they were found significant in the univariate analysis. However, we observed that some pathologies such as various types of haemorrhage were not significant in our models unlike the literature. This was thought likely to be due to the given combination of intracranial pathologies in table 23. Therefore, various combinations were investigated by merging/omitting various pathologies. This was performed based on the literature and the results of our multivariate analysis. We differentiated numerically each combination (i.e. combinations 1A to 5A and 1B to 3B where A and B represent models A and B).

6.4. Results

The results are presented in 3 sections to cover: the univariate analysis, the significance of each variable in the multivariate models and lastly their added value to the performance of the models.

6.4.1. Patients characteristics and the univariate analysis

The clinicodemographic characteristics of the population studied are shown in Table 24. The dataset comprised 802 TBI cases. The median age was 39 and males constituted 75.2% of the population. The commonest causes of injury were Road Traffic Collisions (RTC) and falls. The majority of cases (51.1%) had severe TBI i.e. GCS < 9 recorded on admission to the emergency department. Most cases (68.6%) had normal pupillary reactivity followed by

neither reactive at 16.5%. The median ISS was 25 and the frequency of major extracranial injury was 14.5% as determined by extracranial AIS score > 3.

Covariate		*Median (interquartile range)*	*Frequency (percentage)*
Age		39 (22-58)	
Gender	Male	-	603 (75.2%)
	Female	-	199 (24.8%)
Cause of injury	RTC	-	314 (39.2%)
	Fall	-	313 (39.0%)
	Assaults	-	143 (17.8%)
	Others	-	32 (4%)
GCS(categorical)	Mild (13-15)	-	278 (35.8%)
	Moderate (9-12)	-	99 (12.7%)
	Severe (3-8)	-	399 (51.5%)
Pupillary reactivity	Normal	-	446 (68.6%)
	Abnormal-both reactive	-	69 (10.2%)
	Abnormal-only one reactive	-	28 (4.3%)
	Neither reactive	-	107 (16.5%)
ISS		25 (16-29)	
Extracranial injury (cut-off: AIS=4)		-	116 (14.5%)
Survival		-	599 (74.7%)

Table 24 Clinicodemographic characteristics of the population sample studied.

Table 25 demonstrates the frequency of each AIS score, Marshall Class and various intracranial pathologies along with the results of univariate

analysis. Most cases had the highest AIS score of 4 (39.4%) and only 6 cases were recorded with the highest AIS of 6. Likewise, haemorrhage (of any type) was present in 76.4% of cases with SDH the most frequent type of haemorrhage (22.1%) followed by SAH (18.6%). Similarly, 8.9% and 5.1% of cases had brain stem and cerebellar injury respectively (of any kind such as haemorrhage or contusion). Nevertheless, non-hemorrhagic brain injuries such as contusion and swelling were present in 39.9% and 34.2% of the patients respectively. When using the Marshal CT Classification, the most frequent pathology is class II (50.5%) followed by class V/VI (20.7%) and III (10.6%). All covariates including AIS scores and the Marshal Classification were significantly associated with survival apart from contusion and SDH. Therefore contusion and SDH were not investigated further in the multivariate analyses.

Covariate		*Frequency (percentage)*	*Odds ratio for survival (CI)*	*p value by Chi square test*
Highest AIS scores	3	189 (23.6%)		< 0.005
	4	316 (39.4%)	0.73 (0.42-1.27)	
	5	291 (36.2%)	0.15 (0.09-0.25)	
	6	6 (0.7%)	0.02 (0.00-0.22)	
Cerebral contusion		320 (39.9%)	1.25 (0.90-1.74)	0.18
Cerebral brain swelling		275 (34.2%)	0.31 (0.22-0.43)	< 0.005
Cerebral haemorrhage		533 (66.5%)	0.61 (0.42-0.87)	< 0.005
Cerebral EDH		95 (11.8%)	2.85 (1.49-5.45)	< 0.005
Cerebral SDH		178 (22.1%)	1.17 (0.79-1.73)	0.43
Cerebral SAH		149 (18.6%)	0.56 (0.38-0.82)	< 0.005
Brain stem injury		72 (8.9%)	0.08 (0.04-0.14)	< 0.005
Cerebellar injury		41 (5.1%)	0.30 (0.16-0.56)	< 0.005
Marshal Classification	I	65 (8.1%)		< 0.005
	II	405 (50.5%)	1.17 (0.54-2.50)	
	III	85 (10.6%)	0.39 (0.17-0.90)	
	IV	74 (9.2%)	0.19 (0.08-0.44)	
	V/VI	166 (20.7%)	0.160 (0.08-0.35)	
	Brain stem/cerebellar injury	4 (0.5%)	0.05 (0.005-0.57)	
	Penetrating injury	3 (0.4%)	0.32 (0.03-3.92)	

Table 25 Frequency of various AIS score, Marshall Classes and intracranial pathologies. The *p* value represents the significance of association with survival at discharge.

6.4.2. The significance of each variable in the model

Table 26 and Table 27 show the effect of AIS scores, the Marshall Classification and intracranial pathologies on outcome (survival at discharge) using multivariate analysis. The combination 1A and 1B were the ones presented in tabel 23 (except for contusion and SDH which were not significant in the univariate analysis). In model A (Table 26), AIS score 4 was not significantly associated with outcome whereas AIS score 5/6 were marginally significant (<0.1 but > 0.05). Regarding the Marshal Classification, no Marshal Class showed significant association with outcome although the Marshall Class II was marginally significant (p = 0.06). Moreover, among all intracranial pathologies, only brain stem injury was significantly associated with discharge survival whereas brain swelling and cerebellar injury were marginally significant (*p* = 0.08 and 0.09 respectively).

		Baseline model A	*Model A + AIS scores*	*Model A + Marshal Classification*	*Model A + combination 1A*
			Model A		
Age		0.96 (0.95-0.97)**	0.96 (0.95-0.97)**	0.95 (0.94-0.96)**	0.95 (0.94-097)**
GCS	Mild	-	-	-	-
	Moderate	0.41 (0.20-0.82)*	0.40 (0.20-0.81)*	0.37 (0.18-0.77)**	0.40 (0.19-0.82)*
	Severe	0.22 (0.12-0.39)**	0.22 (0.12-0.41)**	0.22 (0.12-0.41)**	0.23 (0.12-0.43)**
Pupillary reactivity	Normal	-	-	-	-
	Abnormal-both reactive	0.40 (0.21-0.74)**	0.40 (0.21-0.75)**	0.37 (0.19-0.71)**	0.40 (0.21-0.76)**
	Abnormal-only one reactive	0.28 (0.12-0.65)**	0.26 (0.11-0.60)**	0.26 (0.11-0.61)**	0.26 (0.11-0.61)**
	Neither reactive	0.04 (0.02-0.09)**	0.05 (0.03-0.09)**	0.05 (0.02-0.1)**	0.05 (0.02-0.10)**
ISS		0.18 (0.10-0.33)**	0.29 (0.15-0.56)**	0.26 (0.14-0.5)**	0.27 (0.14-0.51)**
AIS	3	-	-	-	-
	4	-	1.28 (0.63-2.60)	-	-
	5/6	-	0.52 (0.25-1.08)	-	-

Table 26 Prognosis associated with AIS scores, the Marshall Classification and intracranial pathologies in the multivariate model A (*: $p < 0.005$, **: $p < 0.05$)

		Baseline model A	*Model A + AIS scores*	*Model A + Marshal Classification*	*Model A + combination 1A*
Marshal Class	I	-	-	-	-
	II	-	-	2.70 (0.97-7.5)	-
	III	-	-	0.1 (0.31-3.16)	-
	IV	-	-	0.70 (0.22-2.21)	-
	V/VI	-	-	0.74 (0.26-2.12)	-
	Brain stem/cerebellar injury	-	-	0.39 (0.00-42.11)	-
	Penetrating injury	-	-	1.2 (0.02-82.91)	-
Combination 1A	Brain swelling	-	-	-	0.64 (0.38-1.05)
	EDH	-	-	-	1.80 (0.76-4.28)
	SAH	-	-	-	0.22 (0.10-0.50)
	Brain stem injury	-	-	-	0.45 (0.18-1.14)**
	Cerebellar injury	-	-	-	0.94 (0.52-1.69)

Table 26 Prognosis associated with AIS scores, the Marshall Classification and intracranial pathologies in the multivariate model A (*: $p < 0.005$, **: $p < 0.05$) *(continued)*

In model B (Table 27), no significant association between AIS score 4 and outcome was observed whereas this association was significant for score 5/6. Unlike model A, Marshal Classes IV and V/VI demonstrated significant effects on outcome. Moreover, among intracranial pathologies, only brain swelling and brain stem injury significantly influenced survival at discharge in model B.

		Baseline model B	*Model B + AIS scores*	*Model B + Marshal Classification*	*Model B + combination 1B*
		Model B			
Age		0.96 (0.95-0.97)**	0.96 (0.95-0.97)**	0.96 (0.94-0.97)**	0.96 (0.95-0.97)**
GCS	Mild	-	-	-	-
	Moderate	0.39 (0.19-0.80)*	0.37 (0.18-0.76)**	0.34 (0.16-0.72)**	0.39 (0.18-0.81)*
	Severe	0.18 (0.10-0.32)**	0.23 (0.12-0.42)**	0.22 (0.12-0.42)**	0.22 (0.12-0.41)**
Pupillary reactivity	Normal	-	-	-	-
	Abnormal-both reactive	0.40 (0.21-0.74)**	0.40 (0.21-0.77)**	0.38 (0.2-0.75)**	0.40 (0.21-0.76)**
	Abnormal-only one reactive	0.21 (0.09-0.49)**	0.23 (0.10-0.55)**	0.23 (0.09-0.56)**	0.21 (0.09-0.50)**
	Neither reactive	0.03 (0.02-0.07)**	0.04 (0.02-0.07)**	0.04 (0.02-0.08)**	0.04 (0.02-0.08)**
Extracranial injury		0.22 (0.12-0.40)**	0.21 (0.11-0.39)**	0.18 (0.10-0.35)**	0.24 (0.13-0.44)**
Cause of injury	RTC	-	-	-	-
	Fall	0.60 (0.34-1.05)*	0.64 (0.36-1.17)	0.71 (0.39-1.3)	0.60 (0.33-1.10)
	Assaults	1.71 (0.77-3.82)	1.91 (0.83-4.42)	2.09 (0.90-4.85)	1.80 (0.78-4.15)
	Others	1.01 (0.29-3.55)	1.11 (0.31-3.98)	0.97 (0.25-3.77)	0.98 (0.27-3.52)

Table 27 Prognosis associated with AIS scores, the Marshall Classification and various intracranial pathologies in the multivariate model B (*: $p < 0.05$, **: $p < 0.005$)

		Baseline model B	*Model B + AIS scores*	*Model B + Marshal Classification*	*Model B + combination 1B*
AIS	3	-		-	-
	4	-	0.85 (0.41-1.75)	-	-
	5/6	-	0.23 (0.11-0.46)**	-	-
Marshal Class	I	-	-	-	-
	II	-	-	1.85 (0.64-5.36)	-
	III	-	-	0.7 (0.23-0.49)	-
	IV	-	-	0.26 (0.08-0.86)*	-
	V/VI	-	-	0.33 (0.11-0.99)*	-
	Brain stem/cerebellar injury	-	-	0.24 (0.00-39.06)	-
	Penetrating injury	-	-	0.11 (0.00-4.46)	-
Combination 1B	Brain swelling	-	-	-	0.48 (0.29-0.80)**
	EDH	-	-	-	1.67 (0.70-4.01)
	SAH	-	-	-	1.00 (0.56-1.81)
	Brain stem injury	-	-	-	0.17 (0.08-0.40)**
	Cerebellar injury	-	-	-	0.50 (0.20-1.35)

Table 27 Prognosis associated with AIS scores, the Marshall Classification and various intracranial pathologies in the multivariate model B (*: $p < 0.05$, **: $p < 0.005$) *(continued)*

Since in the analysis of types of hemorrhage, none showed significant association with outcome in neither model A nor B despite this being observed in other studies [24, 36, 92-94, 140, 143, 148-153], other combinations of intracranial pathologies were investigated. These combinations were based on (1) hemorrhage commonly thought to be predictive of adverse outcome in TBI from a clinical viewpoint, underpinned by literature [23, 93], (2) brain stem being a significant predictor in both models A and B, (3) brain swelling being a significant predictor in model B and marginally significant in model A and (4) cerebellar injury being marginally significant in model A. Table 28 presents various combinations of intracranial pathologies, created according to the above observations. All combinations, which were formed based on the results of model A (i.e. combinations 2A, 3A, 4A, 5A), contained brain stem injury as this variable was significant in model A. The different combinations for model A related to the presence/absence of marginally significant covariates (brain swelling and cerebellar injury) and the clinically important variable: hemorrhage. Further, combinations 2B and 3B were originated based on the results of model B. Both combinations contain the significant covariates of model B (brain stem injury and brain swelling). The only difference of these two combinations related to presence or absence of hemorrhage i.e. hemorrhage is included in combination 2B but not in the combination 3B.

	Combination	*Brain stem injury*	*Cerebellar injury*	*Brain swelling*	*Hemorrhage*
Based on results from model A	**Combination 2A**	+	+	+	+
	Combination 3A	+	+	+	-
	Combination 4A	+	-	-	+
	Combination 5A	+	-	-	-
Based on results from Model B	**Combination 2B**	+	-	+	+
	Combination 3B	+	-	+	-

Table 28 Categories within combinations of various intracranial pathologies (+: present, -: not present)

Following the trial of these various combinations of intracranial pathologies, it was observed that, in model A, brain stem injury remained significant in all combinations, with haemorrhage never being significantly associated with outcome. Furthermore, in combinations where the brain swelling and the cerebellar injury were present i.e. combination 2A (brain stem injury, cerebellar injury and brain swelling) and combination 3A (combination 2A plus haemorrhage i.e. brain stem injury, cerebellar injury, brain swelling and haemorrhage), these covariates showed a marginally significant association with discharge survival. Similarly, in model B, the brain stem injury and the brain swelling were significantly associated with outcome in both combinations whilst haemorrhage demonstrated no significant association in combination 2B (brain stem injury, brain swelling and haemorrhage). We also added each intracranial pathology individually i.e. without combining with other intracranial pathologies and observed the same results.

6.4.3. Added value of each variable to the model performance

Table 29 shows the added value of AIS scores, the Marshall Classification and various intracranial pathologies to models A and B. The AUC and Nagelkerke R^2 of the models A and B prior to and following addition of each new variable (AIS scores, Marshall Classification or intracranial pathologies) are presented. The baseline models A and B had AUC and Nagelkerke R^2 of respectively 0.91 and 0.57. Addition of AIS scores and the Marshall Classification to model A resulted in a significant decrease in the deviance of the model along with increase in AUC from 0.91 to 0.92. Further, addition of various intracranial pathologies to this model demonstrated similar increase in AUC. In terms of Nagelkerke R^2, the degree of increase varied. Similar to model A, both AIS and the Marshall Classification significantly decreased the deviance of model B whilst increasing AUC from 0.91 to 0.92. Moreover, adding various intracranial pathologies to this model resulted in the same increase of AUC as to AIS score or the Marshall Classification, apart from the combination 1B which raised AUC slightly higher as to 0.93. Similar to model A, the increase in Nagelkerke R^2 varied.

	AUC	*Decrease in deviance (p value)*	*Nagelkerke R^2*		*AUC*	*Decrease in deviance (p value)*	*Nagelkerke R^2*
Baseline model A	0.91 (0.89-0.93)	-	0.57	**Baseline model B**	0.9 (0.88-0.93)	-	0.57
AIS	0.92 (0.89-0.94)	<0.001	0.58	**AIS**	0.92 (0.90-0.94)	<0.001	0.61
Marshal Classification	0.92 (0.90-0.94)	<0.001	0.60	**Marshal Classification**	0.92 (0.90-0.94)	<0.001	0.62
Combination 1A	0.92 (0.90-0.94)	-	0.60	**Combination 1B**	0.93 (0.90-0.95)	-	
Combination 2A	0.92 (0.90-0.94)		0.60	**Combination 2B**	0.92 (0.90-0.94)	-	0.61
Combination 3A	0.92 (0.90-0.94)		0.60	**Combination 3B**	0.92 (0.90-0.94)	<0.001	0.61
Combination 4A	0.92 (0.90-0.94)		0.59				
Combination 5A	0.92 (0.90-0.94)	<0.001	0.59				

Table 29 The added value of AIS scores, the Marshall Classification and various intracranial pathologies to the performance of models to predict survival at discharge.

6.5. Discussion

In this study, the predictive power of AIS severity scores, the Marshal Classification and various traditionally descriptive intracranial pathologies has been investigated. Using a subset of TBI patients submitted to TARN, two reference prognostic models to predict the discharge survival were constructed with important TBI prognosticators namely age, GCS, pupillary reactivity, ISS, cause of injury and extracranial injury as covariates. Then the association of each classification system (AIS severity score or the Marshall Classification) or descriptive pathology with outcome and also its contribution to the performance of the reference models were assessed in univariate and logistic regression analyses respectively. AIS score 5/6 appears to have a significant effect on outcome with AIS score 3 as the reference. Regarding the Marshal Classification, various Marshal Classes do not appear to have significant influence on outcome prediction in model A with only classes V and V/VI being significant in model B. Moreover, haemorrhage does not seem to be important although brain swelling and brain stem injuries (of any kind including haemorrhage and contusion) may be the important predictors of outcome among all structural damages. This analysis shows that including these classifications or intracranial pathologies slightly enhances the predictive power of baseline prognostic models as per AUC and Nagelkerke R^2.

6.5.1. Implications of the study

Although various AIS scores on univariate analysis have odds ratios of less than one for survival, which decreases in line with the increase of the score, this effect seems not to be maintained in multivariate analysis. This is because

score 4 does not hold significant association with outcome despite score 5 being significantly more predictive of death than score 3. This implies that the prognostic difference between AIS scores 3 and 4 is not significant. We cannot explain this finding because both scores contain various types and severities of injuries. However, it is possible that some injuries which are coded 3 should be perhaps placed under AIS score 4 or vice versa.

The Marshall Classification does not appear to be a reliable classifier for TBI patients based on their probable outcome. Although, addition of the Marshal Class to the baseline predictive models improved the performance, this does not add anything to the notion that structural brain damage (of any kind) is an indicator of a worse outcome than no intracranial pathology [147]. A reliable classification should be able to effectively provide relative predictive strength for each category of pathological findings which can be referred to as the capability of the classification system to score each individual category for outcome prediction. The Marshall Classification appears to fail in this matter in our analysis. However, we believe our results do not undermine the validity of the Marshall Classification for descriptive purposes or to indentify those TBI cases who are at high risk of developing raised ICP. In fact, not only this classification has not been proposed from the prognosis viewpoint, it may still be appropriate as a *CT* classification. This is because we used AIS codes to assign one Marshall Class to a TBI case in our data. AIS coding, however, employs the information not only from CT reports but also any source which can provide information on intracranial pathology (such as MRI, operation notes, clinical diagnosis etc.). Thus, according to results, one may assume this

classification is not good for categorising intracranial pathologies but valid for pathologies observed by CT.

This study in part permitted us to assess the prognostic importance of the type and anatomical location of brain damage following TBI. Contusion appears not to have a negative prognostic value because it does not influence the survival status. This is also the case for haemorrhage, when it is used in multivariate analysis. However, brain swelling still seems to be an important factor for outcome prediction. Regarding the location of the lesion, it appears that the most important location is the brain stem with no importance for cerebellar and cerebral lesions apart from brain swelling. These findings may have some therapeutic implications as there is currently no definitive or appropriate therapy/intervention for brain stem injury. Although this finding appears intuitive for clinicians involved in treating these patients, this poor prognosis is observed in comparison with other factors influencing the outcome. This means victims of brain stem injury still have a chance of survival since the brain stem injury is only one factor for calculating the probability of outcome in our prognostic models. Furthermore, the negative effect of brain swelling on outcome compared with haemorrhage or other injuries may imply the current therapeutic strategies are not sufficient for averting the poor outcome. This is highlighted further as our dataset is current and so modern therapeutic approaches have been taken into account. It is important to note that lack of significance for traumatic cerebral haemorrhage does not indicate the full efficiency of existing therapeutic approaches to intracranial pathologies. There are still many patients who sustain traumatic

cerebral haemorrhage and do not survive or subsequently end up with severe disability.

6.5.2. Comparison with the literature

Gennarelli *et al.*reported a non-linear relationship between AIS scores and mortality in that mortality consistently decreases as AIS score increased especially for scores 3 and above [144]. In our study, we observed that for AIS scores of brain injury, although there may be a decrease in the odds ratio of survival as AIS score increases (model B), the effect of score 4 on outcome is not significant as compared to score 3. Two important differences between our study and Gennarelli's may indicate more reliable results in our study. Firstly, we performed a multivariate analysis taking other important predictors into account where this was not performed in Gennarelli's study. It might be that the additional prognosis of AIS 4 versus 3 is covered by inclusion of GCS and pupillary reactivity in the models as GCS and the pupillary reactivity are more likely to be low/abnormal as AIS increases. Secondly, our analysis was exclusively performed on a subgroup of head AIS scores whereas Gennarelli's study is about AIS scores from all body regions.

The association of the Marshall Classification with outcome has been investigated in a number of other studies [24, 36, 92, 94, 140, 143]. Maas el. al. demonstrated that classes V and VI had a lower mortality rates than class IV but they observed that overall the Marshal Classification has reasonable discriminative power (AUC= 0.669) [140]. This appears similar to our finding that adding the Marshal Classification to the predictive models would improve the performance. No adjustment for confounders was made in this study. Moreover, Servadei *et al.* performed a univariate analysis on the outcome

predictability of the Marshal Classification with a similar result to our univariate analysis [143]. Among the studies which performed a multivariate analysis, the Marshal Classification was reported as a significant prognosticator when this classification was not used in its original form [24, 36, 92]. Apart from classes V and VI in these studies, various other Marshal classes were merged together such as class III merged with class IV [24, 92] or class I merged with class II [36]. We did not perform such mergence as we intended to use the Marshall Classification in its original form although mergence of classes V and VI was unavoidable due to lack of reliable information. One study found significant influence of each Marshall Class on outcome with only merging classes V and VI following adjustment with age, motor GCS and pupillary reactivity [94]. However, in this study the reference category was class II whereas in ours class I was the reference category.

The difference in severity of TBI may explain the difference of our findings with regards to haemorrhage, SAH and EDH with many other studies consistently reporting the predictive significance of haemorrhage [93], SAH [24, 36, 92, 140, 148-152], or EDH [24, 92, 94, 140, 148]. For example, Fearnside *et al.* observed a significant prognostic importance for SAH by multivariate analysis of a dataset in which all patients had GCS of 8 or less whereas in our dataset 35.8% of patients had mild GCS (i.e. >12) [149]. Similarly, Azian *et al.* excluded those cases without intracranial haemorrhage and reported the significant predictive strength of SAH and EDH [148]. It may be that haemorrhage, SAH or EDH is a less important predictor of outcome in patients with less severe TBI unlike those who sustained more serious injuries. However, one important strength of our research is that our dataset is much

more recent than those used in other such studies. Moreover, the other explanation of the different results may relate to using AIS codes which rely also on MRI, operation notes or even mere clinical diagnosis and not only CT images. Regarding this, it is important to consider our study as a prognostic analysis of intracranial pathologies and not merely CT findings. As such, although it somewhat gives prognostic insight about various CT finding as they count in AIS coding, we believe the results of our study are more relevant for clinical decision makings with regards to the ultimate patient's diagnosis rather than only CT findings.

6.5.3. Limitations

Although our dataset contains severe cases of trauma per TARN submission criteria, our study may be considered to suffer from strong selection bias. This is because death is one criterion to be included into the TARN trauma registry (a bias towards inclusion of patients who die). Whilst this can potentially pose a bias, in our analyzed dataset only 5 cases stayed at hospital for less than 3 days and of these only 3 cases died. We can assume that the remaining 2 who survived entered the dataset due to either intensive care or inter-hospital transfer (the other two criteria for submission to TARN). This also highlights that majority of cases in the data (99.37%) sustained brain injuries severe enough to stay at hospital for longer than 3 days. As such, it can be assumed that the inclusion criteria for this study is longer than 3 days stay at hospital since 99.37% of cases entered the dataset because of this criterion. One may, however, consider a selection bias towards excluding those cases who sustain intracranial pathology but not severe enough for more than 3 days in-hospital care and they do not enter the registry since they survive. Whilst this is an issue

with our current dataset, we believe the number of such cases must not be high to pose a significant bias to our study. In fact, the analyzed dataset is more representative of TBI victims who sustain intracranial pathologies as 35.8% of cases had admission mild GCS (i.e. > 13). This subgroup of patients represents those patients who 'talk and die'. Overall, the survival rate in our data is close to the average survival rate in other severe TBI populations (70%). We believe although the selection bias is an issue with our analysis, it is not strong to significantly undermine the results.

The other limitation of our study is that we classified various intracranial pathologies described by AIS codes according to the Marshall Classification [99]. This classification was designed to identify risk of raised ICP based on CT image whereas AIS coding is performed from any source of information including MRI, operating notes etc.. Whilst AIS coding mostly relies on CT reports, the effect of other sources of information on the Marshall Classification obtained with AIS intermediation is unclear [99]. Furthermore, AIS coding is reflective of the dynamic nature of the brain injury as it employs multiple sources complied over time whereas the Marshall Classification is from a certain point in time (oftentimes on admission). Servadei *et al.* demonstrated that the evolution of intracranial pathologies per se is a TBI prognosticator [143]. The temporal change of brain injury is consequently inherited in the Marshall Classification when obtained from AIS descriptions but not when obtained from observing the actual CT image. Overall, there may be some overestimation of negative prognostic strength of the Marshall System when it is performed via AIS coders.

In this study, the highest AIS score was taken as a single variable for the analysis of association with outcome. However, patients may sustain multiple intracranial injuries and thus attract many AIS codes for each injury, especially if the brain injury is of severe type. Thus choosing only one AIS score out of several severity scores allocated might not be appropriate. This is an area for future research to determine how the predictive strength of intracranial injuries changes in the event of multiple intracranial injuries. It is still unclear which patient is at higher risk of experiencing unfavourable outcome, for instance, if one of them has multiple injuries with severity of 3 and the other has one brain injury scored as 4.

In this study, the original hypothesis was that prognostic analysis of combinations 1A or B would enable us to propose a reliable classification with significant prognostic value. However, apart from the brain stem injury and brain swelling, all other intracranial pathologies appeared non-significant in this analysis. Although this finding adds valuable information to the current literature with respect to the relative efficiency of therapeutic approaches to each intracranial pathology, at this stage we can not propose a classification apart from using the common conventional terms such as SAH, EDH, SDH etc.

6.5.4. Future direction

As some TBI patients, especially those who sustain severe injuries, have a high chance of multiple types of injury, the future research on development of a classification of intracranial pathologies for prognostic purposes should examine the impact of the number of intracranial injuries. Other important factors might be the anatomical location because of the

importance of brain stem injury, types of injury because of the importance of brain swelling and also the extent or severity of each injury.

The results of our study still require further validation in a different series of TBI. This particularly relates to the follow-up interval with regards to the outcome assessment and also the type of outcome (disability versus survival). Whilst in this study, the endpoint of outcome was discharge survival (well-recognised in prognostic analysis of trauma registries data [22, 120, 129, 142, 154]), this analysis should be replicated for disability and also long term outcome such as 6 months following injury.

6.6. Conclusion

Within a subset of severe TBI patients, not all AIS scores or Marshall Classes have prognostic significance when taken into account along with clinicodemographic prognostic factors. This suggests AIS scores and the Marshall Classification may not be appropriate to classify intracranial pathologies. The significant association of brain swelling and brain stem injury implies the need to improve therapeutic approaches to those patients who have sustained these injuries. Furthermore, development of any new classification which can be employed for predictive purposes entails considerations given to the type, location; extent and multiplicity of injuries prioritising which one of these factors are of more importance.

Conflict of interest statement

None

Acknowledgements

We would like to thank TARN members of staff and participating hospitals for the collection and submission of the data. This work was in part funded by the Trauma Audit and Research Network (TARN) and Oversees Research Students (ORS) Award Scheme, University of Manchester.

7. Paper 6: Models of Mortality Probability in Severe Traumatic Brain Injury: Results of TARN Modelling

Authors

- Mehdi Moazzez Lesko
- Tom Jenks
- Hester Lingsma
- Pablo Perel
- Sarah O'Brien
- Charmaine Childs
- Omar Bouamra
- Maralyn Woodford
- Fiona Lecky

7.1. Abstract

7.1.1. Background

Prognostic models in traumatic brain injury (TBI) are employed to design clinical trials, to assess/compare trauma care systems and to adjust trauma care for an individual patient. The current available prognostic models are rather old (the IMPACT models) or derived from non-homogenous datasets in terms of the trauma care delivered (the CRASH models).

7.1.2. Aim

To construct prognostic models to predict outcome in recent UK TBI patients.

7.1.3. Method

Records of patients with brain injury since January 2005 were extracted from the Trauma Audit and Research Network (TARN) database. TARN holds the records of patients with severe injuries i.e. longer than 3 days stay at hospital, inter-hospital transfer, critical care in hospital or death. Following a literature review, the variables age, cause of injury, GCS, pupillary reactivity, Injury Severity Score (ISS), CT classifications and various intracranial pathologies, systolic and mean blood pressure, hypoxia and the presence of extracranial injury were tested with survival at discharge as outcome. Variables with no significant correlation on univariate analysis were excluded. Stepwise logistic regression analysis was performed.

7.1.4. Results

Two models were derived on 802 patients with significant brain injury (models A and B). Age, GCS, pupillary reactivity, hypoxia and brain stem injury are

significant predictors in both. However, model A contains ISS in contrast to model B with the presence of brain swelling, cause of injury and major extracranial injury. Both models have high predictive performance (Model A; Area Under the ROC Curve (AUC) =0.92 (95% CI: 0.90-0.95), Nagelkerke R^2: 0.62 and HL test: P value = 0.20, Model B; AUC = 0.93 (95% CI: 0.91-0.95), Nagelkerke R^2: 064 and HL test: P value= 0.19).

7.1.5. Conclusion

We have developed two prognostic models applicable to UK patients recently hospitalised after traumatic brain injury.

7.2. Introduction

Traumatic Brain injury (TBI) is a global public health issue and is the cause of a substantial number of deaths and disabilities each year [155]. Tackling the problem involves devising and implementation of several preventive measures ranging from legislations on speed limit and seat belt for primary and secondary prevention [156] to the provision of the appropriate acute trauma care and rehabilitative strategies as tertiary prevention to avert the negative consequences of brain damage when it has already occurred. To this aim, prognosis is one of the several factors which can potentially help clinicians with decision making. Further, availability of a prognostic tool would assist trauma registries to benchmark local care systems by comparing their performance to a national standard performance. It has been suggested that accurate prognostic tools can also improve patient selection in trials of new interventions in TBI [157].

It is well-established that prognostication of TBI can not be based on a single clinical measure. For example, although in general, brain-injured patients with low or moderate Glasgow Coma Scale (GCS) are more prone to unfavourable outcome but a significant proportion of these patients can have a reasonable outcome with appropriate care [158]. Similarly, many patients with admission GCS scores of more 14 or 15 may deteriorate to experience unfavourable outcome [158]. This is because the prognosis in TBI depends on several clinical factors such as age, level of consciousness, pupillary reactivity or Computed Tomography (CT) features [92]. The interaction among these factors is such that the effect of one factor on outcome may be influenced or eliminated in the presence or absence of the other. To address this complex

interaction, prognostic models have been employed to construct a tool which provides the probability of various outcome measures for a given victim of TBI over time taking multiple predictors into account.

In two recent systematic reviews, Perel. *et al.* [31] and Mushkudiani *et al.* [159] observed that the literature contains numerous TBI prognostic models but many of these models are methodologically flawed in that the derivation samples are too small to provide powerful results, the models are not externally validated and some studies lack measures of model performance. These are important considerations with regards to the reliability of the presented prognostic models. However, there are currently two large initiatives which have constructed accurate and reliable prognostic models available online; these projects being International Mission for Prognosis and Analysis of Clinical Trials (IMPACT) [28] and Corticosteroid Randomization After Significant Head injury (CRASH) studies [23, 24].

The derivation of the CRASH and IMPACT models from clinical trials can pose a selection bias as only cases who met the inclusion criteria for the sake of given intervention were contained in the dataset. Furthermore, it is important that prognostic analysis is accounted for as an ongoing procedure in updated and contemporary TBI series [88]. As such, the IMPACT models may be considered somewhat historic. Similarly, the regional diversity in trauma care should be taken into account as countries exercise different policies with regards to many factors which can affect the outcome such as pre-hospital care, intervention versus conservative approach or provision of neuro-intensive care [86]. The Trauma Audit and Research Network (TARN) is a trauma registry based in the University Of Manchester, UK which holds records of trauma

patients admitted to participating hospitals across England and Wales [89]. Therefore, the TARN dataset contains cases who were cared under the British trauma care system whereas the British data in either CRASH or the IMPACT prognostic studies were merged with data from other parts of the world with different trauma care policy than that in Britain.

With regards to the importance of regional and temporal differences in trauma care and the importance of observational studies for prognostic analysis, the objective of our study is to construct a prognostic model by using TBI cases from TARN which can be reliably applied to TBI outcome prediction in the UK.

7.3. Methods

To obtain a well-developed prognostic model, the quality assessment tool proposed in Perel's systematic review was used as a guide [31]. Briefly, the internal validity of a model involves adequate rationale to include clearly-defined predictors, employment of appropriate imputation strategies for missing information, performance of an adequate strategy to construct the model, appropriate management of interactions, appropriate management of continuous variables and lastly the inclusion of 10 outcome events per variable. Moreover, the derivation sample should be adequately described and the model should be presented with confidence intervals along with measures of discrimination and calibration. A well-developed model is also expected to sustain its performance in data different to the derivation dataset.

7.3.1. Selection of predictors and population sample

The predictors were selected based on the literature review; these being age [23, 31, 35, 36, 92], cause of injury [92], GCS [23, 31, 35, 36, 92], pupillary reactivity [23, 31, 35, 36, 92], Injury Severity Score (ISS) [31, 35], systolic and mean blood pressure [31, 36, 92], presence/absence of hypoxia [31, 36, 92], CT findings [23, 31, 35, 36, 92] and presence/absence of extracranial injury [23]. The dataset was selected from TARN. The criteria for submission to TARN are more than three days stay at hospital, reception of intensive care, inter-hospital transfer for specialist care or death due to injury after admission. Hospitals submit data to TARN via a web based data collection system. Trained coders at TARN would then code the injuries of each trauma case using the Abbreviated Injury Scale (AIS) dictionary (currently update 1998) [90]. The inclusion

criteria for this study were head AIS codes under the 'internal organ' [90, 91] which held severity score of 3 or above including basal and compound/depressed/open skull fracture and availability of pupillary reactivity at any time point of trauma care in TARN. The latter criterion arose as pupillary reactivity has been shown to be correlated with outcome in the IMPACT and CRASH models and TARN commenced recording this variable only recently i.e. from September 2005 onward. The outcome measure for the analysis was survival at discharge and where applicable, the time point of measurement of variables was on arrival at the first hospital (such as for GCS, pupillary reactivity, blood pressure etc.).

7.3.2. Univariate analysis

The correlation of each covariate with survival was assessed utilizing Mann Whitney U test for continuous variables and Chi square test for categorical variables with $p < 0.05$ indicating significance. Age and ISS were considered continuous with cause of injury, pupillary reactivity, presence/absence of hypoxia (O2 saturation < 90 mmHg), CT findings and the presence/absence of extracranial injury as categorical variables. However, GCS, systolic and mean blood pressures were assessed both categorically and continuously. GCS was categorized into mild i.e. GCS 13 -15, moderate i.e. GCS 9-12 and severe i.e. GCS scores of < 9. The cut-offs for systolic blood pressure were 120 mmHg and 150 mmHg categorizing that into low blood pressure (<120 mmHg), normotension (120-150 mmHg) and hypertension (>150 mmHg) [95]. These were the cut-offs proposed by the IMPACT group which are the 25^{th} and 75^{th} centiles of systolic blood pressure in their data and were observed to have

significant influence on outomce in the multivariate anslsyis. Similarly, the cut-offs of 85 and 110 mmHg were used for mean blood pressure [95].

The descriptions of AIS codes available from AIS dictionary were used as "substitutes" for CT findings. This variable was categorized according to the Marshal CT Classification [99]. Descriptions of AIS codes were also grouped based on commonly-used descriptive terms namely contusion, brain swelling, intracranial haemorrhage, Epidural Haemorrhage (EDH), Subdural Haemorrhage (SDH), Subarachnoid Haemorrhage (SAH), brain stem injury and cerebellar injury. In this manner, if the contusion or haemorrhage occurred in the brain stem or cerebellum, it was grouped as the brain stem or cerebellar injury. Moreover, we examined the association of each head AIS severity score (ranging from minimal 1 to maximal 6) to the outcome.

To address the linear relationship of continuous variables with $\log_e$ (odds of survival) as a requirement for logistic regression analysis (commonly referred to as linearity assumption) [134], fractional polynomials functions of continuous variables were employed [135]. Briefly, in this method, power transformation(s) of the variable is selected out of the power candidates of -3, -2, -1, 0, 1, 2, 3 where 0 is $\log_e$ transformation. The fractional polynomial analyses showed that whilst age can be included in the model without any transformations; GCS, ISS, systolic and mean blood pressure require transformation to correct their non-linear relationship. This simply implies that, for instance instead of the crude GCS or ISS values, the following transformations are required to be made prior to addition to the model: $\left(\frac{10}{GCS+1}\right)^2 - 0.76$, $\left(\frac{10}{GCS+1}\right)^2 \times \log_e\left(\frac{GCS+1}{10}\right) - 1.02$ and $\log_e\left(\frac{ISS}{10}\right) - 0.91$

7.3.3. Model derivation

Based on univariate analyses and with the significant level of 5%, the covariates identified suitable for inclusion in the modelling procedure were: age, cause of injury, GCS (continuous/categorical), pupillary reactivity, ISS, extracranial injury, systolic blood pressure (categorical), mean blood pressure (categorical), hypoxia, brain swelling, intracranial haemorrhage, EPH, SAH, brain stem injury, cerebellar injury and the Marshal Classification. Forward stepwise logistic regression was used "manually". Initially a model was constructed with age, GCS, pupillary reactivity, ISS and extracranial injury and subsequently other variables (CT finding, systolic and mean blood pressure and presence/absence of hypoxia) were added. Age, GCS, pupillary reactivity and extracranial injury are the covariates in the basic CRASH models. Model A of IMPACT also contains these covariates apart from extracranial injury which is not recorded in IMPACT [92]. We added ISS to this list as the extent of extracranial injury can affect it and thus ISS contains information on extracranial injury too. The next step was adding CT features which is the case in the CRASH models and the IMPACT model B. Lastly, addition of blood pressure and hypoxia was based on the order in the IMPACT model C. Each step involved addition of the next variable and evaluation of the new model in terms of the decrease in the deviance, the multicolinearity effect with other covariates and improvement in the models performance. If the decrease in the deviance compared to the model of the previous step was not significant, the variable was excluded from the rest of the modelling. In case of multicolinearity effect; the modelling procedure was branched with two

‘parallel’ models to avoid containment of variables with this effect in the same model.

During the modelling procedure, it was observed that categorical GCS is similar to its continuous form in terms of the added predictive value to the model and thus, the modelling was continued with categorical GCS due to its simplicity to use as compared to complicated fractional polynomials transformations. Likewise, two types of classification of pupillary reactivity were tested in terms of the predictive value in the model (4 categories: normal, abnormal both eyes reactive, only one eye reactive and bilaterally unreactive *versus* 3 categories: both eyes reactive, only one eye reactive and bilaterally unreactive). It was observed that the prognostic strength is better when this variable is used with 4 categories rather than with 3 categories.

Neither systolic nor mean blood pressure demonstrated significant effect on outcome when they included in the same model. However, when systolic and mean blood pressures were tested in the separate models. Furthermore, it was observed that between hypertension and low blood pressure only low blood pressure demonstrated significant association with outcome when normotension was the reference category. Thus the hypertension category was merged with the normotension category leaving these variables with only two categories of normotension (including hypertension) versus low blood pressure. Moreover, as the model with mean blood pressure did not show acceptable goodness of fit as per HL statistics (i.e. p value < 0.05), the model with systolic blood pressure was selected.

7.3.4. Interactions

This happens when the effect of one covariate on outcome is influenced by the presence of the other covariate. Based on the literature, these interactions were investigated: age with cause of injury [160], systolic blood pressure with hypoxia [95] and mean blood pressure with present/absent hypoxia [95].

7.3.5. Imputation

In multivariate analyses, every case with one missing value is discarded as if it does not exist in the dataset. Proper 'guess' on missing values in a dataset is superior to such loss of cases. In this study, all the univariate analyses were performed without imputation where in the multivariate analysis, the missing information was sequentially imputed with values recorded en route or at scene. However, this strategy failed to fill some missing values on systolic blood pressure and presence/absence of hypoxia. Therefore, the remaining unrecorded data on these variables were all placed into a separate category as 'missing'.

7.3.6. Model validation

For the external validation, the final models were run on a different TBI dataset from TARN (from May 2008 to April 2010) and also the IMPACT dataset.

Fractional polynomials transformation was performed in Stata and all other univariate analyses, model derivation and validation procedure were performed in Statistical Package for the Social Sciences (SPSS version 15).

7.4. Results

The inclusion criteria retrieved 802 TBI cases from the TARN trauma dataset. The admission date of all cases was from September 2005 (which was the start date of recording pupillary reactivity in TARN) until April 2008 (with only one case recorded in April). However, there were some occasional records on pupillary reactivity prior to start date of recording pupillary reactivity in TARN which included in this analysis (127 (15.8%) cases in total).

Table 30 compares the characteristics of patients whose profiles were submitted to TARN after September 2005 and were included in the modelling procedure to those who were excluded during the same period. The reason for this exclusion was lack of information on pupillary reactivity on every time point of measurement in TARN. As seen, the figures related to admission GCS and ISS are equal or close in both groups. Similarly, although male percentage and survival rate in included cases are different from those in excluded cases; this is not statistically significant ($p > 0.05$). The included dataset, however, seems to consist of significantly younger patients than the excluded dataset.

Variable	*Excluded submissions (1558)*	*Included submissions (675)*	*p* value
Age	43.4 (25.8-67.4)	38.2 (23.1-56.5)	0.01
Male	71.1%	74.6%	0.20
Admission overall GCS	12 (6-14)	11 (4-14)	0.09
ISS	25 (17-30)	25 (17-30)	-
Survival rate	72.8	73.8	0.72

Table 30 Comparison of demographic and injury characteristics of cases which were excluded from the model derivation to those included (i.e. submissions to TARN after September 2005) (brakets: 75% interquatile range).

Table 31 shows the clinicodemographic characteristics of the population sample and the results of univariate analysis. The median age is 39 with interqaurtile range of 22 to 58 with significant correlation with outcome. Males constitute 75.2% of the population with no influence of gender on outcome. The majority of the study population (91.9%) are recorded as being British with 40% missing information on this variable. No correlation is observed between nationality and outcome. Median GCS is 13 (interquartile range: 5-15) and the frequency of mild, moderate and severe GCS categories

are 35.8%, 12.7% and 51.4% respectively. GCS is significantly correlated with outcome both continuously and categorically. Most patients have pupillary reactivity recorded as brisk-brisk (68.6%) followed by bilaterally no reaction (16.5%). Additionally, the effect of pupillary reactivity on outcome prediction is observed to be significant. The patients' ISS scores has a median value of 25 (interquartile range: 16-29) with significant correlation with outcome. 14.5% of patients have concomitant extracranial injury as determined by extracranial AIS score of 4 or above. This significantly subjects the patient to a lower likelihood of survival.

Covariate		*Median (interquartile range)*	*Frequency (percentage)*	*Odds ratio for survival*	*p* **value** *by Man Whitney U test/Chi square test*
Age		39 (22-58)		0.98 (0.98-0.99)	0.00
Gender	**Male**		603 (75.2%)		0.75
	Female		199 (24.8%)	0.94 (0.65-1.36)	
Nationality	**British**		442 (91.9%)		0.24*
	European		15 (3.1%)	2.10 (0.47-9.46)	
	Others		24 (4.9%)	1.22 (0.45-3.37)	
Cause of injury	**RTC**		314 39.2%)		0.00
	fall		313(39.0%)	0.93 (0.66-1.32)	
	Assaults		143(17.8%)	2.662 (1.53-4.62)	
	others		32(4%)	1.369 (0.57-3.28)	

Table 31 Clinicodemographic characteristics of population sample and results of univariate analysis *The p value indicates the correlation of all categories of covariate with outcome and not only one category.

Covariate		*Median (interquartile range)*	*Frequency (percentage)*	*Odds ratio for survival*	*p value by Man Whitney U test/Chi square test*
GCS(categorical)	**Mild**		278 (35.8%)		0.00
	Moderate		99 (12.7%)	0.35 (0.19-0.66)	
	Severe		399 (51.4%)	0.08 (0.05-0.12)	
Pupillary reactivity	**Brisk-brisk**		446 (68.6%)		0.00*
	Sluggish-sluggish		53 (8.1%)	0.21 (0.11-0.41)	
	Brisk-sluggish		16 (2.4%)	0.17 (0.06-0.49)	
	None-brisk		16 (2.4%)	1.52 (0.20-11.79)	
	None-sluggish		12 (1.8%)	0.02 (0.00-0.10)	
	None-none		107 (16.5%)	0.03 (0.02-0.05)	
ISS		25 (16-29)		0.93 (0.92-0.94)	0.00
Extracranial injury (cut-off: AIS score =4)			116 (14.5%)	0.51 (0.37-0.71)	0.00
Survival			599 (74.7%)		

Table 31 Clinicodemographic characteristics of population sample and results of univariate analysis *The p value indicates the correlation of all categories of covariate with outcome and not only one category. *(continued)*

With regards to physiological measures (Table 32), the median systolic and mean blood pressure are 136.5 (interquartile range: 120-75) and 145 (interquartile range: 126-160) respectively. According to systolic blood pressure, 40.4% of cases are normotensive with 29.1% and 26.1% remaining records of hypertension and low blood pressure respectively. However, per mean arterial blood pressure, majority of cases are hypertensive at 82.3% with only 8.1% and 3.7% records of normotension and hypotension respectively. Neither systolic nor mean blood pressure have any influence on outcome prediction when used continuously whereas this influence is significant categorically. Furthermore, 6.5% patients are hypoxic with a significantly higher chance of death compared to non-hypoxic patients.

Covariate		*Median (interquartile range)*	*Frequency (percentage)*	*Odds ratio for survival*	*p value by Man Whitney U test/Chi square test*
Systolic blood pressure		136.5 (120-75)		1.008 (1.00-1.01)	0.4
Systolic blood pressure (categorical)	**Hypotension (< 120 mmHg)**		209 (26.1%)	0.37 (0.25-0.56)	0.00
	Normtension (120-150 mmHg)		324 (40.4%)		
	Hypertension (> 150 mmHg)		233 (29.1%)	0.41 (0.27-0.62)	
Mean blood pressure(continuous)		145 (125-160)		1.01 (1.00-1.01)	0.37
Mean blood pressure (categorical)	**Hypotension (< 85 mmHg)**		30 (3.7%)	0.11 (0.04-0.31)	0.00
	Normtension (85 – 110 mmHg)		65 (8.1%)		
	Hypertension (> 110 mmHg)		660 (82.3%)	1.73 (1-2.99)	
Hypoxia (O2 Saturation < 90 mmHg)			51 (6.5%)	0.17 (0.09-0.31)	0.00

Table 32 Vital signs of the population studied and results of univariate analysis.

With regards to AIS scores, the Marshall Classification and various intracranial pathologies (Table 33), most cases have the highest brain injury AIS score of 4 at 39.4% and only 6 cases are recorded with the highest AIS of 6. Likewise, intracranial haemorrhage (of any type) constitutes 76.4% of cases with SDH being the most frequent type of haemorrhage (22.1%) followed by SAH (18.6%). Similarly, 8.9% and 5.1% of cases have brain stem and cerebellar injury respectively (of any kind such as haemorrhage or contusion). Nevertheless, non-hemorrhagic brain injuries such as contusion and swelling comprise 39.9% and 34.2% of the sample population respectively. Regarding, the Marshal CT classification, the most frequent class is class II (50.5%) followed by class V/VI (20.7%). The Marshal Classification, various AIS scores and CT findings are significantly correlated with survival apart from contusion and SDH.

Covariate		*Frequency (percentage)*	*Odds ratio for survival*	*p value by Man Whitney U test/Chi square test*
Highest AIS scores	3	189 (23.6%)		0.00
	4	316 (39.4%)	0.73 (0.42-1.27)	
	5	291 (36.2%)	0.15 (0.09-0.25)	
	6	6 (0.7%)	0.02 (0.00-0.22)	
Contusion		320 (39.9%)	1.25 (0.90-1.74)	0.18
Brain swelling		275 (34. 2%)	0.31 (0.22-0.43)	0.00
Intracranial haemorrhage		613 (76.4%)	0.55 (0.36-0.83)	0.00
Epidural haemorrhage		95 (11.8%)	2.85 (1.49-5.45)	0.00
Subdural haemorrhage		178 (22.1%)	1.17 (0.79-1.74)	0.43

Table 33 AIS score, CT findings and the Marshal Class of the population studied and results of the univariate analysis

Covariate		*Frequency (percentage)*	*Odds ratio for survival*	*p value by Man Whitney U test/Chi square test*
SAH		149 (18.6%)	0.56 (0.38-0.82)	0.00
Brain stem injury		72 (8.9%)	0.08 (0.04-0.14)	0.00
Cerebellar injury		41 (5.1%)	0.30 (0.16-0.56)	0.00
The Marshal CT classification	**I**	65 (8.1%)		0.00
	II	405 (50.5%)	1.17 (0.54-2.50)	
	III	85 (10.6%)	0.39 (0.17-0.90)	
	IV	74 (9.2%)	0.19 (0.08-0.44)	
	V/VI	166 (20.7%)	0.160 (0.08-0.35)	
	Brain stem/cerebellar injury	4 (0.5%)	0.05 (0.005-0.57)	
	Penetrating injury	3 (0.4%)	0.32 (0.03-3.92)	

Table 33 AIS score, CT findings and the Marshal Class of the population studied and results of the univariate analysis *(continured)*

7.4.1. Proposed models

Table 34 (a: model A, b: model B) presents two models derived from the dataset. Each model is presented with the covariates, their odds ratios (plus confidence intervals), the significance level of the effect on outcome and the constant. The reason for the construction of two models is mainly related to the correlation of ISS and extracranial injury observed in the correlation matrix during the model construction. This implies that these two covariates can not be contained in the same model due to multicolinearity effect. Furthermore, it was observed that in the multivariate analysis the cut-off point of AIS severity score ≥ 3 for the presence of extracranial injury does not yield significant correlation with outcome. Nevertheless, when this cut-off was increased to extracranial AIS severity score ≥ 4, the significance level of this correlation decreased to less than 5%. The other difference between model A and model B relates to the predictive power of cause of injury and brain swelling which, unlike in model A, has importance in model B.

		Coefficient	*Odds ratio*	*95. 0% C. I for odds ratio*		*Sig.*
Age		-0.05	0.95	0.94	0.96	< 0.005
GCS	**mild**					
	moderate	-0.86	0.42	0.21	0.87	0.02
	severe	-1.42	0.24	0.13	0.44	< 0.005
Pupillary reactivity	**Normal**					
	Abnormal-both reactive	-0.98	0.38	0.2	0.72	< 0.005
	Only one reactive	-1.27	0.28	0.12	0.68	< 0.005
	None reactive	-2.87	0.06	0.03	0.11	< 0.005
$\log_e\left(\frac{ISS}{10}\right) - 0.91$		-1.36	0.26	0.14	0.47	< 0.005
Brain stem injury		-1.71	0.18	0.08	0.39	< 0.005
Hypoxia	**Yes**	-1.31	0.27	0.13	0.57	< 0.005
	Missing	-0.5	0.61	0.27	1.36	0.225
Systolic blood pressure	**Normotension**					
	low blood pressure	-0.63	0.53	0.31	0.91	0.02
	Missing	-0.77	0.46	0.1	2.17	0.329
Constant		5.63				

Table 34 (a) results of multivariate analysis of outcome prediction: model A

		Coefficient	*Odds ratio*	*95. 0% C. I. for odds ratio*		*Sig.*
Age		-0.04	0.96	0.94	0.98	<0.005
GCS	**mild**					<0.005
	moderate	-0.99	0.37	0.17	0.79	0.01
	severe	-1.42	0.24	0.13	0.46	<0.005
Pupillary reactivity	**Normal**					<0.005
	Abnormal-both reactive	-1.01	0.36	0.18	0.71	<0.005
	Only one reactive	-1.54	0.21	0.08	0.54	<0.005
	None reactive	-3.13	0.04	0.02	0.09	<0.005
Injury cause	**RTC**					0.37
	Fall	0.83	2.31	0.56	9.43	0.24
	Assaults	-1.02	0.36	0.05	2.57	0.31
	Others	0.76	2.14	0.04	114.99	0.71

Table 34 (b) results of multivariate analysis of outcome prediction: model B

		Coefficient	*Odds ratio*	*95. 0% C. I. for odds ratio*		*Sig.*
Extracranial injury		-1.26	0.28	0.15	0.55	<0.005
Brain stem injury		-1.85	0.16	0.07	0.35	<0.005
Brain swelling		-0.89	0.41	0.24	0.7	<0.005
Hypoxia	**Yes**	-1.42	0.24	0.11	0.54	<0.005
	Missing	-0.38	0.68	0.30	1.54	0.36
Systolic blood pressure	**Normotension**					
	low blood prusse	-0.60	0.55	0.31	0.97	0.04
	Missing	-0.71	0.49	0.10	2.48	0.39
Interaction of age and injury cause	**Age and RTC**					
	Age and fall	-0.02	0.98	0.95	1	0.05
	Age and assault	0.04	1.05	0.99	1.10	0.08
	Age and others	-0.02	0.98	0.90	1.07	0.69
Constant		5.99				

Table 34 (b) results of multivariate analysis of outcome prediction: model B *(continued)*

7.4.2. Models performance

Table 35 shows various dimensions of models performances in the derivation dataset and two external datasets from IMPACT and TARN. The two models have approximately similar figures on every measure of performance in the derivation set. The commonly used measure of performance, i.e. AUC, is 0.92 (95% CI: 0.90-0.95) for model A and AUC = 0.93 (0.91-0.95) for model B in the prediction set. With regards to calibration, HL test of model A and B had a p value of 0.20 and 0.19 respectively. The TARN external dataset contains TBI cases from May 2008 till May 2010 with the same inclusion criteria as those for the derivation set although the AUC of model A remains the same in the TARN external validation set, the AUC of model B drops from 0.93 to 0.82. However, it was not possible to run model A in the IMPACT data as ISS was not recorded by IMPACT and thus model B was run for survival and favourable outcome prediction (Glasgow Outcome Scale (GOS) $\geq$ 4). The AUC of model B drops from 0.93 to 0.68 and 0.69 respectively for survival and favourable outcome prediction in the IMAPCT external dataset.

	Model A		*Model B*			
	Prediction set (n=802)	***External validation set from TARN (n=990)***	***Prediction set (n=802)***	***External validation set from TARN (n=792)***	***External validation set from the IMPACT (n=5476)***	
					survival	***favourable outcome***
Positive predictive value	68.5%	91.5%	68.5%	90.4%	18.6%	39.8%
Negative predicative value	94.5%	73.1%	94%	85%	56.6%	28.1%
Sensitivity	80.8%	94.6%	79.4%	99.3%	73.2%	77.1%
Specificity	89.8%	64.6%	89.8%	25.3%	11.8%	7.1%
Classification accuracy	87.9%	89.1%	87.5%	90.1%	24.3%	38.2%
HL statistics (*p* value)	0.20		0.19			
Brier score	0.09	0.02	0.08	0.1		
AUC	0.92 (0.90-0.95)	0.92 (0.89-0.94)	0.93 (0.91-0.95)	0.82 (0.78-.0.86)	0.68 (0.67-0.70)	0.69 (0.68-0.70)
Nagelkerke R^2	0.62		0.64			

Table 35 Performance of models A and B across various measures of performance (n: number of cases).

Figure 13 and Figure 14 show the Receiving Operating Curve (ROC) curve of respectively model A and B. For model A, the decrease in AUC in the TARN external dataset as compared to that in the derivation set is slight (constant line versus dashed line). This is also the case for model B as depicted by Figure 14. However, the decline is relatively huge in IMPACT data for either survival or favourable outcome prediction.

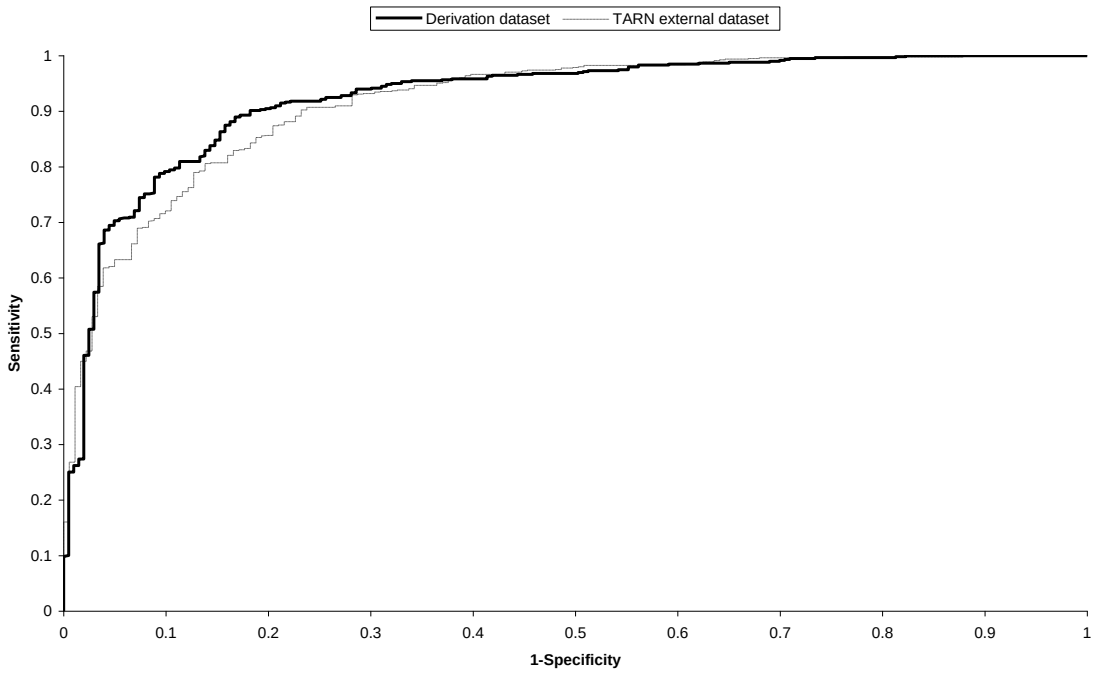


Figure 13 ROC curves of model A

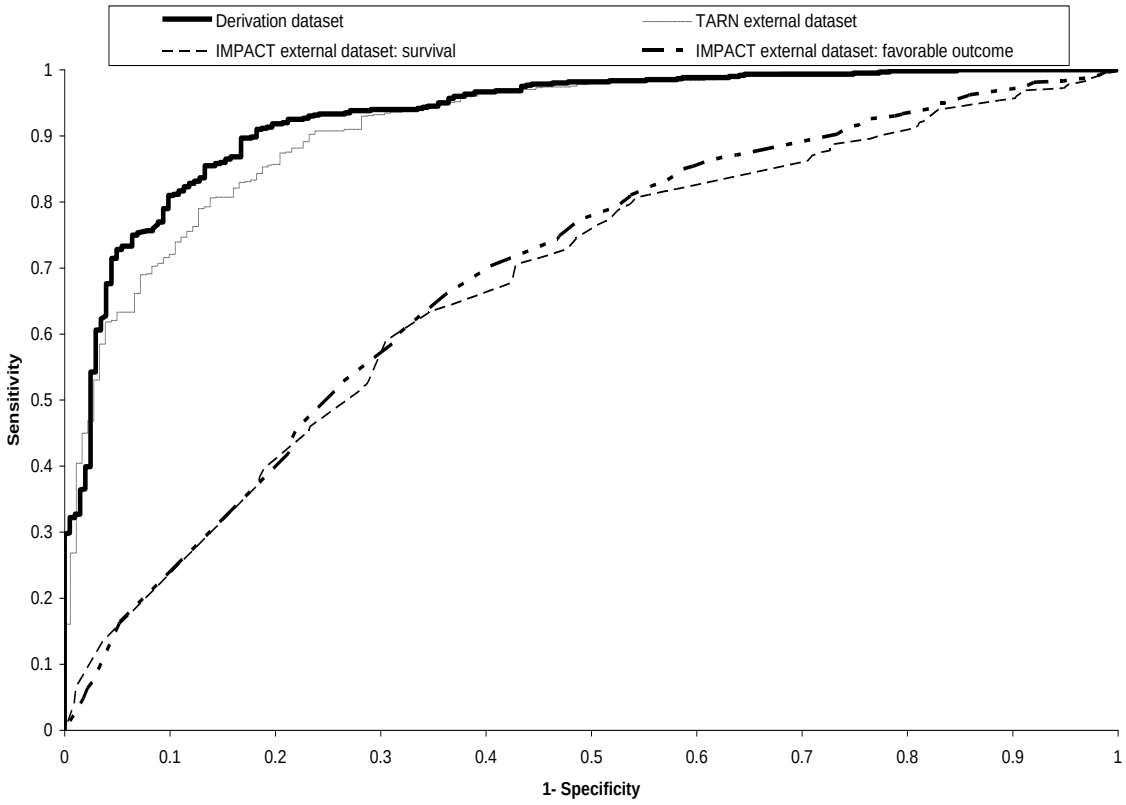


Figure 14 ROC curves of model B

Figure 15, Figure 16 and Figure 17 show the degree of agreement between the predicted probability of survival made by the models and the probability observed (calibration plot) (Figure 15: the derivation set, Figure 16: the TARN external validation set, Figure 17: the IMPACT external validation set). The line reflects the ideal situation in which every prediction is equal to the observed probability. As seen, the dispersion from such ideal situation is reasonable for both models A and B in either the derivation set or the TARN external validation set. However, the calibration appears poor at some points in the IMPACT external validation set. We compared the IMPACT dataset with the derivation dataset across various patients characteristics and observed that the two datasets are significantly different as per every covariate in the model such as age, GCS, pupillary reactivity etc. apart from the survival rate.

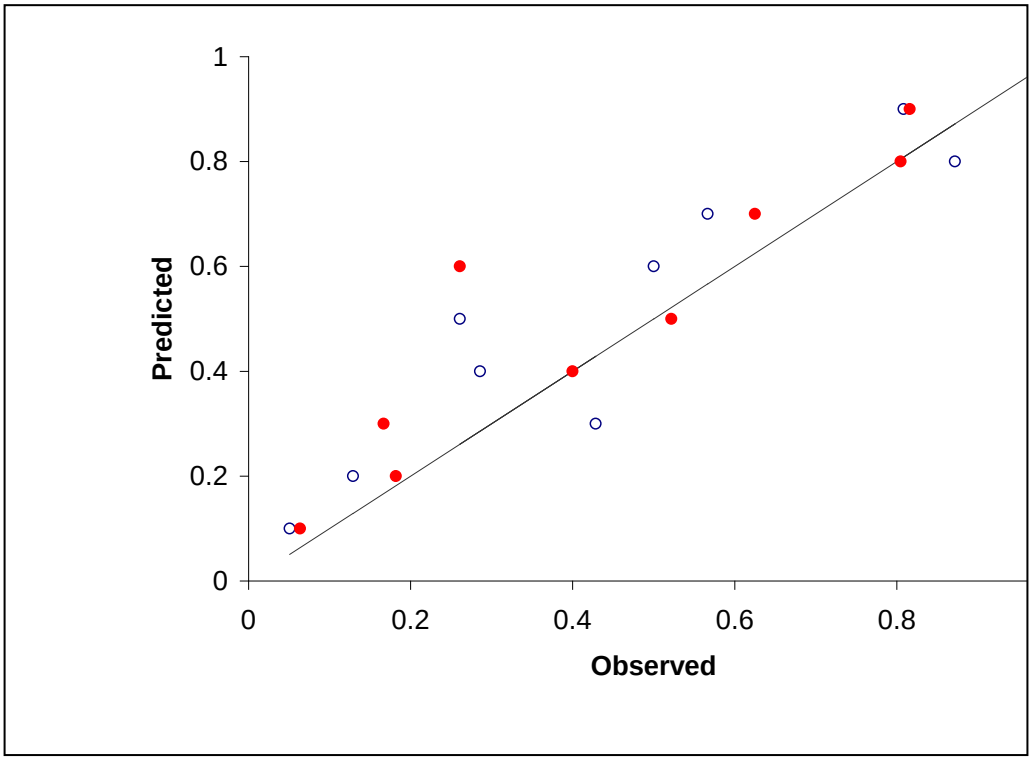


Figure 15 Calibration plot of Models A (○) and B (●) in the TARN derivation dataset

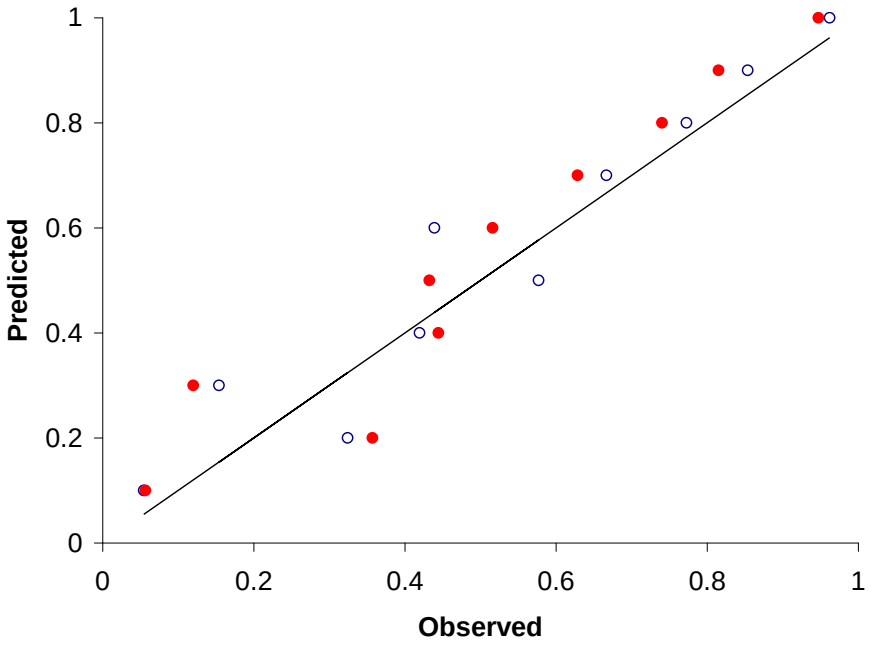


Figure 16 Calibration plot of Models A (○) and B (●) in the TARN external validation dataset.

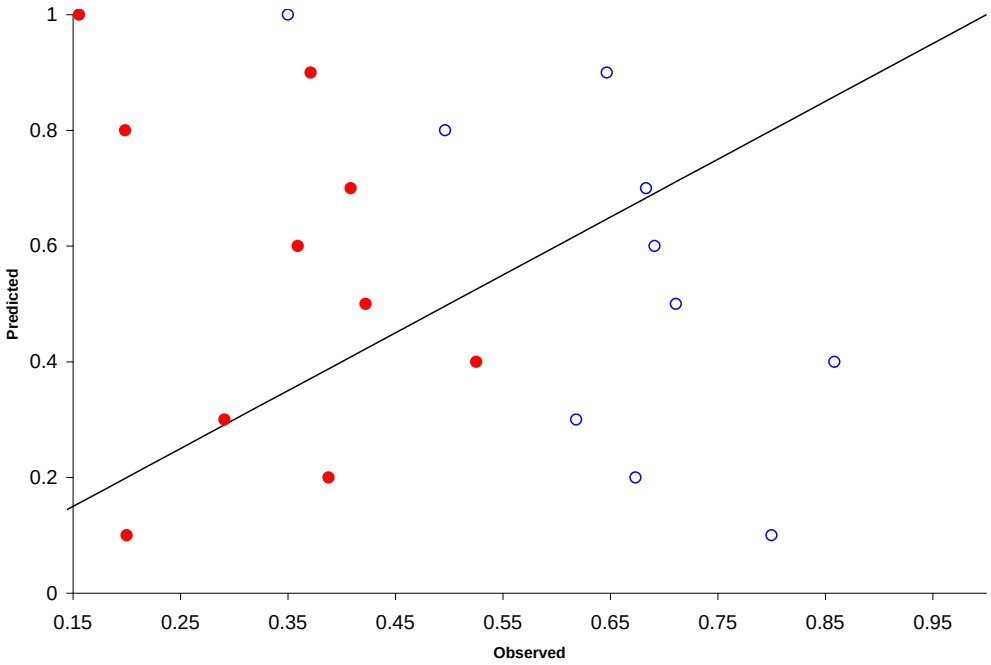


Figure 17 Calibration plot of Model AB(○: favourable outcome prediction, ●: survival prediction) in the IMPACT external validation dataset.

7.5. Discussion

Two predictive models of survival to discharge (model A and B) have been developed in this study utilizing a recent dataset of severe TBI patients, the majority whom received their trauma care in the UK. These two models share the covariates: age, GCS, pupillary reactivity, hypoxia and brain stem injury. However, model A contains ISS in contrast to model B which includes extracranial injury, brain swelling, cause of injury and its interaction with age instead. The discrimination powers (AUCs) are 0.92 (95% CI: 0.90-0.95) for model A and 0.93 (0.91-0.95) for model B. Both models have acceptable calibration per HL test (p value > 0.05). In addition to good performances of these models in terms of their discrimination and calibration, the derivation method applied in this study conforms to the criteria for a well-developed model in TBI [31, 159].

7.5.1. Limitations

We performed a selection bias analysis to compare the characteristics of patients who were excluded from the modelling due to lack of information on their pupillary reactivity to those who were included in the modelling. To this aim, the characteristics of 675 (84.2%) of cases in the dataset (included cases) who were submitted to TARN from September 2005 to Aril 2008 were compared to 1558 TBI cases who were not included in the dataset (excluded cases) during the same time period. It was observed that, these two groups do not hold statistically significant differences with respect to their gender, admission GCS, ISS and survival rate whereas it appears the excluded cases

are older than the included cases (43.4 versus 38.2). However, this difference between the ages, although being statistically significant, may not be clinically significant

One of the limitations of this study relates to the population sample which can not be a truly representative of all TBI patients. This is because the sample was extracted from TARN which requires certain submission criteria for a given trauma patient such as longer of 3 days stay at hospital or admission to intensive care unite. However, although the presented prognostic models are not applicable to every TBI patient, it may be easy to ascertain the possible applicability to a given TBI patient since the prediction of long stay at hospital or transfer to ICU can often be made on admission based on initial clinical situation or CT findings.

The further limitation of this study may be the inclusion of children in the analysis (7.5% were less than the age of 15). It has been shown the effect of age on outcome in TBI may be different in adults and children in that increasing age in children renders a better functionality than in adults [161]. In our models, age holds an odds ratio of less than one for survival and as age increases, the odds ratio decreases irrespective of the age value representing childhood or adulthood. This may suggest that the models presented here are best used for adult patients since the majority of the derivation dataset comprised adults (median: 39 with interquartile range of 22 to 58). However, it is unclear whether or not, despite the differences in functional measures, the trends in mortality due to TBI differ in children and adults as well.

7.5.2. Comparison with the literature

The AUCs of our models are higher than those of the IMPACT and CRASH models which have AUCs of respectively 0.87 and 0.88. This may primarily relate to the differences of characteristics between the IMPACT and CRASH datasets to that in this study. In the IMPACT dataset, the majority of patients had severe head injury (GCS<9), and no mild brain injuries were included. This results in a smaller spread of baseline risks which automatically decreases the AUC. Moreover, the outcome prediction in the IMPACT study was long term (6 months mortality and unfavourable outcome). The CRASH study recruited cases with a lower degree of severity with the inclusion criteria of presenting GCS of < 15 on arrival at hospital within 8 hours of injury. However, the time point for survival as the outcome measure in our study (discharge) is similar to that of the CRASH study (14 days or earlier).

With regards to application of prognostic models to benchmark the trauma care, we believe our TBI models are better than the IMPACT and CRASH models for this purpose. This is because the IMPACT and CRASH models were derived from clinical trials and hence are less likely to be a good representative of the trauma care recipients than data saved in trauma registers. On the other hand, for the purpose of stratification of TBI severity in trials, the CRASH and IMPACT models are better options as our models have been derived from an ongoing observational project. Furthermore, if one wishes to apply a prognostic model in a clinical setting, this may depend on the specific purpose of such use. Our models predict the discharge survival which is useful to provide insight on the acute care of TBI victims. During the acute course management, one important strategy can be the provision of neuro-intensive

care which has been shown to improve the outcome in TBI [129]. Using our models may facilitate balancing decisions on admission/transfer to neuro-intensive care against appropriate allocation of such resources. On the other hand, the IMAPCT and CRASH models can be used to long-term outcome prediction which, despite having value during acute care, there are of unique value for decisions on delivery of chronic care such as rehabilitative or community care.

Although ISS is contained in none of the CRASH and IMPACT prognostic models, it is one of the covariates which is contained in prognostic models for general trauma patients such as the TRISS [120] and TARN models [142]. Apart from our TBI model A, this variables is also included in the Signorini's model [93] of TBI prognosis which was reported as one of the best-developed models in the literature by the Perel's systematic review [31]. The ISS partially reflects the effect of extracranial injury on outcome in TBI and is calculated by summing the square of three highest AIS severity scores in different body regions. Thus extracranial injuries which are severe enough to attract high AIS severity scores in multiple trauma would affect ISS and thus indirectly the outcome in a model which contains ISS. This may explain why ISS and extracranial injury can not be contained in the same model as like our model A, Signorini's TBI model which included ISS, did not contain extracranial injury either.

7.5.3. Implications of the study

Among the two models, the advantage of one model over the other mainly depends on the purpose of using a prognostic model since both have similar performances. In a setting where access to trained coders for coding the

injuries are not available, model B may be a better option as it does not involve calculation of fractional polynomial transformation of ISS ($(\log_e(\frac{ISS}{10}) - 0.91$). On the other hand, trauma registries may opt for model A which contains less covariates to run specifically where missing values are an issue. Conspicuously, the selection of the appropriate model in a retrospectively accrued TBI dataset depends on the availability of values on each covariate used by the model.

We examined the importance of various intracranial pathologies from the literature in our modelling procedure and found that only brain stem injury (as in either model A or B) and brain swelling (as in model B) are significant. Although we based this piece of analysis on observations by other studies about prognostic strength of CT, unfortunately, we did not have access to the actual CT images in our registry and thus descriptions to brain injury AIS codes in the dictionary were used as substitutes for CT reports. However, CT is not the only source for AIS coding as the coding can be done based on the results of MRI, operational notes and etc. Therefore, whilst majority of information for AIS coding is obtained from CT reports, caution should be taken with regards to using the term 'CT' abnormality to refer to brain stem injury and brain swelling in our models. As such, it is important to note the diagnosis of these pathologies in our models should not exclusively depend on CT and it can be according to MRI results, intracranial operation or even merely clinical basis. Consequently, using the term 'intracranial pathology' is preferable to 'CT finding' when referring to these pathologies in our models.

We have no certain explanation as to why cause of injury and brain swelling lose their significance if the extracranial injury in Model B is replaced

with ISS in model A. This may, however, imply that extracranial injury, injury causes and brain swelling contains the same predictive strength as ISS does on its own since the performances of the two models do not significantly differ. Considering that extracranial injury influences ISS, it can be assumed that each ISS value represents partly extracranial injury with the remaining part being dependent on intracranial injury. Therefore, in reality, this 'intracranial' part of ISS may be equivalent to cause of injury and brain swelling in its predictive strength. This may lead to the conclusion that due to the direct possible effect of brain swelling on ISS, there must be some degree of correlation between cause of injury and the type of intracranial injury sustained which would affect AIS coding and subsequently ISS [94].

We have externally validated our models on two other TBI datasets from TARN and the IMPACT collaboration. Our models performance seems promising in the TARN dataset which is from a different time periods but otherwise with the same inclusion criteria (historical validation) [113]. This implies the models are valid for trauma benchmarking in Britain. However, the external validation in the IMPACT dataset does not appear satisfactory despite AUCs still being far above the cut-off of random guess (0.50). This may be due to different case-mix or time point of outcome prediction which is long term in IMPACT (6 months). The other explanation may be that our derivation dataset is more up to date.

Trauma registries use prognostic models derived for general trauma patients including patients with traumatic brain injury to benchmark trauma care systems. The most well-known of these prognostic models are the TRISS methodology [120] but they are several other models tailored for different

subsets of patients such as intubated patients [154], patients with penetrating injury [120] or children [162]. TARN has also developed its own model of prognosis to suit the UK trauma population [142]. Benchmarking of neuro-trauma care may have to be performed separately from benchmarking general trauma care through models derived from TBI populations. Furthermore, all commonly used general trauma models do not contain pupillary reactivity which in both models A and B holds a large effect on outcome (for example, the category of 'both absent' holds the largest coefficient in both models and other categories of pupillary reactivity also have high coefficients compared to coefficients of other covariates). In fact, pupillary reactivity is the only covariate which is included in the TBI models (such as our proposed TBI models and also the CRASH and the IMPACT models) but is excluded in models for general trauma patients. Furthermore, bearing in mind that GCS not always suggests brain injury (as its impairment may be due to intoxication or secondary brain damage following hypoxia or hypotension), abnormal pupillary reactivity is the clinical finding which amongst the factors included in either general trauma or TBI prognostic models, may be the most indicative of underlying brain damage. This suggests that recording the pupillary reactivity for all victims of TBI must be declared mandatory in trauma registries including TARN.

7.5.4. Future directions

Beside the historic external validation which was performed on a TARN TBI dataset from a different time period, the proposed models still require further external validation to examine their 'universal' generalisibiity. Ideally, a prognostic model should hold geographic validation (on datasets from different

localities), methodological validation (on datasets with a different study design or method of data collection), spectrum validation (on datasets with different severity of injuries) and follow-up interval validation (on datasets with different time point of outcome assessment) [113]. Although one may consider the performance of our models unsatisfactory in the IMPACT dataset to rule out various aspects of their external validity such as methodological, spectrum or follow-up interval, this drop of performance can be due to our contemporary derivation dataset as compared to the IMPACT data or different case-mix and thus our models still have to be run in other recent TBI series to examine their validity.

In this study, the coefficient of each regressor is provided which, if presented on a web-calculator could assist clinicians in quantifying the probability of survival for a TBI patient with given characteristics. Nevertheless, the safety of prognostic models in predicting the outcome for an individual patient is a controversial topic and requires further investigation. On the one hand, provision of this model may lead to withdrawal of active therapy to a patient with a poor predicted prognosis. On the other hand, with growing consensus on the various components of TBI management such as intubation or use of osmotic diuretics, the concern over the negative effect of models prognosis on appropriate therapy may be allayed especially if probability of outcome is considered along with other factors influencing TBI management. In fact, senior doctors often make predictions on the potential outcome of individual TBI patients and the role of prognostic models here may be solely to quantify this prediction.

The models presented in this study are used to predict survival at discharge. However, TBI management which includes therapeutic measures followed by rehabilitative schemes should target full recovery to the same level of physical and mental health as prior to the occurrence of injury and not only survival. Regarding this, in TBI management, knowledge on the probability of survival is not the only required factor in terms of prognosis since it is also important to know the risk of disability in survivals. Derivation of models to assist with prediction on outcomes apart from survival on a dataset from more homogeneous trauma care systems/policies is a matter of future research.

7.5.5. Conclusion

Two well-constructed prognostic models have been derived and internally validated which can be used to predict the survival of severe TBI patients based on simple clinical characteristics particularly applicable to British trauma care system.

Acknowledgement

We would like to thank all TARN participating hospitals

Conflict of interest

None

PROGNOSIS IN TRAUMATIC BRAIN INJURY (TBI)

A thesis submitted to the University of Manchester for the degree of the Doctorate of Philosophy (PhD): Medicine in the Faculty of Medical and Human Sciences

2011

By:
Mehdi Moazzez Lesko

School of Medicine

Volume II of II

Table of contents

List of tables

List of figures

7.6 Expansion on methods

The main stages of model development as presented in the paper were covariates selection, dataset selection, selection bias analysis, univariate analysis, model derivation (i.e. manual stepwise logistic regression), imputation of missing information and model validation (internal and external). Some further details about performance of these steps follow. Moreover, not all of these stages were performed independently or in the above-mentioned sequence. For example, the imputation of missing information was performed during the model derivation.

7.5.6. Dataset selection

This was performed in two steps

Step I

The format of recording data in TARN is such that each hospital holds a distinct `*submission* to TARN. This means if a patient has been transferred from the first admitting hospital to a second hospital or further to a third hospital, he/she holds a number of TARN submissions equal to the number of attended hospital (assuming all attended hospitals submit data to TARN as otherwise the attended non TARN hospital is missing). Therefore, TARN saves information of *submissions* and not cases. In this first step of dataset selections, all *submissions* which met the criteria (i.e. brain injury AIS score of 3 or more plus basal and compound/depressed/open skull fracture AND also availability of pupillary reactivity at any time point (at scene, en route, at Emergency

Department (ED))) were extracted from the TARN 'mother' dataset through running the appropriate syntax in Microsoft SQL Server. This procedure returned 1009 *submissions*.

Step II

This step involved firstly matching all submissions which belonged to one patient (case) and then to exclude those patients who did not have submission either from the first attended hospital or the last hospital from which the patient was discharged. The first submission was required for the record of admission variables such as GCS, pupillary reactivity, blood pressure etc. and the last submission was required for the record of the discharge survival. After matching procedure, there were 735 cases which had not been transferred and thus information on admission and discharge was available. Of the remaining 274 transferred cases, only 67 cases had all the information available which meant 207 (274 - 67) cases did not have information from either the first hospital or the last hospital. To sum up, the final dataset contained 802 cases of TBI.

7.5.7. Univariate analysis

This stage involved the following analyses:

Describing the distribution of each variable: the one-sample Kolmogorov-Smirnov test was used to test whether the values of continuous variables were normally distributed or not. The p value of this test for all continuous variables was less than 0.05 (i.e. indicating non-normal distribution) and thus the median and the interqaurtile range were calculated to describe the distribution of each

covariate. For categorical variables, the frequency of each category was calculated.

Describing distribution of each variable in survivors and non survivors: the Mann Whitney U test was used for this purpose to compare the median and IQR of continuous variables and Chi Squared test was used for categorical variables. The logistic regression was also run to calculate the unadjusted odds ratio for survival for each covariate.

Fractional polynomial analysis: this was performed to assess the linear relationship of the continuous covariates with $log_e\ (odds\ of\ outcome)$ and, in case of non-linearity, to identify the appropriate power transformations. This refers to the linearity assumption by the logistic regression. If the continuous variable does not demonstrate this linearity (between the variable and $log_e\ (odds\ of\ outcome)$), then it has to be categorised or transformed. Fractional polynomial transformation is the power transformation of variables out of 7 power candidates of -3, -2, -1, 0, 1, 2, 3. [135]. This method is used to transform those continuous variables which do not meet the linearity assumption. Fractional polynomial analysis has to be performed in Stata since SPSS does not have this feature. The output of this analysis in Stata can be used not only to assess various transformations of the variable but also to assess the linearity assumption in the first place. The output of Stata shows a table containing three models: *the linear model, the first degree model, and the second degree model*. The linear model includes the continuous variable as it is (i.e. with no transformations or power transformation 1). The first degree model includes the variable with only one power transformation and the second degree model is with 2 power transformations. Based on the changes in the

deviance, one model has to be selected which demonstrates what power transformations should be used. Selection of the model is based on the significance of 'gain' in the deviance as compared to the previous model. For example, if the change (or gain) in the deviance is significant in the first degree model as to the linear model (the model at the previous stage), then this means one power transformation should be applied to meet the linearity assumption. The power transformation is presented for each model derived in Stata. However, the final transformation(s) is the power transformation as suggested by the selected model plus/minus a constant which is provided separately by Stata. This implies that the formula to transform the continuous variable takes a different form to merely being a power transformation. For example, the GCS transformation takes the following form:.

$$\left(\frac{10}{GCS+1}\right)^2 - 0.76$$

$$\left(\frac{10}{GCS+1}\right)^2 \times \log_e\left(\frac{GCS+1}{10}\right) - 1.02$$

The above formula clearly show that the power transformation is not the only mathematical calculation. For example, the first formula contained a constant (-0.76) alongside the power transformation of 2.

Examples of the type of tables given by Stata to select the model of transformation presented later in this section with respect to each continuous variable analysed (Table 36 , Table 39, Table 41, Table 42 and Table 43).

Furthermore, there is a way to visualise the results of Fractional polynomial analysis in a graph. This graph is drawn to assess how close the

observed probabilities were to the predicted probability for various values of the continuous variables (a discussion on predicted versus observed probablity is presented in Paper 1; section 2.7). For this purpose, the logistic regression is run on the dataset with the transformation indentified by fractional polynomials as covariates and the survival as outcome. The calculated predicted probability for each case in the dataset was saved (SPSS has this feature to save this probability in front of each case included in the data). The observed probability is also calculated for each value of the continuous variable. For example various cases with the same age of 23 may experience different outcomes as one may die and the other may survive. As such, the observed probability for each value of age is the frequency of cases who experience the outcome. The final graph can be drawn with x axis being the continuous variable and y axis being the probability. In this section, various examples of this type of graph are presented with respect to each continuous variable analysed. (Figure 18, Figure 19, Figure 20, Figure 21, Figure 22, Figure 23, Figure 24, Figure 25 and Figure 26). The constant line represents the predicted probability with circles (○) standing for the observed probability. It is expected that the circles (i.e. the predicted probabilities obtained through fractional polynomial transformation) follow the line (observed probability). This will be referred to as whether these two follow the same pattern or not. The image which shows the pattern of the predicted and observed probabilities visualises whether or not the fractional polynomial transformation meets the linearity assumption satisfactorily. A perfect transformation is when the predicted and observed probabilities (y axis) follow absolutely the same pattern as the continuous variable changes (x axis).

For sake of presentation, the *y* axis of the graph was scaled using the logit of the observed/predicted probability (i.e. Log (probability/(1-probability)). This, however, caused a problem when the probability was 1 or 0. In order to rectify this problem, the frequency of the cases who were all dead or all alive were taken into account by using the following formula: 0.5 – frequency/ frequency. This calculation is referred to as 'correction for the frequency'.

Age

Table 36 shows the results of fractional polynomial analysis for age. According to this table, the linear model had p value of 0.00. This significant p value for the linear model indicates there is no need for transformation or the age meets the linearity assumption as it is. Figure 18 and Figure 19 demonstrate the plot of predicted and observed probability against age (with no transformations) before and after correction for the frequency. Overall, predicted and observed probability followed the same pattern apart from some cases who were less than the age of 20 or around the age of 60 which this has been corrected in Figure 20 following the correction for frequency.

	df	*Deviance*	*Gain*	*P*	*Powers*
Not in the model	0	907.437			
Linear	1	882.981	0.000	0.000	1
First degree	2	879.814	3.167	0.075	3
Second degree	4	878.215	4.766	0.450	-2 3

Table 36 Comparison of models of fractional polynomial transformations for age

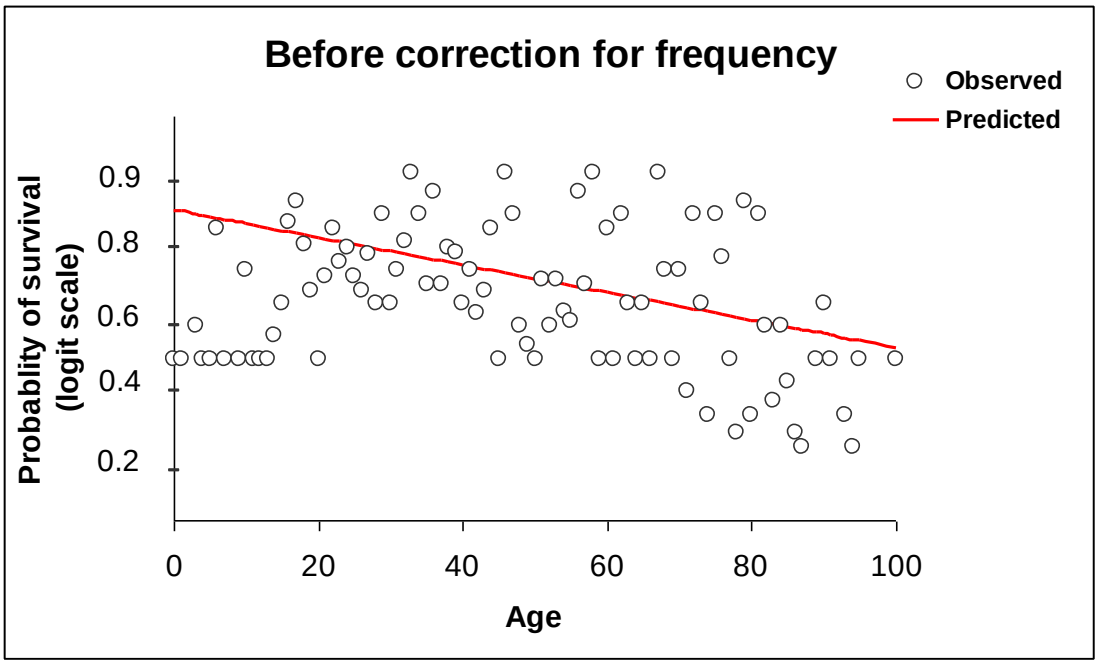


Figure 18 The plot of age against predicted and observed probability without transformation before correction for the frequency.

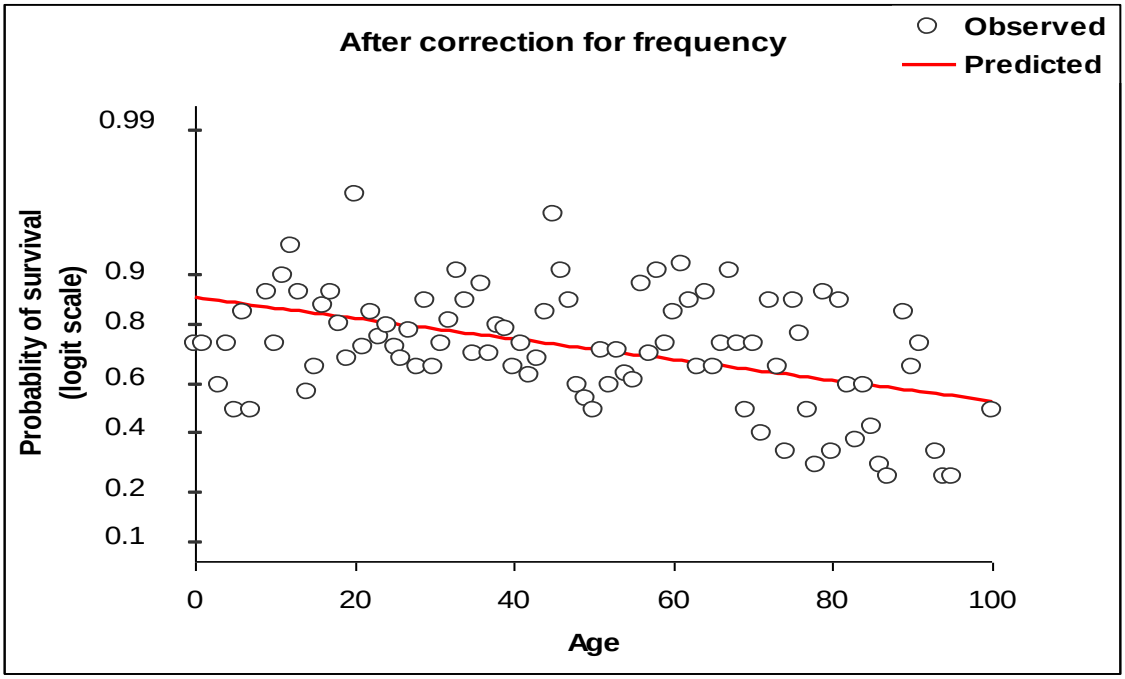


Figure 19 The plot of age against predicted and observed probability without transformation after correction for the frequency.

Cause of injury

TARN does not hold a distinct variable with this name. Instead it has two other variables with the names of ‘mechanism of injury’ and ‘level of intent’ which were used to generate the cause of injury in this study. Categories included under level of intent in TARN were: *non-intentional, alleged assault, suspected child abuse, suspected self-harm, sports, suspected high risk behaviour, alleged intent (non-assault) and others*. Injury mechanism included *vehicle incident/collision, fall more than 2 meters, fall less than 2 meters, shooting, stabbing, blast, blow, and burn, skeletal/organ/vessel destruction and others*.

The categories to include in the cause of injury were selected from the IMPACT study which contained *RTC, work-related, assault, sports/recreation and others*. There were two major differences between TARN formatting of cause of injury and that in IMPACT. This related to work-related injuries which are not recorded in TARN but recorded in IMPACT and ‘fall’ which is recorded in TARN but not in IMPACT. Thus, the final categories to be included as cause of injuries were RTC, fall, assaults, sports and others (work-related was discarded from the IMPACT categorisation and instead fall from the TARN categorisation was added).

With the above modified IMPACT categorization, a strategy was required to allocate a cause of injury (out of RTC, fall, assaults, sports and others) to each case of TBI in the dataset from the information available on “injury mechanism” and “injury intent”. Table 37 demonstrates the cross-tabulation of ‘injury intent’ with ‘injury mechanism’. The strategy to allocate an injury cause to each case based on this cross-tabulation was decided as follows:

		Injury Mechanism				**Total**
		RTC	***Fall***	***assault***	***others***	
Injury Intent	**assault**	3	11	97	3	114
	sports	4	8	0	2	14
	others	314	313	29	18	674
Total		321	332	126	23	802

Table 37 The cross-tabulation of injury mechanism and injury intent.

All patients who have their level of intent recorded as sports should go under ‘sports’ regardless of their injury mechanism. Similarly, all patients who have their injury mechanism recorded as assault or level of intent as assault should go under ‘assault. This means, for instance, if the patient has RTC under injury mechanism and assault under injury intent, assault was allocated irrespective of RTC as injury mechanism (there were overall 3 cases with this combination). With regards to RTC (as a cause of injury), as seen in Table 37, there were 321 cases with this injury mechanism. Of these, 3 were assault and 4 were sports with remaining 314 being “others” under injury intent. To exclude the assaults and sports patients, RTC needed to be defined as when injury mechanism is RTC and level of intent is “others”. This was because patients with injury intent of assault or sports were respectively allocated to assault and sports. The same thing applies for falls. Table 38 summarises how each category of cause of injury were defined from injury mechanism and injury intent.

Cause of injury	*Definition*
RTC	injury mechanism = RTC **AND** level of intent = others
Fall	injury mechanism = Fall **AND** level of intent = others
Assault	injury intent = assault **OR** level of intent = assault
Sports	Level of intent = sports
Others	remaining cases

Table 38 Allocation of injury cause based on injury mechanism and level of intent.

GCS

The IMPACT models use motor GCS instead of total GCS whereas the CRASH models use total GCS. The analysis presented in Paper 3 was performed to assess whether total GCS is preferable or its motor component. The results demonstrated that these two do not have a significant difference in outcome prediction. Furthermore, in the TARN dataset, the admission GCS score had only 26 missing values where this was much more for motor GCS (246 missing cases). Regarding this and the same predictive strength of total GCS to motor GCS, the decision was made to include the total GCS rather than motor GCS in the modelling.

Total GCS was treated as both continuous and categorical. With the cut-offs of 12 and 9 (inclusive), this variable was categorised into three categories of mild, moderate and severe brain injury. For the continuous form, Table 39 shows the result of its fractional polynomials transformation. According to this table, the second degree model provided significant gain in the deviance (as determined by the P value) compared to the first degree and the linear model. The column representing gain is in fact the subtraction of the

deviance of the respective model from the previous model. The output of STATA contained the following transformations for GCS: $\left(\frac{10}{GCS+1}\right)^2 - 0.76$ and $\left(\frac{10}{GCS+1}\right)^2 \times \log_e\left(\frac{GCS+1}{10}\right) - 1.02$. Figure 20 shows the observed and predicted probabilities follow reasonably similar patterns with these transformations.

	df	*Deviance*	*Gain*	*P*	*Powers*
Not in the model	0	876.751			
Linear	1	704.556	0.000	0.000	1
First degree	2	684.970	19.586	0.000	0
Second degree	4	650.716	53.839	0.000	-2 -2

Table 39. Comparison of models of fractional polynomial transformations for GCS.

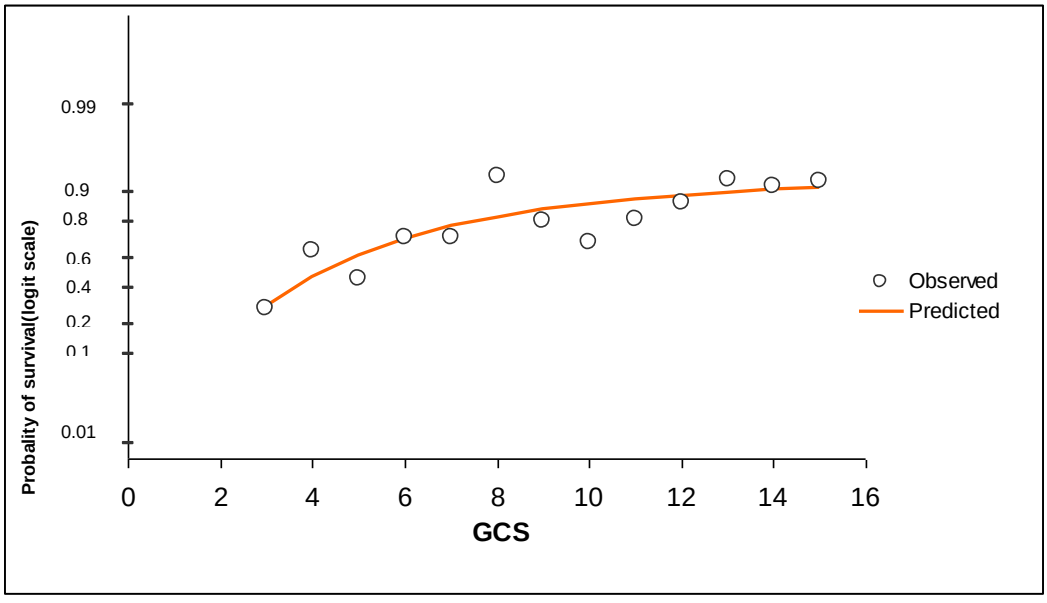


Figure 20 The plot of predicted and observed probability of survival without correction for the frequency.

Pupillary reactivity.

TARN records pupillary reactivity for each individual right and left eye as brisk, sluggish, absent or otherwise missing. A new variable was created which combined the records of each eye. Table 40 shows the frequency of each category following combining the right and left eye recordings. Chi square test for this pupillary reactivity categorisation demonstrated a significant difference between survivors and non-survivors ($P < 0.0005$).

	Total	***Survival***	***Non-survival***
Brisk-brisk	446(55.6%)	405(82.5%)	41(25.8%)
Sluggish-sluggish	53(6.6%)	36(7.3%)	17(10.7%)
Brisk-sluggish	16(2%)	10(2%)	6(3.8%)
None-brisk	16(2%)	15(3.1%)	1(0.6%)
None-sluggish	12(1.5%)	2(0.4%)	10(6.3%)
None-none	107(13.3%)	23(4.7%)	84(52.8%)
Missing	152(19%)	-	-

Table 40. Frequency of various categories of pupillary reactivity after combination of right and left eyes in survivals and non-survivals.

ISS

The Injury Severity Score (ISS) is a measure on an ordinal scale of anatomical severity of injury. It is the sum of the squared of the 3 highest AIS severity scores allocated to the patient. Table 41 shows the results of fractional polynomial analysis. According to this table, the first degree model is the best model since it provided a significant gain over the linear model. The formula for this presentation as given in the STATA output was $\log_e\left(\frac{ISS}{10}\right) - 0.91$

Figure 21 demonstrates the plot of ISS with observed and predicted probability without correction for frequency. According to this graph, some ISS values (just below 20) have similar observed probabilities whereas their predicted probabilities differ. However, this is not the case after the correction

for frequency was performed as depicted in Figure 20. According to this figure, observed and predicted probabilities follow almost the same pattern in an acceptable way.

	df	*Deviance*	*Gain*	*P*	*Powers*
Not in the model	0	907.437			
Linear	1	802.639	0.000	0.000	1
First degree	2	794.982	7.657	0.006	0
Second degree	4	793.691	8.948	0.524	-2 -1

Table 41 Comparison of models of fractional polynomial transformations for ISS.

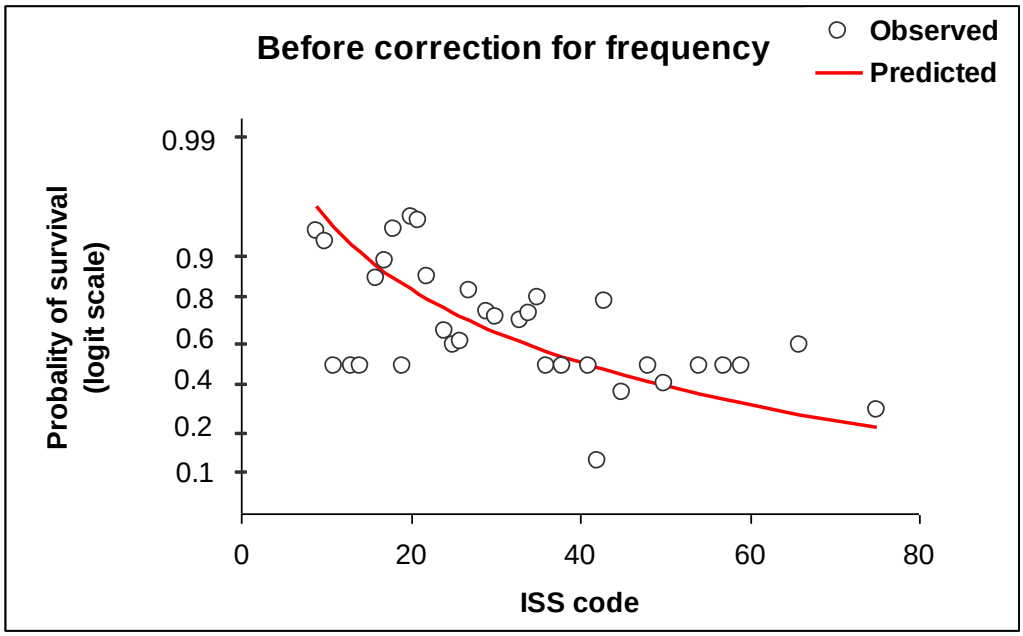


Figure 21 The plot of predicted and observed probability of survival before correction for the frequency.

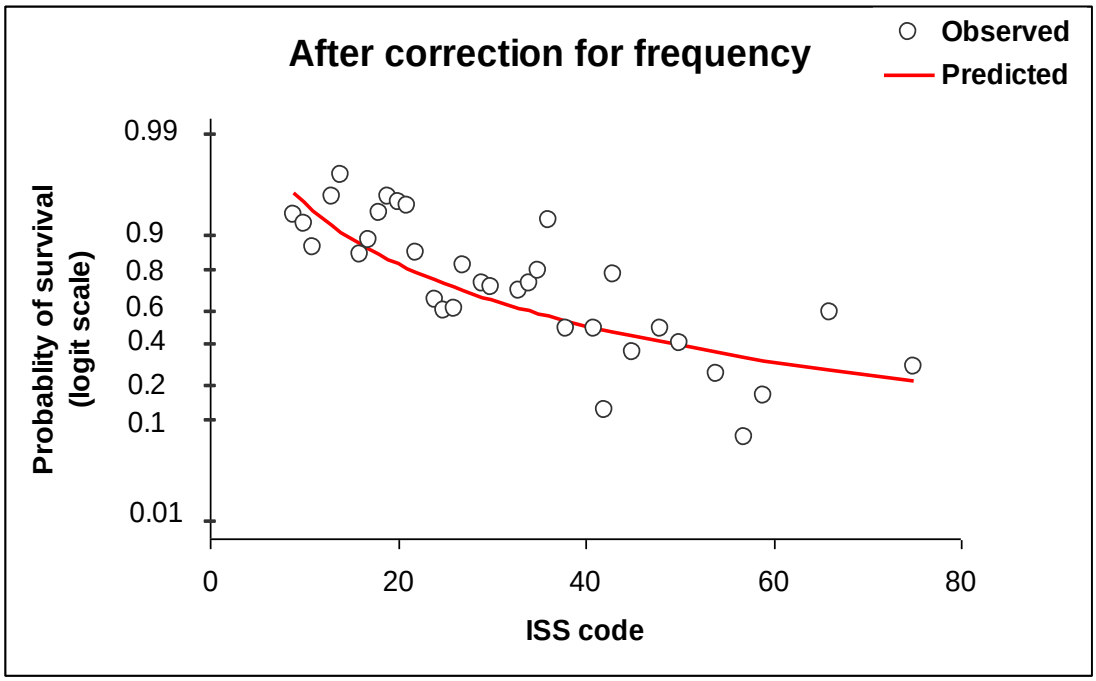


Figure 22 The plot of predicted and observed probability of survival after correction for frequency.

Extracranial injury

In TARN, AIS codes and severities are not recorded together under the same variables. This means there are a number of variables representing AIS codes with some others representing severity scores. AIS codes are saved under C (C1 to C21) with AIS severities under S (S1 to S21). For example, if a patient receives AIS codes and scores of 7164006.3 and 630299.2, he/she holds C1 and C2 (AIS codes) recorded as 7164006 and 630299 with S1 and S2 (severity scores) recoded as 3 and 2. The ending number of C and S is used to match the AIS code with the severity score. As in the above example, C2 and S2 represent 630299.2.

To create the variable representing the extracranial injury, a new nominal variable was created to count the number of cases with extracranial injury AIS codes, whose AIS score was above 3. To do this, the syntax was used which first searched C1 to C21 to spot extracranial AIS codes and then subsequently checked the matched AIS score (i.e. S1 to S21) for scores above 3.

Using this cut-off, the frequency of extracranial injury was 277 (34.5%).

Systolic blood pressure

This variable was treated both continuously and categorically with the cut-offs for categorisation being 120 and 150 from the IMPACT [95]. Table 42 displays the results of fractional polynomial analysis of systolic blood pressure. As seen, the second degree model provides the highest significant gain in deviance

as compared to the linear model. The formula for the transformation in the output of STATA was $SysBP1 = \left(\frac{100}{SysBP + 1}\right) - 0.731$ and $SysBP2 = \left(\frac{100}{SysBP + 1}\right) \times \ln\left(\frac{100}{SysBP + 1}\right) - 0.229$. Figure 23 and Figure 24 display the plot of systolic blood pressure with observed and predicted probabilities respectively before and after correction for the frequently. Even after the correction for the frequency (Figure 24), the patters of distribution of the predicted and the observed probabilities do not appear to be satisfactorily the same.

	df	***Deviance***	***Gain***	***P***	***Powers***
Not in model	0	555.333			
Linear	1	546.005	0.000	0.000	1
First degree	2	518.632	27.373	0.000	-2
Second degree	4	496.449	49.556	0.000	-1 -1

Table 42 Comparison of models of fractional polynomial transformations for systolic blood pressure.

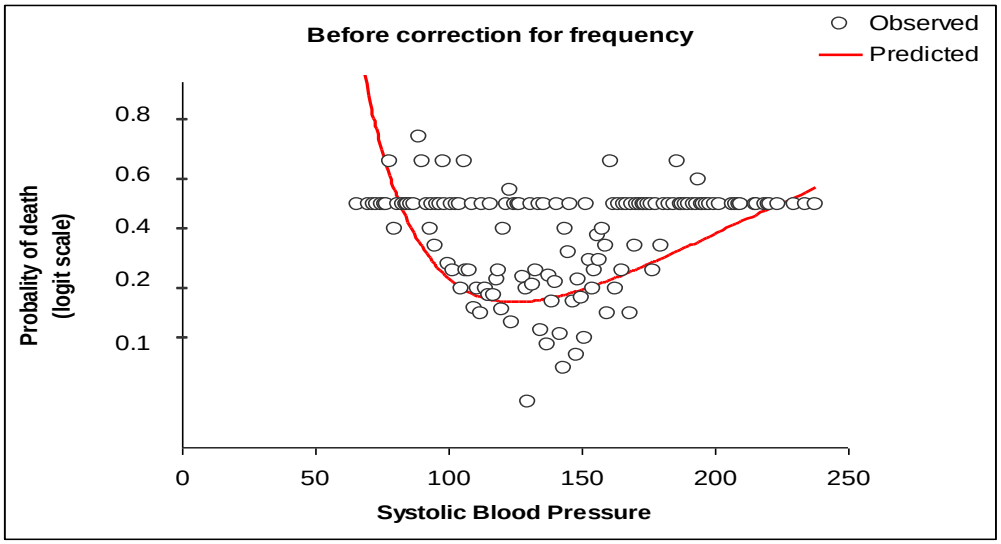


Figure 23 Plot of predicted and observed probability of survival for systolic blood pressure before correction for frequency.

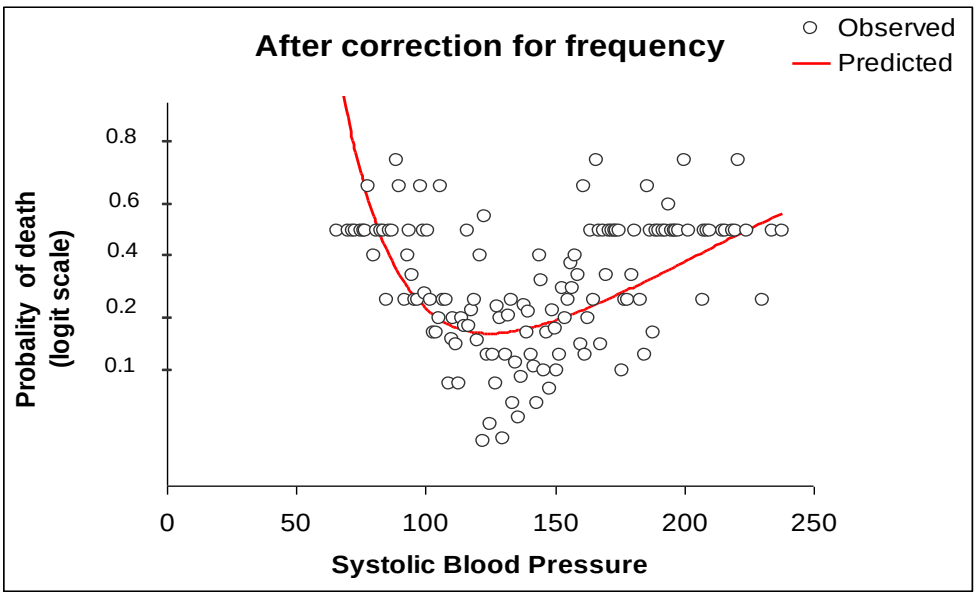


Figure 24 Plot of predicted and observed probability of survival for systolic blood pressure after correction for frequency.

Mean blood pressure

This variable was also treated both continuously and categorically with the cut-offs for categorisation being 85 and 110 from the IMAPCT [95]. Table 43 displays the results of fractional polynomials analysis of mean blood pressure. According to this analysis, the second degree model held the best gain according to the p value in the deviance as compared to the linear model. The formula for the transformation in the output of the STATA was $MeanBP1 = \sqrt{\frac{MeanBP + 0.333}{100}} - 0.981$ and $MeanBP2 = \sqrt{\frac{MeanBP + 0.333}{100}} \times \ln\left(\frac{MeanBP + 0.333}{100}\right) + 0.386$. Figure 25 and Figure 26 show the plot of mean blood pressure with predicted and observed probability of survival. Similar to systolic blood pressure, even after correction for the frequency, the patters of distribution of the predicted and the observed probabilities do not appear to be satisfactorily the same. Furthermore, mean blood pressure demonstrated a significant association with outcome by logistic regression whereas Mann Witney U test did not find any significant difference in mean blood pressure between survivors and non-survivors.

	df	***Deviance***	***Gain***	***P***	***Powers***
Not in model	0	544.192			
Linear	1	538.157	0.000	0.014	1
First degree	2	520.952	17.206	0.000	-2
Second degree	4	511.815	26.342	0.010	0.5 0.5

Table 43 Comparison of models of fractional polynomial transformations for mean blood pressure.

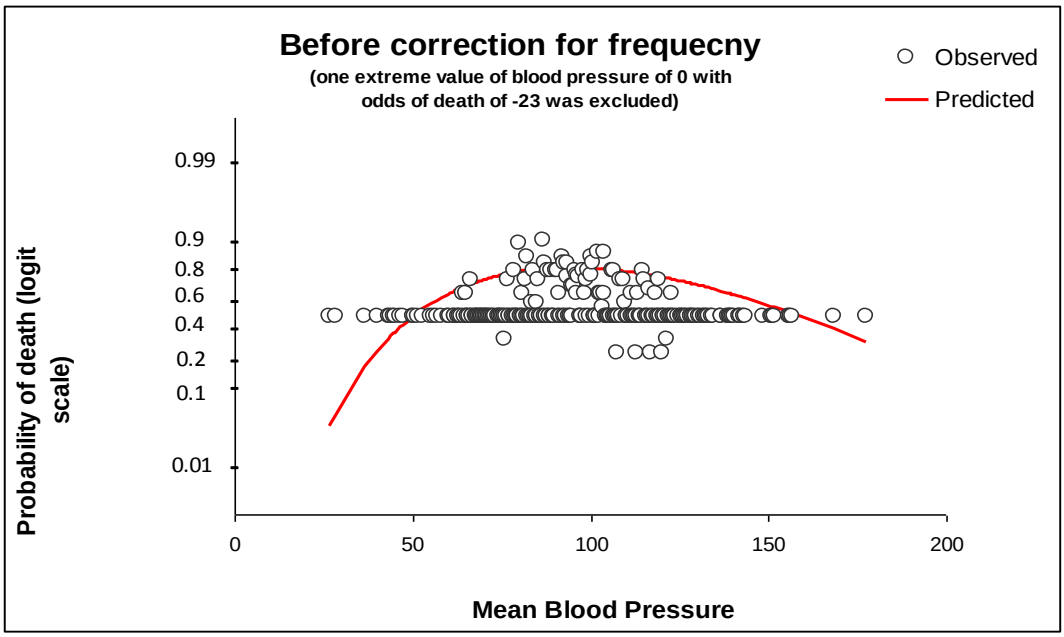


Figure 25 Plot of predicted and observed probability of daeth for mean blood pressure before correction for frequency.

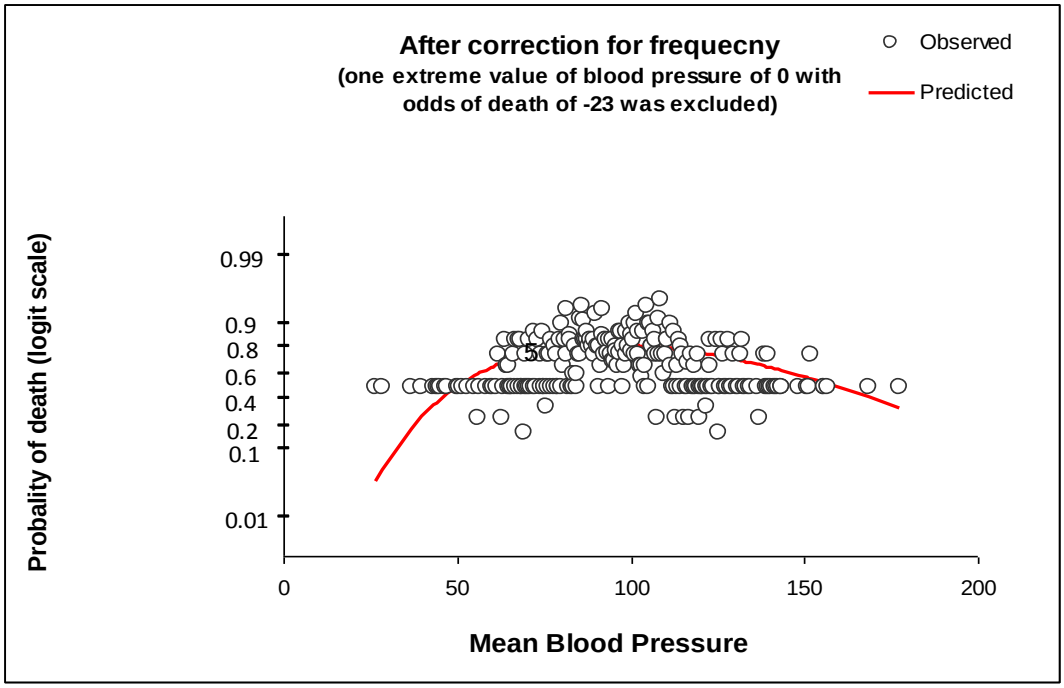


Figure 26 Plot of predicted and observed probability of daeth for mean blood pressure after correction for frequency.

AIS scores, the Marshall Classification and various intracranial pathologies

AIS severity scores are saved in TARN under S1 to S21. Moreover, 34.4% of cases had more than 3 brain injury AIS codes/scores allocated to them. Due to this, a syntax was written to search all brain injury AIS scores for the highest one.

TARN does not have record of CT images/reports and thus AIS descriptions in the AIS dictionary were used for the Marshall Class allocation. The method applied is presented in Paper 4.

Similarly, AIS codes were used to define various types of intracranial pathologies. The pathologies investigated are presented in Paper 5 (cerebral contusion, cerebral brain swelling, cerebral haemorrhage, cerebral EDH, cerebral SAH, brain stem injury and cerebellar injury). Table 44 provides AIS codes allocated to each intracranial pathology. Apart from SAH, all injuries have more than one code which represents taking severity of injury based on size or being unilateral versus bilateral into account. However, haemorrhage (Not Specified: NS) encompasses cases with unknown type of cerebral haemorrhage or intracerebral haemorrhage. Furthermore, the brain stem and cerebellar injuries encompass various types of injuries including contusion, haemorrhage, infarction etc.

Brain injury	*AIS codes*
Contusion	140204,140402,140403,140404,140405,140602,140604,140606,140608,140610,140612,140614, 140616,140618,140611,140620,140622, 140624,140626
Brain swelling	140662, 140664, 140666, 40660, 140628
Epidural haemorrhage	140414,140418,140422,140630,140632,140634,140636
Subdural haemorrhage	140438,140442,140446,140650,140652,140654,140656
SAH	140684
Haemorrhage (Not Specified: NS)	140629, 140638,140640,140642,140644,140646,140648,140678,140686
Penetrating injury	140216, 140478, 140690
Brain stem injury	140202,140204,140206,140208,140210,140212,140214,140216, 140218,140299
Cerebellum injury	140402,140403,140404,140405,140406,140410,140414,140418,140422,140426,140430,140434,140438, 140442,140446,140450,140458,140462,140466,140470,140474,140478,140499

Table 44 Various AIS codes allocated to each intracranial pathology.

7.5.8. Model derivation

The overall stages of model derivation are described in the Paper 6. What follows is more detailed explanation of the modelling procedure. In order to avoid confusion with the 'stepwise' logistic regression, each step is referred to as 'level'. In each level one model or more were derived for the purpose of that very level and the observations obtained during performance of that level are provided. The model(s) derived is (are) presented in the Appendix labelled according to the relevant level.

Some levels involved adding of new variable(s) to the model derived at the previous level and observing the significance of change in the model deviance, in AUC and in the p value of the HL statistics. Multicolinearity among variables was also examined in the correlation matrix given by the output of SPSS at each level to avoid containment of two variables with this effect in the same model (*multicolinearity effect:* when two variables included in the same model are associated with each other. This should not occur and one of the two variables should be excluded from the model. SPSS output contains a table with the name of the Correlation Matrix which gives the degree of the associations among all the variables included.). Furthermore, based on the literature, the interaction between variables was specified in the SPSS syntax and the significance of the interaction was examined in the model output. In order for a variable to stay in the model, that variable had to decrease the deviance significantly and demonstrated significant association with survival with an acceptable significance and direction of odds ratio. Failure to meet any of these three requirements resulted in the removal of the variable.

It should be noted that although in the Paper 6 the imputation is described as a separate stage for the sake of presentation suitable for a journal article, this was in fact performed during model derivation. Furthermore, model derivation is a ‘try and error’ procedure with regards to how a covariate may be included in the model (i.e. continuous or categorical, if categorical with what/how many categories), the order in which variables are included (i.e. which one is the first, the second, and the last) and the order of the levels of modelling. Although an overall strategy of model derivation was decided prior to modelling, however, ‘adaptations’ were necessary due to the given observations obtained during the progress of the procedure. This section describes what prospectively occurred during the levels of the modelling. A perspective on the approach taken below in terms of efficiency and effectiveness is given in the main thesis discussion.

The model derivation was performed over 15 levels and table 45 presents the title of each level along with the main observation(s) obtained at that very level. The title of the level somewhat shows the main aim of that level which is based on either the observation(s) on the previous level or the progress of the modelling procedure by adding/considering new variables. In fact Table 45 can reflect on the inherent ‘try and error’ nature of modelling by demonstrating the lack of a clear coherence in the procedure. This is because some levels could have been done prior or following to their current position or some repetitions could have been avoided. For example, following performing 8 levels, the various cut-offs for extracranial injury (extracranial AIS severity score) were explored. This level could have been done at a much earlier level such as level 3. The current position of this level then resulted in reconstruction

of all models which had been derived up to level VIII (i.e. unnecessary repetition in retrospect). The following section supplies the way the model derivation was performed with respect to Table 45. Table 57 situated after this explanation provides main points and decisions of the modelling procedure important for understanding the way the final models were proposed and the levels which addressed these points.

Level	*Title*	*Observation(s)*
I	Automatic stepwise modelling of age, continuous GCS, pupillary reactivity, ISS and extracranial injury	• One of GCS transformations was discarded by SPSS which was not acceptable.
II	Manual stepwise modelling of age, continuous GCS, pupillary reactivity, ISS and extracranial injury	• Pupillary reactivity: absent-brisk had odds ratio of above one for survival (not acceptable) • ISS and extracranial injury can not be included in the same model due to multicolinearity effect.
III	New categorisation of pupillary reactivity/decision to employ ISS or extracranial injury	• Pupillary reactivity with three categories of both reactive, only one reactive and none reactive is better. • Extracranial injury (extracranial AIS score > 2) did not have a significant association with outcome in the model.
IV	Model IIIA including categorical GCS (12 categories)	• GCS scores of 9, 4 and 3 had odds ratio of above one for survival (not acceptable)
V	Model of level IIIA including new categorisation of GCS (3 categories)	• GCS with 3 categories had significant association with outcome.
VI	Testing the value of cause of injury	• Cause of injury is significant only when GCS is continuous.

Table 45 Various levels of model derivation

Level	*Title*	*Observation(s)*
VIII	Trial of the new categorisation of extracranial injury	• Extracranial injury (extracranial AIS > 3) can be included in the model without ISS (with GCS either continuous or categorical).
IX	Imputation of missing GCS and pupillary reactivity	• All missing GCSs were imputed with en-route and then at scene scores without significant changes in the model performance. • All missing pupillary reactivities were imputed with en-route and then at scene scores without significant changes in the model performance.
X	Inclusion of intracranial pathology	• AIS scores were not selected as not all of them demonstrated significant association with outcome. • The Marshall Classification was not selected as not all classes demonstrated significant association with outcome. • SAH, EDH and haemorrhage did not show significant association with outcome. • Brain swelling and brain stem injury demonstrated significant association with outcome.

Table 45 Various levels of model derivation (continued)

Level	*Title*	*Observation(s)*
XII	Adding systolic blood pressure, mean blood pressure and hypoxia	• Neither systolic blood pressure nor mean blood pressure were significantly associated with outcome (with 195 missing cases in total) • Hypoxia was significantly associated with outcome.
XIII	Analysis of missingness/imputation of mean and systolic blood pressure and hypoxia	• On univariate analysis, cases with missing systolic blood pressure tended to be younger, have lower GCS, more frequency of brain stem injury and lower survival rate. • On multivariate analysis, age and cause of injury significantly affected the missingness of systolic blood pressure. • On univariate analysis, cases with missing mean blood pressure tended to be younger, have lower GCS, lower frequency of extracranial injury and brain stem injury and lower survival rate. • On multivariate analysis, only age had significant effect on missingness of mean blood pressure.

Table 45 Various levels of models derivation *(continued)*.

Level	*Title*	*Observation(s)*
XIII ***(continued from the previous page)***	Analysis of missingness/imputation of mean and systolic blood pressure and hypoxia	• On univariate analysis, cases with missing O2 Sat. tended to have lower GCS, higher frequency of extracranial injury and brain stem injury and lower survival rate. • On multivariate analysis, only GCS had a significant effect on missingness of O2 Sat.. • All missing cases with systolic or mean blood pressure or O2 Sat. were placed in a separate category as missing.
XIV	Trial of systolic and mean blood pressure individually/2 versus 3 categories	• Hypertension (either as per systolic or mean) was not significantly associated with outcome. • The model with mean blood pressure did not show acceptable goodness of fit.
XV	Assessment of interactions	• Among possible interactions of age with cause of injury and systolic blood pressure with hypoxia, only age with cause of injury was significant.

Table 45 Various levels of model derivation *(continued)*

7.5.9. Level I: Automatic stepwise modelling of age, continuous GCS, pupillary reactivity, ISS and extracranial injury

We selected the variables from the literature that have the most consensus for being of higher prognostic value in TBI. At this first level age, continuous GCS, pupillary reactivity, ISS and extracranial injury were included in the automatic stepwise logistic regression. Pupillary reactivity had 6 categories of brisk-brisk, brisk-sluggish, sluggish-sluggish, sluggish –none, brisk-none, none-none. GCS was in its continuous form with two fractional polynomials transformation of $\left(\frac{10}{GCS+1}\right)^2 - 0.76$ and $\left(\frac{10}{GCS+1}\right)^2 \times \log_e\left(\frac{GCS+1}{10}\right) - 1.02$. Similarly, ISS was transformed into $\log_e\left(\frac{ISS}{10}\right) - 0.91$. These variables are included in the basic IMPACT/CRASH models. However, ISS is not present in these models but is in another well-developed model [93].

The final model discarded one of the GCS transformations during the automatic procedure whereas practically if one transformation is discarded, the other should not be retained. However, the 'automatic' nature of this level offered no control to keep both transformations in the model. Thus the next level was performed to run logistic regression in the 'manual stepwise' method.

7.5.10. Level II: manual stepwise modelling

At this level, a model was initially derived with age (step 1 model). Subsequently four further models were derived by adding continuous GCS

(two fractional polynomial transformations) to “model step 1” to construct ‘model step 2’. Similarly pupillary reactivity (with 6 categories) was added to ‘model step 2’ to construct ‘model step 3’. This process was continued by adding ISS and extracranial injury to the models constructed at the previous steps (Table 46). The deviance of each constructed model at each step was spotted in the SPSS output and the significance of the difference of the drop in the deviance as compared to the deviance of the model in the previous step was assessed through Chi square test with one degree of freedom. To do this the online Chi square calculator was used [163]. From this level onward all logistic regressions were run with the ‘enter’ method of SPSS (i.e. not stepwise).

Table 46 lists the deviance of each model, the difference of this deviance with the model at the previous step and the significance of the difference (p value). As seen, adding each covariate to models resulted in a significant decrease in the deviance indicating that the covariate should be retained in the model. In this ‘manual’ stepwise procedure, two fractional polynomials transformations of GCS were supplied to the model together whereas in the automatic form, the computer supplied them separately.

The final resulting model demonstrated that the odds ratio of the pupil category: absent-brisk did not have a reasonable direction for survival (it was more than one) (Appendix: level II model). Furthermore, ISS and extracranial injury were correlated in the correlation matrix indicating that these two variables can not be retained in the same model.

	deviance	*difference*	*P value of the difference*
Step 1: Age	883.069	-	
Step 2 : Age +GCS	565.905	317.164	<0.005
Step 3: Age +GCS +Pupillary reactivity	411.731	154.174	<0.005
Step 4: Age +GCS +Pupillary reactivity + ISS	397.587	14.144	<0.005
Step 5: Age +GCS +Pupillary reactivity + ISS + Extracranial Injury	392.310	5.277	0.0216

Table 46 Changes in the deviance in the manual stepwise modelling at level II.

7.5.11. Level III: new categorisation of pupillary reactivity/decision to employ ISS or extracranial injury

As one category of pupillary reactivity in model II did not have a reasonable direction for its odds ratio of survival, pupillary reactivity was categorised differently as both reactive, only one reactive and none reactive. Furthermore, since at the previous level, ISS and extracranial injury were correlated, at this level, two models were constructed: model IIIA with age, continuous GCS (fractional polynomials transformations), pupillary reactivity (with new categorisation), ISS (fractional polynomials transformation) and model IIIB with age, continuous GCS (fractional polynomials transformations), pupillary reactivity (with new categorisation) and extracranial injury.

In model IIIA (age, continuous GCS, pupillary reactivity, (3 categories) and ISS) all covariates were significantly correlated with outcome including pupillary reactivity with new categorisation and also ISS. However, in model

IIIB, extracranial injury (extracranial AIS severity score > 2) did not hold a significant association with outcome. The AUC of models IIIA and IIIB were respectively 0.91 and 0.90 with p value of HL statistics being 0.83 and 0.45.

7.5.12. Level IV: model IIIA including categorical GCS

Since GCS was correlated with outcome both in the continuous and categorical form, model IIIA was reconstructed this time with GCS included categorically with each GCS score counted as one individual category (thus 13 categories). The reference category was GCS 15.

The resulting model (i.e. model IV) demonstrated that firstly, only GCS scores 14, 12, 10 and 7 held a significant effect on outcome and secondly, GCS score of 9, 4 and 3 had odds ratios of above one for survival whilst this was expected to be less than one. Therefore, in the next level, GCS was categorised into mild, moderate and severe GCS with the cut offs as 9 and 12.

7.5.13. Level V: model of level IIIA including new categorisation of GCS

GCS was categorised into mild (GCS < 12), moderate (12-9) and severe (< 9). Subsequently model IV was reconstructed with this new categorisation of GCS.

The resulting model (model V) showed that all three categories of GCS had significant association with outcome with an acceptable direction for odds ratios (i.e. the odds ratios were less than one and decreasing as GCS became lower). The AUC of this model was 0.908 with P value of HL statistics being 0.93.

7.5.14. Level VI: testing the value of cause of injury

The cause of injury was added to models IIIA (with continuous GCS) and V (with categorical GCS) with RTC as the reference category. This resulted in a significant decrease in the deviance of the model IIIA (with continuous GCS) but no significant decrease was observed in model V (with categorical GCS). However, the sports category in model VIA has an "astronomical" value (81656351.28) for odds ratio. This was attributed to the small number of cases in this category (only 14 cases in the whole dataset). Therefore, the 'sports category' was merged with the 'others' category and the model was run again. The resulting model (model VIA) demonstrated that all included covariates (i.e. age, continuous GCS, pupillary reactivity (with 3 categories), ISS and cause of injury) had significant association with outcome with acceptable directions for odds ratios. The AUC of this model was 0.916 with p value of 0.55 for HL statistics.

7.5.15. Level VII: trial of the new categorisation of pupillary reactivity

It was felt that the pupillary reactivity could be better described using further information available from the database of abnormal both reactive. New categorisation of pupillary reactivity as normal (brisk-brisk), abnormal-both reactive (brisk-sluggish or sluggish-sluggish), only one reactive (none-sluggish or none-brisk) and none reactive (none-none) were tested on models level VIA (age, continuous GCS, pupillary reactivity, ISS and injury cause) and level V (age, categorical GCS, pupillary reactivity and ISS). With the new categorisation of pupillary reactivity two models were constructed: model

VIIA from model VIA and model VIIB from model V. It was observed that this new categorisation of pupillary reactivity (i.e. with 4 categories) is better than the previous one (i.e. with three categories) since AUC of models VIA and V respectively increased from 0.916 to 0.922 and from 0.908 to 0.914. The p value of HL statistics for models VIIA and VIIB were respectively 0.549 and 0.572 in both models, Furthermore, pupillary reactivity still held its significant association with outcome.

So far two models were constructed. Both models shared age, GCS, pupillary reactivity and ISS. The differences were: GCS being continuous in model VIIA but categorical in model VIIB and cause of injury being present in model VIIA but absent in model VIIB. This meant when GCS was categorical, cause of injury was not significant.

7.5.16. Level VIII: Trial of the new categorisation of extracranial injury

So far it was observed that the extracranial injury with the extracranial AIS cut-off of 3 can not be held in the same model with ISS and also it does not hold a significant effect on outcome in a model which does not contain ISS. It was thought these observations may be different if the cut-off for extracranial AIS increases to 4 (inclusive). Level VIII was performed to assess this. The frequency of extracranial injury based on this new cut-off was 116 (14.5%).

Up to this stage, several models were constructed and the following issues were explored: the importance of ISS versus extracranial injury, various categorisations of pupillary reactivity, continuous versus categorical GCS and the importance of cause of injury. These issues required to be considered in the conduct of this level. To assess this new cut-off of extracranial injury

(extracranial AIS severity score > 3), it was firstly assessed whether or not this new cut-off would enable extracranial injury to be included in the same model with ISS. Thus this variable was added to model IIIA (age, continuous GCS, pupillary reactivity (4 categories) and ISS). Since this model contained continuous GCS, the addition of extracranial injury to model IIIA was also tested with categorical GCS (i.e. three categories). The resulting models did not have a significant decrease in the deviance indicating that extracranial injury still can not be held in a model with ISS. Secondly, it was examined that whether or not extracranial injury with this new cut-off would have a significant association with outcome if the model does not contain ISS. For this purpose, model IIIB (age, GCS and pupillary reactivity) was reconstructed twice once with continuous GCS and then with categorical GCS. It was observed that in both resulting models, the deviance significantly decreased and also extracranial injury demonstrated a significant association with outcome. Thirdly, the value of cause of injury was tested by adding this variable to the newly-constructed models IIIB. It was observed that cause of injury significantly decreased the deviance of both models and also held a significant association with outcome (models VIIIA and VIIIB). The AUC of models VIIIA and VIIIB were respectively 0.917 and 0.910 and HL statistics of both models had a significant p value.

7.5.17. Level IX: Imputation of missing GCS and pupillary reactivity

Four models were constructed so far. Model VIIB (age, categorical GCS, pupillary reactivity and ISS) was deemed better than VIIA model (age, continuous GCS, pupillary reactivity, ISS and cause of injury) because its AUC

is only slightly lower (0.914 versus 0.922) but has simpler classification of GCS by not using fractional polynomials making the model user-friendly. For the same reason, Model VIIIB (age, categorical GCS, pupillary reactivity, extracranial injury and cause of injury) was deemed better than model VIIIA (age, continuous GCS, pupillary reactivity, extracranial injury and cause of injury) despite having only slightly a lower AUC (0.917 versus 0.920). Therefore, models VIIB and VIIIA were chosen to proceed with imputation of missing data and further levels of models derivation. So far all the logistic regression models were run on complete cases which amounted to 645 cases. The excluded cases had missing information of either GCS or pupillary reactivity.

The imputation strategies for GCS was firstly to sum motor, eye and verbal components if they were available as there were occasions when the GCS components were recorded but total GCS was missing; Secondly, to impute the remaining missing information with en route and then at scene GCS recordings and thirdly to identify patients who had been intubated by weighing within GCS similar to TARN general trauma predictive models [142]. There were overall 26 cases with missing total GCS. The first step resulted in imputation of one case. Imputing with the en route records resulted in the reduction of missing cases to 23. Subsequently, all the remaining missing GCSs were imputed with scene scores with no missing value left for the last step of imputation strategy. The model was not rerun at this stage because 21 of the missing GCS cases also had missing pupillary response. Following the imputation of GCS, missing pupillary reactivities were imputed. The imputation was performed on uncombined pupillary reactivity (i.e. separately

for the left and right eye) and with 6 categories. The adopted imputation strategy for each eye was as follow: firstly, if one eye was missing and the other was recorded, the missing eye was regarded as normal. This was initially performed for all admission, en route and at-scene records of pupillary reactivity. Table 47 presents the number of missing pupils prior to and following this first step of imputation. Secondly, the missing admission pupils were imputed with en route records. This decreased the missingness of both right and left eye from 122 to 108. Thirdly, the missing information on each eye was imputed with at-scene records. This final step decreased the number of missing information to 0 for each eye. Following the imputation, the syntax to combine the pupillary reactivity of each eye was run again.

Table 48 compares the performance of the models prior to and following the imputation of missing GCS and pupillary reactivity. As seen, the change in AUC is negligible for both models (drops of 0.003 and 0.007 respectively for models VIIB and VIIIB). Even the p value of HL statistics in model VIIIB became more distant from the significant level when it was 0.063 before imputation and 0.547 after imputation.

Site of record	***Left eye***		***Right eye***	
	Before imputation	***After imputation***	***Before imputation***	***After imputation***
Admission	136	122	138	122
En rout	761	761	762	761
At scene	484	473	477	473

Table 47 Frequency of missing information on pupillary reactivity for each eye prior to and following the first step of imputation

Model	*Before imputation*			*After imputation*		
	Number of cases	*AUC*	*HL test*	*Number of cases*	*AUC*	*HL test*
Level VIIB	650	0.914 (0.889-0.940)	0.573	802	0.911 (0.888-0.934)	0.658
Level VIIIA	650	0.917 (0.891-0.943)	0.063	802	0.908 (0.884-0.932)	0.547

Table 48 Comparison of performances of models VIIB and VIIIA before and after imputation of GCS and pupillary reactivity.

7.5.18. Level X: Inclusion of intracranial pathologies

The results of the level X and the following level (Level XI: adjustment to model X) are presented and discussed in Paper 5. In this paper, the baseline models A and B in the paper were models IXA (model VIIB after imputation of GCS and pupillary reactivity) and IXB (model VIIIB after imputation of GCS and pupillary reactivity) in the actual modelling procedure. Then AIS scores (i.e. 3, 4, and 5/6), the Marshal Classification and various intracranial pathologies (i.e. including brain swelling, intracranial haemorrhage, EDH, SAH, brain stem injury and the cerebellar injury) were added separately to each model. This resulted in the construction of 6 models. However, these 6 models were named with the prefix ‘X’ in the staged approach discussed here. Their equivalent naming is presented in Table 49 (plus the equivalent naming of baseline models). The Appendix does not repeat these 8 models given in Paper 5.

Name of the model in the paper	*Name in the modelling procedure*
Baseline model A	Model IXA
Model A + AIS scores	Model XA1
Model A + Marshal Classification	Model XA2
Model A + combination 1A	Model XA3
Baseline model B	Model IXB
Model B + AIS scores	Model XB1
Model B+ Marshal Classification	Model XB2
Model B + combination 1B	Model XB3

Table 49 The name of each model presented in the paper with the equivalent name in the actual procedure model construction.

This level and the following level (level XI) assess the prognostic value of AIS scores, the Marshall Classifications and various intracranial pathologies. However, these levels were performed within the modelling procedure and thus the 'appropriate' models among the resulting models required to be selected to proceed with the modelling. The selection of the most appropriate classification/intracranial pathologies was based on AUC and the significance of the association of each type of brain injury with outcome.

All models of XA1 to XA3 had the same AUC. However, model XA1 was not selected because in this model, none of AIS scores i.e. scores 4 and 5/6 held significant association with outcome. Similarly, model XA2 was not selected because among all 6 categories of the Marshall Classification, only class II was associated with outcome with marginal significance. Among models XB1 to XB3, model XB1 was not selected since in this model AIS score 4 did not hold significant association with outcome. Similarly, model XB3 was not selected since among all 6 categories of the Marshal Classification included in this model, only classes IV and V/VI (merged) were significantly associated with outcome. Overall models XA3 and XB3 were selected for further analysis i.e. trial of various combinations of intracranial pathologies at level XI.

7.5.19. Level XI: Adjustments to model X

At this level, six models were constructed. The rational for various combinations of intracranial pathologies and the AUC of constructed models is presented in Paper 5. Briefly, this was based on the haemorrhage being significantly associated with outcome from either clinical or evidence viewpoint and brain swelling, brain stem injury and cerebellar injury being

marginally (0.05 < p value < 0.10) or significantly associated with outcome in models constructed at the previous level. However, since in Paper 5 the odds ratios and significance (p values) are not given, the Appendix, unlike level X, contains all models at this level. Table 50 lists the equivalent name of each model in the modelling procedure to that in the Paper 5.

Name of the model in the paper	*Name in the modelling procedure*
combination 2A	Model XIA1
combination 3A	Model XIA2
combination 4A	Model XIA3
combination 5A	Model XIA4
combination 2B	Model XIB1
combination 3B	Model XIB2

Table 50 The name of each model presented in the paper with the equivalent name in the actual procedure of TBI prognosis model construction.

After the analysis of the model performance, models XIA4 and XIB2 were selected; these being the models in which all intracranial pathologies described held significant association with outcome. The reason for this selection was that, apart from model XIA4, in none of the models XIA1 to XIA4, all intracranial pathologies included held significant association with outcome. All these models had equal AUCs. Similarly among models XIB1 and XIB2, model XIB2 was selected because unlike model XIB1, all intracranial pathologies included were significantly associated with outcome.

7.5.20. Level XII: Adding systolic blood pressure, mean blood pressure and hypoxia

Systolic blood pressure (categorical), mean blood pressure (categorical) and hypoxia (categorical) were added to models XIA4 and XIB2 altogether.

Systolic and mean blood pressures were not used continuously, since firstly although they were associated with outcome by logistic regression on univariate analysis, this is not supported by Mann Whitney U test. Secondly, the plot of predicted and observed probabilities against the fractional polynomials transformation of both variables did not demonstrate that these two probabilities would follow the same pattern of distribution in an acceptable way (Figure 24 and Figure 26).

In the resulting models (models XIIA and XIIB), it was observed that neither systolic blood pressure nor mean blood pressure were significantly associated with outcome with normotension as the reference category. However, hypoxia demonstrated a significant association with outcome. The logistic regression was run on 607 cases.

7.5.21. Level XIII: missingness analysis/imputation of missing mean and systolic blood pressure and hypoxia

At the previous level, only hypoxia turned out to be significant in the multivariate models. However, the analysis was run only on 607 cases since the remaining 195 cases had missing information on hypoxia, mean or systolic blood pressure. Thus it was thought that if the missing cases are included in the modelling, the results might change. This level was performed to include all cases in the analysis of hypoxia and blood pressure following imputation of missing values. However, unlike GCS and pupillary reactivity, imputing missing values with en route and scene scores did not result in imputation of all missingness for mean blood pressure, systolic blood pressure and hypoxia. Thus an analysis was performed to assess what variables could affect the missingness for each of the above 3 variables. This analysis provides the

answer as to whether or not if some variables changes in the dataset, the missingness would change. For example, it may that if the patient ages, the chance of missingness of hypoxia decreases. This analysis is important with regards to ‘multiple imputation’ as a method to impute missing values based on the existing data. By multiple imputation the missingness is predicted from the values of other none-missing variables. To perform this statistical method, it is first necessary to know if other variables can affect the missingness of the variable of interest.

Systolic blood pressure

There were overall 36 cases with missing systolic blood pressure. Replacing missing admission scores with en route and scene values resulted in the reduction of missing information to 31 and 21 respectively. Subsequently, an analysis was performed to assess the effect of other variables on the systolic blood pressure missingness. These ‘other’ variables were decided to be those which had been included in the models so far. Initially, the characteristics of population with missing systolic blood pressure were compared with those who had this variable recorded. Chi square test was run for categorical variables (extracranial injury, injury cause, brain stem haemorrhage, brain swelling and survival rate) with Mann Whitney U test for the continuous variables (age, GCS and ISS). Also logistic regression was run with age, GCS, ISS, extracranial injury, brain stem haemorrhage, brain swelling and survival as covariates and systolic blood pressure missingness as outcome. This latter

multivariate analysis would show the effect of each variable on the missingness of systolic blood pressure.

Table 51 demonstrates the comparison of the two populations of missing and non-missing systolic blood pressure. It appears those subjects with missing systolic blood pressure tended to be younger, have lower GCS and more frequency of brain stem injury and lower survival rate (these are the variables with significant p values). Similarly, the missingness was significantly associated with pupillary reactivity per univariate analysis (p value = 0.015). Table 52 demonstrates the results of logistic regression analysis. As seen, only age and cause of injury (assault) were significantly associated with missingness in the multivariate analysis. This means there were only two variables which can predict the missingness of systolic blood pressure. Due to this small number of 'predictors of missingness' in the data, multiple imputation was deemed inappropriate. Therefore, following imputation of missing systolic blood pressure with en route and scene scores the decision was made to group all the missing values as a separate category. Subsequently, the continuous systolic blood pressure was categorised into low blood pressure (< 119), normotension (120-150), hypertension (> 151) and missing.

		Missing (n= 21)	*Available (n=781)*	*P value*
age		26	39	0.016
GCS		3	13	0.001
Pupillary reactivity	**Normal**	0.9	99.1	0.015
	Abnormal-both reactive	2.9	97.1	
	Only one reactive	3.6	96.4	
	Neither reactive	5.6	94.4	
ISS		29	25	0.196
Extracranial				0.063
Injury cause	**RTC**	3.2	96.8	0.544
	Falls	1.6	98.4	
	Assault	3.5	96.5	
	Others	3.1	96.9	
Brain stem injury		23.8	1	0.016
Swelling		38.1		0.710
Survival rate		38.1	75.7	0.00

Table 51 Comparing characteristics with missing systolic blood pressure with those without missing systolic blood pressure.

		Sig.	*odds ratio*
Age		.002	.929
GCS		.710	.965
Pupillary reactivity	**Normal**	.238	
	Abnormal-both reactive	.112	4.702
	Only one reactive	.116	8.656
	None reactive	.083	6.866
ISS		.485	.975
Extracranial injury		.942	1.086
Brain stem injury		.787	1.250
Brain swelling		.385	.547
Cause of injury	**RTC**	.224	
	Fall	.114	3.834
	Assaults	.041	5.250
	Others	.998	.000
survival		.383	.433

Table 52 Logistic regression analysis to predict systolic blood pressure missingness.

Mean blood pressure

The dataset contained 47 cases with missing mean blood pressure (this variable was calculated from systolic and diastolic blood pressure and since there were already 36 missing systolic blood pressure, 11 cases with missing diastolic blood pressure rendered in total 47 cases with missingness on mean blood pressure). Initially, scene and en route mean blood pressure (in the continuous from) was calculated from respective scene and en route systolic and diastolic blood pressure. Then the missing admission mean blood pressure (continuous) was imputed first with en route and then with scene values. This resulted in the decrease of missingness to 43 and 32 respectively. Then the analysis was performed to compare the characteristics of the population with missing mean blood pressure to that with non-missing mean blood pressure. This comparison was made on age, GCS, pupillary reactivity, ISS, extracranial injury, cause of injury, brain stem injury, brain swelling and survival rate using Chi square and Mann Whitney U test. Similarly, logistic regression was run with these variables and the missingness as the outcome.

Table 53 demonstrates that the population with missing mean blood pressure was significantly younger, had lower GCS, had higher frequency of extracranial injury, brain stem injury and survival rate. Moreover, it appeared by univariate analysis, pupillary reactivity also affected the missingness (p value = 0.003). According to the multivariate analysis, only age had a significant effect on the missingness of mean blood pressure (Table 54). As such, multiple imputation was deemed inappropriate. Subsequently and following the imputation of missing mean blood pressure with en route and

scene records, the remaining missingness was categorised in a separate category. Thus, mean blood pressure was re-categorised as hypotension (< 84 mmHg), normotension (85-110 mmHg), hypertension (> 111 mmHg) and missing.

		Missing (n=32)	*Available (n=770)*	*P value*
age		26	39	0.002
GCS		3	13	0.00
Pupillary reactivity	**Normal**	1.3	98.7	0.003
	Abnormal- both reactive	2.9	97.1	
	Only one reactive	7.1	92.9	
	Neither reactive	7.5	92.5	
ISS		27.5	25	0.092
Extracranial		37.5	13.5	0.00
Injury cause	**RTC**	8	92	0.224
	Falls	4.5	95.5	
	Assault	4.2	95.8	
	Others	6.3	93.8	
Brain stem injury		25	8.3	0.001
Swelling		31.3	34.4	0.712
Survival rate		37.5	76.2	0.00

Table 53 Comparing characteristics of cases with missing mean blood pressure to those without missing mean blood pressure.

		Sig.	*odds ratio*
Age		.000	.923
GCS		.768	.973
Pupillary reactivity	**Normal**	.244	
	Abnormal-both reactive	.163	3.743
	Only one reactive	.246	4.900
	None reactive	.052	7.668
ISS		.204	.955
Extracranial injury		.343	2.627
Brain stem injury		.220	2.440
Brain swelling		.299	.510
Cause of injury	**RTC**	.551	
	Fall	.351	2.033
	Assaults	.152	2.865
	Others	.998	.000
survival		.476	.519

Table 54 Logistic regression analysis to predict missingness of mean blood pressure.

O2 sat. (hypoxia)

There were overall 114 cases with O2 saturation recording missing. The strategy of imputation was similar to other variables which was replacing the missing information initially with en route and then scene records. Replacement with en route values resulted in 108 cases with missing O2 Sat. Further replacement of missing values with scene values resulted in 78 missing cases. An analysis was carried out to compare the characteristics of the population with missing O2 Sat. to that with recorded (available) hypoxia. Similarly, logistic regression was run to assess the association of each variable with missingness on O2 Sat.

According to Table 55, patients with missing O2 Sat. tended to have lower GCS, larger frequency of major extracranial injury and brain stem injury and a lower survival rate. Furthermore the effect of pupillary reactivity on missingness was significant accordingly. Moreover, according to Table 56 which demonstrates missingness in a multivariate analysis, only GCS had a significant association with missing O2 Sat. Therefore, similar to mean and systolic blood pressure, following the imputation of missing O2 Sat. with en route and scene values, this variable was re-categorised as hypoxia (O2 sat < 90%), non-hypoxia and missing without using multiple imputation.

		Missing O2 Sat. (n=78)	*Available O2 Sat. (n=723)*	*P value*
Age		36	39	0.081
GCS		7	13	0.00
Pupillary reactivity	**Normal**	7.8	92.2	0.005
	Abnormal-both reactive	8.5	91.5	
	Only one reactive	12.8	87.2	
	None reactive	18	82	
ISS		26	25	0.067
Extracranial		28.2	13	0.00
Cause of injury	**RTC**	11.8	88.2	0.307
	Fall	8.9	91.1	
	Assaults	8.4	91.6	
	Others	3.1	96.9	
Bran stem injury		17.9	8	0.004
Swelling		35.9	34.1	0.753
Survival rate		59	76.4	0.001

Table 55 Comparing characteristics of cases with missing O2 sat. to those without missing O2 Sat.

		Sig.	*odds ratio*
Age		.112	.988
GCS		.017	.908
Pupillary reactivity	**Normal**	.523	
	Abnormal-both reactive	.492	.711
	Only one reactive	.632	1.334
	None reactive	.293	.584
ISS		.967	.999
Extracranial injury		.566	1.363
Brain stem injury		.396	1.461
Brain swelling		.544	.823
Cause of injury	**RTC**	.477	
	Fall	.182	1.638
	Assaults	.382	1.430
	Others	.592	.567
Survival		.694	.845

Table 56 Logistic regression analysis to predict missingness of O2 sat.

7.5.22. Level XIV: trial of systolic and mean blood pressure individually/2 versus 3 categories

This level was performed to assess the importance of blood pressure on the complete data. It was initially necessary to assess the effect of hypoxia on the deviance of the models prior to inclusion of blood pressures. This was because hypoxia was a significant predictor at level XII where neither systolic nor mean blood pressures were significant. Furthermore, since there were two types of models to be run (models with suffix A and models with suffix B), this level was divided into two sections. At each section, systolic and mean blood pressure were added separately as these variables are apparently correlated with each other (multicolinearity effect). This means if systolic blood pressure changes, mean blood pressure would change as well. It is the pre-requisite for logistic regression that all the variables included should not affect each other in that if one variable changes, all others remain constant.

Models XIVA

Model XIVA1 (model XIA4 + hypoxia)

Adding hypoxia to model XIA4 resulted in a significant decrease in the deviance of model XIA4 (495.663 versus 481.898, $P < 0.005$). Furthermore, hypoxia showed a significant association with outcome. AUC of the resulting model was 0.932.

Model XIVA2 (model XIVA1 + mean blood pressure with 3 categories)

Adding mean blood pressure resulted in a significant decrease in the deviance of model XIVA1 (From 481.898 to 468.188, $p < 0.0005$). The category low blood pressure demonstrated a significant association with outcome with no significant association for the hypertension category. The AUC of the resulting model was 0.927.

Model XIVA3 (model XIVA1 + mean blood pressure with 2 categories)

Categories of hypertension and normotension were merged together leaving mean blood pressure with only 2 categories of hypotension versus no-hypotension plus one category of missing information. This resulted in a significant decrease in the deviance of model XIVA1 (from 481.898 to 468.333, p value < 0.0005). Furthermore, hypotension was significantly associated with outcome and AUC of the model was observed to be 0.927.

Model XIVA4 (model XIVA1 + systolic blood pressure with 3 categories)

Adding systolic blood pressure resulted in a significant decrease in the deviance of the model (from 481.898 to 476.259, $p < 0.0005$). Category: low blood pressure demonstrated a p value of close to significance (0.057) and AUC of the resulting model was 0.924.

Model XIVA5 (model XIVA1 + systolic blood pressure with 2 categories)

Categories of hypertension and normotension were merged together leaving systolic blood pressure with only 2 categories of low blood pressure versus no-low blood pressure plus one category of missing information. This resulted in a significant decrease in the deviance of model XIVA1 (from 481.898 to 476.318, p value < 0.0005). Furthermore, Hypotension was significantly associated with outcome and AUC of the model was 0.924.

Selection of the best model among Models type A

Overall models with blood pressure (models XIVA2-5) were better than the model without blood pressure (model XIVA1) in terms of the effect on the deviance. This was because adding blood pressure (systolic or mean, with 2 or 3 categories) resulted in a significant decrease in the deviance. Among models with mean blood pressure (models XIVA2 and XIVA3), the model with 2 categories of mean BP is better since the category hypertension has no association with outcome in model XIVA2. Furthermore, among models with systolic BP, the model with 2 categories of systolic BP (model XIVA5) is selected as the category: hypertension has no association with outcome (model XIVA4). Lastly, among models XIVA3 and XIVA5, model XIVA5 was selected as the HL statistics of the other was observed to have a p value of less than 0.05 (0.111 versus 0.048 for model XIVA5).

Models XIVB

Model XIVB1 (model XIB2 + hypoxia)

Adding hypoxia to model XIB2 resulted in a significant decrease in the deviance (478.725 versus 465.888, p value < 0.005). Moreover, the association of hypoxia with outcome was significant and the AUC of the model was 0.928.

Model XIVB2 (model XIVB1 + mean blood pressure with 3 categories)

Adding mean blood pressure resulted in a significant decrease in the deviance of the model (from 481.898 to 451.936, p < 0.0005). The category hypotension demonstrated a significant association with outcome. However, the association of hypertension was not significant. The AUC of this model was AUC = 0.932.

Model XIVB3 (model XIVB1 + mean blood pressure with 2 categories)

Categories of hypertension and normotension were merged together leaving mean blood pressure with only 2 categories of hypotension versus no-hypotension plus one category of missing information. This resulted in a significant decrease in the deviance of model XVB1 (from 481.898 451.940, p value < 0.0005). Furthermore the hypotension category was significantly associated with outcome. The AUC of the model was 0.932.

Model XIVB4 (model XIVB1 + systolic blood pressure with 3 categories)

Adding systolic blood pressure resulted in a significant decrease in the deviance of the model (from 481.898 to 461.154, p < 0.0005). None of systolic

blood pressure categories (neither hypotension nor hypertension) demonstrated significant association with outcome. The AUC of the model was 0.930.

Model XIVB5 (model XIVB1 + systolic blood pressure with 2 categories)

Categories of hypertension and normotension were merged together leaving systolic blood pressure with only 2 categories of low blood pressure versus no-low blood pressure plus one category of missing information. This resulted in a significant decrease in the deviance of model XIVB1 (from 481.898 to 461.349, p value < 0.0005). The hypotension category significantly associated with outcome and the mode had an AUC of 0.929.

Selection of the best model among models XVB1-5

Overall models with blood pressure (models XIVB2-5) are better than the model without blood pressure (model XIVB1). This is because adding blood pressure (systolic or mean, with 2 or 3 categories) resulted in a significant decrease in the deviance. Among models with mean blood pressure (models XIVB2 and XIVB3), the model with 2 categories of mean BP is better since the category hypertension has no association with outcome in model XIVB2. Furthermore, among models with systolic BP, the model with 2 categories of systolic BP (model XIVB5) is selected as the categories hypertension and hypotension have no association with outcome in model XIVB4. Lastly, among models XIVB3 and XIVB5 (systolic versus mean blood pressure), model

XVB5 was selected as this model contains systolic blood pressure which is also included in model XIVA5.

7.5.23. Level XV: assessment of interactions

According to the literature, there are two interactions which needed to be investigated in the so-far-constructed models (models XVA5 and XVB5): interaction of age with cause of injury [160] and interaction of systolic blood pressure with hypoxia [96]. Each syntax of models XIVA5 and XIVB5 was run again with declaration of the aforementioned interactions.

Model XVA1(model XIVA5 plus interaction of hypoxia and low blood pressure)

Adding the interaction of hypoxia with hypotension resulted in a significant decrease in the deviance of the model from 476.318 to 470.917 (p value < 0.005). Furthermore, among the interaction of various categories of hypoxia (yes, no and missing) with various categories of systolic blood pressure (normotension, hypotension and missing), only the interaction of missing hypoxia with hypotension appeared significant. This appeared to be not clinically significant. Thus, this interaction was not included in the model. The AUC of this model was 0.925 (0.903-0.946).

Model XVB1 (model XIVB5 plus interaction of age and cause of injury and interaction of hypoxia and systolic blood pressure)

Following addition of the above interactions to the model, the deviance of the model XIV5 significantly decreased from 461.349 to 446.764 (p value <

0.005). Furthermore, among interactions of various categories only interaction of age and fall was significant with no significant interaction between hypoxia and systolic blood pressure.

Model XVB2 (model XIVB5 without interaction of hypoxia and systolic blood pressure)

At this stage, model XIVB5 was run only with interaction of age and cause of injury. It was observed that the deviance of model XIVB5 significantly decreased from 461.349 to 451.830 (p value < 0.005). Age demonstrated a significant interaction with fall. The AUC of this model was 0.931.

7.5.24. Main points and decisions in model derivation

Table 57 presents the main points/decisions which were made during the modelling procedure based on the observations obtained at each level/model. This table assists with clarification of the ‘hidden’ coherence of the modelling procedure.

Main point/decision	*Relevant levels*	*Involved models*
Pupillary reactivity with 4 categories: normal, abnormal both reactive, only one reactive, none reactive	II, III, VII	II, IIIA, IIIB, VIIA, VIIB
Categorical GCS rather than continuous GCS	IX	VIIA, VIIB, VIIIA, VIIIB
Categorical GCS with 3 categories rather than 13 categories	IV, V	IIIA, IV, V
ISS and extracranial injury not to be in the same model	II, VII	II, IIIA,IIIB, VIIIA, VIIIB,
Cause of injury in the model with extracranial injury but not with ISS	VI, VIII	IIIA, V, VIA, IIIB, VIIIA, VIIIB
Cut-off for extracranial injury AIS severity score 4 and not 3	VIII	IIIA, IIIB, VIIIA, VIIIB
Only hypotension important (hypertension merged with normotension)	XIV	XIVA2, XIVA3, XIVA4, XIVA5, XIVB2, XIVB3, XIVB4, XIVB5
Why systolic blood pressure and not mean blood pressure	XIV	XIVA2, XIVA3, XIVA4, XIVA5, XIVB2, XIVB3, XIVB4, XIVB5
Only brain stem injury and brain swelling as important intracranial injuries	X, XI	IXA, XA1, XA2, XA3,IXB,XB1,XB2,XB3,XIA1,XIA2,XIA3,XIA4,XIB1,XIB2

Table 57 Main points/decisions made during the model derivation with their respective levels and models.

7.5.25. Calibration plot/Brier Score

To draw the calibration plot, survival probability of each individual as predicted by the model (predicted probability) was saved. The predicted probabilities were then divided into bands with intervals of 0.1. The mean predicted probability and the observed probability were calculated within each band. Then, the scatter plot of the mean predicted probability and the mean observed probability were drawn in Excel.

The brier score is not given in the SPSS output of logistic regression and thus this index was calculated by creating the appropriate syntax using the following formula (section 2.7):

$$\sum (observed\ probability - predicted\ probability)^2 / n$$

7.5.26. External validation

The IMPACT dataset of TBI was sent through email by HL. It included 11023 cases. Firstly, the IMPACT dataset was 'prepared' for the validation according to the variables and their format in the TARN models. Secondly, since the IMPACT dataset contained missing information, an analysis was performed to compare the characteristics of patients with full variables recorded to those with one or more missingness. Thirdly, the characteristics of the TARN and the IPMACT were compared across various variables included in the TARN models. Fourthly, the TARN model B was run on the IMPACT data. The validation of TARN model A was deemed not feasible since the IMPACT dataset did not contain record of ISS.

Preparation of the IMPACT dataset

This involved adjustments for the different format of recording of pupillary reactivity and cause of injury in the IMPACT dataset to that in the TARN dataset and the lack of extracranial injury and brain stem injury in the IMPACT dataset.

In TARN TBI models, pupillary reactivity contains four categories (normal-both reactive, abnormal-both reactive, only one reactive, none reactive) whereas in IMPACT, the pupillary reactivity contain three categories (reacting, one reacting and neither reactivity). In order to alleviate this mismatched format of pupillary reactivity, the categories of normal and abnormal-both reactive were merged into 'both reacting' in the TARN dataset.

Furthermore, cause of injury in IMPACT holds 11 categories (Table 58) whereas the TARN models contain only 4 categorises. Following the advice of a TARN member of staff (TJ), the equivalent category of cause of injury in TARN models was defined for each category in IMPACT. This is presented in Table 58. This mapping was based on the closest possible equivalent injury cause among RTC, assault, fall or others to that in the IMPACT classification. However, one may argue that the 'bike/skate ect.' should have been considered as RTC whereas in this mapping it is considered as others. In TARN data recording, RTC is recorded when vehicle incidence or a collision occurs. However, in the IMPACT dataset it was not clear what percentage of cases with 'bike/skate etc' as their cause of injury sustained their injury following collision and not skating.

IMPACT classification	*TARN model classification*
Motor/vehicle occupant	RTC
Pedestrian	RTC
Motor/pedestrian/RTA/ other	RTC
Motorbike/Moped	RTC
Assault	Assault
domestic/fall	Fall
fall/alcohol	Fall
work-related	Others
Sports	Others
bike/skate ect.	Others
Other	Others

Table 58 The mapping of categories of injury cause in IMPACT to TARN.

There are two variables in the TARN models which were not recorded in IMPACT; these being extracranial injury and brain swelling. In order to rectify this problem, extracranial surgery and Marshal Class III in IMPACT were used as proxies respectively for extracranial injury and brain swelling.

IMPACT missingness

The IMPACT dataset supplied to TARN had complete data on pupillary reactivity, hypoxia and systolic blood pressure. However, there was a number of missing information on other variables presented in Table 59. In total, 5542 cases had a missing value on one variable or the other. Thus only 5481 cases remained with all the data available. Chi square test was run to assess the significance of the difference between the complete cases and those with one or more missing values on GCS, cause of injury and extracranial surgery.

Variable	*Missingness*
GCS	3230 (58.2%)
Cause of injury	17(0.3%)
Extracranial surgery	3090 (55.7%)

Table 59 The number of missing information in the IPMPACT dataset.

Comparison of the TARN and the IMPACT datasets

This was performed after merging the TARN and the IMPAT dataset in SPSS and across age, GCS, pupillary reactivity, cause of injury, extracranial surgery/injury, brain swelling and hypoxia using Mann Whitney U test for continuous variables (age) and Chi square test for categorical variables (GCS, pupillary reactivity, cause of injury, extracranial surgery/injury, Marshall Class III/brain swelling and hypoxia).

Running of TARN model B on the IMPACT dataset

This was performed on the merged TARN and IMPACT dataset. Since IMPACT holds the record on favourable versus unfavourable outcome, model B was also run for prediction of favourable outcome. The logistic regression was run on the complete cases of IMPACT dataset i.e. with no imputation.

External validation in TARN dataset

For this purpose, the derived models were run on another dataset of TBI cases from TARN which contained submissions from May 2010 till May 2010. Both internal and external datasets had the same inclusion criteria. First the two datasets (prediction and the external validation sets) were compared across various patients' characteristics.

7.6. Further results

The results of univariate analysis, final derived models and their validation are presented in Paper 6 (Table 31, Table 32, Table 33, Table 34 and Table 35). What follows is the further results of the univariate analysis and the external validation.

Table 60 presents the frequency/median of each variable across the groups of survivals and non-survivals. The significance of the differences are presented in Paper 6 (Table 31, Table 32 and Table 33).

Variable		*Survival*		*Non-survivals*	
		Median	*Frequency (percentage)*	*Median*	*Frequency (percentage)*
Age		40		49	
Gender	**Male**		452 (75%)		151 (25%)
	Female		147 (73.9%)		52 (26.1%)
Nationality	**British**		334 (75.6%)		108 (24.4%)
	European		13 (86.7%)		2 (13.3%)
	Others		19 (79.9%)		5 (20.8%)
Cause of injury	**RTC**		227 (72.3%)		87 (27.7%)
	Fall		222 (70.9%)		91 (29.1%)
	Assaults		125 (87.4%)		18 (12.6%)
	Others		25 (78.1%)		7 (21.9%)
GCS (continuous)		12		6	
GCS (categorical)	**Mild**		377 (92.2%)		32 (7.8%)_
	Moderate		84 (81.6%)		19 (18.4%)
	Severe		138 (47.6%)		152 (52.4%)
Pupillary reactivity	**Brisk-brisk**		405 (82.5%)		41 (25.8%)
	Sluggish-sluggish		36 (7.3%)		17 (10.7%)
	Brisk-brisk		10 (2%)		6 (3.6%)
	None-brisk	.	15 (3.1%)		1 (0.6%)
	None-sluggish		2 (0.4%)		10 (6.3%)
	None-none		23 (4.7%)		84 (52.8%)

Table 60 The frequency/median of various variables across survivals and non-survivals *(continued)*

Variable		*Survival*		*Non-survivals*	
		Median	*Frequency (percentage)*	*Median*	*Frequency (percentage)*
Extracranial injury (AIS > 2)			546 (91.2%)		140 (69%)
Extracranial injury (AIS > 3)					
Systolic blood pressure *(continuous)*		138		129	
Systolic blood pressure (categorical)	**low bllod pressure (< 120 mmHg)**		142 (67.9%)		67 (32.1%)
	Normotension (120-150 mmHg)		290 (85%)		51 (15%)
	Hypertension (> 150 mmHg)		151 (69.9%)		65 (30.1%)
Mean blood pressure (continuous)		145.68		135.8	
Mean blood pressure (categorical)	**Hypotension (< 85 mmHg)**		6 (20%)		24 (80%)
	Normotension (85-110 mmHg)		47 (69.1%)		21 (30.9%)
	Hypertension (> 110 mmHg)		522 (79.5%)		135 (20.5%)
Hypoxia		98		93S	
Highest AIS score	**3**		21 (11.1%)		168 (88.9%)
	4		46 (14.6%)		270 (85.4%)
	5		131 (45%)		160 (55%)
	6		5 (83.3%)		1 (16.7%)
Contusion			247 (77.2%)		73 (22.8%)
Intracranial haemorrhage			443 (72.3%)		170 (27.7%)
EDH			84 (88.4%)		11 (11.6%)
SDH			137 (77%)		41 (23%)
SAH			97 (65.1%)		52 (34.9%)

Table 60 The frequency/median of various variables across survivals and non-survivals *(continued)*

Variable		*Survival*		*Non-survivals*	
		Median	***Frequency*** ***(percentage)***	***Median***	***Frequency*** ***(percentage)***
Marshal Classification	**I**		56 (86.2%)		9 (13.8%)
	II		356 (87.9%)		49 (12.1%)
	III		60 (70.6%)		25 (29.4%)
	IV		40 (54.1%)		34 (45.9%)
	V/VI		84 (50.6%)		82 (49.4%)
	Brain stem/cerebellar injury		1 (25%)		3 (75%)
	Penetrating injury		2 (66.7%)		1 (33.3%)

Table 60 The frequency/median of various variables across survivals and non-survivals *(continued).*

Table 61 presents the comparison between those cases in the IMPACT dataset which had one or more missing value (excluded cases) and those which had all the data available. As seen the differences are significant across all variables (age, GCS, pupillary reactivity, cause of injury, brain swelling) except extracranial surgery, hypoxia and systolic blood pressure (hypoxia showed a marginal difference i.e. p value < 0.10 and more than 0.05).

		Excluded cases (5542)	*Included cases (5481)*	*p value*
Age		32 (22-48)	30 (21-43)	0.00
GCS	**Mild**	1.8% (100)	6.9% (378)	0.00
	Moderate	6% (333)	10.7% (585)	
	Severe	34% (1883)	82.4 (4514)	
	Missing	58.2% (3230)	-	
Pupillary reactivity	**Both reacting**	64.5% (3575)	68.7% (3765)	0.00
	One reacting	16.4% (915)	12.1% (660)	
	Neither reacting	19.1% (1059)	19.2% (1052)	
Cause of injury	**RTC**	57.3% (3176)	63.2% (3459)	0.00
	Assault	5.2% (289)	7.7% (423)	
	Fall	22.6% (1251)	17.5% (959)	
	Others	14.7% (813)	11.6% (636)	
	Missing	0.3% (17)	-	
Extracranial surgery	**yes**	8.7% (481)	20.2% (1105)	0.543
	Missing	55.7% (3090)	-	
Brain swelling		17.5% (972)	21.9% (1197)	0.00
Systolic blood pressure	**Hypotension**	49% (1291)	51% (1344)	0.418
	Normotension	49.9% (4185)	50.1% (4202)	
Hypoxia		22.4% (1243)	21.1% (1153)	0.083

Table 61 Comparison of the cases which had all data recorded (included) to those cases which had one or more missing value across various variables in IMPACT.

Table 62 presents the comparison between the IMPACT and the TARN datasets. According to this table, the differences between all variables (age, GCS, pupillary reactivity, cause of injury, extracranial injury, brain swelling, hypoxia and systolic blood pressure) are significant as per the p values apart from survival rates. The IMPACT dataset appears to contain slightly younger population of TBI cases. Furthermore, based on GCS, the TARN dataset included less severe cases since 35.8% of cases were recorded as mild GCS in contrast to this being 6.1% in IMPACT. This is also the case for pupillary reactivity as more cases in TARN dataset had both reactive pupils (79.2% versus 66.6%). Furthermore, according to hypoxia, the cases in IMPACT appear to have been more physiologically disturbed than those in the TARN dataset.

		The IMPACT dataset	*The TARN dataset*	*P value*
Age		31 (22-46)	39 (22-58)	<0.005
GCS	**Mild**	478 (6.1%)	278 (35.8%)	<0.005
	Moderate	918 (11.8%)	99 (12.7%)	
	Severe	6397 (82.1%)	399 (51.7%)	
Pupillary reactivity	**Both reacting**	7340 (66.6%)	515 (79.2%)	<0.005
	One reacting	1572 (14.3%)	28 (4.3%)	
	Neither reacting	2111 (19.2%)	107 (16.5%)	
Cause of injury	**RTC**	6949 (58.8%)	314 (39.2%)	<0.005
	Assault	2353 (8.7%)	313 (39.0%)	
	Fall	1025 (19.9%)	143 (17.8%)	
	Others	1481 (12.5%)	32 (4%)	
Extracranial injury		1702 (19.5%)	116 (14.5%)	<0.005
Brain swelling		2169 (19.7%)	275 (34.3%)	<0.005
Hypoxia		2396 (21.7%)	51 (6.5%)	<0.005
Systolic blood pressure	**Hypotension**	2635 (92.5%)	215 (7.5%)	<0.005
	Normotension	8387 (93.7%)	566 (6.3%)	
Survival		8149 (73.9%)	599 (74.7%)	0.635

Table 62 Comparison of the IMPACT and the TARN datasets across various variables included in the TARN model B.

Table 63 presents the differences between the TARN internal (derivation) dataset and the TARN external dataset. The two datasets appear significantly different across every variable apart from ISS as per the p values.

		Internal TARN dataset	*External TARN dataset*	*P value*
Age		39 (22-58)	43.9 (24.2-67)	<0.005
GCS	**Mild**	278 (35.8%)	758 (57%)	0.021
	Moderate	99 (12.7%)	159 (12%)	
	Severe	399 (51.7%)	412 (31%)	
Pupillary reactivity	**Normal**	541 (67.5%)	821 (80.7%)	<0.005
	Abnormal-both reactive	94 (11.7%)	17 (1.7%)	
	Only one reactive	39 (4.9%)	51 (5%)	
	None reactive.	128 (16%)	128 (12.6%)	
Cause of injury	**RTC**	314 (39.2%)	448 (40%)	<0.005
	Assault	313 (39.0%)	1 (0.1%)	
	Fall	143 (17.8%)	664 (59.3%)	
	Others	32 (4%)	6 (0.5%)	
ISS		25 (16-29)	25 (16-29)	0.452
Extracranial injury		116 (14.5%)	123 (8.9%)	<0.005
Brain swelling		275 (34.3%)	415 (29.9%)	0.034
Brain stem injury			92 (6.6%)	0.045
Hypoxia		51 (6.5%)	59 (4.7%)	<0.005
Systolic blood pressure	**Hypotension**	215 (7.5%)	304 (22.5%)	<0.005
	Normotension	566 (6.3%)	1050 (77.5%)	<0.005
Survival		599 (74.7%)	1113 (80.2%)	<0.005

Table 63 Comparing characteristics of patients between the TARN internal and external datasets.

8. Paper 7: Comparing the Prognostic Performance of S100B with Prognostic Models in Traumatic Brain Injury

Authors

- Mehdi Moazzez Lesko
- Timothy Rainey
- Omar Bouamra
- Sarah O’Brien
- Charmaine Childs
- Fiona Lecky

8.1. Abstract

8.1.1. Introduction

There are currently two prognostic tools available for predicting outcome in Traumatic Brain Injury (TBI). The first involves prognostic models combining clinico-demographic characteristics of patients for outcome prediction, whilst the second employs serum brain injury biomarkers. S100B is a widely-acknowledged biomarker of brain injury.

8.1.2. Objective

To identify which method has better prognostic strength and explore how combining these methods might improve the prognostic strength.

8.1.3. Methods

We analysed data from 100 TBI patients, all of whom were admitted to the intensive care unit and had arterial S100B levels recorded at 24-hours after injury. TBI prognostic models A and B, constructed in Trauma Audit and Research Network (TARN), were run on the dataset and then S100B was added as an independent predictor to each model. Furthermore, another model was developed containing only S100B and subsequently, other important TBI predictors were added to assess their ability to enhance the predictive power of this model. The outcome measures were survival and favourable outcome at three months.

8.1.4. Results

Among all the prognostic variables (including age, cause of injury, GCS, pupillary reactivity, Injury Severity Score (ISS) and CT features); S100B has the highest predictive strength on multivariate analysis. No difference between performance of prognostic models or S100B in isolation is observed. Addition of S100B to the prognostic models improves the performance (e.g. Area Under the roc Curve (AUC), R^2 Nagelkerke and classification accuracy of TARN model A to predict survival increase respectively from 0.64, 0.10 and 71% to 0.72, 0.20 and 74.7%). Similarly, the predictive power of S100B increases by adding other predictors to S100B (e.g. AUC (0.69 versus 0.78), R^2 Nagelkerke (0.15 versus 0.30) and classification accuracy (73% versus 77%) for survival prediction).

8.1.5. Conclusion

S100B appears to be the strongest prognostic variable in TBI. A better prognostic tool than those which are currently available may be a combination of both clinic-demographic predictors with S100B.

8.2. Introduction

Outcome prediction in Traumatic Brain Injury (TBI) is one of the many factors taken into account by clinicians for the provision of acute and rehabilitative care to the sufferers. The prediction is made based on the relationship between the outcome and indicators of injury severity and should provide the likelihood of an individual patient experiencing various types of outcomes such as survival versus death. There are currently two prognostic tools available to make such prediction: serum biomarkers of brain injury that relate to outcome or prognostic models that incorporate various clinico-demographic factors (routinely measured characteristics such age, Glasgow Coma Scale (GCS), pupillary reactivity, Computed Tomography (CT features) etc.) to calculate the probability of a given outcome. So far several biomarkers have been proposed which show higher concentrations in blood of those TBI patients experiencing an unfavourable outcome [164]. However measurement of S100B is one of the brain injury biomarkers which has received more attention. This astroglial protein [47] has been demonstrated to be associated with outcome in TBI [43, 69, 75, 78, 80, 83, 98, 126, 128, 165, 166].

The association of S100B with outcome has prompted the researchers to determine serum cut-off levels, which can be used as a 'diagnostic' test for a given outcome of interest such as survival or disability. We have recently reported our study on the effectiveness of serum S100B measured 24-hours after injury to predict unfavourable outcome i.e. Glasgow Outcome Scale (GOS < 4) or death at 3-months after injury in a subset of TBI patients who were all admitted in the Intensive Care Unit (ICU) at Salford Royal NHS Hospital, UK [98]. In this study, S100B levels above the cut-off point 0.53 µg/l

had a sensitivity of 82% and 83% to respectively predict unfavourable outcome and mortality. However, the specificity was moderate at 60% for unfavourable outcome and 49% for mortality (Area Under the Roc Curve (AUC) for unfavourable outcome prediction: 0.77, AUC for mortality prediction: 0.69). Vos *et al.* also found similar results in a cohort of severe TBI patients, the cut off of 1.13 µg/l being higher as an earlier admission sample [43]. They reported an AUC of 67.7 to predict poor outcome 6 months after severe TBI. Similarly, other cut-offs for TBI outcome prediction has been suggested by other researchers [69, 75, 76, 78, 167]. It may be that considering other TBI prognosticators such as age or GCS would enhance the prognostic power of S100B.

Prognostic models are often constructed through logistic regression run on large datasets of TBI cases which hold information on various clinico-demographic characteristics or laboratory measurements. A "model" refers to an equation which calculates the probability of a given outcome by summing scores attributed to each predictor found to be important in the multivariate analysis. Thus the prognostic models are in fact equations providing the probability of outcome using given patient's characteristics. Two important such models proposed for international application are the International Mission for Prognosis and Analysis of Clinical Trials (IMPACT) models [24] and the Corticosteroid Randomisation After Significant Head injury (CRASH) [23]. The IMPACT model is presented in a user friendly fashion in that it attributes a score to each important prognosticator and then the probability of death or unfavourable outcome at the time point of 6 months can be calculated. The CRASH model is pertinent to low and middle income countries as the data

from TBI patients in these nations are contained in the derivation dataset and separate models applicable to these nations have been presented. An online calculator for either of these models is available [29, 34]. Similar to these international models, the Trauma Audit and Research Network (TARN) [89] has proposed its own TBI models (models A and B) suitable for the British trauma population at a national level [168].

Due to the limited literature on comparing prognostic models with brain injury biomarkers, it is still unclear which TBI prognostic tool (a brain biomarker or a prognostic model) is superior. Whilst prognostic models appear to have received more popularity than biomarkers (and S100B) in their applicability, the AUC of these models may still not be adequate for clinical purposes. For example, the highest AUC of the IMPACT models is 0.87. Comparing this AUC with a more familiar diagnostic tool such as mammography for breast cancer diagnosis which has AUC of about 0.85 [110], may indicate prognostic models are still required to improve in their performance. However, there are studies which have obtained very high AUCs of close to 0.90 [75, 78] for S100B unlike some other studies on S100B which found much lower AUCs of less than 0.70 [76, 79]. This can be due to the difference in case-mix across studies which per se highlights the importance of considering other TBI outcome predictors in assessing S100B prognostic strength. In fact, this is the point where prognostic modelling through multivariate analysis meets research into S100B.

The overall objective of this study is threefold: to compare the performance of these two prognostic tools (prognostic models and a brain injury biomarker), to investigate how adding S100B to TBI prognostic models

improves their performance and to investigate how combinations of S100B with other predictors may improve the predictive strength of S100B.

8.3. Methods

8.3.1. Data collection

For the purpose of this study, we used an available dataset from a study of 100 TBI patients conducted previously. The study dataset had been accrued prospectively to assess the prognostic performance of S100B on its own in a univariate analysis. Patients, who had been admitted within 24 hours after the injury, were recruited from ICU at Salford Royal Hosptial, UK. The dataset contained information on gender, age, Injury severity Score (ISS), cause of injury, Computed Tomography (CT) descriptions, serum S100B levels and outcome as measured by GOS at 3 months following injury. Serum blood samples had been taken at 24 hours after the injury time and were analysed using a one-step immunoassay (enzyme-linked immunosorbant assay, ELIZA; Sangtec 100^{TM} Diasorin, Wokingham, UK) [98].

The inclusion criteria were all TBI patient admitted to Hope ICU who were older than 16 years old {Rainey, 2009 #210}. Exclusion criteria were all patients who arrived at ICU more than 24 hours after their injury and in case the consent was not performed. After obtaining the blood sample at 24 hours following the injury, the legally acceptable relative was consented as the patient was still unconscious. Had the patient died meanwhile then the relative would not be approached. Three months later the patient or the consenting relative were contacted to assess the patient's outcome. Prior to this contact, the casenote of the patient was reviewed to ascertain if the patient had died during his stay at hospital. All patients were at home at 3 months after their injury, had they been still alive.

Important predictors of outcome in TBI were selected following a literature review: *clinico-demographic characteristics* namely age [23, 24], GCS [23, 24], pupillary reactivity [23, 24], cause of injury [92], ISS [31, 92, 93], CT features [23, 24, 94], extracranial injury [23], *vital signs* including mean Blood Pressure (BP) [92, 95], systolic BP [92, 95], temperature [92, 96], and *laboratory measures* including PH [92, 97], Haemoglobin (Hb) [92, 97], Glucose [92, 97], Platelet (Plt) count [92, 97] and Prothrombin Time (PT) [92, 97] along with O_2 saturation (O_2 Sat.) [92, 96] and Intracranial Pressure (ICP) [35]. Although some of these variables were available in the dataset, a search of patients' records was needed to fill the gaps. Ethical approval for this was obtained from the Salford and Trafford Research Ethics Committee (reference 05/Q1404/157(amendment 2007)). The potential sources for this extra information were case notes, regional electronic records of patients (either the hospital Electronic Patient Record (EPR) or ICU) and TARN, which is a trauma registry based at Salford Royal NHS Hospital, UK. The time point for measurement of clinico-demographic characteristics was admission records (GCS and pupillary reactivity). For vitals signs, laboratory values, O_2 Sat. and ICP, the nearest observation to 24 hours after injury was used. The time point of 24 hours post injury was when the blood sample was taken for S100B assay in the original study.

8.3.2. Univariate analysis

The effect of each covariate on survival and favourable outcome was assessed using Mann Whitney U test for continuous variables and Chi square test for categorical variables with the significant level of 5%. Unlike some variables

which were clearly categorical, e.g. pupillary reactivity, the decision to use other variables categorically or continuously was based on whether or not there was a linear relationship with log odds of survival or favourable outcome (linearity assumption). This was a preliminary requirement for the multivariate statistical analysis selected for this study i.e. logistic regression [134]. Fractional polynomials analysis [135] was applied to assess continuous variable linearity. If there was no linearity, the relevant variable was then used categorically with the best cut-offs adopted from the literature. Subsequently the significance of association with outcome was tested on the dataset by Chi square test. If the variable demonstrated significant association in its continuous form with no linearity with log odds of outcome, trials of various cut-offs were performed to obtain the categorical significance as well.

8.3.3. Multivariate analysis

The performance of TARN TBI prognostic models (models A and B) [168] was assessed on the study dataset since these models are suitable for our dataset in that all cases have received British trauma care and were nursed in the ICU. Overall 7 models were constructed. *Model A without S100B* and *model B without S100B* were constructed using the covariates from TARN TBI models A or B through logistic regression. Following the derivation of these models, S100B was added to each model to construct two further models (*model A with S100B* and *model B with S100B*). Similarly, a model was derived only including S100B *(S100B model).* Subsequently, predictors from TARN TBI models A and B and those variables which were found significant per univariate analysis but not contained in the TARN models were added to the S100B model. This resulted in construction of two more models: *expanded*

S100B model A and *expanded S100B model B*. Each model was run twice; once for survival prediction and once for favourable outcome prediction.

The performance of each model was assessed using three measures i.e. AUC, Nagelkerke R^2 [136] and classification accuracy. Then for each research objective, the performance of pairs of models was compared for each measure of performance. The pair of models compared are presented in Table 64. For example, the performance of the model A without S100B was compared with the model A with S100B per three measures of performance to address the research objective as to the value of adding S100B to a TBI prognostic model. The difference between each measure of performance was considered high (or clinically significant) if it was more than 0.05 for AUC and Nagelkerke R^2 and more than 10% for classification accuracy. These cut-offs were arbitrary and were chosen in order not to take small increases in performance measures into account.

Objective	*Models compared*
Comparing performance of S100B with multivariate prognostic models	• Model A without S100B *versus* S100B model • Model B without S100B *versus* S100B model
Assessing the added value of S100B to prognostic models	• Model A without S100B *versus* model A with S100B • Model B without S100B *versus* model B with S100B
Assessing S100B performance after adjustment with other predictors (the added value of other prognosticators to S100B)	• S100B model *versus* expanded S100B model A • S100B model *versus* expanded S100B model B

Table 64 Various pairs of models compared according to research objectives. Each model was run twice; once for survival prediction and once for favourable outcome prediction.

Models A and B (with and without S100B) were derived through logistic regression (enter method) in Statistical Package for the Social Sciences (SPSS) 15 for windows. However, the expanded S100B models A and B were derived through stepwise method.

8.4. Results

Table 65 presents the results of the univariate analyses for clinico-demographic characteristics and S100B (median/frequency, odds ratio and significance of association with survival or favourable outcome). The dataset contains 100 TBI patients with median age of 31 and male to female ratio of 81/19. Neither age nor sex is significantly associated with survival or favourable outcome. However, GCS is significantly associated with either outcome (i.e. survival and favourable outcome) both continuously and categorically when the cut-off for categorization is 9 (i.e. severe versus non-severe brain injury). Nevertheless, GCS with the cut-offs of 9 and 12 (i.e. severe versus moderate versus mild brain injury) does not demonstrate significant association with outcome. The other variables which are significantly associated with survival are ISS (in its categorical form) and S100B. For favourable outcome, apart from continuous and categorical GCS, age and S100B are significant.

				Median or Frequency (%)	*Odds ratio*		*P value (Survival)*	*P value (Favourable outcome)*
					Survival	*Favourable outcome*		
Age				31	0.99 (0.97-1.02)	0.96 (0.94-0.98)	0.45	**<0.001**
Male/female				81%/19%	1.95 (0.69-5.48)	1.48 (0.54-4.06)	0.2	0.44
GCS	**Continuous**			8	1.14 (0.98-1.33)	1.14 (1.01-1.29)	**0.02***	**0.02**
	Categorical	**Cut-offs 9 and 12**	**mild**	17%			0.11	0.11
			moderate	19%	0.71 (0.10-4.86)	1.18 (0.29-4.73)		
			severe	54%	0.27 (0.05-1.29)	0.47 (0.15-1.45)		
		Cut-off 9 (Severe GCS)		54%	0.32 (0.11-0.97)	0.43 (0.18-1.03)	**0.03**	**0.05**
Pupillary reactivity		**Both reacting**		53%			0.25	0.09
		Only one reacting		6%	0.26 (0.05-1.48)	0.12 (0.01-1.11)		
		Neither reacting		16%	1.13 (0.27-470)	1.01 (0.32-3.20)		

Table 65 Clinicodemographic characteristics of the study patients. *Despite the significant P value, the CI for odds ratio does include 1 which implies a non-significant association. The reason for this discrepancy is that odds ratios were obtained through logistic regression whereas the P value was derived through Mann Whitney U test.

			Median or Frequency (%)	*Odds ratio*		*P value (Survival)*	*P value (Favourable outcome)*
				Survival	*Favourable outcome*		
Cause of injury	**RTC**		43%			0.47	0.15
	Fall		42%	0.62 (0.24-1.57)	0.79 (0.33-1.85)		
	Assault		11%	1.55 (0.29-8.29)	4.71 (0.91-24.42)		
	Sports		4%	0.34 (0.03-2.74)	1.05 (0.13-8.13)		
ISS	**Continuous**		25	1.00 (0.94-1.05)	1.00 (0.95-1.05)	0.74	0.85
	Categorical	**3-24**	27%			**0.01**	0.26
		25-75	73%	0.21 (0.06-0.77)	0.60 (0.24-1.47)		
Extracranial injury			5%	1.83 (0.66-5.12)	2.25 (0.91-5.54)	0.24	0.07
S100B			0.7	0.49 (0.30-0.80)	0.21 (0.10-0.44)	**<0.005**	**<0.005**

Table 65 Clinicodemographic characteristics of the study patients. *Despite the significant P value, the CI for odds ratio does include 1 which implies a non-significant association. The reason for this discrepancy is that odds ratios were obtained through logistic regression whereas the P value was derived through Mann Whitney U test.

Table 66 presents the results of univariate analysis for various CT findings. Among all CT features only the presence of compressed cisterns is significantly associated with survival. Similarly, for favourable outcome prediction, only the presence of mass lesion is significantly associated with outcome.

		Median or Frequency (%)	*Odds ratio*		*P value (Survival)*	*P value (Favourable outcome)*
			Survival	*Favourable outcome*		
Marshal CT classification	**II**	6%			0.45	0.08
	V	33%	0.62 (0.06-6.17)	0.87 (0.14-5.51)		
	VI	61%	0.38 (0.04-3.48)	0.35 (0.06-2.04)		
Mass lesion		89%	0.21 (0.02-1.69	0.19 (0.04-0.93)	0.1	**0.02**
Haemorrhage		94%		0.18 (0.02-1.63	0.1	**0.09**
Contusion		42%	0.76 (0.32-1.804)	0.61 (0.27-1.35)	0.54	**0.22**

Table 66 CT findings of the population studied.

	Median or Frequency (%)	*Odds ratio*		*P value (Survival)*	*P value (Favourable outcome)*
		Survival	*Favourable outcome*		
SAH	26%	0.95 (0.36-2.51)	0.81 (0.33-1.99)	0.92	0.65
SDH	43%	0.452 (0.19-1.08)	0.48 (0.21-1.06)	0.07	0.07
EDH	20%	0.57 (0.20-1.58)	0.78 (0.29-2.08)	0.27	0.62
Brain swelling	15%	0.83 (0.26-2.69)	1.17 (0.39-3.51)	0.76	0.78
Midline shift	22%	0.69 (0.253-1.87)	0.79 (0.31-2.05)	0.46	0.63
Compressed Cisterns	6%	0.19 (0.03-1.11)	0.18 (0.02-1.63)	**0.04**	0.09

Table 66 CT findings of the population studied. ***(continued)***

Table 67 (Page 370, Page 371, Page 372 and Page 373) presents the results of univariate analyses for vital signs and laboratory measurements. Among all the variables (mean and systolic BP, temperature, PH, Hb, Glucose, Plt count, PT, hypoxia and ICP), only ICP in both categorical and continuous form holds a significant relationship with survival but is unrelated to favourable outcome.

			Median or Frequency%	*Odds ratio*		*P value (Survival)*	*P value (Favourable Outcome)*
				Survival	*Favourable outcome*		
Mean BP [95]		**Hypotensive (< 85 mmHg)**	31%	1.5 (0.40-2.76)	2.69 (1.09-6.65)	0.82	0.1
		Normotensive (85-110 mmHg)	49%				
		Hypertensive (> 110 mmHg)	14%	2.04 (0.21-20.05)	1.82 (0.31-10.58)		
Systolic BP [95]		**Hypotensive (< 120 mmHg)**	57%	0.59 (0.22-1.57)	0.99 (0.40-2.44)	0.48	0.19
		Normotensive (120-150 mmHg)	31%		-		
		Hypertensive (> 150 mmHg)	6%	1.19 (0.28-4.98)	0.33 (0.09-1.18)		
Temperature [35]		**Normothermia**	55%			0.2	0.08
		Hyperthermia (> 38)	19%	0.42 (0.11-1.63)	0.38 (0.13-1.16)		
PH		**Continuous**	7.4	0.23 (0.00-56.99)	2.19 (0.01-489.22)	0.6	0.78
	categorical[§]	**normal**	34%				
		Alkalosis (> 7.45)	24%	0.69 (0.23-2.11)	1.33 (0.47-3.79)	0.518	0.594

Table 67 Vital signs and laboratory measurements of the study patients. The citations refer to the literature where the cut-offs were obtained. [§] There were no cases recorded as acidosis (PH < 7.35).

				Median or Frequency %	*Odds ratio*		*Survival*	*Favourable outcome*
					Survival	*Favourable outcome*		
Hb		**Continuous**		99.5	1.01 (0.99-1.03)	1.00 (0.98-1.02)	0.41	0.87
	Categorical	**Anemia (< 135 mmg/dl in makes and < 116 mmg/dl in females) [169]**		88%	0.40 (0.07-2.10)	0.44 (0.08-2.51)	0.26	0.34
		normal		6%				
Glucose		**Continuous**		6.7	1.02 (0.96-1.08)	1.00 (0.96-1.05)	0.97	0.32
	Categorical	**Cut-offs of 2.5 and 11 [170, 171]**	**Hypoglycaemia**	-			0.84	0.34
			Normoglycemia	88%				
			Hyperglycaemia	6%	0.84 (0.14-4.87)	0.44 (0.08-2.51)		

Table 67 Vital signs and laboratory measurements of the study patients. The citations refer to the literature where the cut-offs were obtained. [§] There were no cases recorded as acidosis (PH < 7.35). *(continued)*

				Median or Frequency%	*Odds ratio*		*Survival*	*Favourable outcome*
					Survival	*Favourable outcome*		
Glucose *(Continued from the previous page)*	**Categorical**	**Cut-off of 6 [97]**	**Hypoglycaemia**	28%			0.085	0.278
			Hyperglycaemia	66%	1.17 (0.45-3.05)	0.61 (0.25-1.50)		
Plt	**Continuous**			150	1.00 (0.99-1.01)	1.00 (0.99-1.00)	0.46	0.97
	Categorical	**Cut-offs of 100 and 450 [172] ¥**	**Thrombocytopenia**	16%	2.11 (0.70-6.39)	1.500 (0.51-4.43)	0.18	0.46
			Normal	78%				
		Cut-off of 200 [97]	**thrombocytopenia**	74%	1.35 (0.44-4.17)	0.69 (0.26-1.88)	0.6	0.52
			normal	20%				

Table 67 Vital signs and laboratory measurements of the study patients. The citations refer to the literature where the cut-offs were obtained. § There were no cases recorded as acidosis (PH < 7.35).¥: there were no cases with thrombosis (i.e. Plt count > 450) ***(continued)***

			Median or Frequency%	*Odds ratio*		*Survival*	*Favourable outcome*
				Survival	***Favourable outcome***		
	PT		17.7	0.91 (0.75-1.11)	1.02 (0.86-1.22)	0.36	0.8
	Hypoxia		0			-	-
ICP	**Continuous**		13	0.91 (0.85-0.97)	0.94 (0.89-0.99)	**<0.005**	0.06
	Categorical	**Normal**	67%			**<0.005**	0.08
		Increased (> 20 mmHg) [35]	19%	0.22 (0.08-060)	0.43 (0.17-1.11)		

Table 67 Vital signs and laboratory measurements of the study patients. The citations refer to the literature where the cut-offs were obtained.*(continued)*

8.4.1. Importance of prognosticators in multivariate models

Table 68 (Page 375 and Page 376) and Table 69 (Page 377 and Page 378) display the odds ratios along with the significance of association on each covariate with survival (Table 68) and favourable outcome (Table 69) in the models which were constructed.

		Model A without S100B	*Model A with S100B*	*Model B without S100B*	*Model B with S100B*	*S100B Model*	*Expanded S100B model (A)*	*Expanded S100B model (B)*
				Survival				
Age		0.99 (0.96-1.01)	1.01 (0.98-1.04)	0.99 (0.95-1.02)	1.01 (0.97-1.06)			1.01 (0.97-1.06)
Severe GCS		0.33 (0.09-1.27)	0.42 (0.11-1.69)	0.32 (0.08-1.28)	0.42 (0.10-1.74)			
Pupillary reactivity	**Both reactive**							
	Only one reactive	0.44 (0.11-1.75)	0.46 (0.17-2.01)	0.4 (0.10-1.66)	0.42 (0.09-1.97)		0.43 (0.10-1.94)	0.35 (0.07-1.66)
	Neither reactive	0.98 (0.32-2.94)	1.65 (0.47-5.76)	0.98 (0.31-3.06)	1.73 (0.47-6.37)		1.92 (0.50-7.28)	2.12 (0.55-8.11)
ISS		0.49 (0.10-2.31)	0.88 (0.18-4.30)					

Table 68 Odds ratios and significance of relationships of each covariate in various models investigated for survival and favourable outcome prediction. The numbers in the parentheses are 95% confidence intervals. The significant associations are starred (*: marginally significant i.e. p value <0.10 but >0.05, **: significant i.e. p value < 0.05, ***: p value < 0.005).

		Model A without S100B	*Model A with S100B*	*Model B without S100B*	*Model B with S100B*	*S100B Model*	*Expanded S100B model (A)*	*Expanded S100B model (B)*
					Survival			
Cause of injury	**RTC**							
	Fall			0.47 (0.05-4.46)	0.7 (0.06-8.19)			0.8 (0.07-9.83)
	Assault			0.64 (0.00-235)	1.34 (0.003-581)			0.61 (0.001-255.)
	Sports			0.03 (0.00->10^3)	<0.005 (0.00->10^3)			0.003 (0.00->10^3)
Interaction of cause of injury and age	**RTC**							
	Fall			1.01 (0.96-1.07)	0.99 (0.94-1.05)			0.99 (0.93-1.05)
	Assault			1.03 (0.86-1.23)	0.99 (0.83-1.20)			1.02 (0.84-1.22)
	Sports			1.11 (0.44-2.8)	1.19 (0.47-2.98)			1.22 (0.48-3.08)
Brain swelling				0.63 (0.17-2.28)	0.69 (0.17-2.87)			1.22 (0.48-3.01)
Increased ICP							0.29** (0.09-0.88)	
Compressed cisterns							0.08 (0.01-0.59)**	0.1 (0.01-0.67)**
S100B			0.44 (0.25-0.80)**		0.39 (0.21-0.73)**	0.49 (0.30-0.80)***	0.42 (0.24-0.75)***	0.34 (0.18-0.66)***

Table 68 Odds ratios and significance of relationships of each covariate in various models investigated for survival and favourable outcome prediction. The numbers in the parentheses are 95% confidence intervals. The significant associations are starred (*: marginally significant i.e. p value <0.10 but >0.05, **: significant i.e. p value < 0.05, ***: p value < 0.005).

		Model A without S100B	*Model A with S100B*	*Model B without S100B*	*Model B with S100B*	*S100B model*	*Expanded S100B model (A)*	*Expanded S100B model (B)*
				Favourable outcome				
Age		0.95 (0.93-0.98)***	0.97 (0.94-0.10)**	0.94 (0.89-0.98)**	0.95 (0.89-1.01)*		0.97 (0.94-0.98)**	0.95 (0.89-1.01)*
Severe GCS		0.33 (0.10-1.10)*	0.37 (0.09-1.48)	0.28 (0.07-1.08)*	0.32 (0.07-1.51)		0.37 (0.09-1.49)	0.32 (0.07-1.52)
Pupillary reactivity	**Both reactive**							
	Only one reactive	0.11 (0.01-1.00)**	0.1 (0.01-0.94)**	0.07 (0.007-0.79)**	0.06 (0.004-0.79)**		0.1 (0.01-0.94)**	0.06 (0.004-0.79)**
	Neither reactive	0.61 (0.21-1.75)	1.14 (0.33-3.95)	0.63 (0.21-1.95)	1.26 (0.32-5.05)		1.15 (0.33-3.98)	1.27 (0.32-5.05)
ISS		0.41 (0.09-1.93)	0.81 (0.15-4.45)					
Cause of injury	**RTC**							
	Fall			0.48 (0.04-5.14)	0.41 (0.02-7.35)			0.41 (0.02-7.31)
	Assault			0.18 (0.00-104.82)	0.31 (0.00-296.42)			0.33 (0.00-294.278)
	Sports			0.02 (0.00->10^3)	<0.005 (0.00->10^3)			0.001 (0.00->10^3)

Table 69 Odds ratios and significance of relationships of each covariate in various models investigated for survival and favourable outcome prediction. The numbers in parantheses are confidence intervals. The significant associations are starred (*: marginally significant i.e. p value <0.10 but >0.05, **: significant i.e. p value < 0.05, ***: p value < 0.005) ***(continued)***

		Model A without S100B	*Model A with S100B*	*Model B without S100B*	*Model B with S100B*	*S100B model*	*Expanded S100B model (A)*	*Expanded S100B model (B)*
				Favourable outcome				
Interaction of cause of injury and age	**RTC**							
	Fall			1.03 (0.97-1.10)	1.03 (0.95-1.11)			1.03 (0.95-1.11)
	Assault			1.12 (0.91-1.38)	1.1 (0.88-1.36)			1.09 (0.88-1.36)
	Sports			1.15 (0.46-2.88)	1.32 (0.53-3.30)			1.32 (0.53-3.30)
Brain swelling				0.83 (0.23-3.07)	0.91 (0.21-4.08)			
Mass lesion								
S100B			0.22 (0.10-0.49)***		0.18 (0.07-0.45)***	0.21 (0.10-0.44)***	0.21 (0.10-0.48)***	0.18 (0.07-0.45)***

Table 69 Odds ratios and significance of relationships of each covariate in various models investigated for survival and favourable outcome prediction. The numbers in parantheses are confidence intervals. The significant associations are starred (*: marginally significant i.e. p value <0.10 but >0.05, **: significant i.e. p value < 0.05, ***: p value < 0.005) *(continued)*.

Models A and B with and without S100B: The blank cells related to these models in Table 68 and Table 69 are those variables which are not contained in the respective TARN TBI model. In both models A and B none of the covariates show significant associations with survival status and this still holds after S100B was added, whilst S100B is the only significant predictor in both models. Also adding S100B to these models resulted in a significant decrease in the deviance according to Chi square test. For favourable outcome prediction (Table 69), only pupillary reactivity competes with S100B in models A and B holding a statistically significant association. Age also appears significant in model A without S100B, Model A with S100B and model B without S100B but this is marginal in model B with S100B (5% < *p* value < 10%). The other marginally important factor is severe GCS in models without S100B.

S100B model and expanded S100B models A and B: Construction of expanded S100B models A and B was with the inclusion of covariates from the TARN TBI models A and B along with those predictors found significant in univariate analysis (Table 65 and Table 67); being presence/absence of compressed cisterns and ICP for survival and presence/absence of mass lesion for favourable outcome. The blank cells are those variables which were absent in TARN TBI models A and B, lack significance in univariate analysis or discarded during stepwise logistic regression. As seen, S100B on its own has a significant influence on outcome be it survival or favourable outcome and addition of other factors to S100B models does not change the significance of this influence. Amongst those covariates displayed for survival prediction, S100B is followed by compressed cisterns (in both expanded S100B models A

and B) and then increased ICP (in the expanded S100B model A). For prediction of favourable outcome, S100B still holds its dominant significance with pupillary reactivity as the second best predictor followed by age (significant in expanded S100B model A but marginally significant in expanded S100B model B).

8.4.2. Models performance

Table 70 presents the measures of performance of each model to compare the constructed models for survival or favourable outcome prediction. As seen, per AUC or R^2 Nagelkerke the worst performing models are those which do not include combinations of S100B with other TBI predictors (model A without S100B, model B without S100B and S100B model). According to this table, also the best models appear to be expanded S100B models A and B (S100B plus one or two other predictors).

	Classification accuracy	*AUC*	*Nagelkerke R^2*	*HL statistics*
		Survival		
Model A without S100B	71%	0.64 (0.52-0.76)	0.10	0.18
Model A with S100B	74.7%	0.72 (0.61-0.83)	0.20	0.56
Model B without S100B	70%	0.66 (0.55-0.76)	0.11	0.33
Model B with S100B	75%	0.77 (0.67-0.87)	0.25	0.54
S100B	73%	0.69 (0.57-0.80)	0.15	0.15
Expanded S100B model (A)	77%	0.78 (0.66-0.88)	0.30	0.59
Expanded S100B model (B)	77%	0.79 (0.70-0.89)	0.30	0.42

Table 70 Various measures of performance for each constructed model (survival).

.

	Classification accuracy	*AUC*	*Nagelkerke R^2*	*HL statistics*
		Favourable outcome		
Model A without S100B	67%	0.76 (0.67-0.85)	0.27	0.24
Model A with S100B	76%	0.84 (0.76-0.92)	0.46	0.25
Model B without S100B	70%	0.78 (0.70-0.87)	0.11	0.85
Model B with S100B	77%	0.86 (0.79-0.94)	0.52	0.94
S100B	67%	0.77 (0.68-0.86)	0.32	0.02
Expanded S100B model (A)	76%	0.84 (0.76-0.91)	0.46	0.73
Expanded S100B model (B)	77%	0.87 (0.80-0.94)	0.52	0.95

Table 70 Various measures of performance for each constructed model (favourable outcome).

Comparing the performance of S100B model with model A without S100B and model B without S100B (the first objective of the study): This can address the question as to which prognostic tool among S100B or a prognostic model is better. Figure 27 and Figure 28 show the ROC curves of these models to respectively predict survival and favourable outcome. According to the graphs,

the differences in AUCs do not appear significant (the constant line versus the dashed lines).

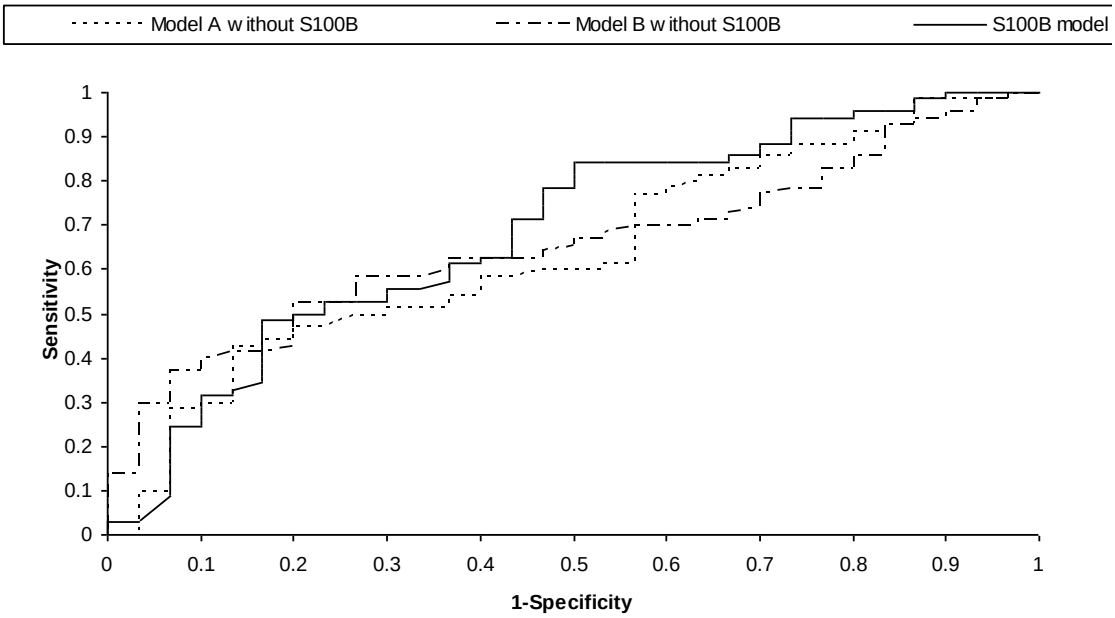


Figure 27 The ROC curves of models A and B without S100B and S100B model for survival prediction

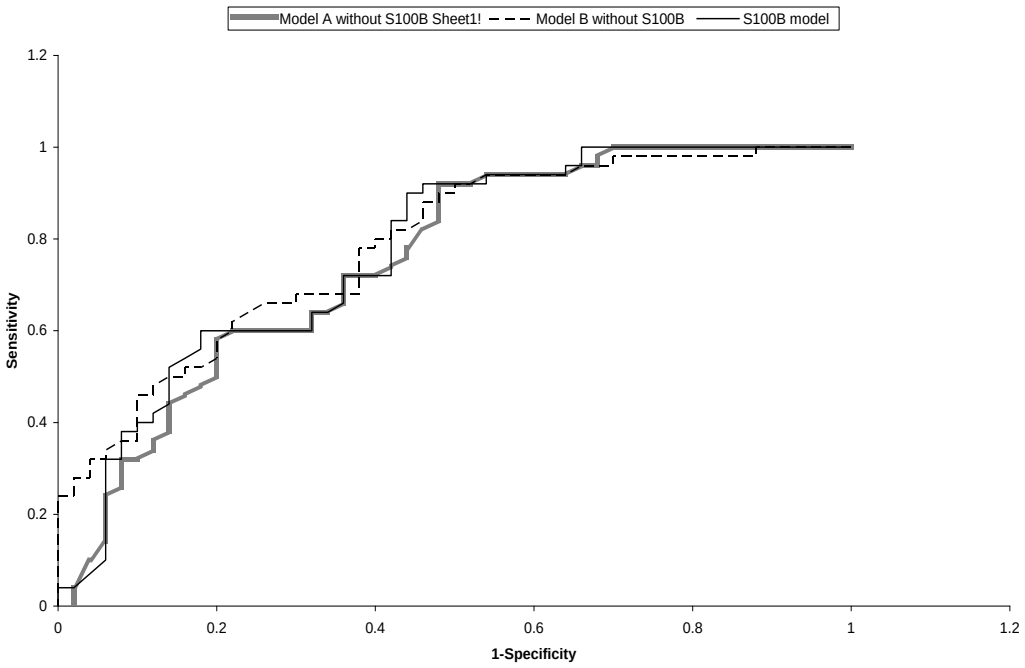


Figure 28 The ROC curves of models A and B without S100B and S100B model for favourable outcome prediction.

Table 71 provides the differences in models performance across various measures (for example: AUC of S100B model minus AUC of model A without S100B = 73% - 71% = 2% for survival prediction). According to this table, AUC of S100B model does not appear to be highly different to that from models A and B without S100B to predict either survival or favourable outcome (the difference less than our clinical significance of 0.05). This is also the case for Nagelkerke R^2 showing a difference of less than 0.05 and also for classification accuracy showing a difference of less than 10%.

	Classification accuracy	*AUC*	*R^2 Nagelkerke*
		Survival	
S100B *versus* model A without S100B	2%	0.05	0.05
S100B *versus* model B without S100B	3%	0.03	0.04
		Favourable outcome	
S100B *versus* model A without S100B	0	0.01	0.05
S100B *versus* model B without S100B	3%	0.01	0.21

Table 71 The performance of S100B model versus models A and B without S100B (the figures demonstrate the difference in the respective measure of performance across the compared models)

Changes in the performance of models A and B following the addition of S100B (the second objective of the study): This would address the added value of S100B to prognostic models. Figure 29 and Figure 30 show the ROCs curves of models to respectively predict survival and favourable outcome. According to graphs, adding S100B model to models A and B results in increase of AUC (thin lines versus the thick lines).

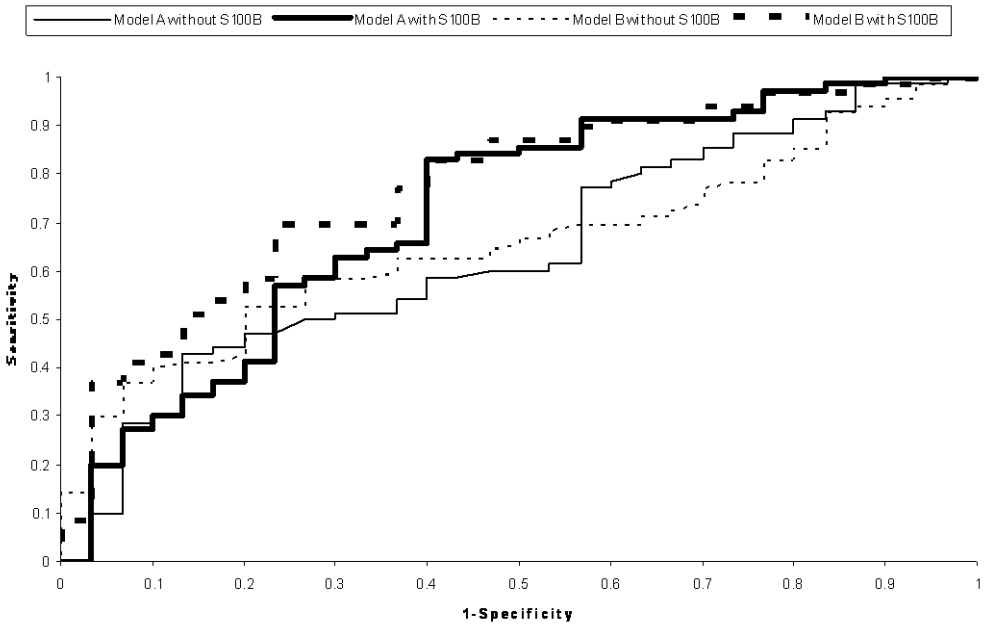


Figure 29 The ROC curves of models A and B with and without S100B for survival outcome prediction.

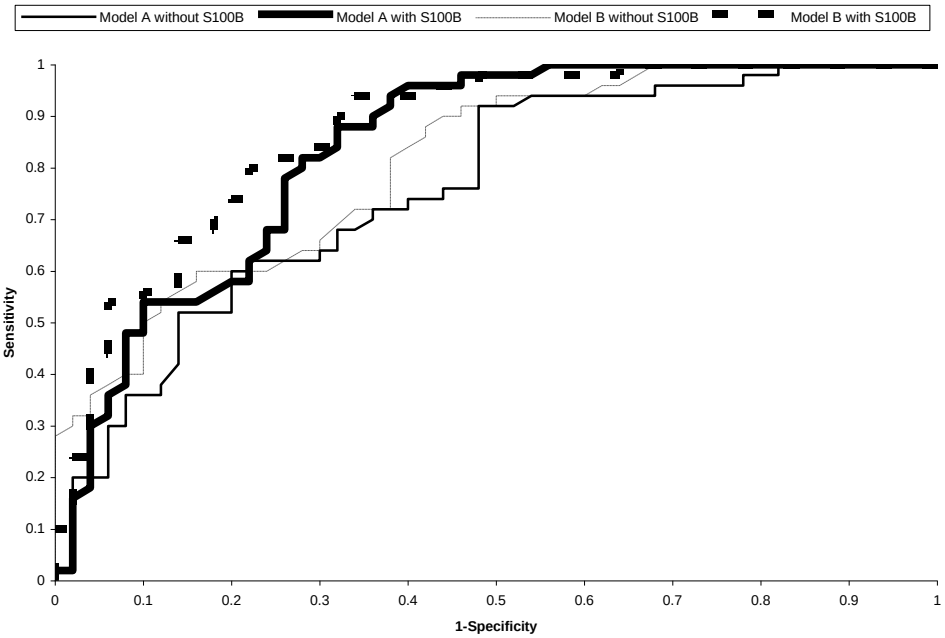


Figure 30 The ROC curves of models A and B with and without S100B for favourable outcome prediction.

Table 72 provides the differences in the performance of models prior and following addition of S100B. According to this table, the increases in model's AUC are clinically significant; being 0.08 for model A after adding S100B to predict survival or favourable outcome and also for model B to predict favourable outcome. For survival prediction, the increase in AUC of model B is even greater still at 0.11. Similarly, the increases in Nagelkerke R^2 with S100B are high in both models A and B for either outcome prediction. With regards to classification accuracy, this index of performance increases after addition of S100B, although the degree of increase is varied and not close to the clinical significance (i.e. an increase of 10%) at all times.

	Classification accuracy	*AUC*	*R^2 Nagelkerke*
		Survival	
Model A without S100B *versus* Model A with S100B	3.7%	0.08	0. 10
Model B without S100B *versus* Model B with S100B	5%	0.11	0.14
		Favourable outcome	
Model A without S100B *versus* Model A with S100B	9%	0.08	0.19
Model B without S100B *versus* Model B with S100B	7%	0.08	0.41

Table 72 Comparing the performance of Models A and B with and without S100B (the figures demonstrate the difference in the respective measure of performance across the models compared)

Changes in the performance of the S100B model following construction of expanded S100B model A or B (the third objective of the

study): This should be considered to assess how the prognostic strength of S100B would improve after other TBI prognosticators are taken into account. Figure 31 and Figure 32 demonstrate the ROC curves of these models for respectively survival and favourable outcome prediction. It can be observed that the AUC of S100B model appears to rise following addition of other TBI prognosticators (the dot line versus the constant lines).

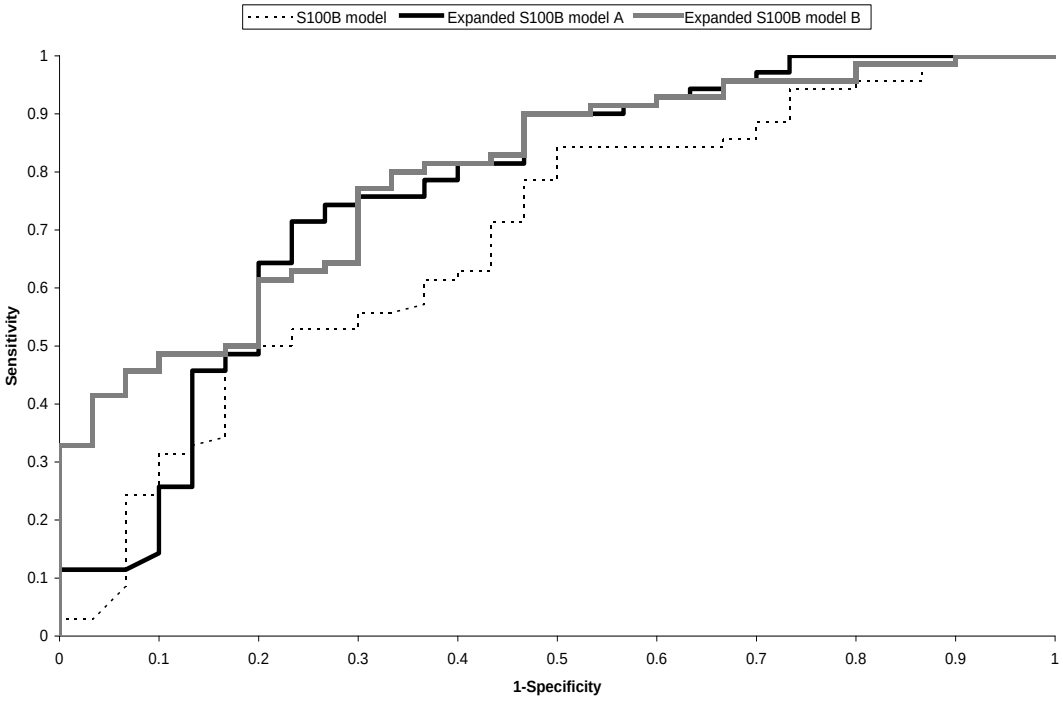


Figure 31 The ROC curves of S100B model and expanded S100B models A and B for survival prediction.

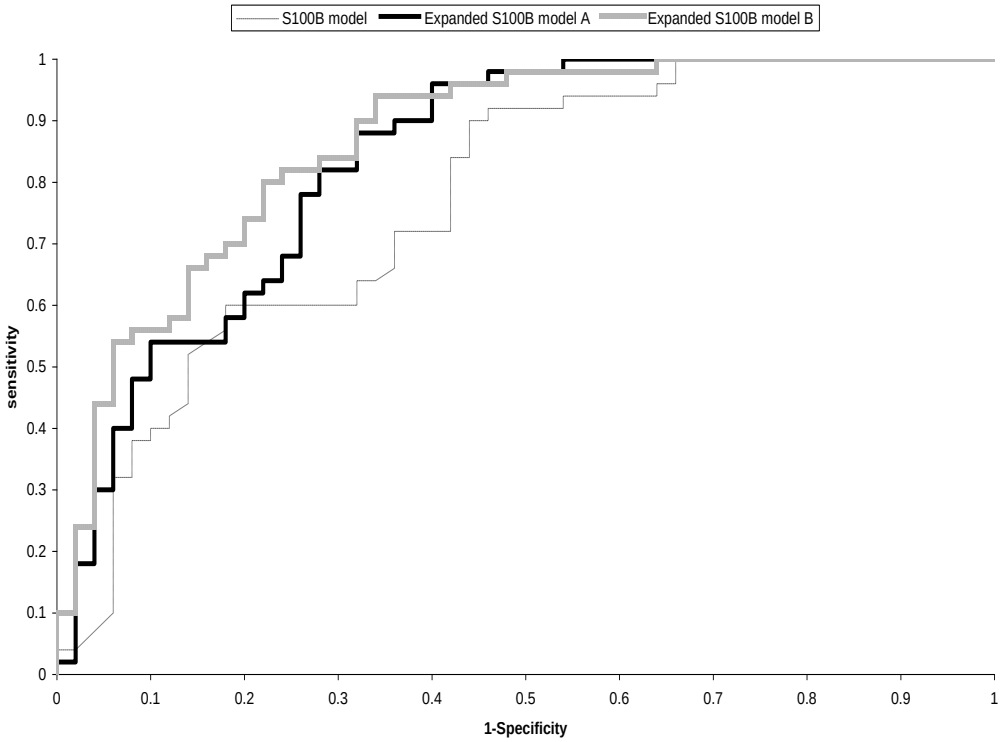


Figure 32 The ROC curves of S100B model and expanded S100B models A and B for favourable outcome prediction.

Table 73 gives the differences in model performance following adding other variables to the S100B model. The AUC of S100B model increases by 0.09 and 0.07 respectively following construction of expanded S100B model A for survival and favourable outcome prediction. This increase is also of similar magnitude for expanded S100B model B at 0.1 for both types of outcome prediction. Similarly, Nagelkerke R^2 increases by 0.15 and 0.14 in expanded S100B model A respectively for survival and favourable outcome prediction and this increase is still significant for the expanded S100B model B being 0.14 and 0.20. However, the change in classification accuracy for survival prediction is low (less than our clinical significance of 10%) in both expanded models whereas this is high for favourable outcome prediction.

	Classification accuracy	*AUC*	*R^2 Nagelkerke*
		Survival	
S100B model *versus* expanded S100B model A	4%	0.09	0.15
S100B model *versus* expanded S100B model B	4%	0.10	0.15
		Favourable outcome	
S100B model *versus* expanded S100B model A	9%	0.07	0.06
S100B model *versus* expanded S100B model B	10%	0.10	0.20

Table 73 The performance of S100B versus expanded S100B model A and B (the figures demonstrate the difference in the respective measure of performance across compared modesl).

8.5. Discussion

In this study, the two common prognostic tools (prognostic models and a brain injury biomarker) in TBI were compared in a multivariate analysis by constructing two types of models: prognostic models containing a composite of important clinical prognosticators and a model which only contains S100B. The results show that these two prognostic tools do not differ in their performance per AUC, Nagelkerke R^2 and classification accuracy. However, following addition of S100B to TBI prognostic models (TARN TBI models A and B), the performance of the models improve per AUC and Nagelkerke R^2. Furthermore, taking other important TBI prognosticators into account along with S100B in a prognostic model improves the S100B prognostic strength. In all models which contain S100B and other predictors, S100B appears to be the most reliable predictor, showing a significant effect on outcome at all times.

8.5.1. Limitations

Our series suggests that including S100B with two or three other important TBI prognosticators would provide a stronger prognostic tool than either alone. It was hoped that the results of the multivariate analysis of the S100B dataset would provide the precise information on the relative prognostic importance of each variable contained in the models. However, in this study the findings for survival prediction are significantly different to that for favourable outcome prediction and also to the literature [23, 24, 31, 92]. For example, only compressed cisterns (as in expanded models A and B) and the increased ICP (as in expanded S100B model A) were found to be significant, with all other variables being insignificant. For favourable outcome prediction,

despite pupillary reactivity being present in all constructed models for outcome prediction with age being present in 5 out of 7 models, compressed cisterns and increased ICP were never found to be significant. Overall, our results can not suggest the most important TBI prognosticators to be included in a 'small' prognostic model. This is because of the differences of our results regarding most important prognosticators between the two types of outcome and also because of the differences of the results with the literature which suggests important prognostic strength for age, GCS and pupillary reactivity. However, we believe our data suggest that S100B combined with 1 or 2 predictors out of age, pupillary reactivity and GCS is able to provide a strong prognostic tool in that these three variables are shared by currently well-developed prognostic models i.e. CRASH and IMPACT models.

Sample size

It should be noted that the objective of our study was not to construct a prognostic tool which could be reliably reproduced in other population of TBI cases. We performed a comparative analysis of various models in our dataset of 100 cases to investigate which of the current prognostic tools might be a stronger predictor of outcome and due to the small population sample the characteristics of the final models are not reliable. However, despite the small dataset, we believe our findings are valuable for comparing the performance of the various models with each other. The comparisons were made, firstly, across three measures of model performance i.e. AUC, Nagelkerke R^2 and classification accuracy, secondly, on more than one pair of models and thirdly,

for the prediction of more than one type of outcome i.e. survival and favourable outcome. From all these aspects the results of comparisons are consistent. However, the number of cases prevents model differences being statistically as well as clinically significant hence the overlaps in AUC confidence intervals (Table 70).

Furthermore, although the increase in classification accuracy is not close to the our cut-off of clinical significance (i.e. > 10%) at all times, the clinically significant changes in AUC or Nagelkerke R^2 would support the study findings. For example, although adding S100B to models A and B or adding other TBI predictors to S100B model does not yield an increase of 10% in classification accuracy for survival prediction, the increases in AUC and Nagelkerke R^2 are large enough to be considered clinically significant per the cut-off of 0.05. In fact, AUC and Nagelkerke R^2 are more reliable to compare competing models in that if a model holds higher AUC and Nagelkerke R^2, this model holds a better performance even if the classification accuracy is not different.

8.5.2. Comparison with the literature

With regards to the combination of a brain injury biomarker with other TBI prognosticators to obtain a more reliable predictive power, the results of our study are consistent with other studies. Diminopoulou *et al.* observed that a prognostic model which contains age, GCS and S100B performs better than one without S100B [165]. Similarly, Vos *et al.* constructed a prognostic model which only contained clinical characteristics namely GCS and CT Marshal Classification [139] and demonstrated the improvement in the model performance following adding S100B, Glial Fibriliary Protein Acidic Protein

(GFAP) and Neuron Specific Protein (NSP) [43]. The two latter serum proteins have been shown in some studies to be raised in TBI which, similar to S100B, may have prognostic value as well. However, we observed a stronger predictive power can be obtained with only S100B in combination with other predictors routinely measured. To the best of our knowledge, no researchers have, so far, attempted to compare the two current prognostic tools (proposed prognostic models versus brain injury biomarkers), in terms of their relative predictive strength. This is an important comparison since we found no significant difference in the performance of these two current prognostic tools. This goes against the conclusion by many authors that S100B may not hold enough prognostic strength in TBI [67, 77, 82, 98, 173]. Should other prognosticators be taken into account, S100B could be a good tool perhaps even superior to current multivariate models.

8.5.3. Implications of the study

It is currently acknowledged that the most reliable of the available TBI prognostic models are those developed from the CRASH and the IMPACT datasets. Both of these models have been derived from a large cohort of TBI cases (> 10,000 cases). Conspicuously, collection of such large numbers of TBI cases occurs over a long period of time and requires devotion of a huge amount of funding and resources. We observed that S100B on its own has the same prognostic strength as the prognostic models when identification of its prognostic characteristics in a univariate analysis does not entail accruing a large dataset. For instance, the pilot study in our centre demonstrated that a dataset of 100 TBI patients should suffice to investigate the prognostic characteristics of S100B. One may argue that collection and recording of data

from TBI cases may also occur within ongoing procedures such as in trauma registries. However, it is unclear how often a TBI prognostic model requires updating to take account of improved trauma care policies, new advances in management and therapeutic approaches. The time frame to update a prognostic model may also not provide an adequate number of TBI cases for construction of a reliable prognostic model. The major advantage of prognostic models over S100B appears to be the value of each covariate in the model being easily obtainable in the clinical notes because they are routinely measured. However, measuring S100B serum level is currently not a part of routine clinical practice but should its use increase, cost would decrease. In fact, the costly research into developing a prognostic model which uses commonly measured clinical data should be contrasted with the possible lower cost of measuring S100B in every TBI patient which can be included as a part of routine blood sampling.

Among all constructed models, expanded S100B models A and B are the best per three measures of performance (AUC, Nagelkerke R^2 and classification accuracy). These two models, apart from S100B, include only 1 or 2 other covariates. This may imply that combination of S100B with 1 or 2 other TBI prognosticators could provide a stronger prognostic tool than either multivariate models without S100B or S100B in isolation.

In this study, we investigated the effect of many important prognostic factors (ranging from demographic and clinical observations to CT findings, vital signs and laboratory measures) and observed that S100B shows a significant effect on outcome in all models which contain this covariate. This is not the case with other predictors. This finding suggests an important clinical

implication in that S100B could be the most reliable factor among other measures such as GCS, pupillary reactivity or even ICP to monitor the course of brain injury [65].

8.5.4. Future direction

We embarked on the current project on the basis of limited literature on comparing prognostic models and brain injury biomarkers. As such, it is important to establish which prognostic tool is better and whether or not the combination of the two would be a better option. Whilst examining this requires a large dataset of TBI cases, we decided to first analyse the relatively small TBI data available to us. In this small sample, prognostic models do not appear significantly different to S100B. Similarly, their combinations either by adding S100B to prognostic models or considering one or two other TBI prognosticators along with S100B may be a better option. However, these results have to be validated in a larger TBI series. Similarly, we believe the idea of S100B (or brain injury biomarkers) versus prognostic models needs to be considered and taken forward in investigation on accurate prognosis in TBI. Had the results obtained on our small sample of S100B data been different (such as no significant improvement of performance in combination of S100B and prognostic models), then future research on this would appear less important. However, comparing prognostic models with biomarkers or assessing their combinations is a problem in medicine in general [174] and in TBI prognosis in particular [88]. This is because multivariate analysis requires much large datasets which are costly and time consuming to accrue for S100B. It is stated that the adequate sample size for modelling should include 500 subjects [159].

8.6. Conclusion

A comparisons of performance of the two currently used prognostic tools in TBI; a) multivariate prognostic models and b) laboratory biomarkers (in this study S100B) was shown to have an equivalent performance. However, S100B individually has the strongest influence on outcome. A future prognostic model which may perform better to predict outcome would be one which combines the S100B laboratory biomarker with a minimum of 1 to 2 TBI clinical-demographic prognostic factors. This, however, requires to be examined in a larger cohort of TBI cases.

Acknowledgment

We would like to thank Mr. Rahpael Sacho for his valuable assistance in patients follow-up. This work was in part funded by the Trauma Audit and Research Network (TARN) and Oversees Research Students (ORS) Award Scheme, University of Manchester.

Conflict of interest

None

8.7. Expansion on methods

8.7.1. Data collation/missing information

Due to the retrospective collation of some patients' data, strategies were implemented to minimise the amount of missing information particularly on clinico-demographic values. Whilst the values for age, cause of injury, CT features and ISS were available in the original study dataset, information on GCS and pupillary reactivity was obtained from case notes, Electronic Patient Records (EPR) and TARN. These sources were not uniform in data recording in that, for example, admission GCS might have been available in TARN when missing in EPR. Therefore, for clinico-demographic data, all sources were searched for the following time points: - at scene of injury, en route and on admission to the first hospital (in case of transfer) and subsequently at the earliest time point following admission to ICU (the observations could be recorded by either clinicians or the nursing staff). Subsequently, missing admission values were imputed with en route, at scene and finally from the earliest time point following admission to ICU. However, this strategy still yielded 10% and 15% missing values on GCS and pupillary reactivity respectively. Similarly, with the patients' identifiers, the EPR was searched for laboratory values. This resulted in 42%, 7%, 6%, 6%, 6%, 7% missing values respectively on PH, Glucose, Hb, Plt count and PT. There was no limitation for the length of elapsed time on either side of 24-hours (earlier or later) to spot the closest measurement; had the information been missing at this exact time point. However, if the difference between the closest time of recording was more than 24 hours different to the time point of 24 hours following the injury, the value

was then counted as missing. Since all patients were enrolled from the ICU, it was possible to access the observations on their vital signs along with O2 Sat. and ICP electronically recorded every 10 minutes during the ICU stay. There were only 74% cases with their temperature recorded [175] (38% rectal temperature and 36% brain temperature) which means 26% cases with missing information on this variable (rectal temperature was imputed with brain temperature). However, the amount of missing data was far lower for systolic blood pressure, mean blood pressure and ICP retrospectively at 6%, 6%, and 8%.

8.7.2. Data preparation/continuous versus categorical

Clinicodemographic variables

Fractional polynomials analysis showed that whilst age held linear relationship with logarithmic odds of both types of outcome, this is not the case for ISS and continuous GCS. Nonetheless, unlike ISS, GCS presented a significant association with both outcome measures of favourable outcome or survival in the continuous form by Mann Whitney U test. Therefore, for GCS, initially cut-offs of 9 and 12 were used to categorise this variable into mild, moderate and severe GCS. However, this categorisation resulted in no significant relationship with outcome by Chi square test and thus mild and moderate categories were merged together to group GCS values into severe ($GCS < 9$) and not severe ($GCS \geq 9$) which yielded a significant association. On the one hand, GCS was significantly associated with outcome continuously but could

not be used in the multivariate model without transformation and on the other hand, its categorical form with 3 categories did not show significant association with outcome as per Chi square test. Since transformation of continuous GCS by means of fractional polynomials was expected to yield complicated mathematical calculations on a small sample size of only 100 cases, the decision was made to use GCS categorically rather than continuously. Regarding ISS, initially, cut-offs points of 8, 15. 24, 40 and 75 were used [176]. Nonetheless, due to small sample of cases within each range of ISS: 9 to 15 (only one case) and ISS: 41 to 75 (only seven cases) with no case recorded as ISS less than 8, ISS: 9-15 was pooled into ISS: 16-24 with ISS: 41-75 merged with ISS: 25-40. Therefore, ISS was left with two categories of 3 to 24 and 25-75. Further, for identification of cases with extracranial injury, extracranial AIS scores > 2 were used as the criterion.

S100B

S100B showed a linear relationship with either survival or favourable outcome as per fractional polynomials analysis and therefore was analysed only continuously.

CT features

CT reports of patients were available from the original dataset and therefore using these reports the appropriate Marshal Class using the algorithm suggested by Maas *et al.* [94] were allocated to each case. However, Marshal

Class IV was merged with the Marshal Class V because there were only one case with Marshal Class IV. Furthermore, other traditional descriptive terms for structural brain damage proposed by literature were also applied as nominal (yes/no) variables; being brain swelling, mass lesion, midline shift, cisterns, SAH, SDH, EDH, haemorrhage and contusion..

Vital signs, O2 Sat. and ICP

Cut-off points of 120 and 150 were used to categorise systolic blood pressure into hypotension, normtension and hypertension with 85 and 110 for mean blood pressure as proposed in the IMPACT study [95]. Furthermore, temperature values were first grouped into hypothermia (< 35 ° C; based on IMPACT [96])**,** normothermia (> 35 ° C but < 38 ° C) and hyperthermia (> 38 ° C; based on Signorini's study [35]) and since there was only one case within the hypothermic group, this was merged with normothermia. Furthermore, there were no cases with hypoxia recorded in the dataset. Consequently, hypoxia was discarded at this early stage from further analysis. Regarding ICP, fractional polynomials analysis did not prove a linear relationship with outcome. However, it was observed this variable holds significant correlation with survival in its continuous form. Using the cut-offs of 20, 30 and 40 as suggested by Signorini *et al.* [35] to categorise ICP into mildly, moderately or severely raised ICP did not affect this significance. Using this categorisation, there were no cases within the severely increased group with only 6 cases in the moderately increased group. Because of this, all severities of raised ICP were merged into one group to have only two categories as raised ICP versus

normal ICP. The results demonstrated that both categorisations (either with four categories or with two categories) were significantly associated with survival but not favourable outcome. In the modelling procedure, ICP with two categories was included.

Laboratory values

Although the effect of PH, Hb, glucose, Plt count and PT on outcome are expected to flatten in the normal range values, in the IMPACT dataset, these parameters demonstrate a linear relationship with outcome with no obvious threshold point. Therefore, all laboratory measures were also continuously assessed along with their categorical form. For PH, cut-offs of 7.35 and 7.45 were used to categorise PH values into acidosis, normal and alkalosis. Nevertheless, there were only 4 cases with acidosis and thus these cases were included in the normal group. For Hb, the upper and lower normal limits were employed to categorise this variable. The cut-offs were different per gender (normal range: males: 135-180 mg/dl and females: 115-160 mg/dl) [169]. Using this categorisation, the dataset contained only one male case with Hb of above 181 which was then included in the normal range group resulting in the dataset containing no case with high Hb. Similarly, the upper and lower limits of 2.5 mmol/dl and 11 mmol/dl [171] [170] were used to recode glucose values into hypoglycaemia, normoglycemia and hyperglycaemia. Along with this categorisation, the cut-off of 6 mmol/dl as observed in the IMAPACT study [97] as the changing point in the trend of the influence of glucose on outcome was also tested. Additionally, since a similar change in the trend was observed

by the IMPACT for the Plt count at the value of 200×10^3 [97], this value was used to categorise Plt count into thrombocytopenic versus normal group. However, along with this categorisation, Plt count was also categorised using the cut-offs of 100 and 450 which conventionally groups the values into thrombocytopenia, normal and thrombocytosis [172, 177]. Regarding PT, using 9 min. and 12 min. as the upper and lower limits of normal range yielded no cases within the normal and below-the-normal-range groups and thus this variable was only continuously analysed.

The performance of TARN TBI prognostic models (models A and B) [168] was assessed on the study dataset. Adaptations of these models to the dataset in this study included using GCS with two categories (severe and non-severe GCS), using pupillary reactivity with three categories (normal, only one reactive and both reactive) and omission of hypoxia, brain stem injury and extracranial injury due to the small number of cases with each of these observations (there were no cases recorded with hypoxia or brain stem injury and there were only 5 cases recorded with extracranial injury which all survived till 3 months after injury and only one experienced unfavourable outcome).

8.7.3. Handling missing information

All the remaining missing information were imputed with multiple imputation strategy in Stata.

9. Discussion

In this PhD, tow prognostic models have been constructed and the prognostic strength of S100B has been compared with prognostic models and also the prognostic strength of combination of these two tools has been examined. These next few pages summarise the approach and results in each project.

TARN project

The TARN TBI dataset (n=802) was used for the model derivation. Initially, the data were retrieved from TARN based on head AIS codes under internal organ in the AIS dictionary plus those skull AIS codes with AIS severity of 3 or above. Those submissions which did not have their pupillary reactivity recorded at any time point were not selected.

Further TBI data were retrieved from TARN with the same inclusion criteria irrespective of pupillary reactivity being recorded or not (n= 21657). Then the prognostic strength of various time points of GCS measurement and also various sub scores were compared in a multivariate analysis (i.e. following adjustment with age and ISS). The results demonstrated that total GCS and the motor subscore hold the similar prognostic power with admission scores/subscores being stronger than scene scores/subscores. This implied that in the modelling admission GCS/motor score may be better than scene GCS/motor score and since the number of missing information of motor subscore was much higher than that of total GCS in the derivation dataset (167 versus 26), the admission total GCS was used in the subsequent modelling procedure.

Parallel to the above analysis, a method was devised to translate the brain injury AIS codes to the Marshall Classification facilitating allocation of a Marshall Class when AIS codes are at hand. Initially a cross-tabulation was devised which was agreed upon from both clinical and AIS coding viewpoints. Furthermore, an algorithm was proposed which can allocate one single Marshall Class to a TBI patient in a trauma registry who has various AIS codes of brain injury recorded. Following this, the cross-tabulation and algorithm were programmed in SPSS and thus a Marshall Class was assigned to each case in the derivation dataset.

Subsequently, the univariate analysis was performed using Mann Whitney U test for continuous variables i.e. age, GCS, ISS, systolic and mean blood pressures and Chi Square test for categorical variables i.e. GCS, gender, nationality, cause of injury, extracranial injury, systolic and mean blood pressure, hypoxia, various AIS severity scores, the Marshall Classification and various intracranial pathologies (T test was never used since none of the continuous variables were found normally distributed by one-sample Kolmogorov-smirnov test). The univariate analysis demonstrated all variables significantly associated with outcome apart from gender, nationality, contusion and SDH. Therefore, these variables were excluded from the modelling. Furthermore, fractional polynomial analysis was performed to assess the assumption of linearity for logistic regression. This analysis demonstrated that age has a linear relationship with logit odds of survival and thus can be included in the model as ‘it is’ (i.e. with no transformation or categorisation). However, for other variables the power transformations were required.

Following the univariate analyses, logistic regression was run on the TARN dataset. First age, GCS, pupillary reactivity, ISS and extracranial injury were supplied to one single model with automatic regression. However, it was observed that the automatic regression discarded one of the two fractional polynomial transformations of GCS. This was not acceptable. Thus this stage was run again 'manually' in that each variable was supplied to the model at separate stages. Furthermore, it was observed that extracranial injury is significantly associated with outcome when firstly, the cut-off is extracranial AIS severity score of 4 and above (and not 3) and secondly, when it is not included in the model which contains ISS. This resulted in branching the modelling procedure with two parallel models one model including ISS and excluding extracranial injury and the other excluding ISS but including extracranial injury. Additionally, it was observed that cause of injury is significant only in the model which contains extracranial injury.

During the modelling procedure, the prognostic value of various AIS severities (3, 4 and 5/6), the Marshall Classification and various intracranial pathologies were assessed. It was observed that whilst adding the AIS severity and the Marshall Classification improves the prognostic strength of multivariate models, these may not be accurate in grouping injuries based on prognostic merit. This is because not all AIS severities or Marshall Classes demonstrated significant association with outcome. Regarding various intracranial pathologies namely haemorrhage, SAH, EDH, brain swelling, brain stem injury and cerebellar injury, only the brain stem injury and brain swelling were significantly associated with outcome. Several models were constructed to perform this piece of analysis. However, only those models in which all

intracranial pathologies were significant were chosen for the modelling procedure. These are two models: one only containing the brain stem injury and the other containing the brain stem injury along with the brain swelling.

The prognostic value of hypoxia, mean and systolic blood pressures were also examined. It was discerned that whilst hypoxia is significantly associated with outcome, hypertension (as per either mean blood pressure (i.e. > 110 mmHg) or systolic blood pressure (i.e. > 150 mmHg) was never significant. Thus only hypotension versus normotension (including hypertension) was considered. Finally, between systolic and mean blood pressure, systolic blood pressure was selected since these two variables could not be included in the same model (as they lost their significance) and also the model with mean blood pressure did not demonstrate acceptable calibration as per HL statistics.

There were overall 26, 136 and 138 missing cases with admission GCS, left and right pupillary reactivity respectively. This missingness was filled with the en route and then the scene records which left the dataset with no missing values of these variables. However, for hypoxia and systolic blood pressure (each with 36 and 114 missing cases), this strategy failed to fill all the missing values and thus the cases with missingness were all grouped in a different category as ‘missing’.

In the end two models were derived: model A and B. Both models contain age, GCS, pupillary reactivity, the brain stem injury, hypoxia and low blood pressure. However, model A contains ISS which is absent in model B whereas model B contains cause of injury, extracranial injury and brain swelling instead. The performance of these models is presented across various

indices. They have AUCs of 0.92 (C.I.: 0.90-0.95) and 0.93 (C.I.: 0.91-0.95) respectively for models A and B. Model A was externally validated on a different serious of TBI from TARN and maintained its performance as per AUC (0.92). Model B also demonstrated reasonably good performance in the same external dataset (AUC: 0.82). However, the drop in AUC for model B as it was re-run in the IMPACT data was somewhat huge (AUC: 0.68 for survival prediction and 0.69 for favourable outcome prediction at 6 months). From various aspects the proposed models can be considered well-developed regarding the strategies taken during the model derivation.

S100B project

The analysis of the first prospective S100B project which the investigator contributed to its completion, demonstrated that S100B, 24 hours after injury, has a high sensitivity to predict unfavourable outcome or death (more than 80%) but its specificity is somewhat low (60% for unfavourable outcome and 49% for death).

With regards to the second part of S100B project, initially, the patients' case notes, EPR and TARN were searched for the variables which were not recorded in the original S100B study. In fact, the first S100B project only had records of age, cause of injury, CT findings and ISS. Subsequently, the association of each TBI prognosticator with outcome was assessed with the outcome (survival and favourable outcome at 3 months). The TARN TBI models A and B derived in the other part of PhD were run on the S100B dataset of 100 cases and their performance was compared with the model

which included only S100B. The performance was compared across various measures namely AUC, Nagelkerke R^2 and classification accuracy. In order not to consider small differences in these indices, the differences of more than 0.05 were considered 'clinically' significant for AUC and Nagelkerke R^2. Similarly, the difference of more than 10% was considered clinically significant for classification accuracy.

The results demonstrated that, particularly as per AUC, the difference between performance of prognostic models and S100B is not clinically or statistically important. Regarding the importance of combinations of S100B with prognostic models, S100B was added to TARN models A and B to evaluate their change in performance. It was observed that this addition results in clinically significant increase in the models performance as per AUC and R^2 Nagelkerke. However 95% confidence intervals for AUCs overlapped implying inability to show the statistical significance. Similarly, those TBI prognosticators which were present in either model A or B or were found significant in the univariate analysis were added to the S100B model. Following to this, a clinically significant improvement in performance was observed as per AUC and Nagelkerke R^2. However, this improvement was not statistically significant.

9.1. Limitations

9.1.1. TARN project

Summary from the respective papers

Some limitations of this project have been already discussed in Papers 3, 4, 5 and 6.

Paper 3: in the analysis of prognostic value of various GCS subscores and the admission scores/subscores versus the scene scores/subscores, it was unclear that whether the immeasurable values (i.e. due to intoxication or intubation) were assigned as missing (i.e. leaving it blank during data entry) or were received the lowest GCS subscore based on the local hospital policy. This implies that some lowest GCS subscores in the analysis might represent missing values unlike the GCS criteria for the lowest subscore. Similarly, in this analysis the adjustment was performed with age and ISS although pupillary reactivity is an important confounder too. Furthermore, the definition of 'admission to the emergency department' (the record which was taken from TARN as representing admission GCS) varies across hospital in terms of being prior to or following resuscitation. This is important as it has been suggested that post-resuscitation values may be better for prognostic analysis [137].

Paper 4: regarding the method which was proposed to assign a Marshall Class to a case of TBI in trauma registries, this suffers from some limitations as to AIS scores being substitutes to CT reports and the assumptions made about the mass lesion and brain swelling. The source of information for AIS coding is not only CT and can be MRI or operational notes as well. This indicates the dynamic nature of brain injury which is inherited in AIS coding

but not the Marshall Classification which is obtained via direct observation of CT image. Furthermore, the cut-offs for sizewise categorisations of mass lesions in AIS dictionary are different to that in the Marshall Classification. In the same way, the criteria to assess the degree of brain swelling are different in the two systems (being midline shift and cisterns status in Marshall Classification versus only the status of cisterns in the AIS dictionary).

Paper 5: using AIS codes as substitutes to CT reports is also a limitation for the prognostic analysis of various intracranial pathologies in this paper as the CT findings/classifications from the literature were used as a basis for this analysis. Moreover, brain injury AIS codes are commonly multiple and taking only the highest code to assess the prognostic value of AIS scores may not be an appropriate approach.

Paper 6: regarding the proposed final TBI models, the derivation dataset may not be a true representative of severe TBI according to TARN inclusion criteria. Even within the TARN context and according to the selection bias analysis, the derivation dataset was slightly younger than excluded TBI cases due to lack of pupillary reactivity recorded. Furthermore, children were not excluded from the multivariate analysis although it is thought the course of TBI differs in children versus adults.

******

The approach taken in model derivation may be ‘holistic’. This applies to the involvement in overly detailed observations which may not only lack importance with regards to their enhancement in the model performance; they may in fact disturb the model performance in other populations of TBI. This

overly detailed approach in modelling may be reflected in the difference between AUC of levels IIIÁ and VIIIB models and the final TARN models A and B which is only 0.01 (for model A) and 0.02 (for model B). Despite this slight difference, we still proceeded with the modelling since at the very early level it was unclear how far the performance of the model(s) would increase. This slight degree of difference occurs when levels IIIA and VIIIB models respectively contain only 4 or 5 variables (age, GCS, pupillary reactivity and ISS in level IIIA model and age, GCS, pupillary reactivity, extracranial injury and cause of injury in level VIIIB model) whereas TARN model A contains only 3 more variables (brain stem injury, hypoxia and systolic blood pressure) and TARN model B contains only 4 more variables (brain swelling, brain stem injury, hypoxia and systolic blood pressure). However, although in terms of AUC the difference between models IIIA to final models is small, it is unclear whether or not this degree of difference is in fact unimportant in the clinical context. For example, the AUC of mammography to diagnose breast cancer is statistically high (85%) [110] whereas this diagnostic tool has much less value in the diagnosis of breast cancer than other tools such as Fine Needle Aspiration (FNA) or biopsy. Furthermore, models A and B contain extra variables than model IIIA which puts emphasis on the prognostic value of these extra variables such as CT findings, hypoxia or hypotension. This is important especially with regards to tertiary preventive measures to consider interventions to treat CT abnormalities, hypoxia or hypotension in TBI patients. However, if the models are to be used to assess and compare the quality of care, this degree of difference in AUC may be negligible and models IIIA or IIIB may sufficiently serve the purpose.

Our series of TBI cases may not be representative of severe TBI cases to refute the hypotheses 1 and 2. Although according to TARN inclusion criteria, only the profile of those trauma patients who sustained severe injuries are submitted to TARN, comparing our dataset of TBI to the IMPACT dataset demonstrates that the severity of TBI in TARN is less as per GCS and pupillary reactivity. For example, 19.2% of cases in the IMPACT dataset had neither reactive pupils whereas this figure was significantly lower in the TARN data at 16.5%. Even if one wishes to consider severe brain injury as GCS $\leq$ 8, 35.8% of our sample sustained mild brain injury (i.e. admission GCS 15, 14 or 13). However, the cases in our dataset had severe injuries enough to stay at hospital for more than 3 days or to receive ICU care (or to fulfil other criteria of TARN submission reflecting severe injuries). Similarly, all cases had one or more head AIS code representative of either intracranial pathology or compressed/depressed/open skull fracture.

One other limitation of the models relate to survival being assessed at discharge rather than assessing this at a certain point in time such as 30 days after injury. A time-fixed outcome assessment has two important advantages over discharge outcome. Firstly, it is well-known that as time passes, the probability of death decreases after TBI. As such, varied length of time of outcome assessment plays a confounding role as it influences what the prognostic models predict which is the probability of survival. Secondly, there is a bias to consider some patients with severe TBI who are discharged from hospital but end up in rehabilitative centers and subsequently die there. These patients are in fact non-survivors of TBI but have been discharged because of the severity of injury being deemed unresponsive to any therapeutic

intervention. The bias is that this group of patients with severe injuries are considered as alive whereas they have not survived their severe injuries.

It is an obvious disadvantage of TARN TBI model B that its performance significantly drops in the IMPACT external TBI population and one may consider that the same observation would have been observed, had the IMPACT had ISS records to externally validate model A (model A was not validated on the IMPACT data since ISS was not recorded). The reason for this poor external performance may be either significant difference of case-mix between the IMPACT and the TARN dataset or the inherent lack of transportability in TARN models. The IMPACT dataset pooled various datasets from observational and clinical trial studies conducted in varied geographical regions (across Europe and the North America) and the endpoint of outcome measure was at 6-months after the injury. The IMPACT dataset also contained more severe cases of TBI as per GCS and pupillary reactivity. Moreover, TARN models hold historic validation as there is not a significant drop in performance across various indices where the models are validated in a TARN TBI series from a different time period as to the derivation set. Similarly, the decrease in performance in the IMPACT dataset does not reflect our failure in constructing models suitable for the British trauma care system. So far, it is clear that the external performance of the models may be less than some IMPACT and CRASH TBI models in terms of some types of transportability related to geography, methodology, spectrum and fellow-up interval (Paper 1, section 2.8).

Handling the continuous variables is one aspect of model development. We applied fractional polynomial transformation to address the linearity between the continuous variables and the logit odds of survival. During the modelling procedure we decided to discard continuous GCS (due to the same performance of the model with continuous GCS to the model with categorical GCS). Using categorical GCS with three categories of mild, moderate and severe may not be an appropriate approach since in this way GCS is not treated as a continuous entity. For example, two patients with different GCSs of 4 and 8 are both put in the same category as severe GCS whereas GCS 8 and 4 are numerically different and thus may have different prognostic values. It may be that using other methods to address the linearity of GCS with logit odds of survival such as spline functions [178] be superior than fractional polynomials especially if the resulting model is expected to sustain its performance in other external datasets. Although, it is unclear which method to address linearity is better from the statistical point of view, using complicated mathematical formula proposed by fractional polynomials does not have clinical appeal.

There are some downsides related to TBI models A and B compared to the CRASH and the IMPACT models. First and the utmost is the large population sample in which the CRASH and IMPACT models have been derived. The CRASH dataset contained more than 2000 TBI cases from high income countries and more than 7000 cases from the low to middle income courtiers. The IMPACT dataset contained 8509 cases (all from high income countries). However, our sample size in TARN was still large enough for derivation of a model as the number of patients with the outcome (i.e. survivals) was more than 10 times the number of predictors. Furthermore, both

the CRASH and the IMPACT models, unlike the proposed model in PhD, do not experience a huge drop in performance as per AUC (the IMPACT models were validated in the CRASH dataset and vice versa). The other upside of the IMPACT and the CRASH models over the TARN models may be related to the amount of human resources and expertise devoted to the development of such models.

The derivation of prognostic models on the TARN TBI data was based on the argument that the IMPACT and CRASH models may not be valid for the British trauma care system. Therefore, one may argue that at the beginning these models should have been run on a sample of TBI cases in TARN to demonstrate whether or not the models actually lack the validity. Whilst this is a clearly missed step in the approach taken in this PhD, the IMPACT group have run their model on a subset of TBI cases from TARN and have observed an AUCs of between 0.80 and 0.85 for various IMPACT models. Despite this degree of performance is much higher than the random guess (i.e. AUC of 0.50) but there is still scope to obtain a better performance closer to 1. Furthermore, this degree of performance is lower than that of the TARN general trauma model used for benchmarking of trauma care (Ps07 model; AUC = 0.94) [142]. Unfortunately, the CRASH model has not been run on the TARN data so far. However, it is not likely that the results of such analysis would have changed the procedure to construct a prognostic model for the British TBI population. This is because the performance of the CRASH models as per AUC dropped when externally validated. In fact, the maximal external AUC of the various CRASH models is 0.77. This is much lower than the AUC of the TARN general trauma model (AUC = 0.94) [142]. It is not likely that the

external validation of the CRASH models would increase in a TARN dataset to become close to the current AUC of the Ps07 model. Even in the internal validation, the highest AUC of the CRASH models is far below that of the Ps07 model (being 0.87). These issues are important with regards to the trauma benchmarking since the models used for this purpose overall have much higher AUCs than what has been obtained by the IMPACT and CRASH. Whilst from clinical perspective the AUCs by the IMPACT and CRASH appear acceptable, they are not suitable to offer a national standard which the local trauma care can be compared to (given the current TARN model performance for general trauma).

In the same way, the validation of the TARN general trauma model in a subset of TBI cases was not performed in this PhD prior to embarking on the modeling. This is important since if general trauma models maintain their performance in TBI patients, then construction of models specifically designed for TBI patients may not be necessary. However, despite in the actual approach taken in this PhD, the general trauma models were not validated in the TBI dataset (s), the Ps07 model was once run in the derivation dataset of TBI models A and B (n=802) following the modeling procedure having been already completed. The results demonstrated that the AUC of this model underwent a drop in performance from 0.94 to 0.85. This can justify the derivation of prognostic models for TBI patients different to those built for general trauma patients.

Although, for the objective of PhD, the prospective study design was preferable (as discussed in section 1.7.1), the PhD by design is retrospective. This in itself poses problems on those parts of study relevant to the protocol

and data collection. Whilst a prospective study offers the possibility that the investigator tailors the protocol and data collection to the research objective, in a retrospective study only the existing data is analysed. In TARN project, it was not possible to assess the prognostic value of some variables such Hb [24], temperature [92] or ICP [35] whereas these might have been possible to record in a prospective design. In fact, some limitations in TARN project could have been eliminated, had the study been prospective such as less strong selection bias as per the inclusion criteria and assessing the prognostic value of various CT findings.

9.1.2. S100B project

Summary from the respective papers

Some limitations of this project have been discussed in papers 2 and 7 so far.

Paper 2: there are issues with the blood sampling at 24 hours which was merely based on pragmatic reasons and not evidence. However, the patient is expected to be more stabilised at this time point than admission. In the same way, the flexibility of 2 hours on either side of 24 hours following the injury is important as the half life of S100B is thought to be around 2 hours and as such significant changes in blood levels can be expected for example from 24 hours to 26 hours after injury (a gap of 2 hours in the sampling time which is close to the half life).

Paper 7: the main limitation is the sample size which, despite being large enough for the univariate analysis of S100B prognosis, does not offer powerful results for the multivariate analysis in refuting the hypotheses 3, 4, and 5 (type II error). Similarly, due to this small sample, it was also not

possible to suggest which TBI prognosticators can be considered along with S100B to predict the outcome in TBI.

******

For a number of reasons further caution should be taken in interpretation of the results of the study having recruited only adult cases. Apart from the possible different trajectories of brain injury in children and adults, S100B concentration also tends to be higher in healthy children than adults. It has been shown that as age increases, S100B serum levels decreases up to the age of 20 following which it plateaus [52]. Also, Geyer *et al*., in a sample of 148 children with mild TBI observed that serum levels decrease up to the age of 8 followed by an increase [179]. However, it is yet unclear how the serum levels of S100B differ in children versus adults following TBI and how this possible difference could affect the prognostic strength of S100B. Furthermore, S100B is a protein which is synthesized in astroglial cells and then reaches the blood stream after passing the blood brain barrier. The disposal of this protein is thought to be through renal excretion [180, 181]. Thus the kinetic of S100B with regards to its passage through the blood brain barrier and the excretion in the kidneys may be different in children and adults (this may be similar to the different kinetic of drugs in children versus adults).

Similarly, one may argue that running the IMPACT or the CRASH models might have been superior to TARN models as the IMPACT and CRASH models have been externally validated. Nevertheless, the IMPACT, CRASH and TARN models share the core TBI prognosticators as age, GCS (or

motor GCS as in IMPACT) and pupillary reactivity. The differences of these models in terms of their covariates are the type of intracranial pathologies used or the inclusion of glucose and Hb in the IMPACT model. It appears unlikely that running the IMPACT or the CRASH models on the datasets would lead to substantially different results.

In this study, the multivariate models as one type of prognostic tools in TBI were compared to S100B as a biomarker of brain injury. For this analysis, S100B was selected based on a search in PubMed to determine the relative amount of research/evidence for three commonly known brain injury biomarkers in TBI literature: S100B, GFAP and NSE. This approach does not necessarily reflect the better prognostic performance of S100B than the other two biomarkers. Vose *et al.* compared the ROC curves of GFAP, NSE and S100B and observed GFAP and NSE have higher AUC to predict 6-month GOS < 4 than that for S100B (GFAP: 79.4, NSE: 78.2, S100B: 67.7). Although the confidence intervals are not supplied in this article, it may indicate that other brain injury biomarkers could be better TBI prognosticators than S100B.

This study may not be accurate in the definition of severe TBI in refuting the hypotheses 3, 4 and 5. Since, cases were enrolled from a neruo-ICU, this sample of TBI population represents perhaps only the extreme cases within the severity spectrum of TBI. Thus the dataset does not contain cases who sustained severe brain injury according to some factors such as GCS or pupillary reactivity but did not meet the requirements for admission/transfer to neuroICU (for example they did not need neurosurgical intervention). This may be an important selection bias as those patients who do not end up in nuero-ICU (or do not receive neurosurgical care) may not have necessarily a

more favourable outcome than those who are supplied with such specialist care. Furthermore, with respect to CT findings, the majority of cases (61%) sustained brain injury matched with Marshall Class VI followed by Marshall Class V (33%) with only 6% of cases holding Marshall Class II (the least severe Marshal class in the dataset). Similarly, 94% of cases had haemorrhage detectable by CT. This pattern of CT findings conspicuously is not representative of the heterogeneity in TBI severity.

Similar to the TARN project, some limitations of the S100B project arose due to its retrospective nature. This is particularly important with regards to the sampling time being 24 hours after injury. Although one may assume that at this time point the patient is stabilised and hence effect of secondary insult on the brain has settled, the admission sampling may be more important for management decisions which may need to be made at an earlier time point based on the prognosis.

9.1.3. Summary

The approach taken in the modelling procedure might have been holistic in that the first constructed models do not seem to be significantly different to the final models as per AUC. Furthermore, although the proposed models are targeted at severe cases of TBI, the derivation dataset contained milder cases as per admission GCS (as being > 13) or as compared to the IMPACT data across GCS or pupillary reactivity. Regarding the external validity of the models, model B experienced a huge drop in performance when validated in the IMPACT data. Additionally, due to the retrospective nature of PhD, it was not possible to examine the value of Hb, ICP or temperature as these factors are included in some TBI prognostic models but not recorded by TARN.

The small sample size of the S100B study poses an obvious limitation which undermines the power of the study. Furthermore, since the IMPACT and CRASH models have been externally validated, it might have been a better option to run the IMPACT and CRASH models on this data as well, although it is not likely that the obtained results would have been substantially significantly different. Regarding comparing the prognostic value of brain injury biomarkers with prognostic models, S100B may not be the best representative of biomarkers because GFAP and NSE are proposed to perform better than S100B by some studies. Lastly, the S100B dataset included the extreme cases of TBI who were all cared in ICU and 94% of them sustained brain haemorrhage.

9.2. Comparison with the literature

9.2.1. TARN project

Summary from the respective papers

So far each paper provides some aspects of comparing the results with the literature:

Paper 3: the results with regards to GCS prognostic analysis is in consistence with Healey *et al.* findings, however their dataset was from general trauma patients and no adjustments with confounders was made [22].

Paper 5: with regards to prognostic value of various AIS severity scores, Gennarelli. *et al.* [144] reported as AIS severity scores increase the chance for survival decrease but in the TARN dataset not every AIS severity score of brain injury (out of 3, 4 and 5/6) showed significant association with outcome. However, the Gennarelli's study was only on general trauma patients with no adjustments for confounders. Similarly, the prognostic value of the Marshall Classification has been assessed in many studies but similar to the results of this PhD, not every Marshall Class is significantly associated with outcome unless some classes are merged together [24, 36, 92, 140]. The results of PhD regarding various intracranial pathologies including SAH is not consistent with many other studies [23, 24, 36, 92, 140, 143, 148, 149, 151, 152] which can be due to different case-mix as the TARN dataset includes relatively milder cases of TBI.

Paper 6: the proposed prognostic models in PhD can be added to the current list of TBI prognostic models including the CRASH and IMPACT models. The performance of TARN TBI models are slightly better than that of

the IMPACT and the CRASH models which may be due to the type or time of outcome prediction (disability at 6 months). However, since the TARN models have been derived from trauma registry data and been historically validated in a different dataset from TARN, the TARN models may be a better option for British trauma care assessment. However the CRASH and IMPACT models may better suit clinical trials as their derivation datasets were mainly from clinical trials. Regarding application of prognostic models in clinical setting, TARN models predict an acute outcome (i.e. survival at discharge) compared to the CRASH and the IMPACT models and therefore these models may offer different applications during trauma care (i.e. acute care versus chronic care).

******

AIS scores, the Marshall Classification and various intracranial pathologies were added to the reference models in Paper 5 (reference models did not have intracranial pathologies but did contain other important TBI prognosticators such age, GCS or pupillary reactivity) and subsequently only a slight improvement in the performance (from 0.91 to 0.92) was noticed. This finding is in consistence with that by the CRASH collaboration when they added various intracranial pathologies detected by CT to the multivariate models and noticed a slight increase in AUC of only 0.02 (for models constructed for the third world) or no changes (for models constructed for the developed world). However, the IMPACT prognostic models demonstrated significant degrees of improvement in performance when CT findings were added to the baseline models (up to 0.08 increase in AUC). Regarding the time

of data collection in IMPACT as dating back to more than 13 years ago and more up to date data in the CRASH and the TARN datasets, the difference in results may be attributed to the improvement in trauma care policy for TBI victims. The other explanation may be more severe cases in the IMPACT dataset as they all had admission GCS of less than 9 whereas the CRASH and the TARN datasets include cases with higher GCS as well.

TARN TBI datasets share age, GCS, pupillary reactivity with both CRASH and the IMPACT models. These are in fact the three core TBI predictors [88]. Further, among all the predictors in each model, *pupillary reactivity: none reactive* has the highest impact on outcome prediction according to odds ratio or the coefficient or its score in the scoring system (i.e. the IMPACT models). Extracranial injury was found important in the TARN TBI model B which is the case in the CRASH models but not in the IMAPCT models since this variable was not investigated in the IMPACT modelling. Low blood pressure is also included in the TARN and IMPACT models but not the CRASH models perhaps because this variables was not recoded in the CRASH data. However, with regards to CT findings, the results appear fairly heterogeneous. For example, whilst SAH is significant in both the CRASH and the IMPACT models, this variable was found non significant in our data. This is also the case for the Marshall Classification which is contained in the IMPACT model but not in TARN TBI models. There can be several explanations for this. It may be due to different case-mix of the studies as the IMPACT data contain more severe cases of TBI or the AIS intermediation in assigning the CT findings to TBI cases in TARN. However, with regards to brain swelling, it appears the three models (IMPACT, CRASH and TARN) are

in agreement in that the CRASH models include midline shift and obliteration of the 3rd ventricles/basal cisterns (which can in part be due to brain swelling) and the IMPACT models include Marshall Class III/IV (merged) which partially represents brain swelling.

9.2.2. S100B project

Summary from the respective papers

Paper 2: The cut-off and the prognostic characteristics found in the univariate analysis of S100B dataset are different to other studies. For example, whilst the cut-off by Nylen *et al.* (0.55 μg/l) [79] was close to that obtained in the S100B dataset (0.53 μg/l), the specificity observed by Nylen *et al.* was much higher (100% versus 60% for disability prediction or versus 49% for mortality prediction in our data). Similarly, whilst the sensitivity and specificity proposed by Vos *et al.* [43] was close to what was observed in the S100B dataset, the cut-off was much higher (1.13 μg/l by Vos *et al.* versus 0.53 in our study). These differences may be explained by differences in the case mix.

Paper 7: the results of S100B study with regards to combination of S100B with other TBI prognosticators are inagreement with the findings by Dimiopoulou *et al.* [43, 165] and Vos *et al.* [43] as discussed in Paper 7. However, no study so far has attempted to compare the performance of prognostic models with a brain injury biomarker.

******

It was observed that in our sample of TBI, the median S100B not only lacks significant difference in patients with and without extracranial injury, it is

in fact higher in isolated brain injury. This is in contrast with Savola's finding that major extracranial injury further increases the levels of S100B in TBI [66]. The reason for this difference may relate to the severity of brain injury. In Savola's study severe brain injury was defined as head trauma with amnesia, longer-than-24-hours unconsciousness, intracranial injury observed on CT or focal neurological deficit. Not all cases with these criteria of brain injury severity would end up in ICU which the subjects in our study were recruited from. Unfortunately, measures such as GCS or pupillary reactivity are not supplied in the Savola's study in order to make a rough comparison of the brain injury severity in the two studies. It may be that in more severe cases of TBI, the effect of extracranial injury on S100B declines. However, in moderate severities of TBI, extracranial injury may be able to increase S100B levels. In such case, it may be anticipated that severe brain injury has the same S100B concentration as moderate brain injury with extracranial injury.

9.2.3. Summary

Whilst TARN models appear better for trauma care benchmarking, the IMPACT and CRASH models are particularly suitable for clinical trials. The results on various intracranial pathologies are different to the CRASH and IMPACT's finding (such as the importance of haemorrhage and SAH) which can be due to different case-mix or the fact that intracranial pathologies in TARN are obtained from AIS codes rather than actual CT images. However, it appears that all these models (TARN, IMAPCT and CRASH) share the core TBI prognosticators i. e. age, GCS and pupillary reactivity. It also appears that these models share brain swelling or pathologies which are likely to be

accompanied by brain swelling such as obliteration of the 3rd ventricle/cisterns (CRASH) or Marshall Class III/IV (IMAPCT).

According to univariate analysis of S100B data, the results are different to other studies either on cut-off or the prognostic characteristics i. e. sensitivity and specificity. However, similar to our study, it has been shown by Dimiopoulou et al. and Vos et al. that S100B is one of the most important TBI prognosticators in multivariate analysis. However, to the best knowledge of the investigator there is no studies so far which have attempted comparing S100B with prognostic models. Regarding the effect of extracranial injury, the results are different to that by Savola's group as unlike our study, they found a significant effect of extracranial injury on serum S100B. It may be that in severe cases of TBI, the effect of extracranial injury on S100B diminishes.

9.3. Implications

9.3.1. TARN project

Summary from the respective papers

Some discussion on the interpretation and implication of the results is presented in papers 3, 4, 5, and 6.

Paper 3: in the prognostic analysis of various GCS subscores, it was observed that motor and total GCS may have similar prognostic strength. This may indicate that measurement of eye and verbal subscores are not necessary as the motor subscores would be easier to learn with less inter- and intra-observer disagreement. On the other hand, total GCS has more content information and may be still superior to motor GCS for day-to-day monitoring of patient's course of of conciousness level. Moreover, it was observed that the admission scores/subscores are more predictive than scene scores/subscores which can be due to lingering effect of inebriation by the arrival at hospital or the level of skill to measure GCS at scene.

Paper 5: considering prognostic value of various intracranial pathologies, the results of Paper 5 are important as it highlights the brain stem injury and brain swelling as strong predictors of outcome in TBI. This implies the necessity of research to improve current therapeutic approaches to patients who have sustained these types of injuries.

Paper 6: regarding the two prognostic models constructed (models A and B), the models employ various charactristics to predict survival at discharge. Model A may be better for trauma registries as it uses ISS which is not commonly measured in clinical practice. On the other hand, model B may

be a better choice for clinicians. Moreover, Model A contains ISS instead of cause of injury, extracranial injury and brain swelling which are included in Model B. As both brain swelling and extracranial injury can influence ISS, there may be some relationship with intracranial injury and cause of injury which prohibits the two models holding exactly the same covariates. Further, as the proposed models contain pupillary reactivity as an important TBI predictor, it seems that recording this variable should be declared mandatory for all TBI submissions in trauma registries.

******

For a number of reasons, the constructed models can be referred to as 'well-developed' models. According to Perel's criteria [31] some indications of a well-developed model in our study are:

- ***The patients had adequate follow-up*** as there was no missing information on the discharge survival status.
- ***The predictors are included based on a reasonable rationale*** i.e. following the literature review and with consideration given to the clinical setting.
- ***The variables were clearly defined.*** This is reflected on the time point of measurement of the variables such as admission GCS or pupillary reactivity and clear definition of each category with regards to pupillary reactivity, cause of injury, extracranial injury, intracranial pathology (the brain stem injury and brain swelling) and hypoxia. For example, the category abnormal-both reactive in pupillary reactivity is when both eyes are reactive but one of them is

sluggish compared to the other. Similarly, the brain stem injury can include any types of injury such as compression, contusion, diffuse axonal injury etc whilst brain swelling excludes swelling in the brain stem or cerebellum. Furthermore, the diagnosis of these pathologies can be made based on clinical ground or any diagnostic modality including CT, MRI, operation notes etc. With regards to cause of injury, Table 38 (on page 287) should address any confusion. For example, if the patient has fallen following an assault, injury mechanism would be fall with the level of intent being assault. Using Table 38, the cause of injury would then be assault.

- ***The missing information is handled with imputation strategies rather than complete case analysis.*** Complete case analysis refers to when any case with a missing value on even one variable is excluded from the modelling. This did not occur during model construction as strategies were implemented to impute missing values.
- ***Interactions between the variables were examined.*** A number of interactions were examined based on the literature: age with cause of injury [160], systolic blood pressure with hypoxia and mean blood pressure with hypoxia [95]. However, the mean blood pressure/hypoxia interaction was not assessed since mean blood pressure was excluded from the modelling.
- ***More than 10 outcome events (i. e. survival) were included per each predictor.*** Model A contains 7 covariates and model B contains 10 variables. According to this criteria, the dataset must include at least 7 × 10 survivals for model A and 10 × 10 survivals for model B. The number of survivals in the dataset was 599. Furthermore, this criteria was met at all times when various

models were constructed for comparison or assessment of the variables during various levels of the modelling procedure.

- ***How to estimate the prognosis is explained.*** The models are presented with the coefficients and the constant and the way to make the estimate on prognosis in presented in Paper 1 (section 2.6).
- ***The models have high AUC (discrimination) (0. 92 and 0. 93 respectively for models A and B) and a significant p value for HL statistics (calibration) (0. 32 and 0. 29 respectively for models A and B).***
- ***The confidence interval is given for the odds ratio of each covariate in the model.***

Mushkudiani et.al. [159] made some recommendations for developing and validating prognostic models. These recommendations are divided into study population, predictors, outcome, model development and model validation and overall are the same as the criteria proposed by Perel *et al.* . With regards to the study population, our sample size (i.e. > 500) and its representation of the current practice (since majority of cases sustained the injury following Sep. 2005) are some indicators of the strength of our study according to Mushkudiani *et al.* . Moreover, the predictors used in our models are readily available. For example, GCS and pupillary reactivity are commonly measured in TBI patients on admission and CT scan is widely in use in severe TBI to diagnose intracranial pathology. The recommendations regarding the model development and validity is overall the same as above-mentioned indicators according to Perel's criteria.

Multivariate analysis in medical research is commonly used to account for the confounders. This is also referred to 'adjustment for confounders'. For

example, one may be interested to investigate the association of factor A with factor X when there is possibility that factor B or C may also affect factor X. This is important since the effect of factor A may be 'nullified' in the presence of factors B and C. To address this, a model is constructed where all factors A, B and C are supplied to this model and factor X is taken as outcome. Then the significance of association of each factor A, B or C with factor X is explored. Whilst this is a common approach in adjustment for confounders, in this PhD, a different method was devised as employed in Paper 4 (to assess the prognostic strength of various GCS subscores and combinations) and Paper 5 (to assess the prognostic strength of various intracranial pathologies). In the 'common approach of adjustment for confounders', all the prognosticators (including those of interest such as GCS subscores in Paper 6 along with the confounders as age and ISS) could have been supplied to a single model and then the significance of their associations with outcome could have been assessed. This approach was not taken in the PhD and instead 'baseline models' were first constructed which only included the confounders. Then the factor(s) under investigation were added to the baseline model to firstly assess the degree and significance of the change in the model performance (as per deviance, AUC, R^2 Nagelkerke etc.) and secondly to assess the significance of association with the outcome of interest in the resulting model. The disadvantage with the 'common approach of adjustment for confounders' is that firstly, it does not permit assessing the added value of the factor of interest when the outcome of interest can be predicted by other factors. Secondly, with the 'baseline model approach', comparing the predictive performance of various factors is possible

such as comparing the prognostic value of various intracranial pathologies as it was done in Paper 6.

9.3.2. S100B project

Summary from the respective papers

The results of this part were partially discussed in Papers 2 and 7.

Paper 2: briefly, S100B showed low specificity for outcome prediction at 3 months (60% for unfavourable outcome and 49% for death) despite having a high sensitivity of over 80% for either unfavourable or death prediction. This degree of prognostic performance may not be suitable for clinicians.

Paper 7: it was observed that the difference in S100B and prognostic models' performance is not significant. The importance of this relates to S100B as a simple blood test which is more familiar to clinicians to interpret (through knowing sensitivity and specificity) whereas prognostic models are unfamiliar tools in clinical context and the interpretation of their performance is more complicated. However, prognostic models use those characteristics of patients which are routinely measured and thus do not incur extra cost whereas S100B is not a part of routine laboratory tests. Moreover, a better prognostic tool may be the combination of brain injury biomarkers with other TBI prognosticators.

******

In order to compare the prognostic strength of S100B with multivariate models, TARN TBI models were run on 100 sample of TBI patients and their performance was measured for 3-month outcome prediction. TARN TBI models have AUCs and Nagelkerke R^2s of respectively more than 90% and

60% whereas these measures decline in the S100B dataset (For example, AUC of model A and B declines respectively by 0.28 and 0.20 for survival prediction and by 0.16 and 0.18 for favourable outcome prediction). There are a number of possible reasons for this. It may be due to small sample size of S100B dataset. Furthermore, TARN models are yet to be externally validated in other larger samples of TBI recruited in the British trauma care system (methodological, spectrum, and follow-up interval validation) and the drop in performance can be due to inherent lack of external validity. However, TARN TBI models have been developed to predict the outcome in short-run i.e. at discharge and they may perform poorly for the long-term outcome prediction such as 3-month survival or favourable outcome. This in fact can be one reason that the external validation of TARN model B on the IMPACT dataset yields a huge drop in performance. During the study design of this project, the drop in performance of TARN models in S100B was expected in part due to the sample size, and thus it was decided that, rather than using the exact coefficients of the models, the logistic regression be run again to obtain the coefficients from the dataset. Had this method not been applied, the performance of the TARN models would have perhaps experienced further drop.

9.3.3. Summary

According to many criteria suggested by Perel et al. and Mushkudiani et al. such as sample size, model presentation and performance ect. , the constructed models A and B can be considered well-developed. Moreover, the drop of these models performance in the S100B dataset may be due to small sample size of S100B data, long term outcome prediction (3 months versus

discharge) or inherent lack of validity. Furthermore, as S100B is the only TBI prognosticator which was found significant, its daily measurement may be a better choice than GCS or pupillary reactivity to monitor TBI progress.

9.4. Conclusion

9.4.1. PhD Hypotheses

Hypothesis 1: the probability of survival is not influenced by the patient characteristics in severe TBI.

Both TARN TBI models A and B contain information on patients characteristics. These characteristics are demographic (age), descriptive of severity and type of injury (ISS, major extracranial injury and CT findings) or physiological (GCS, pupillary reactivity, hypoxia and low blood pressure).

Using these characteristics in the logistic regression formula provides the probability of death (or survival) for a given TBI patient. This means if one characteristic changes while the others remain constant, it is expected the probability of outcome changes too. This is because various patient's characteristics demonstrate a significant p value in logistic regression analysis. As presented in Paper 1 (Table 10), the p value in logistic regression assists in assessing the null hypothesis that the coefficient of the respective variable is zero. A zero coefficient means the variable has no effects on the probability of outcome. Since these models demonstrate that the probability of outcome cannot be independent of various patient's characteristics, the above hypothesis is refuted.

The importance of this refuting is that if the probability of survival is influenced by the patients characteristics; using the patients characteristics can then help to make prediction about the subsequent outcome.

Hypothesis 2: the logistic regressing does not explain the pattern of mortality in severe TBI

Explaining the pattern in mortality can be taken as equivalent to the ability to predict the mortality. In this manner, random death can be taken as equivalent to 'absolute lack of pattern' and predicting the mortality with absolute accuracy can be taken as 'ability to explain the pattern'.

In our dataset of 802 TBI cases, two multivariate models were constructed through logistic regression. According to several measures of performance, these models can predict the survival (or mortality which is 1 minus probability of survival). However, none of the measures of performance demonstrate absolute accuracy. For example, both constructed models have Nagelkerke R^2 of more than 60%. This means that each model is able to explain 60% (majority) of the variability in mortality. This is also the case for AUC and classification accuracy which are on the one hand above 0.50 (random event) and on the other hand less than 1 (absolute accurate prediction or definitive event).

Regarding the performance of the models, the above hypothesis can be refuted in that these models do not suggest 'absolute lack of pattern' although they are not able to provide a definitive pattern for mortality as well.

Being able to explain the pattern of mortality by logistic regression, despite not being absolutely accurate, suggests that this prediction is not random and different patients can be regarded as with various risks of mortality.

The prediction of mortality can be made by using the patient's characteristics as per the covariates included in either model A or B. However, it is important to note that the models constructed in Paper 3 which use only age, ISS and GCS (Table 17) can help with refuting of the above hypotheses

used in prognostication of outcome in severe TBI and the lack of difference between the performances of these models with S100B model fails to refute the above hypothesis.

Hypothesis 5: There is no difference in prognostic performance between multivariate models which do/do not contain S100B as a predictor in severe TBI.

The performance of TARN TBI models A and B do not experience a statistically significant increase following addition of S100B to these models as per AUC (due to the overlap of CIs). Since TARN TBI models A and B can represent multivariate models, the study fails to refute the above hypothesis.

The power of the S100B project: type II error in assessing hypotheses 3, 4 and 5 is fairly strong due to the small sample size for logistic regression analysis. This means that the probability that the above hypotheses are failed to be rejected when they are not true in reality is high.

However, despite from statistical perspective, the hypotheses 3, 4 and 5 are failed to be refuted, there is tendency for hypotheses 3 and 5 to be refuted on the following basis: despite non-statistical significant difference between S100B and expanded models (for hypothesis 3) and between TARN TBI models with and without S100B (for hypothesis 5) as per AUC, the changes in AUCs are above the clinical significant cut-off point as discussed in Paper 7. The important of this relates to the future research on this topic (sections 8.5.4, 9.4.2, 9.4.3 and 10).

9.4.2. PhD objectives

Objective 1: to develop a prognostic tool to predict the survival in TBI applicable to the British trauma care.

Two prognostic models (model A and B) have been developed which enable calculation of the survival probability at discharge for a given TBI patients. Both models share age, GCS, pupillary reactivity, the brain stem injury, hypoxia and low blood pressure. However, model A contains ISS which is absent in model B and instead Model B uses extracranial injury, cause of injury and brain swelling. The models are presented with the odds ratio and the coefficient is given for each covariate. Similarly the models performance is presented according to several measures of performance. The models hold a high AUC (0.92 and 0.93 respectively for model A and model B). Similarly, the models have acceptable calibration per HL statistics (p values 0.20 and 0.19 respectively for model A and model B). Furthermore, the criteria for a well-developed model as proposed by Perel *et al.* [31] and Mushkudiani *et al.* [159] were followed during the modeling which implies the final models can be considered well-developed. The models are externally validated and they still hold their high performance.

As these models perform well in different TBI series from TARN, it is reasonable to suggest these models for the purpose of benchmarking of TBI care delivery. However, in order for clinical validity they still need further

validation in other TBI datasets. Similarly, their use in clinical trials is still unclear as the derivation dataset was from a trauma registry.

Unfortunately, S100B can not be suggested as a prognostic tool at this stage as its specificity was found low in a sample of 100 cases of TBI despite high sensitivity.

Objective 2: to ascertain among a multivariate model and a blood test which one is better to be used for prognosis in TBI.

TARN TBI models A and B and S100B (in isolation) were run on the data of 100 cases and it was observed that their performance is not significantly different (either clinically or statistically). Due to the small sample size this finding is not conclusive. However, it highlights the importance of taking this issue further in large datasets of S100B to obtain more powerful results. It is only after demonstrating that prognostic models outperform S100B, any further research on this topic may be deemed less necessary.

Objective 3: to determine whether a combination of multivariate models and a blood test can significantly improve the prognosis in TBI.

In this PhD, it was observed that either addition of S100B to current prognostic models or considering other TBI prognosticators alongside S100B causes a clinically significant increase in the performance of each tool in isolation. This may lead to achieve a better prognostic tool than those currently available. Unfortunately, the small sample size was unable to offer neither powerful

results (which can provide statistical significance) nor a prognostic tool which employs a combination of S100B with other patients' characteristics.

Despite lack of statistical significance, the clinical significance in change of models' performance highlights the importance of future research on this topic in larger datasets (adding S100B to prognostic models or considering other TBI prognosticators along with S100B). Furthermore, as a rise in S100B performance was observed according to AUC, this may explain the variation in AUCs reported in the literature as well. However, the issue as to clinical versus statistical significance stills holds in this matter as well.

9.4.3. Aim of PhD: to improve our understanding of prognosis in TBI

In this PhD, two prognostic models have been proposed for TBI prognosis. These models can compete with the IMPACT and the CRASH models in terms of the methodology employed for their derivation and also the performance. However and more importantly, they are suitable for the British trauma care system and take account of the cotemporary changes in trauma care. Furthermore, they are derived from an ongoing observational project.

Although the results of S100B study are subject to significant type II error, this PhD also adds a new insight about brain injury biomarkers compared to prognostic models which have to be tested in the future. As the performance of these two prognostic tools do not appear significantly different, their usage may depend only on how they are accepted in various settings (clinical or trials) based on factors related to their practicality or popularity. However, it was observed the combination of these two tools may offer a better performance. This opens possibility for future research in construction of

prognostic tools in TBI. This might be in fact the reason why some studies achieved a much higher S100B predictability than prognostic models in contrast to other studies. It may be due to different case-mix according to various factors affecting TBI prognosis.

9.4.4. Summary

The PhD refutes hypotheses 1 and 2 but fails to refute hypotheses 3, 4 and 5. Regarding the objectives, two well-developed prognostic models have been proposed for the British TBI population. These models have a high performance. Moreover, the idea of comparing the prognostic performance of S100B with that of prognostic models and also their combination needs to be taken forward in the future research on TBI prognosis.

10. Future directions

Summary from the respective papers

Various future directions were discussed in each paper tailored to the particular results of the paper and its suitability for a journal article:

Paper 3: it is important that the GCS prognostic analysis be performed in a different dataset of TBI (based on locality or time) and also with adjustment with pupillary reactivity. Furthermore, survival at discharge is not the only endpoint of outcome as longer-term outcome and also disability are important to consider as well.

Paper 4: the proposed method to use AIS codes to perform the Marshall CT Classification requires further validation to test how accurate the proposed method is.

Paper 5: with regards to prognostic analysis of various intracranial pathologies or CT findings, it is important to investigate the effect of the number of pathologies on the outcome and also on the relationship of each pathology with outcome.

Paper 6: the final TBI prognostic models proposed (models A and B), both require further external validation from various dimensions such as geographical, methodological, spectrum ect.. Moreover, as the models only predict survival at discharge, there is still scope for models which can predict disability or long-term outcome. Overall, in the future, it is important to assess the safety of prognostic models as they may lead to early withdrawal of therapy.

Paper 7: regarding the S100B study, since this study suffers from lack of enough power and the results demonstrate that there may not be significant difference between S100B and prognostic models in outcome prediction, it is

important that the results be validated in a lager cohort of TBI cases and the idea of comparing prognostic models with S100B (or brain injury biomarkers) and the combination of these two, be taken forward.

******

In Paper 3 we constructed a baseline model with only age and ISS to reach an AUC of 0.84. This degree of performance per AUC is somewhat similar to the CRAH and the IMPACT models although the CRASH and IMPACT models contain more variables such as pupillary reactivity and CT findings. Moreover, in this paper, following addition of GCS to the baseline model, a high AUC of 0.91 was reached with only three variables (age, GCS and ISS). Similarly Healey *et al.* [22] constructed a survival prediction model with only GCS and achieved an AUC of 0.89 which is well above the AUCs in the CRASH and the IMPACT dataset (the models constructed by Healey *et al.* were for the purpose of assessing the prognostic value of various GCS components and not to suggest a prognostic tool. However, the models demonstrated that GCS on its own may offer a high AUC for outcome prediction). The reason for these differences is not clear but they highlight that less complicated models which contain less predictors may be able to provide the same degree of predictive performance. This requires a further investigation and consideration should be given to measures of model performance other than AUC. Although AUC is the most common index to measure the discrimination power of a model, other measures of discrimination (Nagelkerke R^2) and also other measures of model calibration apart from HL

statistics such as brier score may be important. For instance this is reflected in the HL statistics of the GCS models in Paper 3 which, despite having high AUCs and Nagelkerke R^2, have p values of close to 0 (a good model is expected to have a p value of above 0.05 by HL statistics). It may be that depending on the purpose of the model i.e. using in clinical settings, in clinical trials or for trauma care benchmarking, the desired performance of a model differ i.e. for instance, a high AUC of more than 0.85 of a model for a clinician may not be necessarily required for a model which is to be used in clinical trials.

Investigating these issues requires joint clinical and statistical knowledge and expertise. On the one hand, multivariate statistics involves high level and meticulous knowledge of mathematical statistics which knowing all of the details involved may not be necessary for a clinician who is interested in clinical and health care aspects of the results. As such, knowing the concept, essentiality and relevance of each statistical measure may suffice to conduct a valid study. This was one of the skills which the investigator developed. This skill continues to develop more by 'bridging' the medical clinical knowledge with medical research and statistical knowledge. On the other hand, medical statisticians can not be left alone to perform their analysis and interpret the results without considering the clinical and practical context of the research. The simplest example of this reflects in the idea of clinical and statistical significance in that the blood pressures of two groups of patients may be significantly different in a statistical analysis but this difference may not hold clinical meaning with regards to the magnitude of the difference.

There are a number of variables which their prognostic value have been shown in multivariate analysis of other TBI datasets such as sodium [97], PH [97], haemoglobin [24, 97], Glucose[97], platelets count [97] and prothrombin time [97]. Since, TARN does not hold records of these variables and our S100B sample of TBI did not include enough cases with these observations (for example there were 93.6% anaemic patients), investigating the prognostic value of these laboratory variables on their own and also after adjustment with other TBI prognosticators such as age or GCS was not possible. Knowing this is important with regards to the timely and sufficient intervention in case of disturbance in any of these laboratory measures. It is a matter of future work to examine this in a larger up-to-date dataset of TBI cases recruited from a British trauma care centre(s) with an observational study design.

One of the applications of our proposed models is in the clinical context of treatment and management of TBI patients. Although with the help of these models, more accurate prediction on outcome can be made, it is yet unclear how such prediction may or should affect the clinical decision making or allocation of recourses. Firstly, it has to be examined in a clinical trial that how using these models would affect the patient care / outcome. Secondly, as these models enable stratification of patients based on their risk of experiencing the death, more research is needed to investigate how this stratification can be performed. It may be that some cut-offs should be obtained to group TBI patients as low, moderate or high risk [36].

The earliest research into possible S100B role in TBI prognosis may date back to the year 1998. However, despite all the huge amount of research into S100B so far (a Pubmed search for S100B and TBI using MeSH terms

results in 136 journal articles), this biomarker has not received the acceptance in clinical settings. The exact reason for this is unclear but it may be due to the lack of adequate prognostic performance of S100B in terms of sensitivity and specificity. For example, in our series of 100 TBI cases, S100B demonstrates a sensitivity of more than 80% for prediction of 3-month death or unfavourable outcome when the specificity is low (60% for unfavourable outcome prediction and 49% for death prediction). However, when using the same measure of performance as to those used for prognostic models (i.e. AUC or R^2 Nagelkerke), it was observed that the difference between prognostic strength of S100B and multivariate models is not substantial. Thus if lack of clinical acceptance of S100B is due to its lack of prognostic strength, this may apply to multivariate models as well. Since the current reliable prognostic models have been introduced only recently (the introduction years of CRASH and IMPACT models is 2008 and that of TARN TBI models is 2010), the same 'fate' may occur to prognostic models unless their usefulness over clinical judgment or 'guessing' be shown in a clinical trial.

It is important to validate the proposed models in this PhD on CRASH and IMPACT data. This would address the potential benefit of the models in clinical trials as CRASH and IMAPCT data are mainly trial data. In the same way, this validation would address the performance of the models in a different geographic regions as CRASH and IMPACT *are* not mainly British. On the other hand, validating the CRASH and IMPACT models on TARN data explores the performance of these models in a trauma registry or on observation data. The cross-validation of TARN, CRASH and IMPACT models is a matter of future research to explore the above issues.

Patel. *et al.* demonstrated that those TBI patients who receive neurosurgical care, have higher odds of survival [129]. Despite this, there is reluctance to transfer all brain injury patients to neurosurgical centres and currently, the preference is given to patients who have higher chance to benefit from surgical intervention. It appears this approach deprives many other TBI patients from effective trauma care. Using prognostic models to classify the patients in various risk categories may facilitate decisions on admission or transfer of the patient to a neurosurgical centre. This should be investigated in a clinical trial following determination of how TBI prognostic models can define high risk cases. However, supply of neurosurgical care to would-be-then-defined high risk TBI patients is perhaps not the only adaptation which is required in the management and care of brain injury. Similar changes may be made in the rehabilitation of these patients by providing more immediate or aggressive rehabilitative interventions.

The proposed TBI prognostic models are presented in tables of odds ratio with coefficients for each predictor and the constant. Thus, using the logistic regression formula, the survival probability of a TBI case with given characteristics can be calculated (section 2.6.1). There are other ways of model presentation such as scoring systems (by means of tables, graphs, nomogram or linear equation) or web-based calculators. It is unknown which presentation is most user-friendly. A web-based calculator may be an appealingly quick tool for clinicians in that the probability of outcome can be supplied without any onerous calculation by just putting the already-known patients' characteristics in the blank boxes of a web-page. However, there are issues with regards to

inter-observer and intra-observer reliability of each presentation including the web-based calculator. The clinical appeal is only one factor which has to be taken into account in examining which way of presentation is the best. However, the TARN TBI models may deserve to be presented as an on-line calculator in the future similar to current TARN TRISS calculator [89].

In this study, the prognostic value of S100B at a certain time-point after injury (i.e. 24-hours after injury) was taken into account. Firstly, it is unclear what time point of S100B is the best for outcome prediction. On the one hand, 24-hours following admission may be a better approach as the patient is expected to be stabilised following initial resuscitation which could eliminate the confounding effect of secondary insult (such as hypotension or hypoxia). These factors may relate both to S100B rise through causing ongoing brain damage and also outcome. On the other hand, even 24-hours after injury, 31% of our TBI population samples were hypotensive. Future direction should ascertain whether or not better option as to the time elapsed after injury for blood sampling is post-resuscitation irrespective when this is achieved; immediately after admission or some time after. In fact, the time point of sampling is one factor which differ across various studies on S100B prognosis. This factor may be able to explain different results on S100B prognostic performance observed in the literature review.

The other option may be to account for the temporal pattern of S100B serum concentrations rather than reliance on a single blood measurement. Regarding this, one or more of these following factors in temporal course of S100B measured at certain intervals (such as daily) during a certain time period (such as till 7 days after injury) could be considered: (1) higher initial values,

(2) secondary increase over time [76], (3) the peak concentration [79] and (4) the average time for S100B to return to normal values [165, 182] . Similarly, it requires comparing prognostic value of S100B in this manner to multivariate models without S100B or combination of one single S100B measurement with other TBI prognosticators.

It appears that if only the performance of the two prognostic tools (i.e. multivariate models and a brain injury blood biomarker) is taken into account, there is no difference in which one is being used for outcome prediction. However, performance is not the only factor important in using a prognostic tool. If application of a prognostic tool in clinical practice is regarded as application of a diagnostic test or an' intervention', then it is also important to compare the prognostic models with brain injury biomarker in a clinical trial. This off course has to be conducted following demonstrating that overall, using prognostic tools would be beneficial for the patients' outcome rather than exclusive reliance on clinical judgement or intuition in anther initial clinical trial.

It is important that the prognostic strength of S100B and multivariate models be compared in another cohort of TBI patients from a different country/centre which includes more heterogeneous of TBI cases in a larger dataset than ours. The sampling procedure of the S100B study occurred only from the ICU patients which represents the extreme severity of brain damage. Furthermore, all 100 patients had abnormal CT findings. Thus the dataset can not represent those brain injury patients who have normal CT such as diffuse axonal injury (or the Marshal Class I). This is a confounding factor since the

TBI patients with observed abnormality in their CT have a poorer outcome than those with normal CT.

The method of adjustment for confounders using the baseline models which was devised in Papers 3 and 5, requires to be further explored both from statistical and clinical perspective. Adjustment for cofounders can be done in either of the following ways: (1) to supply all factors (confounders and the factor of interest) to one single model and to assess the significance of association with outcome in the resulting model (the common method of adjustment for confounders), (2) to supply the factor of interest to a baseline model which includes the confounders and then to assess the significance of associations and changes in the model's performance. The latter particularly permits a better comparison of the association of several factors with the outcome of interest when confounders are taken into account. However, the statistical or clinical importance of the results is unclear if, for example, one factor appears non-significant in the model but significantly contributes to its performance or vice versa.

10.1.1. Summary

The proposed models are still required to be validated in a TBI dataset of different location, spectrum of severity etc. Furthermore, as these models were not developed for disability or long-term outcome prediction, separate models are required for these predictions which are suitable for the British trauma care. In the same way, the prognostic value of some factors such as Hb, Glucose, Plt count, prothrombin time are still required to be examined in a recent British dataset as this can help with timely therapeutic decisions.

It is a matter of future research to determine why some models in the literature have high performance according to their AUC whereas they include only a few covariates. It may be that these models do not have acceptable performance as per other measures of performance or using less covariates truly does not change the performance.

It is also important to determine how the clinical decision making can be affected by introducing prognostic models to the clinical setting. It is still unclear whether or not these models are useful or they are indeed harmful in clinical practice. However, they may be helpful on stratifying TBI patients in various risk groups to allocate resources such as neuro-ICU.

Moreover, since the analysis of the small S100B dataset showed no significant difference in prognostic performance of S100B and prognostic models and also this analysis suggested the combination of these two tools might enhance the prognostic strength of either alone. It is important that these findings be validated in a larger dataset which is more heterogeneous in terms of TBI severity as well. Despite the substantial type II error, the S100B study demonstrates the importance of future research on brain injury biomarkers versus prognostic models.

11. Appendix

In this appendix, models derived at each level of modelling procedure are presented. The models are presented in tables including the variables supplied to the logistic regression, significance (sig.), odds ratio, and the 95% confidence interval for the odds ratio. In this appendix, wherever ISS is referred to, it is in the form of $\log_e\left(\frac{ISS}{10}\right)-0.91\cdot$

Model I

		Sig.	*Odds ration*	*95. 0% C. I. for odds ratio*	
Age		.000	.952	.940	.965
$\left(\frac{10}{GCS+1}\right)^2 - 0.76$		.000	.679	.588	.785
Pupillary reactivity	**Brisk-brisk**				
	Sluggish-sluggish	.003	.299	.134	.669
	Brisk-sluggish	.106	.324	.083	1.271
	Absent-brisk	.424	2.428	.276	21.321
	Absent-sluggish	.002	.057	.009	.359
	Absent-absent	.000	.076	.034	.167
ISS		.000	.141	.057	.351
extracranial		.041	2.146	1.033	4.460
Constant		.000	93.506		

Model II

		Sig.	*Odds ratio*	*95. 0% C. I. for odds ratio*	
Age		.000	.952	.939	.964
$\left(\frac{10}{GCS+1}\right)^2 - 0.76$		.057	.370	.133	1.030
$\left(\frac{10}{GCS+1}\right)^2 \times \log_e\left(\frac{GCS+1}{10}\right) - 1.02$		.239	.558	.211	1.474
Pupillary reactivity	**Brisk-brisk**				
	Sluggish-sluggish	.008	.330	.145	.753
	Brisk-sluggish	.143	.360	.092	1.414
	Absent-brisk	.367	2.773	.303	25.376
	Absent-sluggish	.003	.062	.010	.380
	Absent-absent	.000	.081	.036	.181
ISS		.000	.147	.059	.365
Extracranial injury		.047	2.102	1.010	4.376
Constant		.000	53.629		

Model IIIA (AUC: 0.91)

		Sig.	*Odds ratio*	*95. 0% C. I. for odds ratio*	
Age		.000	.952	.940	.964
$\left(\frac{10}{GCS+1}\right)^2 - 0.76$		.010	.283	.108	.738
$\left(\frac{10}{GCS+1}\right)^2 \times \log_e\left(\frac{GCS+1}{10}\right) - 1.02$		.092	.456	.183	1.137
Pupillary category	**Both reactive**				
	Only one reactive	.247	.556	.206	1.502
	None reactive	.000	.127	.062	.261
ISS		.000	.272	.139	.529
Constant		.000	47.405		

Model IIIB (AUC: 0.90)

		Sig.	*Odds ratio*	*95. 0% C. I. for odds ratio*	
Age		.000	.955	.943	.967
$\left(\frac{10}{GCS+1}\right)^2 - 0.76$		.003	.240	.094	.615
$\left(\frac{10}{GCS+1}\right)^2 \times \log_e\left(\frac{GCS+1}{10}\right) - 1.02$		.049	.407	.166	.996
Pupillary category	**Both reactive**				
	Only one reactive	.119	.462	.175	1.221
	None reactive	.000	.122	.060	.247
Extracranial injury		.409	.797	.465	1.366
Constant		.000	43.937		

Model IV

		Sig.	*Odds ratio*	*95. 0% C. I. for odds ratio*	
Age		.000	.950	.938	.963
GCS	**15**				
	14	.000	.058	.022	.148
	13	.009	.197	.058	.668
	12	.001	.093	.022	.397
	11	.275	.414	.085	2.014
	10	.028	.233	.063	.857
	9	.613	1.754	.199	15.500
	8	.161	.313	.062	1.586
	7	.004	.158	.044	.561
	6	.122	.353	.095	1.320
	5	.741	.769	.161	3.664
	4	.200	2.875	.571	14.459
	3	.752	1.166	.449	3.026
Pupillary category	**Both reactive**				
	Only one reactive	.318	.596	.216	1.646
	None reactive	.000	.134	.064	.282
ISS		.000	.255	.128	.506
Constant		.000	154.557		

Model V (AUC: 0.908)

		Sig.	*Odds ratio*	*95. 0% C. I. for odds ratio*	
	Age	.000	.955	.943	.967
GCS	**Mild**				
	moderate	.001	.277	.129	.596
	severe	.000	.138	.070	.272
Pupillary category	**Both reactive**				
	Only one reactive	.089	.437	.169	1.133
	None reactive	.000	.077	.039	.152
	ISS	.000	.234	.121	.452
	Constant	.000	141.810		

Model VI A (AUC: 0.916)

		Sig.	*Odds ratio*	*95. 0% C. I. for odds ratio*	
Age		.000	.960	.947	.973
$\left(\frac{10}{GCS+1}\right)^2 - 0.76$		.006	.256	.097	.681
$\left(\frac{10}{GCS+1}\right)^2 \times \log_e\left(\frac{GCS+1}{10}\right) - 1.02$		.068	.421	.166	1.066
Pupillary reactivity	**Both reactive**				
	Only one reactive	.190	.507	.184	1.401
	None reactive	.000	.109	.051	.232
ISS		.000	.241	.118	.493
Cause of injury	**RTA**				
	Fall	.011	.403	.200	.813
	Assault	.256	1.738	.670	4.507
	others	.433	.541	.117	2.507
Constant		.000	46.058		

Model VIIA (AUC: 0.922)

		Sig.	*Odds ratio*	*95. 0% C. I. for odds ratio*	
Age		.000	.959	.946	.973
$\left(\frac{10}{GCS+1}\right)^2 - 0.76$		.034	.336	.123	.922
$\left(\frac{10}{GCS+1}\right)^2 \times \log_e\left(\frac{GCS+1}{10}\right) - 1.02$		.190	.528	.203	1.371
Pupillary reactivity	**Normal**				
	Abnormal-both reactive	.009	.363	.170	.773
	Only one reactive	.063	.372	.131	1.055
	None reactive	.000	.077	.034	.173
ISS		.000	.237	.115	.489
Cause of injury	**RTC**				
	fall	.012	.404	.199	.822
	Assaults	.253	1.742	.672	4.511
	others	.425	.546	.123	2.415
Constant		.000	69.892		

Model VIIB (AUC: 0.914)

		Sig.	*Odds ratio*	*95. 0% C. I. for odds ratio*	
Age		.000	.955	.943	.967
GCS	**mild**				
	moderate	.003	.310	.143	.671
	severe	.000	.179	.088	.360
Pupillary reactivity	**Normal**				
	Abnormal- both reactive	.003	.337	.165	.689
	Only one reactive	.020	.311	.117	.830
	None reactive	.000	.053	.025	.110
ISS		.000	.232	.119	.454
Constant		.000	165.080		

Model VIIIA (AUC: 0.917)

		Sig.	*Odds ratio*	*95. 0% C. I. for odds ratio*	
Age		.000	.961	.948	.974
$\left(\frac{10}{GCS+1}\right)^2 - 0.76$		.019	.301	.111	.822
$\left(\frac{10}{GCS+1}\right)^2 \times \log_e\left(\frac{GCS+1}{10}\right) - 1.02$		.124	.475	.183	1.228
Pupillary reactivity	**Normal**	.			
	Abnormal-both reactive	.007	.358	.170	.755
	Only one reactive	.036	.326	.115	.927
	None reactive	.000	.061	.027	.140
Extracranial injury		.001	.283	.137	.585
Cause of injury	**RTC**				
	fall	.021	.438	.217	.883
	Assaults	.178	1.894	.748	4.797
	others	.531	.631	.150	2.664
Constant		.000	72.977		

Model VIIIB (AUC: 0.910)

		Sig.	*Odds ratio*	*95. 0% C. I. for odds ratio*	
Age		.000	.964	.951	.977
GCS	**mild**				
	moderate	.004	.312	.142	.688
	severe	.000	.155	.077	.314
Pupillary reactivity	**Normal**				
	Abnormal-both reactive	.003	.334	.163	.686
	Only one reactive	.006	.247	.091	.671
	None reactive	.000	.036	.017	.079
Extracranial injury		.000	.241	.121	.482
Cause of injury	**RTC**				
	fall	.017	.436	.221	.863
	Assaults	.180	1.846	.753	4.525
	others	.536	.642	.157	2.619
Constant		.000	215.875		

Model XIA1 (AUC: 0.92)

		Sig.	*Odds ratio*	*95. 0% C. I. for odds ratio*	
Age		.000	.952	.941	.963
GCS	**mild**				
	moderate	.020	.425	.206	.876
	severe	.000	.225	.122	.414
Pupillary reactivity	**Normal**				
	Abnormal-both reactive	.004	.390	.205	.743
	Only one reactive	.002	.253	.108	.592
	None reactive	.000	.048	.024	.095
ISS		.000	.258	.138	.482
Brain stem injury		.000	.220	.100	.484
Brain swelling		.080	.637	.385	1.055
haemorrhage		.152	1.549	.851	2.817
Cerebellar injury		.072	.422	.165	1.081
Constant		.000	171.466		

Model XIA2 (AUC: 0.92)

		Sig.	*Odds ratio*	*95. 0% C. I. for odds ratio*	
Age		.000	.954	.943	.965
GCS	**mild**				
	moderate	.017	.417	.203	.858
	severe	.000	.228	.124	.420
Pupillary reactivity	**Normal**				
	Abnormal-both reactive	.004	.388	.204	.736
	Only one reactive	.002	.251	.107	.590
	None reactive	.000	.049	.025	.098
ISS		.000	.268	.143	.502
Brain stem injury		.000	.232	.17	.507
Brain swelling		.071	.630	.382	1.041
Cerebellar injury		.084	.442	.175	1.117
Constant		.000	220.067		

Model XIA3 (AUC: 0.92)

		Sig.	*Odds ratio*	*95. 0% C. I. for odds ratio*	
Age		.000	.953	.942	.964
GCS	**mild**				
	moderate	.021	.430	.210	.881
	severe	.000	.212	.116	.387
Pupillary reactivity	**Normal**				
	Abnormal-both reactive	.006	.413	.218	.780
	Only one reactive	.003	.278	.120	.647
	None reactive	.000	.049	.025	.097
ISS		.000	.219	.120	.403
Brain stem injury		.000	.185	.085	.402
Haemorrhage		.162	1.526	.844	2.760
Constant		.000	132.121		

Model XIA4 (AUC: 0.92)

		Sig.	*Odds ratio*	*95. 0% C. I. for odds ratio*	
Age		.000	.955	.944	.966
GCS	**mild**				
	moderate	.019	.425	.208	.867
	severe	.000	.216	.118	.393
Pupillary reactivity	**Normal**				
	Abnormal- both reactive	.006	.411	.218	.775
	Only one reactive	.003	.276	.118	.645
	None reactive	.000	.051	.026	.100
ISS		.000	.229	.125	.420
Brain stem injury		.000	.194	.090	.421
Constant		.000	166.557		

Model XIB1 (AUC: 0.92)

		Sig.	*Odds ratio*	*95. 0% C. I. for odds ratio*	
Age		.000	.957	.945	.969
GCS	**mild**				
	moderate	.015	.402	.193	.837
	severe	.000	.217	.117	.402
Pupillary reactivity	**Normal**				
	Abnormal-both reactive	.006	.406	.212	.776
	Only one reactive	.000	.210	.088	.501
	None reactive	.000	.037	.018	.077
Extracranial injury		.000	.227	.122	.422
Cause of injury	**RTC**				
	fall	.100	.605	.333	1.101
	Assaults	.138	1.880	.817	4.329
	others	.973	1.023	.278	3.768
Brain stem injury		.000	.164	.074	.364
Brain swelling		.003	.466	.282	.768
Haemorrhage		.529	1.213	.666	2.209
Constant		.000	295.630		

Model XIB2 (AUC: 0.92)

		Sig.	*Odds ratio*	*95. 0% C. I. for odds ratio*	
Age		.000	.958	.946	.970
GCS	**mild**				
	moderate	.014	.401	.193	.833
	severe	.000	.220	.119	.408
Pupillary reactivity	**Normal**				
	Abnormal-both reactive	.006	.402	.211	.767
	Only one reactive	.000	.210	.088	.504
	None reactive	.000	.038	.018	.077
Extracranial injury		.000	.225	.121	.418
Cause of injury	**RTC**				
	fall	.106	.611	.336	1.110
	Assaults	.136	1.888	.819	4.353
	others	.968	.974	.269	3.525
Brain stem injury		.000	.167	.075	.370
Brain swelling		.003	.468	.284	.772
Constant		.000	329.004		

Model XIIA

		Sig.	*Odds ratio*	*95. 0% C. I. for odds ratio*	
Age		.000	.952	.939	.964
GCS	**mild**				
	moderate	.064	.470	.211	1.046
	severe	.000	.254	.128	.506
Pupillary reactivity	**Normal**				
	Abnormal-both reactive	.001	.303	.149	.617
	Only one reactive	.001	.172	.062	.478
	None reactive	.000	.063	.029	.138
ISS		.001	.283	.135	.594
Systolic blood pressure	**Normtension**				
	Hypotension	.507	.764	.345	1.694
	Hypertension	.956	.983	.527	1.831
Mean arterial blood pressure	**Normtension**				
	Hypotension	.233	.394	.086	1.817
	Hypertension	.700	.809	.276	2.374
hypoxia		.002	.270	.116	.627
Brain stem haemorrhage		.000	.139	.059	.327
Constant		.000	340.535		

Model XIIB

		Sig.	*Odds ratio*	*95. 0% C. I. for odds ratio*	
Age		.000	.954	.941	.968
GCS	**mild**				
	moderate	.037	.411	.178	.947
	severe	.000	.256	.125	.524
Pupillary reactivity	**Normal**				
	Abnormal-both reactive	.001	.292	.141	.606
	Only one reactive	.000	.111	.036	.338
	None reactive	.000	.049	.022	.111
extracranial_AIS4		.002	.316	.150	.666
Cause of injury	**RTC**				
	fall	.044	.488	.243	.981
	Assaults	.833	1.105	.438	2.786
	others	.919	.916	.168	5.004
swelling		.000	.322	.180	.576
Brain stem haemorrhage		.000	.148	.061	.354
Systolic blood pressure	**Normtension**				
	Hypotension	.597	.800	.350	1.827
	Hypertension	.985	1.006	.531	1.908
Mean arterial blood pressure	**Normtension**				
	Hypotension	.154	.307	.060	1.557
	Hypertension	.658	.776	.253	2.383
hypoxia		.001	.208	.085	.510
Constant		.000	922.011		

Model XIVA1(AUC = 0.932)

		Sig.	*Odds ratio*	*95. 0% C. I. for odds ratio*	
Age		.000	.954	.943	.964
GCS	**Mild**				
	Moderate	.012	.401	.196	.818
	Severe	.000	.237	.129	.435
Pupillary reactivity	**Normal**				
	Abnormal-both reactive	.005	.396	.208	.753
	Only one reactive	.006	.292	.122	.701
	None reactive	.000	.052	.026	.103
ISS		.000	.249	.135	.457
Brain stem injury		.000	.179	.082	.391
Hypoxia	**Yes**	.001	.270	.128	.568
	Missing	.067	.501	.239	1.049
Constant		.000	215.422		

Model XIVA2 (AUC=0.927)

		Sig.	*odds ratio*	*95. 0% C. I. for odds ratio*	
Age		.000	.952	.941	.964
GCS	**Mild**				
	Moderate	.030	.447	.216	.924
	Severe	.000	.272	.146	.506
Pupillary reactivity	**Normal**				
	Abnormal-both reactive	.002	.360	.188	.688
	Only one reactive	.002	.253	.104	.617
	None reactive	.000	.060	.029	.123
ISS		.000	.262	.139	.493
Brain stem injury		.000	.159	.072	.351
Hypoxia	**Yes**	.001	.291	.136	.621
	Missing	.413	.702	.300	1.638
Mean blood pressure	**Normotension**				
	Hypotension	.013	.193	.053	.707
	Hypertension	.702	1.173	.519	2.650
	Missing	.159	.343	.078	1.519
Constant		.000	203.806		

Model XIVA3 (AUC=0.927)

		Sig.	*Odds ratio*	*95. 0% C. I. for odds ratio*	
Age		.000	.953	.942	.964
GCS	**Mild**				
	Moderate	.029	.445	.215	.920
	Severe	.000	.272	.146	.506
Pupillary reactivity	**Normal**				
	Abnormal-both reactive	.002	.359	.188	.688
	Only one reactive	.002	.252	.103	.613
	None reactive	.000	.059	.029	.121
ISS		.000	.260	.138	.490
Brain stem injury		.000	.159	.072	.351
Hypoxia	**Yes**	.001	.286	.135	.608
	Missing	.404	.697	.299	1.626
Mean blood pressure	**No-hypotension**				
	Hypotension	.001	.168	.056	.504
	missing	.074	.300	.080	1.123
Constant		.000	233.249		

Model XIVA4 (AUC= 0.924)

		Sig.	*Odds ratio*	*95. 0% C. I. for odds ratio*	
Age		.000	.951	.940	.963
GCS	**Mild**				
	Moderate	.019	.423	.206	.867
	Severe	.000	.238	.128	.442
Pupillary reactivity	**Normal**				
	Abnormal-both reactive	.004	.379	.197	.730
	Only one reactive	.005	.282	.116	.684
	None reactive	.000	.057	.028	.115
ISS		.000	.255	.138	.472
Brain stem injury		.000	.180	.083	.391
hypoxia	**Yes**	.001	.272	.128	.577
	Missing	.225	.609	.273	1.356
Systolic blood pressure	**Normotension**				
	Hypotension	.057	.553	.300	1.019
	Hypertension	.809	1.075	.599	1.928
	Missing	.363	.483	.100	2.322
Constant		.000	275.198		

Model XIVA5 (AUC=0.924)

		Sig.	*Odds ratio*	*95. 0% C. I. for odds ratio*	
Age		.000	.951	.940	.963
GCS	**Mild**				
	Moderate	.019	.423	.206	.866
	Severe	.000	.241	.130	.444
Pupillary reactivity	**Normal**				
	Abnormal-both reactive	.003	.376	.196	.722
	Only one reactive	.005	.281	.116	.683
	None reactive	.000	.057	.028	.115
ISS		.000	.256	.138	.473
Brain stem injury		.000	.180	.083	.392
Hypoxia	**Yes**	.001	.270	.127	.571
	Missing	.225	.609	.273	1.357
Systolic blood pressure	**No-hypotension**				
	Hypotension	.022	.533	.311	.912
	missing	.329	.465	.100	2.167
Constant		.000	279.830		

Model XIVB1 (AUC= 0.928)

		Sig.	*Odds ratio*	*95. 0% C. I. for ratio*	
Age		.000	.956	.945	.969
GCS	**Mild**				
	Moderate	.010	.380	.183	.790
	Severe	.000	.242	.129	.453
Pupillary reactivity	**Normal**	.000			
	Abnormal-both reactive	.005	.388	.201	.748
	Only one reactive	.001	.220	.088	.549
	None reactive	.000	.040	.019	.082
Injury cause	**RTC**				
	Fall	.991	1.008	.269	3.779
	Assaults	.471	.612	.161	2.328
	Others	.393	1.908	.434	8.396
Extracranial injury		.000	.272	.145	.512
Brain stem injury		.000	.163	.073	.363
swelling		.001	.424	.253	.710
Hypoxia	**Yes**	.001	.253	.115	.556
	Missing	.134	.561	.263	1.196
Constant		.000	415.746		

Model XIVB2 (AUC=0.932)

		Sig.	*Odds ratio*	*95. 0% C. I. for odds ratio*	
Age		.000	.955	.943	.968
GCS	**Mild**				
	Moderate	.022	.418	.198	.884
	Severe	.000	.280	.148	.530
Pupillary reactivity	**Normal**				
	Abnormal-both reactive	.002	.344	.178	.666
	Only one reactive	.000	.189	.074	.481
	None reactive	.000	.044	.021	.093
Injury cause	**RTC**				
	Fall	.104	.599	.323	1.110
	Assaults	.109	2.040	.853	4.882
	Others	.846	1.149	.283	4.672
Extracranial injury		.001	.325	.167	.631
Brain stem injury		.000	.150	.067	.337
swelling		.000	.383	.225	.652
Hypoxia	**Yes**	.001	.244	.110	.541
	Missing	.471	.729	.309	1.719
Mean blood pressure	**Normotension**				
	Hypotension	.005	.150	.040	.571
	Hypertension	.947	1.029	.437	2.425
	Missing	.134	.320	.072	1.419
Constant		.000	459.735		

Model XIVB3 (AUC=0.932)

			Odds ratio	*95. 0% C. I. for odds ratio*	
Age		.000	.956	.943	.968
GCS	**Mild**				
	Moderate	.022	.417	.197	.882
	Severe	.000	.280	.148	.529
Pupillary reactivity	**Normal**				
	Abnormal-both reactive	.002	.344	.178	.667
	Only one reactive	.000	.189	.074	.480
	None reactive	.000	.044	.021	.093
Injury cause	**RTC**				
	Fall	.102	.598	.323	1.108
	Assaults	.108	2.043	.855	4.885
	Others	.848	1.146	.283	4.651
Extracranial injury		.001	.324	.168	.624
Brain stem injury		.000	.150	.067	.336
swelling		.000	.384	.226	.652
Hypoxia	**Yes**	.000	.243	.110	.539
	Missing	.469	.729	.309	1.717
Mean blood pressure	**No-hypotension**				
	Hypotension	.001	.147	.048	.454
	missing	.083	.312	.084	1.165
Constant			470.994		

Model XIVB4 (AUC=0.930)

		Sig.	*Odds ratio*	*95. 0% C. I. for odds ratio*	
Age		.000	.954	.941	.966
GCS	**Mild**				
	Moderate	.016	.405	.194	.843
	Severe	.000	.242	.128	.456
Pupillary reactivity	**Normal**				
	Abnormal-both reactive	.004	.372	.191	.723
	Only one reactive	.001	.213	.084	.541
	None reactive	.000	.042	.020	.089
Injury cause	**RTC**				
	Fall	.111	.605	.327	1.121
	Assaults	.132	1.910	.822	4.437
	Others	.875	1.117	.282	4.421
extracranial_AIS4		.000	.311	.162	.594
Brain stem injury		.000	.168	.076	.373
swelling		.001	.401	.238	.676
Hypoxia	**Yes**	.000	.240	.108	.534
	Missing	.328	.665	.294	1.505
Systolic blood pressure	**Normotension**	.190			
	Hypotension		.589	.313	1.107
	Hypertension	.658	1.143	.632	2.068
	Missing	.416	.507	.099	2.600
Constant		.000	517.771		

Model XIVB5 (AUC=0.929)

		Sig.	*Odds ratio*	*95. 0% C. I. for odds ratio*	
Age		.000	.954	.942	.967
GCS	**Mild**				
	Moderate	.016	.405	.194	.843
	Severe	.000	.247	.132	.463
Pupillary reactivity	**Normal**				
	Abnormal-both reactive	.003	.369	.190	.715
	Only one reactive	.001	.214	.084	.542
	None reactive	.000	.043	.020	.090
Injury cause	**RTC**				
	Fall	.116	.611	.330	1.129
	Assaults	.135	1.902	.820	4.412
	Others	.896	1.096	.278	4.318
Extracranial injury		.000	.310	.162	.592
Brain stem injury		.000	.169	.076	.373
swelling		.001	.404	.240	.681
Hypoxia	**Yes**	.000	.238	.107	.528
	Missing	.329	.666	.294	1.507
Systolic blood pressure	**No-hypotension**				
	Hypotension	.039	.552	.314	.970
	missing	.363	.475	.095	2.362
Constant		.000	529.858		

Model XVA1 (AUC=0.925)

		Sig.	*Odds ratio*	*95. 0% C. I. for odds ratio*	
Age		.000	.951	.940	.962
GCS	**mild**				
	moderate	.027	.441	.214	.909
	severe	.000	.242	.130	.449
Pupillary reactivity	**Normal**	.000			
	Abnormal-both reactive	.002	.359	.186	.694
	Only one reactive	.003	.253	.102	.627
	None reactive	.000	.055	.027	.113
ISS		.000	.266	.144	.493
Brain stem injury		.000	.183	.083	.403
Hypoxia	**Yes**	.001	.230	.094	.566
	Missing	.418	1.803	.433	7.509
Systolic blood pressure	**Normotension**				
	Hypotension	.100	.599	.325	1.103
	Missing	.699	.621	.055	6.951
Interaction of hypoxia and systolic blood pressure					
	Hypoxia with hypotension	.533	1.659	.337	8.155
	Hypoxia with missing systolic blood pressure	1.000	.000	.000	.
	Missing hypoxia with normotension	.042	.144	.022	.936
	Missing hypoxia with missing systolic blood pressure	.393	.241	.009	6.331
Constant		.000	278.073		

Model XVB1 (AUC = 0.932)

		Sig.	*Odds ratio*	*95. 0% C. I. for odds ratio*	
Age		.000	.962	.944	.981
GCS	**mild**	.000			
	moderate	.013	.382	.178	.820
	severe	.000	.236	.123	.456
Pupillary reactivity	**Normal**	.000			
	Abnormal-both reactive	.002	.341	.172	.674
	Only one reactive	.001	.192	.074	.496
	None reactive	.000	.043	.020	.090
Cause of injury	**RTC**	.425			
	fall	.263	2.245	.544	9.269
	Assaults	.334	.380	.053	2.705
	others	.844	1.478	.030	73.101
Extracranial injury		.000	.295	.153	.570
Brain stem injury		.000	.161	.071	.366
Brain swelling		.001	.400	.234	.684
Hypoxia	**Yes**	.000	.182	.071	.463
	Missing	.465	1.664	.425	6.519
Systolic blood pressure	**Normotension**				
	Hypotension	.077	.561	.295	1.066
	Missing	.969	.952	.077	11.740
Interaction of age and injury cause	**Age and RTC**				
	Age and fall	.054	.976	.952	1.000
	Age and assault	.105	1.043	.991	1.097
	Age and others	.801	.989	.911	1.075
Interaction of hypoxia and systolic blood pressure					
	Hypoxia with hypotension	.243	2.753	.502	15.100
	Hypoxia with missing systolic blood pressure	1.000	.000	.000	.
	Missing hypoxia with normotension	.108	.219	.035	1.392
	Missing hypoxia with missing systolic blood pressure	.296	.162	.005	4.920
Constant		.000	430.839		

Model XVB2 (AUC= 0.931)

		Sig.	Odds ratio	95.0% C.I. for odds ratio	
Age		.000	.963	.945	.981
GCS	**mild**				
	moderate	.010	.370	.174	.788
	severe	.000	.241	.126	.462
Pupillary reactivity	**Normal**				
	Abnormal-both reactive	.003	.362	.184	.712
	Only one reactive	.001	.214	.084	.545
	None reactive	.000	.044	.021	.092
Cause of injury	**RTC**				
	fall	.245	2.306	.564	9.428
	Assaults	.308	.359	.050	2.569
	others	.708	2.140	.040	114.995
Extracranial injury		.000	.283	.147	.546
Brain stem injury		.000	.157	.070	.353
Brain swelling		.001	.409	.241	.696
Hypoxia	**Yes**	.000	.241	.108	.536
	Missing	.356	.682	.302	1.538
Systolic blood pressure	**Normotension**				
	Hypotension	.040	.548	.309	.972
	Missing	.391	.493	.098	2.481
Interaction of age and injury cause	**Age and RTC**				
	Age and fall	.049	.976	.952	1.000
	Age and assault	.085	1.046	.994	1.102
	Age and others	.692	.983	.904	1.069
Constant		.000	399.736		

References

1. Hernandez TD: **Post-traumatic neural depression and neurobehavioral recovery after brain injury.** *J Neurotrauma* 2006, **23:**1211-1221.
2. Langlois JA, Marr A, Mitchko J, Johnson RL: **Tracking the silent epidemic and educating the public: CDC's traumatic brain injury-associated activities under the TBI Act of 1996 and the Children's Health Act of 2000.** *J Head Trauma Rehabil* 2005, **20:**196-204.
3. Pascrell B, Jr.: **Traumatic brain injury. The silent epidemic.** *N J Med* 2001, **98:**47-48.
4. Sahuquillo J, Poca MA, Amoros S: **Current aspects of pathophysiology and cell dysfunction after severe head injury.** *Curr Pharm Des* 2001, **7:**1475-1503.
5. Stuart M, Zafonte R: **Fighting the silent epidemic: the Florida Brain and Spinal Cord Injury Program.** *J Head Trauma Rehabil* 2004, **19:**329-340.
6. Vigue B: **[Update on medical management of severe head trauma].** *Neurochirurgie* 2003, **49:**583-594.
7. Zitnay GA: **Lessons from national and international TBI societies and funds like NBIRTT.** *Acta Neurochir Suppl* 2005, **93:**131-133.
8. **Medical Subject Heading** [http://www.ncbi.nlm.nih.gov/sites/entrez?Db=mesh&Cmd=ShowDetailView&TermToSearch=68014947&ordinalpos=1&itool=EntrezSystem2.PEntrez.Mesh.Mesh_ResultsPanel.Mesh_RVDocSum]
9. *Standards for Surveillance of Neurotrauma.* Geneva: World Health Organisation; 1995.
10. *Head Injury: triage, assessment, investigation and early mangement of head injury in infants, children and adults.* National Institute for Clinical Excellence : National Collaborating Centre for Acute Care; 2003.
11. Thurman DJ, Alverson C, Dunn KA, Guerrero J, Sniezek JE: **Traumatic brain injury in the United States: A public health perspective.** *J Head Trauma Rehabil* 1999, **14:**602-615.
12. Tagliaferri F, Compagnone C, Korsic M, Servadei F, Kraus J: **A systematic review of brain injury epidemiology in Europe.** *Acta Neurochir (Wien)* 2006, **148:**255-268; discussion 268.
13. Tennant A: **Admission to hospital following head injury in England: incidence and socio-economic associations.** *BMC Public Health* 2005, **5:**21.
14. Borg J, Holm L, Peloso PM, Cassidy JD, Carroll LJ, von Holst H, Paniak C, Yates D: **Non-surgical intervention and cost for mild traumatic brain injury: results of the WHO Collaborating Centre Task Force on Mild Traumatic Brain Injury.** *J Rehabil Med* 2004**:**76-83.

15. Finkelstein EA, Corso, P.S., Miller,T.R.: *Incidence and Economic Burden of Injuries in the United States.* Oxford: Oxford Universit Press 2006.
16. Bastida JL, Aguilar PS, Gonzalez BD: **The economic costs of traffic accidents in Spain.** *J Trauma* 2004, **56:**883-888; discussion 888-889.
17. Cassidy JD, Carroll LJ, Peloso PM, Borg J, von Holst H, Holm L, Kraus J, Coronado VG: **Incidence, risk factors and prevention of mild traumatic brain injury: results of the WHO Collaborating Centre Task Force on Mild Traumatic Brain Injury.** *J Rehabil Med* 2004**:**28-60.
18. Wrona RM: **The use of state workers' compensation administrative data to identify injury scenarios and quantify costs of work-related traumatic brain injuries.** *J Safety Res* 2006, **37:**75-81.
19. Wood RL, Liossi C, Wood L: **The impact of head injury neurobehavioural sequelae on personal relationships: preliminary findings.** *Brain Inj* 2005, **19:**845-851.
20. Hawley CA, Ward AB, Magnay AR, Long J: **Parental stress and burden following traumatic brain injury amongst children and adolescents.** *Brain Inj* 2003, **17:**1-23.
21. Montgomery V, Oliver R, Reisner A, Fallat ME: **The effect of severe traumatic brain injury on the family.** *J Trauma* 2002, **52:**1121-1124.
22. Healey C, Osler TM, Rogers FB, Healey MA, Glance LG, Kilgo PD, Shackford SR, Meredith JW: **Improving the Glasgow Coma Scale score: motor score alone is a better predictor.** *J Trauma* 2003, **54:**671-678; discussion 678-680.
23. 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.
24. 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:**e165; discussion e165.
25. Jennett B, Bond M: **Assessment of outcome after severe brain damage.** *Lancet* 1975, **1:**480-484.
26. 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:**285-293.
27. 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.
28. 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:**232-238.
29. **International Mission for Prognosis and Analysis of Clinical Trials in TBI** [http://tbi-impact.org/]

30. 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:**239-250.
31. 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.
32. 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:**259-269.
33. Edwards P, Arango M, Balica L, Cottingham R, El-Sayed H, Farrell B, Fernandes J, Gogichaisvili T, Golden N, Hartzenberg B, 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.
34. **Head Injury Prognosis: CRASH** [http://www.crash2.lshtm.ac.uk/Risk%20calculator/index.html]
35. 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.
36. Hukkelhoven CW, Steyerberg EW, Habbema JD, Farace E, Marmarou A, Murray GD, Marshall LF, Maas AI: **Predicting outcome after traumatic brain injury: development and validation of a prognostic score based on admission characteristics.** *J Neurotrauma* 2005, **22:**1025-1039.
37. Niedeggen A, Adelt D, Berndt R, Hopf T: **Creatine-kinase-BB after severe head-injury as an index of prognosis in relation to nature of trauma and patients age.** *Acta Neurochir (Wien)* 1989, **101:**117-120.
38. Mao Q, Chen J, Li N, Li C, Mao B, Wang R, Wu G: **[The value of serum myelin basic protein in assessment of severity of acute closed head trauma].** *Hua Xi Yi Ke Da Xue Xue Bao* 1995, **26:**135-137.
39. Yamazaki Y, Yada K, Morii S, Kitahara T, Ohwada T: **Diagnostic significance of serum neuron-specific enolase and myelin basic protein assay in patients with acute head injury.** *Surg Neurol* 1995, **43:**267-270; discussion 270-261.
40. Gopcevic A, Mazul-Sunko B, Marout J, Sekulic A, Antoljak N, Siranovic M, Ivanec Z, Margaritoni M, Bekavac-Beslin M, Zarkovic N: **Plasma interleukin-8 as a potential predictor of mortality in adult patients with severe traumatic brain injury.** *Tohoku J Exp Med* 2007, **211:**387-393.
41. Minambres E, Cemborain A, Sanchez-Velasco P, Gandarillas M, Diaz-Reganon G, Sanchez-Gonzalez U, Leyva-Cobian F: **Correlation between transcranial interleukin-6 gradient and outcome in patients with acute brain injury.** *Crit Care Med* 2003, **31:**933-938.
42. Nylen K, Ost M, Csajbok LZ, Nilsson I, Blennow K, Nellgard B, Rosengren L: **Increased serum-GFAP in patients with severe traumatic brain injury is related to outcome.** *J Neurol Sci* 2006, **240:**85-91.

43. Vos PE, Lamers KJ, Hendriks JC, van Haaren M, Beems T, Zimmerman C, van Geel W, de Reus H, Biert J, Verbeek MM: **Glial and neuronal proteins in serum predict outcome after severe traumatic brain injury.** *Neurology* 2004, **62:**1303-1310.
44. Pelinka LE, Kroepfl A, Leixnering M, Buchinger W, Raabe A, Redl H: **GFAP versus S100B in serum after traumatic brain injury: relationship to brain damage and outcome.** *J Neurotrauma* 2004, **21:**1553-1561.
45. [http://www.ncbi.nlm.nih.gov/mesh]
46. Donato R: **S-100 proteins.** *Cell Calcium* 1986, **7:**123-145.
47. Donato R: **S100: a multigenic family of calcium-modulated proteins of the EF-hand type with intracellular and extracellular functional roles.** *Int J Biochem Cell Biol* 2001, **33:**637-668.
48. Torabian S, Kashani-Sabet M: **Biomarkers for melanoma.** *Curr Opin Oncol* 2005, **17:**167-171.
49. Rothermundt M, Peters M, Prehn JH, Arolt V: **S100B in brain damage and neurodegeneration.** *Microsc Res Tech* 2003, **60:**614-632.
50. Santamaria-Kisiel L, Rintala-Dempsey AC, Shaw GS: **Calcium-dependent and -independent interactions of the S100 protein family.** *Biochem J* 2006, **396:**201-214.
51. Gazzolo D, Michetti F, Bruschettini M, Marchese N, Lituania M, Mangraviti S, Pedrazzi E, Bruschettini P: **Pediatric concentrations of S100B protein in blood: age- and sex-related changes.** *Clin Chem* 2003, **49:**967-970.
52. Portela LV, Tort AB, Schaf DV, Ribeiro L, Nora DB, Walz R, Rotta LN, Silva CT, Busnello JV, Kapczinski F, et al: **The serum S100B concentration is age dependent.** *Clin Chem* 2002, **48:**950-952.
53. Ben Abdesselam O, Vally J, Adem C, Foglietti MJ, Beaudeux JL: **Reference values for serum S-100B protein depend on the race of individuals.** *Clin Chem* 2003, **49:**836-837.
54. Tateishi N, Shimoda T, Yada N, Shinagawa R, Kagamiishi Y: **[S100B: astrocyte specific protein].** *Nihon Shinkei Seishin Yakurigaku Zasshi* 2006, **26:**11-16.
55. Harpio R, Einarsson R: **S100 proteins as cancer biomarkers with focus on S100B in malignant melanoma.** *Clin Biochem* 2004, **37:**512-518.
56. Otto M, Holthusen S, Bahn E, Sohnchen N, Wiltfang J, Geese R, Fischer A, Reimers CD: **Boxing and running lead to a rise in serum levels of S-100B protein.** *Int J Sports Med* 2000, **21:**551-555.
57. Stalnacke BM, Ohlsson A, Tegner Y, Sojka P: **Serum concentrations of two biochemical markers of brain tissue damage S-100B and neurone specific enolase are increased in elite female soccer players after a competitive game.** *Br J Sports Med* 2006, **40:**313-316.
58. Stalnacke BM, Tegner Y, Sojka P: **Playing ice hockey and basketball increases serum levels of S-100B in elite players: a pilot study.** *Clin J Sport Med* 2003, **13:**292-302.
59. Stalnacke BM, Tegner Y, Sojka P: **Playing soccer increases serum concentrations of the biochemical markers of brain damage S-100B and neuron-specific enolase in elite players: a pilot study.** *Brain Inj* 2004, **18:**899-909.

60. Kleindienst A, McGinn MJ, Harvey HB, Colello RJ, Hamm RJ, Bullock MR: **Enhanced hippocampal neurogenesis by intraventricular S100B infusion is associated with improved cognitive recovery after traumatic brain injury.** *J Neurotrauma* 2005, **22:**645-655.
61. Kleindienst A, Harvey HB, Rice AC, Muller C, Hamm RJ, Gaab MR, Bullock MR: **Intraventricular infusion of the neurotrophic protein S100B improves cognitive recovery after fluid percussion injury in the rat.** *J Neurotrauma* 2004, **21:**541-547.
62. Li DR, Zhu BL, Ishikawa T, Zhao D, Michiue T, Maeda H: **Immunohistochemical distribution of S-100 protein in the cerebral cortex with regard to the cause of death in forensic autopsy.** *Leg Med (Tokyo)* 2006, **8:**78-85.
63. Li DR, Zhu BL, Ishikawa T, Zhao D, Michiue T, Maeda H: **Postmortem serum protein S100B levels with regard to the cause of death involving brain damage in medicolegal autopsy cases.** *Leg Med (Tokyo)* 2006, **8:**71-77.
64. Townend W, Dibble C, Abid K, Vail A, Sherwood R, Lecky F: **Rapid elimination of protein S-100B from serum after minor head trauma.** *J Neurotrauma* 2006, **23:**149-155.
65. Raabe A, Seifert V: **Protein S-100B as a serum marker of brain damage in severe head injury: preliminary results.** *Neurosurg Rev* 2000, **23:**136-138.
66. Savola O, Pyhtinen J, Leino TK, Siitonen S, Niemela O, Hillbom M: **Effects of head and extracranial injuries on serum protein S100B levels in trauma patients.** *J Trauma* 2004, **56:**1229-1234; discussion 1234.
67. Bazarian JJ, Zemlan FP, Mookerjee S, Stigbrand T: **Serum S-100B and cleaved-tau are poor predictors of long-term outcome after mild traumatic brain injury.** *Brain Inj* 2006, **20:**759-765.
68. Mussack T, Biberthaler P, Kanz KG, Heckl U, Gruber R, Linsenmaier U, Mutschler W, Jochum M: **Immediate S-100B and neuron-specific enolase plasma measurements for rapid evaluation of primary brain damage in alcohol-intoxicated, minor head-injured patients.** *Shock* 2002, **18:**395-400.
69. Savola O, Hillbom M: **Early predictors of post-concussion symptoms in patients with mild head injury.** *Eur J Neurol* 2003, **10:**175-181.
70. Morochovic R, Racz O, Kitka M, Pingorova S, Cibur P, Tomkova D, Lenartova R: **Serum S100B protein in early management of patients after mild traumatic brain injury.** *Eur J Neurol* 2009, **16:**1112-1117.
71. Biberthaler P, Linsenmeier U, Pfeifer KJ, Kroetz M, Mussack T, Kanz KG, Hoecherl EF, Jonas F, Marzi I, Leucht P, et al: **Serum S-100B concentration provides additional information fot the indication of computed tomography in patients after minor head injury: a prospective multicenter study.** *Shock* 2006, **25:**446-453.
72. Poli-de-Figueiredo LF, Biberthaler P, Simao Filho C, Hauser C, Mutschler W, Jochum M: **Measurement of S-100B for risk classification of victims sustaining minor head injury--first pilot study in Brazil.** *Clinics (Sao Paulo)* 2006, **61:**41-46.

73. Biberthaler P, Mussack T, Wiedemann E, Kanz KG, Koelsch M, Gippner-Steppert C, Jochum M: **Evaluation of S-100b as a specific marker for neuronal damage due to minor head trauma.** *World J Surg* 2001, **25:**93-97.
74. Muller K, Townend W, Biasca N, Unden J, Waterloo K, Romner B, Ingebrigtsen T: **S100B serum level predicts computed tomography findings after minor head injury.** *J Trauma* 2007, **62:**1452-1456.
75. Townend WJ, Guy MJ, Pani MA, Martin B, Yates DW: **Head injury outcome prediction in the emergency department: a role for protein S-100B?** *J Neurol Neurosurg Psychiatry* 2002, **73:**542-546.
76. Pelinka LE, Toegel E, Mauritz W, Redl H: **Serum S 100 B: a marker of brain damage in traumatic brain injury with and without multiple trauma.** *Shock* 2003, **19:**195-200.
77. Olivecrona M, Rodling-Wahlstrom M, Naredi S, Koskinen LO: **S-100B and neuron specific enolase are poor outcome predictors in severe traumatic brain injury treated by an intracranial pressure targeted therapy.** *J Neurol Neurosurg Psychiatry* 2009, **80:**1241-1247.
78. Mussack T, Biberthaler P, Kanz KG, Wiedemann E, Gippner-Steppert C, Mutschler W, Jochum M: **Serum S-100B and interleukin-8 as predictive markers for comparative neurologic outcome analysis of patients after cardiac arrest and severe traumatic brain injury.** *Crit Care Med* 2002, **30:**2669-2674.
79. Nylen K, Ost M, Csajbok LZ, Nilsson I, Hall C, Blennow K, Nellgard B, Rosengren L: **Serum levels of S100B, S100A1B and S100BB are all related to outcome after severe traumatic brain injury.** *Acta Neurochir (Wien)* 2008, **150:**221-227; discussion 227.
80. Herrmann M, Curio N, Jost S, Grubich C, Ebert AD, Fork ML, Synowitz H: **Release of biochemical markers of damage to neuronal and glial brain tissue is associated with short and long term neuropsychological outcome after traumatic brain injury.** *J Neurol Neurosurg Psychiatry* 2001, **70:**95-100.
81. Honda M, Tsuruta R, Kaneko T, Kasaoka S, Yagi T, Todani M, Fujita M, Izumi T, Maekawa T: **Serum Glial Fibrillary Acidic Protein Is a Highly Specific Biomarker for Traumatic Brain Injury in Humans Compared With S-100B and Neuron-Specific Enolase.** *J Trauma*.
82. Bechtel K, Frasure S, Marshall C, Dziura J, Simpson C: **Relationship of serum S100B levels and intracranial injury in children with closed head trauma.** *Pediatrics* 2009, **124:**e697-704.
83. Spinella PC, Dominguez T, Drott HR, Huh J, McCormick L, Rajendra A, Argon J, McIntosh T, Helfaer M: **S-100beta protein-serum levels in healthy children and its association with outcome in pediatric traumatic brain injury.** *Crit Care Med* 2003, **31:**939-945.
84. Unden J, Bellner J, Eneroth M, Alling C, Ingebrigtsen T, Romner B: **Raised serum S100B levels after acute bone fractures without cerebral injury.** *J Trauma* 2005, **58:**59-61.
85. Anderson RE, Hansson LO, Nilsson O, Dijlai-Merzoug R, Settergren G: **High serum S100B levels for trauma patients without head injuries.** *Neurosurgery* 2001, **48:**1255-1258; discussion 1258-1260.
86. Murray GD, Teasdale GM, Braakman R, Cohadon F, Dearden M, Iannotti F, Karimi A, Lapierre F, Maas A, Ohman J, et al: **The**

European Brain Injury Consortium survey of head injuries. *Acta Neurochir (Wien)* 1999, **141:**223-236.

87. Heizmann CW: **S100B protein in clinical diagnostics: assay specificity.** *Clin Chem* 2004, **50:**249-251.
88. Lingsma HF, Roozenbeek B, Steyerberg EW, Murray GD, Maas AI: **Early prognosis in traumatic brain injury: from prophecies to predictions.** *Lancet Neurol*, **9:**543-554.
89. **The Trauma Audit and Research Network** [https://www.tarn.ac.uk/]
90. *The Abbreviated Injury Scale, 1990 Revision, Update 98* Barrington, USA: Association for the Advancement of Automatic Medicine; 2001.
91. *Abbreviated Injury Scale 2005, Update 2008.* Barrington, USA: Association for Advancement of Automatic Medicine 2008.
92. 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:**329-337.
93. 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.
94. 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:**303-314.
95. 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:**294-302.
96. McHugh GS, Engel DC, Butcher I, Steyerberg EW, Lu J, Mushkudiani N, Hernandez 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:**287-293.
97. 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:**315-328.
98. 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 (24 h) time-point.** *Resuscitation* 2009, **80:**341-345.
99. Lesko M, Woodford M, White L, O'Brien S, Childs C, Lecky F: **Using Abbreviated Injury Scale (AIS) codes to classify Computed Tomography (CT) features in the Marshall System %U http://www.biomedcentral.com/1471-2288/10/72.** *BMC Medical Research Methodology*, **10:**72.
100. Wells PS, Anderson DR, Rodger M, Ginsberg JS, Kearon C, Gent M, Turpie AG, Bormanis J, Weitz J, Chamberlain M, et al: **Derivation of a simple clinical model to categorize patients probability of**

pulmonary embolism: increasing the models utility with the SimpliRED D-dimer. *Thromb Haemost* 2000, **83:**416-420.

101. Sinclair DR, Chung F, Mezei G: **Can postoperative nausea and vomiting be predicted?** *Anesthesiology* 1999, **91:**109-118.
102. Inouye M: **Predicting Outcomes of Patients in Japan After First Acute Stroke Using a Simple Model.** *American Journal of Physical Medicine & Rehabilitation* 2001, **80:**645-649.
103. Batsis JA, Huddleston JM, Melton LJt, Huddleston PM, Lopez-Jimenez F, Larson DR, Gullerud RE, McMahon MM: **Body mass index and risk of adverse cardiac events in elderly patients with hip fracture: a population-based study.** *J Am Geriatr Soc* 2009, **57:**419-426.
104. Kheirandish P, Seyedalinaghi S, Jahani M, Shirzad H, Seyed Ahmadian M, Majidi A, Sharifi A, Hosseini M, Mohraz M, McFarland W: **Prevalence and Correlates of Hepatitis C Infection among Male Injection Drug Users in Detention, Tehran, Iran.** *J Urban Health* 2009.
105. Biccard Bm Fau - Pooran RR, Pooran RR: **Validation of a model to predict all-cause in-hospital mortality in vascular surgical patients.**
106. Morrow DA, Antman EM, Charlesworth A, Cairns R, Murphy SA, de Lemos JA, Giugliano RP, McCabe CH, Braunwald E: **TIMI Risk Score for ST-Elevation Myocardial Infarction: A Convenient, Bedside, Clinical Score for Risk Assessment at Presentation : An Intravenous nPA for Treatment of Infarcting Myocardium Early II Trial Substudy.** *Circulation* 2000, **102:**2031-2037.
107. Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, Sirio CA, Murphy DJ, Lotring T, Damiano A, Harrell FE: **THE APACHE-III PROGNOSTIC SYSTEM - RISK PREDICTION OF HOSPITAL MORTALITY FOR CRITICALLY ILL HOSPITALIZED ADULTS.** *Chest* 1991, **100:**1619-1636.
108. Georg CH, Jean-Jacques P, Claudio J, Paul P, Alexandre de La T, Laurent S, Gregory V, Jacques T, Luca C, Vincenzo F, et al: **Patients with distant metastases from renal cell carcinoma can be accurately identified: external validation of a new nomogram.** *BJU International* 2008, **101:**39-43.
109. Chen G, Uryasev S, Young TK: **On prediction of the cesarean delivery risk in a large private practice.** *American Journal of Obstetrics and Gynecology* 2004, **191:**616-623.
110. Swets JA, Getty DJ, Pickett RM, D'Orsi CJ, Seltzer SE, McNeil BJ: **Enhancing and Evaluating Diagnostic Accuracy.** *Med Decis Making* 1991, **11:**9-17.
111. Ho Km Fau - Dobb GJ, Dobb Gj Fau - Knuiman M, Knuiman M Fau - Finn J, Finn J Fau - Lee KY, Lee Ky Fau - Webb SAR, Webb SA: **A comparison of admission and worst 24-hour Acute Physiology and Chronic Health Evaluation II scores in predicting hospital mortality: a retrospective cohort study.**
112. Altman DG, Royston P: **What do we mean by validating a prognostic model?** *Stat Med* 2000, **19:**453-473.
113. Justice AC, Covinsky KE, Berlin JA: **Assessing the generalizability of prognostic information.** *Ann Intern Med* 1999, **130:**515-524.

114. Royston P, Moons KGM, Altman DG, Vergouwe Y: **Prognosis and prognostic research: Developing a prognostic model.** *BMJ* 2009, **338:**b604-.
115. Altman DG, Vergouwe Y, Royston P, Moons KGM: **Prognosis and prognostic research: validating a prognostic model.** *BMJ* 2009, **338:**b605-.
116. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG: **Prognosis and prognostic research: what, why, and how?** *BMJ* 2009, **338:**b375-.
117. Teasdale G, Jennett B: **Assessment of coma and impaired consciousness. A practical scale.** *Lancet* 1974, **2:**81-84.
118. 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:**187-196.
119. Champion HR, Sacco WJ, Copes WS, Gann DS, Gennarelli TA, Flanagan ME: **A revision of the Trauma Score.** *J Trauma* 1989, **29:**623-629.
120. Boyd CR, Tolson MA, Copes WS: **Evaluating Trauma Care: The TRISS Method.** *The Journal of Trauma* 1987, **27:**370-378.
121. Lameire N, Hoste E: **Reflections on the definition, classification, and diagnostic evaluation of acute renal failure.** *Curr Opin Crit Care* 2004, **10:**468-475.
122. Babuin L, Jaffe AS: **Troponin: the biomarker of choice for the detection of cardiac injury.** *Cmaj* 2005, **173:**1191-1202.
123. Vissers RJ, Abu-Laban RB, McHugh DF: **Amylase and lipase in the emergency department evaluation of acute pancreatitis.** *J Emerg Med* 1999, **17:**1027-1037.
124. Townend W, Ingebrigtsen T: **Head injury outcome prediction: a role for protein S-100B?** *Injury* 2006, **37:**1098-1108.
125. Netto CB, Conte S, Leite MC, Pires C, Martins TL, Vidal P, Benfato MS, Giugliani R, Goncalves CA: **Serum S100B protein is increased in fasting rats.** *Arch Med Res* 2006, **37:**683-686.
126. Stalnacke BM, Bjornstig U, Karlsson K, Sojka P: **One-year follow-up of mild traumatic brain injury: post-concussion symptoms, disabilities and life satisfaction in relation to serum levels of S-100B and neurone-specific enolase in acute phase.** *J Rehabil Med* 2005, **37:**300-305.
127. **The Brain Trauma Foundation. The American Association of Neurological Surgeons. The Joint Section on Neurotrauma and Critical Care. Critical pathway for the treatment of established intracranial hypertension.** *J Neurotrauma* 2000, **17:**537-538.
128. Chatfield DA, Zemlan FP, Day DJ, Menon DK: **Discordant temporal patterns of S100beta and cleaved tau protein elevation after head injury: a pilot study.** *Br J Neurosurg* 2002, **16:**471-476.
129. Patel HC, Bouamra O, Woodford M, King AT, Yates DW, Lecky FE: **Trends in head injury outcome from 1989 to 2003 and the effect of neurosurgical care: an observational study.** *Lancet* 2005, **366:**1538-1544.

130. Rowley G, Fielding K: **Reliability and accuracy of the Glasgow Coma Scale with experienced and inexperienced users.** *Lancet* 1991, **337:**535-538.
131. Gill MR, Reiley DG, Green SM: **Interrater reliability of Glasgow Coma Scale scores in the emergency department.** *Ann Emerg Med* 2004, **43:**215-223.
132. Ross SE, Leipold C, Terregino C, O'Malley KF: **Efficacy of the motor component of the Glasgow Coma Scale in trauma triage.** *J Trauma* 1998, **45:**42-44.
133. Baxt WG, Jones G, Fortlage D: **The trauma triage rule: a new, resource-based approach to the prehospital identification of major trauma victims.** *Ann Emerg Med* 1990, **19:**1401-1406.
134. David W. Hosmer SL: *Applied logistic regression* Second edn: A Wiley-Interscience Publication 2000.
135. Royston P, Ambler G, Sauerbrei W: **The use of fractional polynomials to model continuous risk variables in epidemiology.** *Int J Epidemiol* 1999, **28:**964-974.
136. Nagelkerke NJD: **A note on a general definition of the coefficient of determination.** *Biometrika* 1991, **78:**691-692.
137. 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:**270-280.
138. Arbabi S, Jurkovich GJ, Wahl WL, Franklin GA, Hemmila MR, Taheri PA, Maier RV: **A comparison of prehospital and hospital data in trauma patients.** *J Trauma* 2004, **56:**1029-1032.
139. Marshall LF, Marshall SB, Klauber MR, Clark MvB, Eisenberg HM, Jane JA, Luerssen TG, Marmarou A, Foulkes MA: **A new classification of head injury based on computerized tomography.** *Special Supplements* 1991, **75:**S14-S20.
140. Maas AI, Hukkelhoven CW, Marshall LF, Steyerberg EW: **Prediction of outcome in traumatic brain injury with computed tomographic characteristics: a comparison between the computed tomographic classification and combinations of computed tomographic predictors.** *Neurosurgery* 2005, **57:**1173-1182; discussion 1173-1182.
141. Sacket DL, Straus, S.E., Richardson, W.S., Rosenberg, W., Haynes, R.B. : *Evidence-based Medicien: How to Practice and Teach EBM.* Second edn. Edinburgh: Churchill Livingstone 2000.
142. Bouamra O, Wrotchford A, Hollis S, Vail A, Woodford M, Lecky F: **A new approach to outcome prediction in trauma: A comparison with the TRISS model.** *J Trauma* 2006, **61:**701-710.
143. Servadei F, Murray GD, Penny K, Teasdale GM, Dearden M, Iannotti F, Lapierre F, Maas AJ, Karimi A, Ohman J, 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:**70-75; discussion 75-77.
144. Gennarelli TA, Wodzin E: **AIS 2005: a contemporary injury scale.** *Injury* 2006, **37:**1083-1091.

145. Vos PE, van Voskuilen AC, Beems T, Krabbe PF, Vogels OJ: **Evaluation of the traumatic coma data bank computed tomography classification for severe head injury.** *J Neurotrauma* 2001, **18:**649-655.
146. Walder AD, Yeoman PM, Turnbull A: **The abbreviated injury scale as a predictor of outcome of severe head injury.** *Intensive Care Med* 1995, **21:**606-609.
147. Chesnut RM: **Evolving models of neurotrauma critical care: an analysis and call to action.** *Clin Neurosurg* 2000, **46:**185-195.
148. Azian AA, Nurulazman AA, Shuaib L, Mahayidin M, Ariff AR, Naing NN, Abdullah J: **Computed tomography of the brain in predicting outcome of traumatic intracranial haemorrhage in Malaysian patients.** *Acta Neurochir (Wien)* 2001, **143:**711-720.
149. 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:**267-279.
150. Servadei F, Murray GD, Teasdale GM, Dearden M, Iannotti F, Lapierre F, Maas AJ, Karimi A, Ohman J, Persson L, et al: **Traumatic subarachnoid hemorrhage: demographic and clinical study of 750 patients from the European brain injury consortium survey of head injuries.** *Neurosurgery* 2002, **50:**261-267; discussion 267-269.
151. Wardlaw JM, Easton VJ, Statham P: **Which CT features help predict outcome after head injury?** *J Neurol Neurosurg Psychiatry* 2002, **72:**188-192; discussion 151.
152. Kakarieka A, Braakman R, Schakel EH: **Clinical significance of the finding of subarachnoid blood on CT scan after head injury.** *Acta Neurochir (Wien)* 1994, **129:**1-5.
153. Bahloul M, Chelly H, Ben Hmida M, Ben Hamida C, Ksibi H, Kallel H, Chaari A, Kassis M, Rekik N, Bouaziz M: **Prognosis of traumatic head injury in South Tunisia: a multivariate analysis of 437 cases.** *J Trauma* 2004, **57:**255-261.
154. Davis EG, MacKenzie EJ, Sacco WJ, Bain LW, Jr., Buckman RF, Jr., Champion HR, Lees PS: **A new "TRISS-like" probability of survival model for intubated trauma patients.** *J Trauma* 2003, **55:**53-61.
155. Langlois JA, Rutland-Brown W, Wald MM: **The epidemiology and impact of traumatic brain injury: a brief overview.** *J Head Trauma Rehabil* 2006, **21:**375-378.
156. Jagger J: **Prevention of brain trauma by legislation, regulation, and improved technology: a focus on motor vehicles.**
157. Machado Sg Fau - Murray GD, Murray Gd Fau - Teasdale GM, Teasdale GM: **Evaluation of designs for clinical trials of neuroprotective agents in head injury. European Brain Injury Consortium.**
158. Thornhill S, Teasdale GM, Murray GD, McEwen J, Roy CW, Penny KI: **Disability in young people and adults one year after head injury: prospective cohort study.** *Bmj* 2000, **320:**1631-1635.
159. 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:**331-343.

160. Butcher I, McHugh GS, Lu J, Steyerberg EW, Hernandez 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:**281-286.
161. Niedzwecki CM, Marwitz JH, Ketchum JM, Cifu DX, Dillard CM, Monasterio EA: **Traumatic brain injury: a comparison of inpatient functional outcomes between children and adults.** *J Head Trauma Rehabil* 2008, **23:**209-219.
162. Kaufmann CR, Maier RV, Kaufmann EJ, Rivara FP, Carrico CJ: **Validity of applying adult TRISS analysis to injured children.** *J Trauma* 1991, **31:**691-697; discussion 697-698.
163. **Fourmilab Switzerland** [http://www.fourmilab.ch/rpkp/experiments/analysis/chiCalc.html]
164. Qureshi AI: **Evaluation of brain injury using a blood test: is there a role in clinical practice?** *Crit Care Med* 2002, **30:**2778-2779.
165. Dimopoulou I, Korfias S, Dafni U, Anthi A, Psachoulia C, Jullien G, Sakas DE, Roussos C: **Protein S-100b serum levels in trauma-induced brain death.** *Neurology* 2003, **60:**947-951.
166. Ingebrigtsen T, Waterloo K, Jacobsen EA, Langbakk B, Romner B: **Traumatic brain damage in minor head injury: relation of serum S-100 protein measurements to magnetic resonance imaging and neurobehavioral outcome.** *Neurosurgery* 1999, **45:**468-475; discussion 475-466.
167. Woertgen C, Rothoerl RD, Metz C, Brawanski A: **Comparison of clinical, radiologic, and serum marker as prognostic factors after severe head injury.** *J Trauma* 1999, **47:**1126-1130.
168. Lesko MM, Bouamra O, Jenks T, Woodford M, Lecky F: **Models of mortality probability in severe traumatic brain injury.** *Injury Extra* 2009, **40:**212-212.
169. **General Pracice Notebook-a UK medical reference on the world wide web. Anemia.** [http://gpnotebook.com/simplepage.cfm?ID=1436155983]
170. **General Practice Notebook- A UK medical reference on the word wide web. Hyperglycemia.** [http://www.gpnotebook.co.uk/simplepage.cfm?ID=899285013]
171. **General Practice Notebook- A UK medical reference on the word wide web. Hypoglycemia .** [http://www.gpnotebook.co.uk/simplepage.cfm?ID=x20090122054239749131]
172. **General Practice Notebook- A UK medical reference on the word wide web. Thrombocytosis.** [http://www.gpnotebook.co.uk/simplepage.cfm?ID=999948289]
173. Piazza O, Storti MP, Cotena S, Stoppa F, Perrotta D, Esposito G, Pirozzi N, Tufano R: **S100B is not a reliable prognostic index in paediatric TBI.** *Pediatr Neurosurg* 2007, **43:**258-264.
174. Kattan MW: **Judging new markers by their ability to improve predictive accuracy.** *J Natl Cancer Inst* 2003, **95:**634-635.

175. Childs C, Vail A, Protheroe R, King AT, Dark PM: **Differences between brain and rectal temperatures during routine critical care of patients with severe traumatic brain injury.** *Anaesthesia* 2005, **60:**759-765.
176. Skaga NO, Eken T, Steen PA: **Assessing quality of care in a trauma referral center: benchmarking performance by TRISS-based statistics or by analysis of stratified ISS data?** *J Trauma* 2006, **60:**538-547.
177. **General Practice Notebook- A UK medical reference on the word wide web. Thrombocytopenia.** [http://gpnotebook.com/simplepage.cfm?ID=999948289]
178. McHugh GS, Butcher I, Steyerberg EW, Lu J, Mushkudiani N, Marmarou A, Maas AI, Murray GD: **Statistical approaches to the univariate prognostic analysis of the IMPACT database on traumatic brain injury.** *J Neurotrauma* 2007, **24:**251-258.
179. Geyer C, Ulrich A, Grafe G, Stach B, Till H: **Diagnostic value of S100B and neuron-specific enolase in mild pediatric traumatic brain injury.** *J Neurosurg Pediatr* 2009, **4:**339-344.
180. Hallen M, Karlsson M, Carlhed R, Hallgren T, Bergenheim M: **S-100B in serum and urine after traumatic head injury in children.** *J Trauma*, **69:**284-289.
181. Ytrebo LM, Nedredal GI, Korvald C, Holm Nielsen OJ, Ingebrigtsen T, Romner B, Aarbakke J, Revhaug A: **Renal elimination of protein S-100beta in pigs with acute encephalopathy.** *Scand J Clin Lab Invest* 2001, **61:**217-225.
182. Raabe A, Grolms C, Sorge O, Zimmermann M, Seifert V: **Serum S-100B protein in severe head injury.** *Neurosurgery* 1999, **45:**477-483.