CRITICAL CARE

COMMENTARY

Monitoring prognosis in severe traumatic brain injury

Andrew IR Maas[1*] and Ewout W Steyerberg[2]

See related research by Raj *et al.*, http://ccforum.com/content/18/2/R60

Abstract

The choice of disease-specific versus generic scales is common to many fields of medicine. In the area of traumatic brain injury, evidence is coming forward that disease-specific prognostic models and disease-specific scoring systems are preferable in the intensive care setting. In monitoring prognosis, the use of a calibration belt in validation studies potentially provides accurate and intuitively attractive insight into performance. This approach deserves further empirical evaluation of its added value as well as its limitations.

In the previous issue of *Critical Care*, Raj and colleagues [1] report a detailed study on the evaluation of commonly employed general ICU scales to predict outcome in patients with traumatic brain injury (TBI). They compare performance to that of simpler models based on only age and Glasgow Coma Scale (GCS). The authors conclude that the simple prognostic model based only on age and GCS showed fairly good prognostic performance and that the use of more complex general ICU scoring systems added little to this. This manuscript clearly demonstrates that TBI patients in the ICU environment are a highly specific population, in which general ICU scoring systems are of limited value. Second, from a methodological perspective, it presents and discusses essential approaches for quantifying the performance of prognostic models and provides empirical illustration of the use of a new instrument for assessing calibration.

Limited value of general ICU scoring systems in traumatic brain injury patients

Scoring systems such as the Acute Physiology and Chronic Health Evaluation II, Simplified Acute Physiology Score II and Sequential Organ Failure Assessment scores are commonly used in the intensive care setting to quantify the impact of disease severity and to benchmark the quality of delivered health care. These scoring systems are developed for use in the general ICU environment and are not disease specific. TBI is a very heterogeneous disease in terms of cause, pathology, severity and also in expected outcome. Disease-specific prognostic models for moderate and severe TBI have been developed and validated. These include the CRASH (Corticosteroid Randomisation After Significant Head Injury) [2] and the IMPACT (International Mission for Prognosis and Analysis of Clinical Trials in TBI) [3] prognostic models. Both the IMPACT and CRASH models have shown reasonable to good performance at external validation, both for mortality and 6-month Glasgow Outcome Scale. The latter is particularly important as the degree of functional recovery in the long term is perhaps even more relevant than early mortality in TBI patients. The IMPACT studies have shown that most prognostic information is contained within three variables: age, GCS motor score, and pupillary reactivity [4]. The findings on age and GCS in the present study are in line with this observation. Apparently the additional information from many parameters obtained at admission (as in the IMPACT model) or during the first 24 hours of care (as for ICU-specific models) adds little prognostic value compared to core information such as age and admission GCS. Further development and validation of TBI-specific prediction models is required, including disease-specific information that becomes available during the clinical course. The latter will require a dynamic prediction framework [5].

* Correspondence: andrew.maas@uza.be
[1]Department of Neurosurgery, Antwerp University Hospital and University of Antwerp, Wilrijkstraat 10, 2650 Edegem, Belgium
Full list of author information is available at the end of the article

BioMed Central

Measuring performance of prognostic models in the traumatic brain injury population

The performance of prognostic models is commonly evaluated by discrimination and calibration. Discrimination concerns the ability to distinguish between survival and death or favourable and unfavourable outcome. It is generally assessed by calculating the area under the receiver operating characteristic curve (AUC). Discrimination is influenced by both the validity of the model for a specific population (that is, the statistical fit of the model) and the case mix of the validation population [6,7]. If the population includes subsets with a more extreme prognosis (for example, mild versus severe TBI), the discriminative ability will be boosted upwards. For this reason, a case mix adjusted AUC has been proposed [6]. In the present study, a more homogeneous population may have lowered the discriminative ability of the ICU-specific models.

Calibration evaluates the agreement between observed and predicted outcome and can be graphically presented in plots. The often used Hosmer-Lemeshow test considers deciles of patients with similar risk, and reflects the average concordance of expected outcome compared to the observed outcome analyzed. Limitations of the Hosmer-Lemeshow test mentioned by the authors include that it may be non-informative in large data sets (that is, statistically significant for minor miscalibration), and that the division of the patient cohort into deciles does not account sufficiently for the individual patient.

As a relatively new instrument to assess calibration, the authors utilised a calibration belt. This approach was developed and tested within the GiViTI consortium in Italy (Italian Group for the evaluation of interventions in intensive care medicine) and was taken forward in a larger ICU network named Prosafe through EU funding (PHEA 2007 331). These studies are now coordinated in the CREACTIVE Project (prospective longitudinal data collection and comparative effectiveness research for TBI). Within this project, TBI-specific prognostic models will be developed to be used as a benchmark for quality of care assessment in individual ICUs. The calibration belt relates the observed and the expected probability of a dichotomised outcome. Importantly, the calibration belt calculates the 80% confidence interval and the 95% confidence interval surrounding the calibration curve. This instrument thus potentially provides accurate and intuitively attractive insight into calibration performance. As with any new instrument, however, its validity has to be demonstrated in broad settings and validated by other groups. Limitations may only become apparent with greater experience. It is not quite clear how dependent the calibration belt and, in particular, the calculated confidence interval may be upon the relative number of patients with specific prognostic risks.

Adding the distribution of patient numbers across the plotted curves would provide additional insight. Notwithstanding this potential limitation, this approach in which disease-specific aspects are combined in an intuitively attractive novel instrument is worthy of further exploration and validation.

Author details
[1]Department of Neurosurgery, Antwerp University Hospital and University of Antwerp, Wilrijkstraat 10, 2650 Edegem, Belgium. [2]Department of Public Health, Erasmus MC - University Medical Center Rotterdam, PO Box 2040, Rotterdam, CA 3000, Netherlands.

References
1. Raj R, Skrifvars MB, Bendel S, Selander T, Kivisaari R, Siironene J, Reinikainen M: **Predicting six-month mortality of patients with traumatic brain injury: usefulness of common intensive care severity scores.** *Crit Care* 2014, **18**:R60.
2. Trial Collaborators MRCCRASH, 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.
3. 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.
4. 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.
5. Hansen BE, Buster EH, Steyerberg EW, Lesaffre E, Janssen HL: **Prediction of the response to peg-interferon-alfa in patients with HBeAg positive chronic hepatitis B using decline of HBV DNA during treatment.** *J Med Virol* 2010, **82**:1135–1142.
6. Vergouwe Y, Moons KG, Steyerberg EW: **External validity of risk models: use of benchmark values to disentangle a case-mix effect from incorrect coefficients.** *Am J Epidemiol* 2010, **172**:971–980.
7. Roozenbeek B, Lingsma HF, Lecky FE, Lu J, Weir J, Butcher I, McHugh GS, Murray GD, Perel P, Maas AI, Steyerberg EW, International Mission on Prognosis Analysis of Clinical Trials in Traumatic Brain Injury (IMPACT) Study Group; Corticosteroid Randomisation After Significant Head Injury (CRASH) Trial Collaborators; Trauma Audit and Research Network (TARN): **Prediction of outcome after moderate and severe traumatic brain injury: external validation of the International Mission on Prognosis and Analysis of Clinical Trials (IMPACT) and Corticoid Randomisation After Significant Head injury (CRASH) prognostic models.** *Crit Care Med* 2012, **40**:1609–1617.