

Data-driven approaches to reveal the pathobiological heterogeneity in patients with traumatic brain injury

Cecilia Åkerlund¹, Ari Ercole²

¹Section of Perioperative Medicine and Intensive Care, Department of Physiology and Pharmacology, Karolinska Institute, Stockholm, Sweden

²Division of Anaesthesia and Cambridge Centre for AI in Medicine, Department of Medicine, University of Cambridge, Cambridge, United Kingdom

In traumatic brain injury (TBI), the inability of Glasgow Coma Scale (GCS) to capture the inherent heterogeneity of the disease may provide some explanation as to why randomised trials of biologically plausible therapies have largely failed. Such considerations provide strong motivation for the development of precision medicine approaches in this domain to improve outcome.³⁻⁵ Subgroups of patients with distinct pathophysiological or pathobiological mechanisms - so called *endotypes* - can be sought as a step to identifying individualised treatments. This can be done by data-driven approaches such as using unsupervised clustering algorithms.

Two of the most important contributions to identifying endotypes in the ICU population are by Calfee and Seymour. Calfee identified subgroups of patients with acute respiratory distress syndrome (ARDS) showing distinct inflammatory profiles by performing latent class analysis on patient data from two previous studies with negative outcome. These subgroups were found to respond differently to positive end expiratory pressure (PEEP).¹ Seymour did a similar analysis of sepsis patients using consensus *k* means clustering. He could identify distinct subgroups defined by inflammatory response which benefited from different fluid management strategies suggesting a substrate for individualised care.²

Current suggestions of TBI phenotypes

To date, more than twenty-five suggested endotypes and *phenotypes* (patients who share clinical traits irrespective of whether these relate to underlying mechanistic similarities) in TBI have been published,⁶ but none has been tested for treatment responses. Most of these focus on the milder spectrum of the

disease and post-concussion symptoms, with only a few including patients with severe TBI in the acute phase. These will be discussed in more detail below and are summarised in Table 1.

Folweiler *et al.*⁷ have elegantly identified three candidate endotypes across all severities in TBI. Phenotypes defined by the haematological and coagulation factors platelet count, haemoglobin, prothrombin time, INR, hematocrit, and glucose were identified by using partitioning around medoids on a dataset containing more than sixty baseline variables. When stratifying these patients by admission GCS, the clear separation pattern of patients disappeared, suggesting these factors are identifying important disease mechanisms not captured by GCS. Yuh *et al.*⁸ performed hierarchical clustering of computed tomography (CT) features in patients with mild TBI, identifying three clusters of intracranial lesions: Epidural hematoma (EDH); Subdural hematoma (SDH), contusion and subarachnoid haemorrhage (SAH); and intracranial haemorrhage (ICH) and petechial haemorrhage. As 36% of patients admitted to the ICU in an international multicenter study presented with mild TBI, these results may be relevant to the ICU cohort.⁹

Within the Collaborative European Neurotrauma Effectiveness Research in TBI (CENTER-TBI) ICU cohort, Åkerlund *et al.*¹⁰ identified six candidate endotypes based on admission physiology and biochemical markers, using a probabilistic clustering model. These endotypes were distinguished by GCS score and degree of metabolic derangement, including glucose, core temperature, pH, lactate, base excess, arterial partial pressure of carbon dioxide, oxygen saturation, and creatinine. Notably, two different pictures of metabolic derangement emerged, where one was characterised by a general stress response, while the other was associated with extracranial injuries.

By solely focusing on intracranial pressure (ICP) trajectories over time, Jha *et al.*¹¹ identified six phenotypic temporal profiles by applying the longitudinal clustering method group-based trajectory means (GBTM). Not only did the trajectories with high ICP show relations with unfavourable outcomes, but so did two trajectories with low ICP levels. Furthermore, the expression of the gene *ABCC8* (coding for the sulfonylurea receptor-1, a known edema regulator) was different between the identified groups.

The identified candidate endotypes of TBI all go beyond a description by GCS at presentation, indicating underlying important pathobiological mechanisms. Data-driven unsupervised clustering methods have been used to identify and describe the proposed endotypes, and although no information on outcome was used in the models, all described phenotypes were informative on outcome. However, they are not hypothesis-free but largely depending on the features included in the

models. In addition, different methods have been used, which further may explain the differences in presentation.

The proposed subclassifications of TBI patients are biologically plausible subtraits to individualise treatment. The intracranial lesion phenotypes and the haematological endotypes have been validated in external datasets showing good generalizability. Inclusion of brain biomarkers such as GFAP and UCH-L1 may further refine the suggested phenotypes to increase the probability of clinical relevance, as these features have shown to improve outcome predictions.¹²

Methodological considerations

The phenotypes described above have all been developed using clustering algorithms – grouping of patients with similar characteristics with no information on outcome in what are often complex, high-dimensional data. Clustering algorithms are methodologically challenging and identifying a principled number of robust clusters requires substantial effort to do well.

First, one has to decide which clustering algorithm should be used. Relatedly there are a variety of choices for what metric is to be used to define ‘similarity’ between two data points. Whilst to some extent the chosen method may be dictated by the data type, it is important to note that it is not given that two methods will produce the same result. Parameter selection, choice of an ‘optimal’ number of clusters and assuring that the clustering obtained is robust are steps which often receive relatively little attention; yet any inferences are critically dependent on these being done well and sensitivity analysis is vital. Demonstrating robustness to modelling assumptions and external validation are crucial steps. After external validation, restratification according to the proposed subgroups of study cohorts in previous interventional studies should be performed, to investigate treatment effects in the hypothesised groups.

Conclusions

Clustering has been shown to be a promising technique for discovering endotypes that are biologically both plausible and interesting. However clustering is methodologically challenging and both authors and peer reviewers should be cognizant of this. Whether data driven endotypes are clinically useful will depend on whether patients in different clusters respond differently to different treatment choices, as is the case for ARDS and sepsis. This requires prospective assessment (or retrospective evaluation in a prospectively collected interventional dataset) and this has not yet taken place. The

endotypes give interesting biological insights, but further studies are needed before we can know whether they can help to improve TBI outcomes.

Author	Number of clusters	Descriptives	Interpretation	Development cohort	Method
Folweiler ⁷	3	Platelet count, haemoglobin, prothrombin time, INR, haematocrit, and glucose	Two clusters, with mild anemia and relatively normal hematological values, respectively, had similar outcome profiles, while the cluster with thrombocytopenia, anemia, and coagulopathy was associated with worse outcome.	COBRIT study (n=1,213). Externally validated in TRACK-TBI Pilot study.	Partitioning around medoids
Yuh ⁸	3	CT characteristics: contusion, SAH and/or SDH; IVH and/or petechial haemorrhage; and EDH	The EDH cluster was associated with better outcome than the other two clusters.	TRACK-TBI, mild TBI (n=1,935). Externally validated in CENTER-TBI.	Hierarchical clustering
Jha ¹¹	6	ICP trajectories day 0-5	Low, slow rise; persistent low; intermittent spikes; frequent spikes; persistently high; severe intracranial hypertension	A single academic medical center, severe TBI (n=446).	Group-based trajectory model
Åkerlund ¹⁰	6	GCS, lactate, oxygen saturation, creatinine, glucose, base excess, pH,	GCS and normal or deranged metabolism.	CENTER-TBI ICU sub-cohort	Mixture of probabilistic graph models

		PaCO ₂ , and body temperature		(n=1,728)	
--	--	--	--	-----------	--

Table 1: Summary of currently proposed phenotypes in traumatic brain injury. EDH: epidural hematoma, GCS: Glasgow coma scale, ICP: intracranial pressure, IVH: intraventricular haemorrhage, PaCO₂: arterial partial pressure of carbon dioxide, SAH: subarachnoidal hematoma, SDH: subdural hematoma, TBI: traumatic brain injury.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

- 1 Calfee CS, Delucchi K, Parsons PE, *et al.* Latent Class Analysis of ARDS Subphenotypes: Analysis of Data From Two Randomized Controlled Trials. *The Lancet Respiratory Medicine* 2014; **2**: 611–20.
- 2 Seymour CW, Kennedy JN, Wang S, *et al.* Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. *JAMA - Journal of the American Medical Association* 2019; **321**: 2003–17.
- 3 Maas AIR, Menon DK, Adelson PD, *et al.* Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. *The Lancet Neurology* 2017; **16**: 987–1048.
- 4 Maas AIR, Menon DK, Manley GT, *et al.* Traumatic brain injury: progress and challenges in prevention, clinical care, and research. *The Lancet Neurology* 2022; : 1–57.
- 5 Tenovuo O, Diaz-Arrastia R, Goldstein LE, Sharp DJ, van der Naalt J, Zasler ND. Assessing the Severity of Traumatic Brain Injury—Time for a Change? *Journal of Clinical Medicine* 2021; **10**: 148.
- 6 Bragge P, Synnot A, Maas AI, *et al.* A State-of-the-Science Overview of Randomized Controlled Trials Evaluating Acute Management of Moderate-to-Severe Traumatic Brain Injury. *Journal of Neurotrauma* 2016; **33**: 1461–78.
- 7 Folweiler KA, Sandsmark DK, Diaz-Arrastia R, Cohen AS, Masino AJ. Unsupervised Machine Learning Reveals Novel Traumatic Brain Injury Patient Phenotypes with Distinct Acute Injury Profiles and Long-Term Outcomes. *Journal of Neurotrauma* 2020; **1444**: 1431–44.
- 8 Yuh EL, Jain S, Sun X, *et al.* Pathological Computed Tomography Features Associated With Adverse Outcomes After Mild Traumatic Brain Injury: A TRACK-TBI Study With External Validation in CENTER-TBI. *JAMA Neurology* 2021; **78**: 1137–48.
- 9 Steyerberg EW, Wiegers E, Sewalt C, *et al.* Case-mix, care pathways, and outcomes in patients with traumatic brain injury in CENTER-TBI: a European prospective, multicentre, longitudinal, cohort study. *The Lancet Neurology* 2019; **18**: 923–34.
- 10 Åkerlund CAI, Holst A, Stocchetti N, *et al.* Clustering identifies endotypes of traumatic brain injury in an intensive care cohort: a CENTER-TBI study. *Crit Care* 2022; **26**: 1–15.
- 11 Jha RM, Elmer J, Zusman BE, *et al.* Intracranial pressure trajectories: A novel approach to informing severe traumatic brain injury phenotypes. *Critical Care Medicine* 2018; **46**: 1792–802.
- 12 Helmrich IRAR, Czeiter E, Amrein K, *et al.* Incremental prognostic value of acute serum biomarkers for functional outcome after traumatic brain injury (CENTER-TBI): an observational cohort study. *The Lancet Neurology* 2022; **21**: 792–802.