Feasibility of Hidden Markov Models for the Description of Time-Varying Physiological State after Severe Traumatic Brain Injury

Shadnaz Asgari**, Ph.D.[1,2], Hadie Adams, M.D.[3], Magdalena Kasprowicz, Ph.D.[4], Marek Czosnyka, Ph.D.[3,5], Peter Smielewski*, Ph.D.[3] and, Ari Ercole*, M.B. B.Chir., Ph.D.[6]

[1] Department of Biomedical Engineering, California State University, Long Beach, California, USA
[2] Department of Computer Engineering and Computer Science, California State University, Long Beach, California, USA
[3] Division of Neurosurgery, Department of Clinical Neurosciences, Addenbrooke's Hospital, University of Cambridge, Cambridge, UK
[4] Department of Biomedical Engineering, Wroclaw University of Science and Technology, Wroclaw, Poland
[5] Institute of Electronic Systems, Warsaw University of Technology, Poland
[6] Division of Anesthesia, Department of Anesthesia, University of Cambridge, Cambridge, UK

*Joint senior authors

**Corresponding Author: Shadnaz Asgari, Ph.D.
1250 Bellflower Blvd.-MS 8302
Long Beach, CA 90840-8302
Office: (562) 985-5467
FAX: (562) 985-7823
EMAIL: Shadnaz.Asgari@csulb.edu

Address for reprints: Division of Anesthesia, Department of Anesthesia, University of Cambridge, Cambridge, UK. However, reprints will not be required.

Word count (from Abstract to the end of Conclusion): 2999

Abstract

Objective: Continuous assessment of physiology after traumatic brain injury (TBI) is essential to prevent secondary brain insults. The present work aims at the development of a method for detecting physiological states associated with the outcome from time-series physiological measurements using a Hidden Markov Model (HMM).

Design: Unsupervised clustering of hourly values of intracranial/cerebral perfusion pressure (ICP/CPP), the compensatory reserve index (RAP), and autoregulation status (PRx) was attempted using a HMM. A ternary state variable was learned to classify the patient's physiological state at any point in time into three categories ('good', 'intermediate' or 'poor) and determined the physiological parameters associated with each state.

Setting: The proposed HMM was trained and applied on a large dataset (28,939 hours of data) using a stratified 20-fold cross-validation.

Patients: The data was collected from 379 TBI patients admitted to Addenbrooke's Hospital, Cambridge between 2002 and 2016.

Intervention: Retrospective observational analysis.

Measurements and Main Results: Unsupervised training of the HMM yielded states characterized by ICP, CPP, RAP and PRx that were physiologically plausible. The resulting classifier retained a dose-dependent prognostic ability. Dynamic analysis suggested that the HMM was stable over short periods of time consistent with typical timescales for TBI pathogenesis.

Conclusions: To our knowledge, this is the first application of unsupervised learning to multidimensional time-series TBI physiology. We demonstrated that clustering using a HMM can reduce a complex set of physiological variables to a simple sequence of clinically plausible time-sensitive physiological states while retaining prognostic information in a dose-dependent manner. Such states may provide a more natural and parsimonious basis for triggering intervention decisions.

Introduction

Traumatic Brain Injury (TBI) is the leading cause of mortality, morbidity and disability in young adults worldwide (1). The critical care of severe TBI is predicated on the timely diagnosis of energetic crisis and optimization of physiological support to prevent secondary injury (2). TBI is a highly heterogeneous disease and comprehensive monitoring is needed to carefully characterize physiology to individualize therapy (3). Consequently, neurocritical care episodes are probably the most data-intensive in routine clinical practice and offer an excellent test-bed for developing and testing novel analytical machine learning (ML) algorithms. Whilst the physiome of critical care including neurocritical care is fundamentally time-series in description, the development of novel analytical techniques for this paradigm has seen little attention with many epidemiological and observational studies focusing instead on simple aggregate or fixed time-point data instead for analytical simplicity.

TBI biology is stereotyped but varied: Accurate temporal assessments of patient state would be helpful in guiding which patients can be managed with first-line, second-line or rescue therapies whose risk-benefit profiles vary (2, 3). Characterization of the ICU physiome is highly desirable as only then randomized trials can target populations of sufficient homogeneity to achieve adequate power. There is therefore a pressing need to develop methods that capture multivariable time-series data to find biologically meaningful clusters of patients to understand the temporal determinants of the underlying biology. Identification of meaningful patient states is also clinically valuable.

Institutional variation in TBI care is still considerable (5, 6), given the variety of the monitored parameters/indices, their complex interactions, and underlying heterogeneity of various injury patterns. A method for tracking patient state over time by integrating clinical parameters would flag up patients whose marginal physiological derangements in concert represent significant deviations from acceptable physiology. This requires a methodology for tracking the physiology

but also to learn the appropriate thresholds for the multidimensional states, for example with reference to population-level outcomes.

However, such multidimensional assessment is challenging: For example, patients with low cerebral perfusion pressure (CPP) or high intracranial pressure (ICP) are both clearly 'sick'. But how best to appreciate the state of a patient with marginal CPP and ICP together, or in whom these parameters are acceptable, but whose compensatory reserve or autoregulation is exhausted? Whilst physiological perturbations of such parameters are being linked to the outcome at a population level in a 'dose' dependent way, such post-hoc epidemiological assessments have little applicability to the stratification of patient care in real time. Admission demographics on the other hand explain only a small proportion of the variance in outcome (4) and therefore again offer the clinicians limited practical guidance for precision medicine.

The current work is the first attempt at developing a statistical fusion methodology to track physiological states. We apply a hidden Markov model (HMM) to analyze multiple time-series, i.e., ICP, CPP, the cerebrospinal compensatory reserve index (RAP), and autoregulation status (PRx), and detect/classify the hourly state of the patients in terms of the likely instantaneous outcome. Text 1 (Supplemental Digital Content 1) provides an author summary of the presented work.

Materials and Methods

Patients and data

Our dataset consisted of anonymized retrospective digital recordings of ICP and ABP along with 6-month Glasgow Outcome Scale (GOS) (5) from 527 TBI adult patients admitted to the Addenbrooke's Hospital, Cambridge, between 2002 and 2016. Under UK regulations, institutional approval is not required for research using anonymous routine clinical data.

Using the locally developed ICM+ software (http://icmplus.neurosurg.cam.ac.uk), the hourly-averaged values of ICP, CPP, PRx and RAP (6) were obtained. Text 2 (Supplemental Digital

Content 2) provides detailed information about the patient management protocol and exclusion criteria and calculation of RAP and PRx. In summary, 18 patients were excluded due to non-neurological related deaths related, and 17 vegetative patients were excluded due to their small number and greater heterogeneity of their data. Furthermore, 113 subjects were not considered because either their data monitoring initiated more than 3 days after TBI or had data discontinuity for more than 5 hours.

Statistical analysis

HMM was used to obtain the hourly states of TBI patients. Text 3 (Supplemental Digital Content 3) provides justification for such choice and an overview to HMM and its structure. Our HMM had three unobserved states ($N = 3$) and four observation variables ($M = 4$), i.e., hourly means of ICP, CPP, PRx and RAP. The ternary state (representing the cerebrovascular state of the injured brain at each hour) was labeled as 0 ('good'), 1 ('Intermediate'), and 2 ('poor'). The HMM was trained in an unsupervised fashion by stratified 20-fold cross-validation where the subjects with specific GOS outcomes were uniformly distributed among all folds. To enhance the accuracy of the proposed HMM in differentiating between 'good' and 'poor' states, at each fold we only considered data on those subjects in the training dataset who had the best and worst GOS outcomes (good outcome and death, respectively). During the testing phase, the Viterbi algorithm (7) was employed to compute the most probable sequence of states for each patient in the testing dataset (given the estimated model parameters and the sequences of hourly-averaged values of ICP, CPP, RAP and PRx for that patient).

Data Analysis and Validation Protocol

Parameter initialization is described in detail in Text 4 (Supplemental Digital Content 6). For our model to provide a potentially useful representation of the patient, we require the states that are determined by the algorithm to be biologically meaningful in some way. As we lack a ground-

truth for the hourly state of the TBI patients, we took some indirect measures to assess the model states' biological relevance:

1) If the detected states are reflective of an unobserved biological process, then we would expect that the states are *physiologically different* on average (i.e. for each of the four observed physiological variables, the mean values of these variables should be significantly different between the three detected states). The percentage of time that the patient spends with 'good' state (pc_{good}) and 'poor' state (pc_{poor}) should be also significantly different among those who died and survived (equivalent to the assumption of a dose-response effect of secondary injury). Both hypotheses were tested using one-way ANOVA against $p < 0.05$.

2) If the HMM reflects biologically meaningful states, then it should predict outcome in a dose dependent manner in a similar way to the individual physiology (i.e. it should be possible to 'collapse' the time-dependent model and obtain a similar classification to traditional time-independent features). To evaluate this, we calculated the performance of death classification using two static features obtained from HMM, i.e., (pc_{good}) and (pc_{poor}), and compared the results with when one employs another set of static features (averages of observation variables per patient). The goal was to investigate the accuracy of HMM performance on average. Regression with a stratified 20-fold cross-validation was used as a classification method and receiver operating characteristic (ROC) curves were calculated along with 95% confidence interval derived from bootstrapping. Brier scores, and the net reclassification index were calculated to compare the results and evaluate whether our HMM provided a classification that was plausible compared to a traditional average-based features.

We also studied the dynamic changes of the detected patients' states relative to their outcomes using a normalized histogram of ternary words (generated from running a three-hour moving window on the sequence of detected states) and visualized the results via a phylogenetic tree as described in Text 5 (Supplemental Digital Content 7).

Results

Table 1 summarizes the baseline characteristics of 379 subjects (28,939 hours of data) by GOS groups. The average age was 39±17years (79% male). The median admission Glasgow Coma Scale (GCS) was 7 with inter-quartile range (IQR) of 4-9. The average data length was 3.3±2.6days. One-way ANOVA showed no significant difference between groups with respect to the data length ($p=0.98$).

Fig 1 illustrates time-related changes of the observation variables (ICP, CPP, RAP, PRx) and the detected sequences of states by the proposed HMM for three example patients with different GOS values of 1, 4 and 5, respectively. We observe that the inferred hourly states of the patients by the HMM are consistent with the values of the variables and their trends. These results qualitatively validate the feasibility of the application of HMM for hourly state detection in TBI patients.

Table 2 compares the values of ICP, CPP, RAP, and PRx over detected states using one-way ANOVA. For all the four variables, the variable means differed significantly between 'good', 'intermediate' and 'poor' states (all $p<0.0001$). As expected, the mean ICP and mean PRx showed an increasing trend from 'good' to 'intermediate' and then to 'poor' state, while the mean RAP values depicted a decreasing trend. On the other hand, mean CPP (average ± standard error) for the 'intermediate' state (81.06±0.11) was in fact higher than those of the 'good' (76.98±0.07) and 'poor' (69.05±0.16) states. This could be explained by the CPP-oriented protocol implemented in Cambridge, especially in sicker individuals.

Fig 2 presents the result of the comparison of percentage of time that the patients spent within each of three detected states over various GOS. From Fig 2 (A) and Fig 2 (C), we observe that the average percentage of time that the patients spent in 'good' ('poor') state was significantly lower (higher) in the fatal group than the other three groups (severe disability, moderate disability, and good outcome) ($p<0.0001$). Interestingly, no significant difference in the percentage of the

time the patients spent in 'intermediate' state was observed among any of the four GOS groups ($p=0.65$).

The ROC curves for death classification are presented in Fig 3 with area under the curve (AUC) of 0.78±0.07 and 0.75±0.07, respectively. The Brier scores of both classifiers was 0.11. The net reclassification index was -0.01±0.06. These results revealed that the accuracy of death classification using the static measures of HMM output is the same as using the averages of the observation variables.

Fig S3 (Supplemental Digital Content 8) presents the results of ternary words analysis as described in Text 5 (Supplemental Digital Content 7). From Fig S3 (A) we observe that the majority of the generated ternary words were either 000, 111, or 222, regardless of the GOS. However, as one would expect, the probability of the word 000 (222) occurring increased (decreased), with improving patient outcome. Phylogenetic tree results indicate that severe/moderate disability and good outcome clustered together and separately from death (Fig S3 (B)).

Discussion

Our method combined ICP, CPP, PRx and RAP to define and track cerebral dynamics states of TBI patients on hourly basis. The results showed that the percentage of time that the patients spent in 'good' ('poor') state was significantly lower (higher) in the fatal group suggesting that our HMM has learned parameters consistent with conventional predictors of outcome.

Table 2 showed that the 'good' state is associated with low ICP, high CPP, intact autoregulation and preserved compensatory reserve; the 'poor' state is associated with higher ICP, lower CPP, reduced compensatory reserve and loss of autoregulation with the 'intermediate' state lying somewhere between the two. Interestingly, even though the HMM was constructed in an unsupervised way, the mean CPP for the 'poor' state obtained is still within published guidance for CPP management (60-70mmHg) (8) and that for good outcome group is rather above this range. Similarly, the 'poor' state was characterized by higher ICPs than the 'good' state, but this

was on average only just above recommended intervention threshold of 22mmHg (8).These observations not only give biological credibility to the states obtained by the model, but fit with the idea that even quite modest increases in ICP or decreases in CPP can be associated with poor outcome in a dose-dependent way.

Without a ground-truth, we used indirect measures to evaluate our method. Using the time-averaged HMM output, we showed that the performance of HMM in classification of death outcome was as high as when we used the average values of the physiological variables (Fig 3). Low net reclassification index and similar Brier scores demonstrated that the HMM provides a discriminating classification with high consistency and calibration. Of course, what the HMM provides in addition is the identification of the hidden states in a time-varying manner allowing it to flag up deterioration in real-time.

Our results also showed that irrespective of outcome, patients tend to stay in the same state over a moving three-hour window (Fig S3) suggesting that significant changes in cerebral dynamics occur over periods longer than 3 hours, a plausible observation given known pathobiology (2). We observed that the 'intermediate' state may have some inherent bias towards the 'poor' state (the probability of the word 111 was highest for the fatal group). One potential solution to address this issue could be using a quaternary (instead of ternary) state variable to differentiate between 'intermediate'-'good' versus 'intermediate'-'poor' states. The phylogenetic tree in Fig S3 (B) clustered the three-hour dynamic changes in patients with good outcomes and moderate disabilities as being similar and different to those who died. These findings are consistent with other studies showing different patterns of physiological changes in patients with fatal outcomes (9, 10).

Information overload is common in ICU and a reduction in the number of parameters that, at least initially, need to be assessed may improve data-salience. Furthermore, incorporating information from multiple sources may lead to improved alarms (11). Finally, aggregation of

parameters into biologically meaningful states may provide a substrate for better phenotypic characterization and treatment targeting or trial design than a single parameter alone.

Potential limitations of the study

Around 28% of the subjects in the original dataset of 527 subjects were excluded from further analysis mainly because of the late initiation of their data monitoring (relative to the time of the traumatic event) and/or the discontinuity of their data collection for more than 5 hours.

Given the retrospective nature of the study, we could not eliminate the influence of medications or treatments on the extracted physiological parameters, and consequently on the detected sequence of states. Furthermore, our choice of CPP, ICP, PRx and RAP was pragmatic and limited to the availability of these four variables related to intracranial physiology. There is some dependency between these variables, but each introduces additional explanatory information too. They have been previously demonstrated to have independent predictive power and potentially independent biological interpretation (10, 12).

Clearly other parameters would also be of interest—including temperature, arterial carbon dioxide tension and treatment effects. Unfortunately, such data was not available to us with sufficient quality/consistency. Heart-rate derived parameters would have been interesting to study, but they were not included for the sake of parsimony and because we were not primarily interested in systemic pathobiology: Although such parameters are known to be predictive of TBI outcome, the mechanism is unclear. We hope to evaluate more complex models in future work by including additional relevant variables. Similarly, our requirement for anonymous data prevented linkage with the clinical record to determine patients with decompressive craniectomy influencing PRx and RAP behavior.

Although the first-order HMM was a good starting point for our proof-of-concept, the assumption that the state of the patient at time encapsulates what we need to know about the history in order to predict the future state is an oversimplification of process. Furthermore,

stationarity of state transition probability could be invalid over prolonged time periods. Using HMM with higher orders could address this issue.

This study defined the patient's state as a ternary discrete variable ('good', 'intermediate' or 'poor'). A more practical and accurate model for detection of TBI patient's state might allow for having a continuous state variable by employment of other modeling techniques such as state-space modeling or Kalman filtering (13).

Our proposed method only employed four systematic variables to detect the hourly state. Using additional monitoring or clinical variables could potentially enhance the state description: A full dynamic Bayesian network could also incorporate information from interventions and other epidemiological and laboratory or imaging information to improve detection.

Conclusions

We developed and validated a proof-of-concept probabilistic approach using a HMM to continuously detect hourly states of TBI patients as a time-sensitive tensor-variate representation of the neurocritical care physiome of TBI. We have demonstrated that the HMM can identify states that are of clinical relevance in an unsupervised way and that they offer a method for learning clinically meaningful critical thresholds. Further enhancement of the state detection using more sophisticated Bayesian networks remains to be investigated. Such approaches may have broad applicability in understanding critical care physiology and in heavily time-dependent medical applications in general.

Acknowledgements

Funding

References

1. Maas AIR, Menon DK, Adelson PD, et al. Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. *The Lancet Neurology* 2017.

2. Menon DK, Ercole A. Critical care management of traumatic brain injury. Handbook of clinical neurology 2017;140:239-274.

3. Stocchetti N, Carbonara M, Citerio G, et al. Severe traumatic brain injury: targeted management in the intensive care unit. The Lancet Neurology 2017;16(6):452-464.

4. Lingsma HF, Roozenbeek B, Steyerberg EW, et al. Early prognosis in traumatic brain injury: from prophecies to predictions. The Lancet Neurology 2010;9(5):543-554.

5. Jennett B, Bond M. Assessment of outcome after severe brain damage: a practical scale. The Lancet 1975;305(7905):480-484.

6. Kim DJ, Czosnyka Z, Keong N, et al. Index of cerebrospinal compensatory reserve in hydrocephalus. Neurosurgery 2009;64(3):494-501; discussion 501-492.

7. Viterbi A. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. IEEE Transactions on Information Theory 1967;13(2):260-269.

8. Carney N, Totten AM, O'Reilly C, et al. Guidelines for the Management of Severe Traumatic Brain Injury, Fourth Edition. Neurosurgery 2017;80(1):6-15.

9. Balestreri M, Czosnyka M, Steiner L, et al. Intracranial hypertension: what additional information can be derived from ICP waveform after head injury? Acta Neurochirurgica 2004;146(2):131-141.

10. Lu C-W, Czosnyka M, Shieh J-S, et al. Complexity of intracranial pressure correlates with outcome after traumatic brain injury. Brain 2012;135(8):2399-2408.

11. Ruskin KJ, Hueske-Kraus D. Alarm fatigue: impacts on patient safety. Current Opinion in Anesthesiology 2015;28(6):685-690.

12. Czosnyka M, Pickard JD. Monitoring and interpretation of intracranial pressure. J Neurol Neurosurg Psychiatry 2004;75(6):813-821.

13. Chen Z. Advanced State Space Methods for Neural and Clinical Data: Cambridge University Press; 2015.

Figure legends

Figure 1- Examples of the trend of changes for four observation variables (Intracranial pressure, cerebral perfusion pressure, the compensatory reserve index, and pressure reactivity index) and the detected sequence of state for three patients with different Glasgow outcome scales. (A) Death: ICP rises above 25mmHg at hour 38, the CPP (RAP) demonstrated a decreasing trend starting from 53mmHg (0.5) at hour 30 (hour 44). Meanwhile, the PRx value rises above 0.05 at 30th hour. Given these trends, the HMM takes a probabilistic approach and infers the hourly state of this patient as 'good' until hour 30 whereupon the state changes to 'intermediate' for two hours, followed by an immediate change to a 'poor' state for the remaining of the data monitoring (note that this occurs many hours ahead of the fatal refractory intracranial hypertension), (B) Moderate disability: ICP starts at around 30 mmHg and rises to 50 mmHg by the 19th hour, before dropping to 6 mmHg in a three-hour time frame, possibly as the result of some intervention, mainly staying below 20 mmHg for the remainder of the monitoring time. The CPP (RAP) starts at around 80 mmHg (0.8) before dropping to 46 mmHg (0.25) by the 19th hour, and then returning to around the original values for the rest of the monitoring. However, the PRx rises above +0.12 at 14th hour and gets to as high as +0.89, before gradually goes back to normal values from 61th hour. The HMM plausibly classifies the first 14 hours as 'intermediate' state, followed by 47 hours of 'poor' state. The detected state changes permanently to good at 61th hour, (C) good outcome: A sudden drop and spike in the ICP value (between 30th and 37th hours), associated with a sudden spike and drop in CPP, and a drop in RAP has reasonably resulted in a change of state from 'good' to 'intermediate' and then to 'poor' over a seven-hour time frame.

Figure 2- Comparison of the percentage of time that the patients spent within each detected state over various Glasgow outcome scales. Results represent the mean and 95% confidence intervals. (A) percentage of time for 'good' state, (B) percentage of time for 'intermediate' state, (C) percentage of time for 'poor' state. *Group differed from three other groups. In Fig 2 (A) and (C), the average percentage of time that the patients spent in 'good' ('poor') state was significantly

lower (higher) in the fatal group than the other three groups (severe disability, moderate disability, and good outcome) with $p < 10^{-9}$.

Figure 3- Comparison of the death classification results employing two different sets of features. Feature set 1 included the averages of intracranial pressure, cerebral perfusion pressure, the compensatory reserve index, and pressure reactivity index per patient. Feature set 2 included the percentage of time the patient spent in the 'good' and 'poor' states. Results represent receiver operating characteristic curves with 95% confidence interval using bootstrapping.
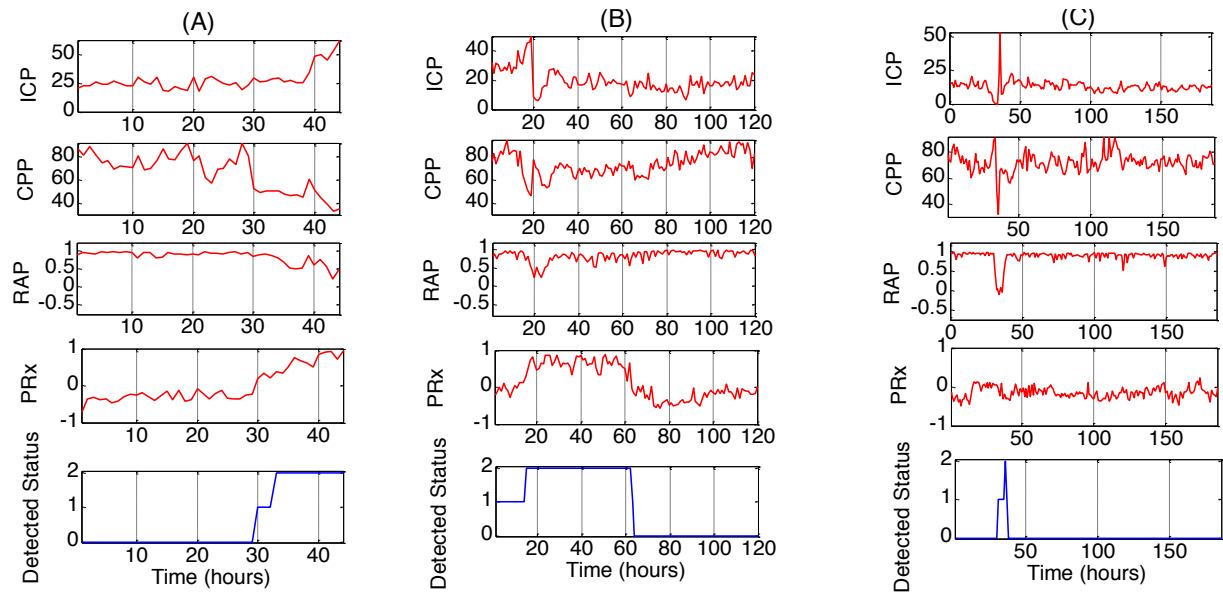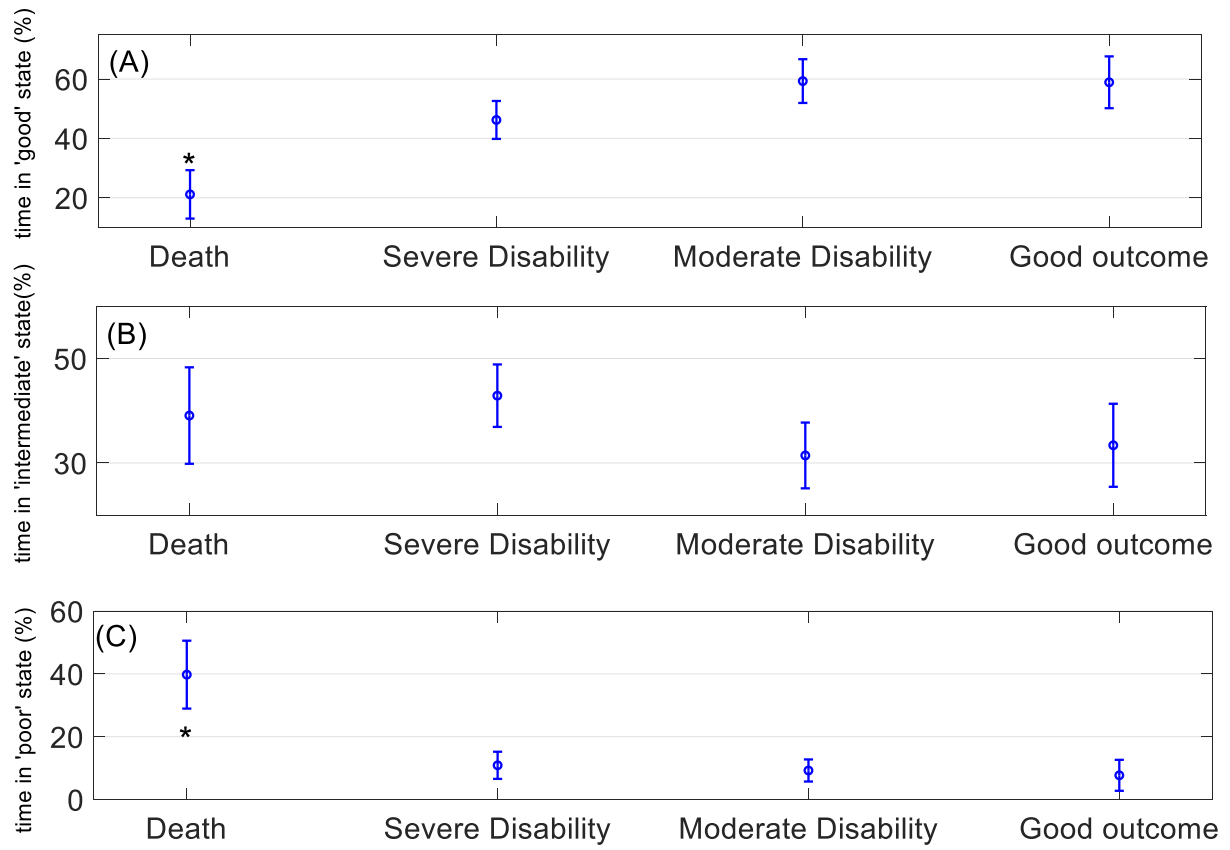
Figure 1

Figure 2

Figure 3
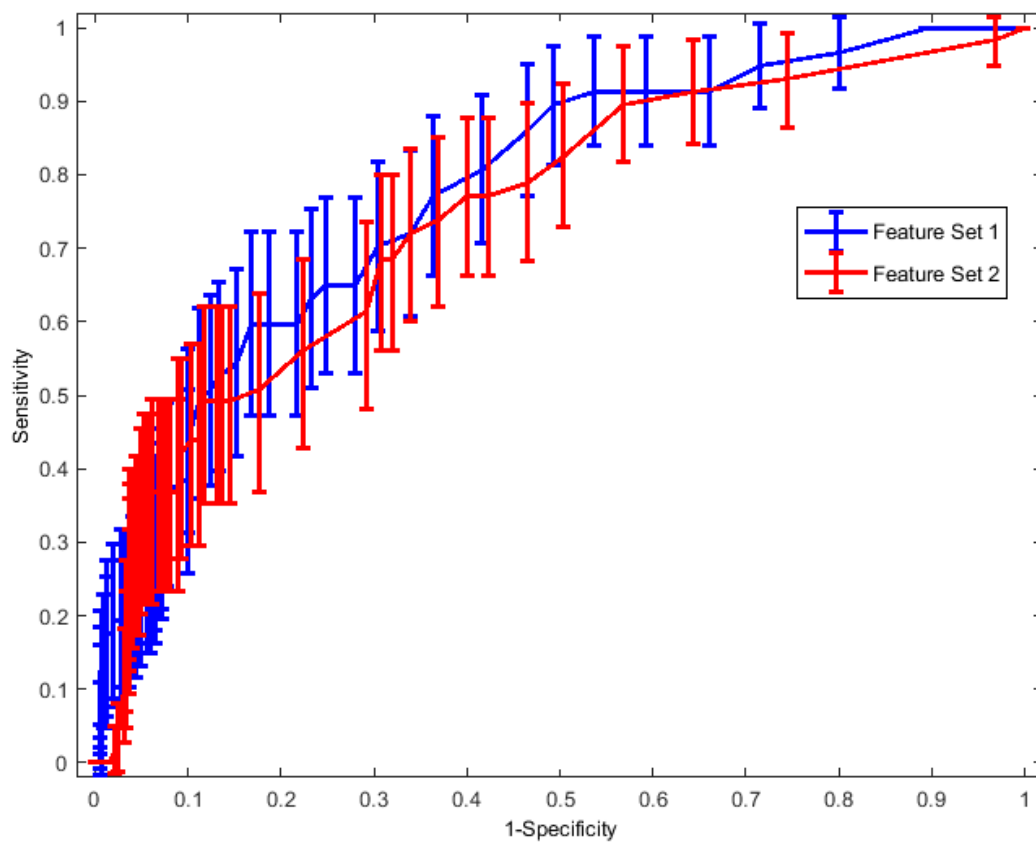
Table1- Demographic data of 379 subjects.

Demographic Parameter	GOS=1 (Dead)	GOS=3 (Severe Disability)	GOS=4 (Moderate Disability)	GOS=5 (Good Outcome)
n	57	138	104	80
Age [a] (year)	45.3±18.3	40.2±15.7	37.5±16.1	34.7±16.9
Male (%)	82.5	76.1	82.7	75.0
Glasgow Coma Scale [b]	4 [3-7]	6 [4-8]	7 [5-10]	8 [5-11]
Data Length [a] (Days)	3.4±2.8	3.3±2.5	3.2±2.5	3.3±2.5

[a] Numerical data are expressed as mean ± standard deviation.
[b] Categorical data are expressed as number (percentage) or median [inter-quartile range].

Table2- Comparison of the mean [95% confidence interval] values of observation variables (Intracranial pressure, cerebral perfusion pressure, the compensatory reserve index, and pressure reactivity index) over three different detected states using ANOVA. All the differences were statistically significant.

Variable	('good' state)	('intermediate' state)	('poor' state)
ICP	14.16[14.08-14.25]	15.88[15.74-16.00]	20.77[20.46-21.08]
CPP	76.98[76.84-77.11]	81.06[80.85-81.28]	69.05[68.73-69.37]
RAP	0.81[0.81-0.82]	0.47[0.46-0.47]	0.39[0.38-0.40]
PRx	-0.10[(-0.11)-(-0.10)]	0.07[0.06-0.07]	0.41[0.41-0.42]

Supplemental Digital Content Legends

1) **Supplemental Text 1**- Author summary of the presented work.

2) **Supplemental Text 2**- Description of patient management and exclusion criteria.

3) **Supplemental Text 3**- An overview of hidden Markov model and its structure.

4) **Supplemental Fig S1-** Temporal architecture of hidden Markov model.

5) **Supplemental Fig S2-**The state transition probabilities and observation likelihoods for a hidden Markov model with 3 states and 4 observation variables. The symbol $\sim \mathrm{N}(\boldsymbol{\mu}_i, \Sigma_i)$ in the figure indicates a Gaussian(normal) distribution with mean vector $\boldsymbol{\mu}_i$ and covariance matrix Σ_i.

6) **Supplemental Text 4**- Description of the initialization of the parameters for the proposed HMM.

7) **Supplemental Text 5**- Investigating the patterns of changes in the detected states of traumatic brain injury patients over a three-hour time frame using ternary word analysis for different Glasgow outcome scales.

8) **Supplemental Fig S3-** Analysis of the patterns of changes of TBI patients' state over a three-hour time frame using ternary word analysis for different Glasgow outcome scales. (A) Distribution of generated ternary words (The clusters correspond to the most common words: 000,111 and 222). Regardless of the Glasgow outcome scales, the majority of the generated ternary words were either 000, 111 or 222. However, as one would expect, the probability of the word 000 (222) occurring increased (decreased), with improving patient outcome. For example, over a course of three hours, those who died had 000 (stable 'good state') pattern only 16% of the time, whereas for those with good outcome this percentage increased to 60%. Interestingly, the probability of the word 111 ('intermediate' state for three hours) was highest for the fatal group. (B) phylogenetic tree of the distances between the ternary word distributions for different Glasgow outcome scale. The dynamic changes of traumatic brain

injury patients' states had the most similarity among those people who had good outcome and moderate disability. Furthermore, since the branch of death outcome is longer relative to other three groups, we conclude that those who survived showed more similar dynamic changes of the hourly state relative to each other than those who died.

Supplemental Text 1

Why was this study done?

- Intensive care admissions are data-dense episodes as they include continuous assessment of physiology. Unlike most other areas of medicine, the patient's clinical course is summarized by a large volume of multivariate time-series data.

- Such data is critical for decision making yet is difficult for clinicians to appreciate due to its multidimensional, high volume and temporal nature.

- Machine learning techniques may enhance time-sensitive detection, diagnosis, prognostication of response to treatment.

What did the researchers do and find?

- We extracted retrospective hourly observations of four key physiological parameters from patients who had sustained severe traumatic brain injury.

- We constructed a Hidden Markov Model of these parameters to reduce them to a novel composite hourly 'state' of the patients understood in terms of the likely instantaneous outcome but retaining original prognostic power in aggregate when compared to conventional epidemiological measures.

- Thresholds for different states learned by the model were found to 'discover' injury thresholds which recapitulated known pathophysiology.

What do these findings mean?

- This is the first use of a Hidden Markov Models to create a clinically meaningful measure of instantaneous 'patient state' which retains prognostic power in aggregate but also reveal biologically relevant thresholds for injury.

- Such model may provide a better insight into patient's condition, within an overall protocolized and evidence-based framework, for individualized real-time TBI patient management that accounts for intra- and inter-patient variability.

Supplemental Text 2

Under UK regulations, institutional approval is not required for research using anonymous routine clinical data. ICP was transduced using an intraparenchymal probe (Codman & Shurtleff, Inc., Raynham, MA, USA). ABP was measured using a standard indwelling intra-arterial catheter. Patients were sedated, mechanically ventilated and managed according to an ICP / CPP-directed protocol (1) and progressive interventions were used to simultaneously attempt to keep ICP < 20mmHg and CPP between 60-70mmHg in correspondence with international consensus. Noradrenaline (norepinephrine) was used as the vasopressor to support CPP in the overwhelming majority of cases.

Between 2002 and 2016, total of 527 adult (≥16 year) TBI patients were admitted to Neurosciences Critical Care Unit (NCCU) of Addenbrooke's Hospital, Cambridge and had available recordings of ICP, ABP, and 6-month GOS. In our study, decompressive craniectomy could not be an exclusion criterion, because our requirement for anonymous data prevented linkage with the clinical record to determine whether patients had received decompressive craniectomy or not. However, this procedure is used as a rare rescue therapy in Addenbrooke's NCCU and therefore is expected to not be an issue for the majority of the recordings. 18 patients were excluded due to death related to non-neurological causes (and therefore potentially not indicative of neurological pathobiology (2, 3)). A further 17 patients with GOS of 2 (vegetative or minimally conscious) were excluded due to their small number and greater heterogeneity compared to the other outcome groups.

Given that the main goal of the study was detection of hourly states of TBI patients, we only focused on the data of those subjects who also satisfied the following conditions:

- Data collection for the majority of the patients (more than 80%) started within 3 days of TBI. Thus, we focused on these patients to ensure comparable biological trajectories since dominant processes vary greatly over the initial days post-injury (4, 5).

- Minimum recording time of 8 hours, since shorter data segments may include too limited information to train the HMM robustly and because patients with only very short recordings may not be clinically representative of the remainder of the cohort (we seek to construct a HMM which will be helpful in patients requiring treatment decisions over a period of days).

- Data continuity such that there were no more than 5 consecutive hours of gaps in the data collection. However, if the duration of the gap(s) in the collected data of a patient was 5 hours or less, the two segments of the data before and after the gap were merged. This criterion was empirically imposed in an attempt to ensure biological continuity since TBI evolution is expected to occur on longer timescales (4, 5).

Applying these constraints, data of 379 patients remained for the training and evaluation of the HMM. We excluded the last 6 hours of the data prior to death or removal of the patient from NCCU to reduce the effect of potential confounding factors of withdrawal of sedation on the reliability of the data for an unbiased detection of the patient state using the proposed HMM.

The collected ICP and ABP data were sampled at 30 to 200Hz. To obtain PRx values, a moving Pearson correlation coefficient was obtained from changes of ABP and ICP in a window of 30 consecutive 10-second time averages. RAP values were calculated using the same moving correlation coefficient between mean ICP and the pulse amplitude (first harmonic) of the ICP waveform (6).

References

1.   Menon DK. Cerebral protection in severe brain injury: physiological determinants of outcome and their optimisation. British medical bulletin 1999;55(1):226-258.
2.   Pfeifer R, Teuben M, Andruszkow H, et al. Mortality Patterns in Patients with Multiple Trauma: A Systematic Review of Autopsy Studies. PLoS One 2016;11(2):e0148844.
3.   Adams H, Donnelly J, Czosnyka M, et al. Temporal profile of intracranial pressure and cerebrovascular reactivity in severe traumatic brain injury and association with fatal outcome: An observational study. PLoS medicine 2017;14(7):e1002353.

4.   Ercole A, Magnoni S, Vegliante G, et al. Current and Emerging Technologies for Probing Molecular Signatures of Traumatic Brain Injury. Frontiers in neurology 2017;8:450.

5.   Menon DK, Ercole A. Critical care management of traumatic brain injury. Handbook of clinical neurology 2017;140:239-274.

6.   Kim DJ, Czosnyka Z, Keong N, et al. Index of cerebrospinal compensatory reserve in hydrocephalus. Neurosurgery 2009;64(3):494-501; discussion 501-492.

Supplemental Text 3

TBI patients have a high variability in their response to the care and treatment. The increased employment of protocolized and evidence-based care have contributed to reduction of this variability and improvement of patient outcome (1). However, there is still a lack of patient-specificity in the existing protocols to address the variability of outcome due to intra- and inter-patient variability (2). Mathematical and computational modeling of TBI with clinical data can provide such patient-specific insight to the patient's condition and consequently new means for clinical decision making and patient management (3).

However, mechanistic models such as lumped-compartment models are usually either oversimplified (of a complicated physiological system such as brain) (4) or prohibitively too complex for practical real-time implementation (3). Statistical modeling, on the other hand, tries to learn about the underlying physiology of the system by stochastically relating several variables that are generated in the process (5). With recent advancements in artificial intelligence and machine learning, sophisticated statistical modeling of big data in medicine have become possible. Intensive care (and neurocritical care in particular) is a highly data-dense and therefore a potentially fertile area for machine learning applications.

This work aimed at employing a machine learning approach to obtain hourly states of TBI patients. As this model is inherently dynamic, as a proof-of-concept, we decided to use the HMM—the simplest form of dynamic Bayesian network. HMMs have been successfully employed for various temporal pattern recognitions, e.g., speech recognition (6, 7), handwriting recognition (8), and bioinformatic analysis of various sequences (9).

HMM consists of a sequence of unobserved/hidden 'states' where transitions between these states are assumed to obey the first-order Markov property (10); the probability of the subsequent state depends only on the current state. The patient moves between states over time with the model 'emitting' values for the observed physiology (in our case ICP, CPP, PRx,

RAP) with a probability distribution that depends only on the current (unobserved) state. Our choice of CPP, ICP, PRx and RAP was pragmatic and just based on the availability of adequate quality data related to intracranial physiology. We decided to include all of these four variables as they (all together) might provide a better description of intracranial physiology reflecting edema via ICP, perfusion via CPP, autoregulatory state via PRx and degree of compensatory reserve via RAP which we know from the literature to be independently predictive of outcome (11, 12).

The HMM finds the sequence of states and corresponding probability distributions that are the most probable description of the observed data. HMM can efficiently decode a sequence of states ($X^T = \{x(1), x(2),...,x(T)\}$) that is not immediately observable, using other sequence of observable data ($\mathbf{Y}^T = \{Y(1), Y(2),...,Y(T)\}$) that is dependent on X^T.

In general, $x(t)$ is a discrete random variable with N values, while the observation $Y(t)$ is a $(M \times 1)$ continuous random vector with Gaussian distribution. Fig. S1 (Supplemental Digital Content 4) depicts the temporal architecture of HMM (Trellis diagram) and its memoryless property. HMM has the following parameters that need to be estimated (during training phase) before decoding the sequence of hidden states (during testing phase):

- $\prod = [\pi_j]$: $(N \times 1)$ prior vector, where π_j is the probability of being at state j, at time $t = 1$,

- $\mathbf{A} = [a_{ij}]$: $(N \times N)$ state transition matrix, where a_{ij} is the conditional probability of moving from state i to state j,

- $\boldsymbol{\mu}_j$: $(M \times 1)$ vector of the mean values of the observation variables for each state j, where $j = 1, 2,..., N$,

- $\sum_j$: $(M \times M)$ covariance matrix of the observation variables for each state j, where $j = 1, 2,..., N$.

Note that for each state j, the likelihood of observation is expressed as a M-dimensional Gaussian distribution with mean vector $\boldsymbol{\mu}_j$ and covariance matrix $\sum_j$. Fig. S2 (Supplemental Digital Content 5) shows the state transition probabilities and observation likelihoods for an HMM with 3 states and 4 observation variables ($N = 3$, $M = 4$).

The HMM has several parameters (prior vector, state transition matrix, mean vector and covariance matrix of the observation Gaussian likelihoods) that need to be estimated during the training phase before one can decode the sequence of hidden states. In supervised training of HMM the sequence of observations along with the given sequence of states (provided as the ground-truth) are used to estimate the HMM parameters. However, in many practical problems the ground-truth for sequence of states is not readily available. In these cases, unsupervised training of HMM is conducted by applying an iterative algorithm such as Baum-Welch algorithm to estimate the HMM parameters. The Baum-Welch algorithm is an instance of a generalized Expectation-Maximization algorithm (6): Starting from some initial values of HMM parameters, the algorithm iteratively updates the expected values of these parameters (taken over the posterior probability of hidden state sequences) until convergence (13).

References

1.   Wendon J. Critical care "normality": individualized versus protocolized care. Critical care medicine 2010;38(10):S590-S599.
2.   Kavanagh BP, Meyer LJ. Normalizing physiological variables in acute illness: five reasons for caution. In: Applied Physiology in Intensive Care Medicine: Springer; 2006. p. 269-275.
3.   Chase JG, Le Compte AJ, Preiser J-C, et al. Physiological modeling, tight glycemic control, and the ICU clinician: what are models and how can they affect practice? Annals of intensive care 2011;1(1):11.
4.   Kashif FM, Verghese GC, Novak V, et al. Model-based noninvasive estimation of intracranial pressure from cerebral blood flow velocity and arterial pressure. Science translational medicine 2012;4(129):129ra144-129ra144.
5.   Neumann GA, Finkler TF, Cardozo NS, et al. Comparison between phenomenological and empirical models for gas-phase polymerization process control. Industrial & engineering chemistry research 2006;45(8):2651-2660.
6.   Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE 1989;77(2):257-286.

7.   Gales M, Young S. The application of hidden Markov models in speech recognition. Foundations and trends in signal processing 2008;1(3):195-304.

8.   Bunke H, Roth M, Schukat-Talamazzini EG. Off-line cursive handwriting recognition using hidden Markov models. Pattern recognition 1995;28(9):1399-1413.

9.   Yoon B-J. Hidden Markov models and their applications in biological sequence analysis. Current genomics 2009;10(6):402-415.

10.   Baum LE, Petrie T. Statistical Inference for Probabilistic Functions of Finite State Markov Chains. Ann Math Statist 1966;37(6):1554-1563.

11.   Czosnyka M, Pickard JD. Monitoring and interpretation of intracranial pressure. J Neurol Neurosurg Psychiatry 2004;75(6):813-821.

12.   Lu C-W, Czosnyka M, Shieh J-S, et al. Complexity of intracranial pressure correlates with outcome after traumatic brain injury. Brain 2012;135(8):2399-2408.

13.   Beal MJ, Ghahramani Z, Rasmussen CE. The infinite hidden Markov model. In: Advances in neural information processing systems; 2002; 2002. p. 577-584.

Supplemental Text 4

The initial values of the HMM parameters were chosen as the following:

- Prior vector: $\Pi = \begin{bmatrix} 0.4 \\ 0.3 \\ 0.3 \end{bmatrix}$, assuming that, right after the TBI moment, the probability of

  starting from a 'good' state is slightly higher than those of the 'intermediate' and 'poor'

  states (The cerebral swelling does not usually happen within the first day after TBI (1,

  2)).

- State transition matrix: $\mathbf{A} = \begin{bmatrix} 0.6 & 0.3 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.1 & 0.3 & 0.6 \end{bmatrix}$. Given the one-hour time interval for moving

  between the states, we can expect that for 'good' (poor) state, the probability of moving

  to the intermediate' state is higher than that of moving directly to the opposite 'poor'

  ('good') state. Meanwhile, we assumed that the intermediate' state can move to a 'good'

  or 'poor' state with equal probability

- Mean vector of 'good' state: $\boldsymbol{\mu}_1 = \begin{bmatrix} 10 \\ 70 \\ 0.8 \\ -0.5 \end{bmatrix}$, Mean vector of intermediate' state: $\boldsymbol{\mu}_2 = \begin{bmatrix} 20 \\ 60 \\ 0.4 \\ 0 \end{bmatrix}$,

  and Mean vector of 'poor' state: $\boldsymbol{\mu}_3 = \begin{bmatrix} 30 \\ 50 \\ 0.4 \\ 0.5 \end{bmatrix}$. These values are chosen randomly based

  on the normal ranges of ICP, CPP, RAP and PRx (3-5) .

The covariance matrices of the 'good' and 'poor' state (Σ_1 and Σ_3) were obtained by calculating

the covariance matrices of the data from those subjects in the training dataset (of each fold) with

GOS 5 and 1, respectively. Then, the average of these two covariance matrices were used as

the covariance matrix for the intermediate' state $\Sigma_2 = 0.5 \times (\Sigma_1 + \Sigma_3)$.

References

1.   Menon DK, Ercole A. Critical care management of traumatic brain injury. Handbook of clinical neurology 2017;140:239-274.

2.   Stocchetti N, Carbonara M, Citerio G, et al. Severe traumatic brain injury: targeted management in the intensive care unit. *The Lancet Neurology* 2017;16(6):452-464.

3.   Bratton SL, Chestnut RM, Ghajar J, et al. IX. Cerebral perfusion thresholds. *Journal of neurotrauma* 2007;24(Supplement 1):S59-S64.

4.   Steiner LA, Czosnyka M, Piechnik SK, et al. Continuous monitoring of cerebrovascular pressure reactivity allows determination of optimal cerebral perfusion pressure in patients with traumatic brain injury. *Critical care medicine* 2002;30(4):733-738.

5.   Balestreri M, Czosnyka M, Steiner L, et al. Intracranial hypertension: what additional information can be derived from ICP waveform after head injury? *Acta Neurochirurgica* 2004;146(2):131-141.

Supplemental Text 5

To investigate the patterns of changes in the detected hourly states, we used a moving window of three hours (with one-hour overlap) over the detected sequence of states for each patient. Since three possible states could be detected for each hour of data (0, 1 or 2), the detected sequence for every three hours could be mapped to a ternary word of 3 trits (similar to binary bits). A normalized histogram of the $27(3^3)$ possible ternary words for different GOS values can provide us with more insight into the dynamic changes of the patients' states (over a course of few hours) relative to each outcome. As a measure of similarity between the obtained distributions, we used the Wasserstein (Earth mover's) distance (1). To visualize the results, a phylogenetic tree arranging different groups on its branches best to fit the pairwise distance measures was constructed based on the average inter-GOS-group distances (2).

1.   Rubner Y, Tomasi C, Guibas LJ. The earth mover's distance as a metric for image retrieval. International journal of computer vision 2000;40(2):99-121.
2.   Fitch WM, Margoliash E. Construction of phylogenetic trees. Science 1967;155(3760):279-284.

(A)

(B)