BMJ Open

Development of a point of care system for automated coma prognosis: a prospective cohort study protocol

John F Connolly,[1,2,3,4,5,6] James P Reilly,[1,2,4,7] Alison Fox-Robichaud,[8,9] Patrick Britz,[10] Stefanie Blain-Moraes,[11] Ranil Sonnadara,[1,2,3,4,5,6,12] Cindy Hamielec,[8,9] Adianes Herrera-Díaz,[4,6] Rober Boshra[1,2,4]

To cite: Connolly JF, Reilly JP, Fox-Robichaud A, *et al*. Development of a point of care system for automated coma prognosis: a prospective cohort study protocol. *BMJ Open* 2019;**9**:e029621. doi:10.1136/bmjopen-2019-029621

► Prepublication history for this paper is available online. To view these files please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2019-029621).

Received 1 February 2019
Revised 6 June 2019
Accepted 12 June 2019

For numbered affiliations see end of article.

Correspondence to
Dr John F Connolly;
jconnol@mcmaster.ca

ABSTRACT

Introduction Coma is a deep state of unconsciousness that can be caused by a variety of clinical conditions. Traditional tests for coma outcome prediction are based mainly on a set of clinical observations. Recently, certain event-related potentials (ERPs), which are transient electroencephalogram (EEG) responses to auditory, visual or tactile stimuli, have been introduced as useful predictors of a positive coma outcome (ie, emergence). However, such tests require the skills of clinical neurophysiologists, who are not commonly available in many clinical settings. Additionally, none of the current standard clinical approaches have sufficient predictive accuracies to provide definitive prognoses.

Objective The objective of this study is to develop improved machine learning procedures based on EEG/ERP for determining emergence from coma.

Methods and analysis Data will be collected from 50 participants in coma. EEG/ERP data will be recorded for 24 consecutive hours at a maximum of five time points spanning 30 days from the date of recruitment to track participants' progression. The study employs paradigms designed to elicit brainstem potentials, middle-latency responses, N100, mismatch negativity, P300 and N400. In the case of patient emergence, data are recorded on that occasion to form an additional basis for comparison. A relevant data set will be developed from the testing of 20 healthy controls, each spanning a 15-hour recording period in order to formulate a baseline. Collected data will be used to develop an automated procedure for analysis and detection of various ERP components that are salient to prognosis. Salient features extracted from the ERP and resting-state EEG will be identified and combined to give an accurate indicator of prognosis.

Ethics and dissemination This study is approved by the Hamilton Integrated Research Ethics Board (project number 4840). Results will be disseminated through peer-reviewed journal articles and presentations at scientific conferences.

Trial registration number NCT03826407.

Strengths and limitations of this study

- This study will be the first to record 24-hour continuous electroencephalogram/event-related potential (EEG/ERP) data in comatose patients across multiple progression points, allowing longitudinal tracking of patient changes to provide evidence for prognosing outcome with unprecedented accuracy.
- A complete hierarchical investigation targeting different levels of sensory, cognitive and language processing will be assessed for predicting emergence and positive coma outcome by using electroencephalography techniques.
- Application of modern machine learning approaches to large continuous EEG/ERP data sets has vast potential in both automating and improving the prediction of coma outcome.
- A limitation of this study is its heterogeneous patient sample due to varying admission and recruitment rates for different aetiologies.

INTRODUCTION

Background and rationale

Coma is a state of prolonged unconsciousness with no eye opening that can be caused by a wide range of clinical conditions, such as traumatic brain injury, cardiac arrest, stroke, brain tumour, and drug or alcohol intoxication.[1] Coma state typically lasts for a few weeks, and transitions into either unresponsive wakefulness syndrome (also known as vegetative state) or minimally conscious state, and is generally the result of diffuse and bihemispheric lesions of the cortex or white matter, bilateral thalamic damage, or focal lesions of the paramedian tegmentum.[2] During coma, patients are shown to be unaware of both self and external surroundings and unable to respond meaningfully to external stimuli. The prediction of functional outcome after coma is of considerable importance for the patients, their relatives, medical care and public health. Another important aspect to consider is that the outcome of coma is related to the aetiology independent of the physical signs, depth of coma or length of coma.[3] In a meta-analysis estimated with the data of 548 comatose and low responsive patients, the prognosis was worst for patients

with anoxia or metabolic encephalopathy and best for trauma or brain surgery.[4]

The current method for determining coma prognosis is the Glasgow Coma Scale (GCS),[5] which is easy to apply but yields coarse or even misleading results as it is mainly based on a set of clinical observations. In brief, the GCS includes three parts: eye opening, verbal response and best motor response, with increasing scores as behavioural responses improve. The Glasgow Coma Scale-Pupils (GCS-P)[6] is an extended version of GCS and includes an additional scoring of a patient's pupillary responses as an indication of injury severity.

Neurophysiological methods have proven of some use,[7] with primary somatosensory responses in the 30 ms range and brainstem auditory evoked potentials (BAEPs) exhibiting high prognostic value of poor outcome in patients with GCS scores of 3.[8 9] Although short-latency evoked potentials have been useful at determining unfavourable outcome in coma survivors, they only estimate the integrity of ascending pathways.[10] Therefore, they are less helpful in prognostic coma recovery. More recently, long-latency event-related potentials (ERPs) have also been introduced as useful predictors of a positive coma outcome.[11 12]

The most common ERP paradigm employed for coma prognosis is traditionally known as the 'passive oddball paradigm'.[13 14] Classically it includes two tones—one occurring frequently (standard) and the other occurring less frequently (deviant).[15 16] The series of standard tones is interspersed with deviant tones (eg, differing in duration, intensity or frequency). Each might elicit two different long-latency ERP waveforms: the standard tone generates a classic auditory sequence consisting of the N1/P2 complex, while the deviant tone elicits the N1 and the mismatch negativity (MMN).[15 17] The presence of the N1 and the MMN (often elicited at about 100 ms and 150 ms poststimulus, respectively) provides evidence of auditory cortical function that in the MMN case may require being in a state of consciousness but not necessarily awareness.[18 19] The N1 is an obligatory sensory response evoked by each tone (ie, both standard and deviant) and highlights the encoding of acoustic input in the auditory cortex. The MMN is an automatic response to deviants and highlights preserved automatic sensory memory processes and what is often called preattentive cognitive processes.[15] These electrophysiological responses are elicited without requiring the subject's active involvement. Clinical studies on coma patients demonstrate that the presence of the MMN component has a good correlation with coma awakening.[13 14] The reported results show that more than 90% of patients who were considered as non-awake showed no MMN (ie, a high specificity), and more than 90% of patients in whom MMN was detected returned to consciousness (ie, a high positive predictive value). But only about 30% of patients who had regained consciousness showed MMN (ie, a low sensitivity). Assessment of the MMN recorded on a single occasion typically is based on the average of ERP signals time locked to each tone recorded over a long recording time period (typically on the order of 30 min),[13 20] in order to increase the effective signal to noise ratio to a suitable level. This longer averaging process can 'smear out' or obscure important, clinically relevant events such as short-duration increases in the level of consciousness. This can occur due to latency variations (or 'jitter') in the individual trial responses that comprise the final average. Also, we have demonstrated that the MMN waxes and wanes when assessed longitudinally across extended time periods (24 hours).[18] We postulate that this 'cycling' of presence/absence is likely the predominant reason for the low sensitivity of the MMN reported in clinical studies.

In addition to MMN, the P300 component has also been reported as a reliable predictor of awakening.[4 13] This component, also elicited by using oddball paradigms, has been related to higher level process such as attention, expectancy, novelty detection, stimulus salience, target recognition, memory and so on.[12 13 20] In particular, the use of 'novel' stimuli (thoroughly unrelated to ongoing stimulus sequence such as a dog barking or a telephone ringing) and the subject's own name (SON) has been demonstrated to increase the chances of recording a response from comatose patients.[13 21 22] In comparison with MMN, novelty P300 has shown as large a specificity (84.6%) but a much higher sensitivity (70.8% for novelty P300 vs 41.6% for MMN).[13 21] Results suggest that novel stimuli activate much larger neuronal networks than the deviants,[16] the signal to noise ratio is much higher for large novelty P300s than for MMNs, and therefore the use of novelty P300 might increase the prognostic value of MMN alone. Interestingly, the N400 component, well known as an ERP index of semantic processing, has also been reported in coma patients,[23 24] particularly in those with intact temporal cortex.[25] Although the integrity of language comprehension seems to be unlikely during coma, these previous studies have shown that assessing higher cognitive levels should not be disregarded.

Typically, detection of all the highlighted components is based on visual inspection of the averaged ERPs by skilled neurophysiologists, a process that is expensive, slow and not always feasible in practice. Therefore, automatic and accurate detection of ERP components over a short time is necessary to provide the most salient clinical information on the current state and prognosis of the patient. The present study is primarily centred around creating a machine learning (ML) paradigm to analyse the electroencephalogram (EEG)/ERPs in order to provide prognosis for patients in coma. ML (also known as data mining or pattern recognition) methods have been previously used in several EEG applications, including the analysis of EEG signals for epilepsy,[26] in evaluating residual functional deficits following concussion,[27 28] in predicting the effect of selective serotonin reuptake inhibitor for treating major depressive disorder and in investigating the effect of clozapine treatments for schizophrenia.[29 30] Additionally, earlier work by the authors has shown ML to be effective at aiding the prediction of coma outcome.[18]

Overview and objectives

Prediction of coma outcome is an important aspect of critical healthcare, since it provides families and their healthcare team with information to guide discussions around goals of care and prolonged life-sustaining therapies. Our primary objective is to apply modern techniques in ML to analyse the patient's EEG and develop a simple and inexpensive point of care system that can significantly improve the accuracy of coma prognosis. We subsequently refer to the resulting product as the Coma Prognosis System (CPS). The intent of the project is to develop a low-cost, easy-to-use prototype CPS device that is ready for clinical trials. Previous work has used ML in investigating cognitive processing in other disorders of consciousness.[31–33] We intend to extend previous work to target coma prognosis using a comprehensive investigation into different ERPs indexing a wide scope of cognitive function as well as resting-state (RS) EEG. We will leverage and augment previous work[18 34 35] of the authors to facilitate the adoption of the work by the critical care community.

METHODS

Study setting

The study will take place at the Hamilton General Hospital, a tertiary/quaternary care centre serving people neurologically injured in an area around Hamilton with a population of 2.6 million. The three specific sites are the following: intensive care unit, neurological step-down unit and coronary care unit.

Eligibility criteria

Inclusion criteria

- Patients admitted to the intensive care unit, neurological step-down unit or coronary care unit at Hamilton General Hospital who are in coma with GCS score of 3–8, between day 4 and day 9 of admission.
- ≥18 years of age.

Exclusion criteria

- Severe liver failure (ie, Child-Pugh class C).
- Severe renal failure (ie, urea ≥40 mg/dL).
- Previous open-head injury.
- Known primary and secondary central nervous system malignancy (ie, tumour).
- Known hearing impairment.
- Previous intracranial pathology requiring neurosurgical interventions in the past 72 hours.
- Patients who are actively being sedated in a medically induced coma.
- Patients who are otherwise deemed medically unsuitable for this study by the attending intensivists (eg, severe anatomical injury preventing electrode application, medical instability requiring frequent interventions).

Patient and public involvement

Patients and the general public were not involved in the design of the present study. Results will be disseminated after completion of the project through public talks, scientific conferences and scientific publications.

Outcomes

Primary outcome measures

Change in multiple electrophysiological measures across specified time points during coma

ERPs and RS periods will be assessed at the specified intervals as a difference between successive time points. The ERP measures will be used to assess different levels of conscious processing and presence of signs of a conscious state predictive of subsequent emergence. Also, resting EEG measures will be obtained at regular intervals. EEG/ERP data will be recorded for 24 consecutive hours at a maximum of five time points spanning 30 days from the date of recruitment to track the participants' progression.

Correlation between behavioural and electrophysiological measures after coma emergence

In the case of patient emergence, the full electrophysiological test procedures will be recorded to correlate with traditional behavioural measures. The electrophysiological measures obtained at this time point (emergence) will be compared with the same measures obtained at the five different time points (outcome 1) to detect both clinically relevant change and possible prognostic markers obtained at an earlier test point.

Sensitivity and specificity changes in prognostic capabilities of electrophysiological measures

Analyses will compare the electrophysiological measures as outcome predictors with traditional behaviourally based tools.

Secondary outcome measures

The study will also collect aetiology, demographics and medical history from the patient's health record, in addition to concurrent physiological assessment during the study period.

Participant timeline

Patients who meet the eligibility criteria will be invited to participate in the study through their substitute decision maker. Informed consent will be obtained from the substitute decision maker either in person or by phone. The initial assessment will take place immediately after the consent. Subsequent assessments will be repeated four times within a 30-day period, which makes a total number of five assessments. Patients' GCS score will be monitored by the nursing staff at least once on a daily basis until they emerge from the coma (as confirmed by an attending physician) or at the end of their 30-day study duration. In cases of discontinued life support, testing will be conducted until declaration of death.

The date of the initial assessment will be denoted as day 0, and the subsequent assessments will take place on day

3, day 10, day 20 and day 30, unless there is a persistent change in the GCS-P scores that is greater or equal to 2 points in either direction. If there is a change ≥2 points that persists over two consecutive GCS-P assessments, an EEG assessment will take place immediately as a substitute for the next scheduled assessment. Any remaining assessments will occur as per the predetermined schedule, unless the change in GCS-P score occurs again. If the patient emerges from the coma prior to the end of the 30-day study duration, the remaining prescheduled EEG assessment(s) will be cancelled and replaced with one 2-hour EEG session. We selected these time intervals to maximise the number of data points being captured while extending the time window to 1 month. Patients in coma tend to recover less than a month after injury.[3] Moreover, the initial date of testing is typically several days after admission, and this provides us with a time restriction before a patient is likely to emerge. Thus, we designed the time windows to have a more testing initially before tapering off in frequency to cover the fifth test at the end of the 30-day interval.

Recruitment

Fifty patients will be recruited for the present study over the course of 2 years. Primarily, the study staff will identify potential patients from the hospital's database. Since approximately 5% of patients present to the emergency department with an altered mental state and 1% of the admissions are due to coma,[36] we predict a similar proportion of comatose patients in the emergency department at the Hamilton General Hospital. Based on the inclusion/exclusion criteria, we estimate two patient recruitments per month. Prior to the commencement of the study, the study team will present the study information, especially the eligibility criteria, to the nursing staff and other front-line healthcare providers. These personnel will, in turn, serve as the secondary source for recruitment. Information pamphlets about the study will also be made available at the three primary recruitment sites at the Hamilton General Hospital. The healthy control participants will be recruited through advertisements.

Data collection, management and analysis

For ERP and RS acquisition, continuous EEG data will be recorded for 24 hours at the patient's bedside (bandpass=0.01–100 Hz and sampled at 512 Hz) using a 64-channel Biosemi ActiveTwo system (Biosemi, Amsterdam, The Netherlands). Electrodes will be placed on the scalp according to the standard 10/20 positioning using a 64-electrode cap. Vertical and horizontal electro-oculogram signals will be monitored by electrodes placed above and over the outer canthus of the left eye. References will be recorded bilaterally from the mastoids and at the nose for offline rereferencing. In the case of a skull fracture or any obstruction to the placement of a regular cap, a customised cap will be used with a reduced number of electrodes. Similarly, data from healthy controls will be recorded using a 64-channel EEG cap. BAEPs and middle latency auditory evoked potentials (MLAEPs) will be recorded using the same equipment but using different acquisition settings. Detailed settings for the two paradigms are discussed below.

All stimuli will be delivered through inserted earphones (Etymotic ER-1) using the Presentation software (Neurobehavioral Systems). A battery of auditory paradigms developed to assess levels of sensory, cognitive and linguistic processing is used to predict emergence from coma. We consider this as a complete hierarchical investigation that includes all levels of information processing, taking into account short-latency evoked potentials (BAEPs, MLAEPs) that help to estimate the integrity of ascending auditory pathways, but also cognitive and late-latency ERP components such as the MMN, P300 and N400. Procedures will include computer-delivered aural instructions for all patients regardless of disorder of consciousness (DOC) category, recognising the possibility of verbal command comprehension (periodically or constantly). A brief description of each paradigm is presented in the following sections.

Brainstem auditory evoked potentials

BAEPs will be recorded monaurally using constructed chirps, which have been shown to generate larger neural responses compared with standard clicks.[37 38] Five intensity levels will be tested.[38] BAEPs will also be recorded using clicks at three stimulation rates.[39] The signal will be bandpass-filtered (100–3000 Hz) and recorded differentially by using an electrode placed on Fz and Cz, in addition to an electrode on the ipsilateral mastoid (M1 or M2), while the electrode Fpz will be used as ground. The BAEPs will be obtained by averaging noise-free signals. This recording will span 5 min. Absolute amplitudes and latencies of the most salient response peak-waves I, III and V and interpeak intervals will be analysed.

Middle latency auditory evoked potentials

MLAEPs will be triggered by monaural clicks with a stimulus rate of 11.3/s. Stimuli intensities and scalp recordings will be similar to the BAEPs, except that the EEG will be bandpass-filtered between 10 kHz and 3 kHz. This recording will span 5 min. Latencies and amplitudes of Pa, Na and Pb waves will be analysed. This recording will span 5 min.

MMN paradigm

Tone and deviant sound features for this task were chosen according to a previous study,[40] and 2400 tones at a regular 450 ms SOA will be recorded. The sequence will comprise 82% standard tones (50 ms, 1000 Hz, 80 dB sound pressure level (SPL)) and three types of deviant tones (6% each): a duration deviant (125 ms), a frequency deviant (1200 Hz) and an intensity deviant (90 dB SPL). The paradigm will last 25 min.

P300 active paradigm

To investigate a patient's ability to follow instructions, we adapted the previous MMN paradigm by adding a task of

identifying standard tones from deviant tones. The number of stimuli was reduced to 492 standard tones and 108 deviant tones (36 per deviant type). Interstimulus interval (ISI) was extended to 1000 ms. Patients will be asked to imagine moving one of their feet every time they hear a sound other than the frequent (standard) tone. Instructions will be played through the earphones automatically before each instance of the paradigm. The paradigm will last 10 min.

Oddball mismatch

This paradigm combines an oddball task of standard and deviant tones with the SON as previously tested in comatose patients[13 21] and additional novel sounds. Stimuli consist of standard (80%) and deviant (11%) tones, a familiar novel sound (SON; 3%), unfamiliar novel sounds that carried no linguistic content (eg, a dog bark, a doorbell; 3%) and non-salient other words (NSOW; 3%). A different unfamiliar novel sound will be delivered every time (60 in total) in order to maintain its novelty. NSOW stimuli will be matched to the SON stimuli in terms of syllable number. For instance, John (SON) would be matched with boat (NSOW). Tones will be digitally generated sine waves of 800 Hz, with a standard tone duration of 75 ms and a deviant tone duration of 30 ms. The familiar novel (SON) will be synthesised by using the ReadSpeaker software (www.readspeaker.com), simulating a female native speaker of American English in a neutral voice. Stimuli will be presented pseudorandomly (no deviant or novel stimulus will be preceded by less than two standard tones) in one block of 2000 items with a stimulus onset asynchrony (SOA) for the tones being 800 ms and 1220 ms for the novels. The paradigm will span a total of 25 min.

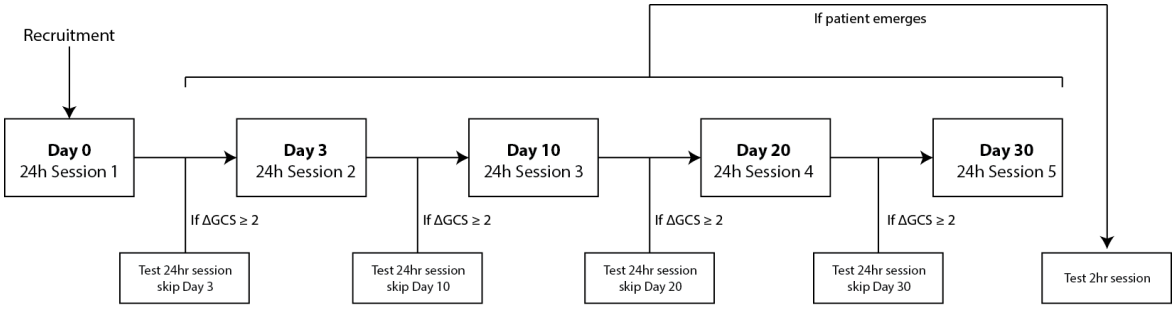
N400 paradigm

This task is an adaptation of a previous terminal-word semantic violation paradigm[41] and consists of 120 sentences recorded from natural speech. All sentences are digital recordings of a speaker of Canadian English reading in a neutral voice. Sixty sentences will have a semantically congruent terminal word (eg, *He drinks his coffee with cream and sugar*) and 60 sentences will have an incongruent terminal word (eg, *The pizza is too hot to sing*). ERPs will be recorded to the onset of the terminal word. Sentences will be presented using an ISI of 4 s. The paradigm will last 20 min.

Resting state

The EEG in RS conditions will be recorded (ie, when no ERP paradigm is being performed). RS intervals of 10 min will be interspersed with the previously mentioned paradigms. Additionally, EEG will be recorded between the main study blocks (see below), yielding large RS recordings that vary in length across sessions and patients.

Patient recordings will span a full day (24 hours). A day's testing will be composed of two types of assessment blocks: A blocks and B blocks. Blocks of type A contain a full GCS-P[6], BAEPs and MLAEPs, with an approximated total time of 20 min. Blocks of type B contain RS and ERP paradigms in the following order: RS, MMN, RS, N400, RS, oddball paradigm, RS, P300 (figure 1). The total time for a single iteration of a B block is 2 hours. An assessment



Block Types

	Paradigm	Duration (min)	Total
Block A	GCS-P	10	20 mins
	BAEPs	5	
	MLAEPs	5	
Block B	RS	10	120 mins
	MMN	25	
	RS	10	
	N400	20	
	RS	10	
	Oddball Mismatch	25	
	RS	10	
	P300	10	

Session Types

	Block order	Duration (hours)
24-hour session	A-B-B-B-B-A-B-B-B-B-A	17 (+ RS padding)
2-hour session	A-B	2

Figure 1 Protocol design starting at recruitment and consenting and including the sequence of a maximum of five 24-hour electroencephalogram sessions. BAEPs, brainstem auditory evoked potentials; GCS, Glasgow Coma Scale; GCS-P, Glasgow Coma Scale-Pupils; MLAEPs, middle latency auditory evoked potentials; MMN, mismatch negativity; RS, resting-state.

will incorporate the following blocks in order: A, B, B, B, B, A, B, B, B, B, A with a total time of 17 hours (see figure 1). Sufficient time is left to facilitate flexibility around clinical staff's schedule and to allocate time for reapplication of the EEG equipment if necessary. RS will be recorded during all instances where EEG is set up, but no block is in effect. In the case of a positive outcome, a patient will proceed through a single assessment consisting of a single block A followed by a single block B. Given the continuous loop of paradigms, we argue that counterbalancing will not be pertinent to the results of this study. Prior pilot work has shown no evidence of carry-over effects in the current sequence, likely due to the 10 min resting state between any two paradigms. Moreover, we argue that if there is an observable carry-over effect, it would be beneficial in detecting cognitive function in comatose patients. The decision to use static ordering is to simplify the application of a tool that will be running for 24 hours with possible interruptions.

Data from healthy controls will be recorded in a manner similar to the outlined design above. Controls will be run through a single assessment spanning 15 hours starting with the BAEP and MLAEP paradigms and followed by six B blocks. Sufficient breaks will be provided as between blocks. Although it could be stated that this control group may not be optimal, we argue that for the purposes of the present study—to develop a clinical tool capable of predicting a patient's outcome—a comparison with a typical healthy, conscious response is the most appropriate. In a basic research study examining the nature of consciousness, comparison of anaesthetised patients with those in a comatose state would be appropriate (see ref [42]); however, this is not the objective of the present project.

Data preprocessing of ERPs will be conducted using Brain Vision Analyzer V.2.1 (Brain Products). All recordings will be visually inspected and epochs containing artefacts (eg, muscle activity, movements) removed. Individual task recordings will be filtered offline with a bandpass of 0.1–30 Hz. Ocular artefacts will be corrected by using the ocular independent component analysis (ICA) transformation. Recordings will be segmented and baseline-corrected into epochs depending on the component of interest in each task: for epochs containing the MMN (in MMN paradigm), 100 ms prestimulus to 600 ms poststimulus; for the P300 (in o*ddball mismatch* and *P300 active* paradigms) and the N400 (in *N400* paradigm), 100 ms prestimulus to 1000 ms poststimulus.

One of the primary outcomes is to assess correlations of measures of coma emergence between behavioural and electrophysiological data. Emergence will be assessed behaviourally by the GCS[5] and the Glasgow Outcome Scale (GOS).[43] The GOS globally rates the functional outcome for patient status into one of five categories: dead, vegetative state, severe disability, moderate disability or good recovery. The GOS result will be confirmed by an attending physician. The study will also collect aetiology, demographics and medical history. Premorbid and baseline data will be collected at the first assessment. Patients' progress information will be collected throughout the study period.

Statistical analyses and ML

The ultimate goal of the study is the creation of an automated prototype for coma prognosis. As ML is a data-driven approach, the full application and viability of the created model(s) are expected to improve over the course of the present study due to more training data being available. Several interim analysis stages are to take place after successful recording of data from 10 patients, forming five main analysis points, the last of which takes place after the termination of data collection.

A two-step process will be carried out to detect the ERPs in each single case. First, in accordance with work demonstrating that visual inspection is still a field standard,[13] visual identification of the components for each of the recorded paradigms and blocks will be performed. Second, the presence of components will be objectively confirmed by conducting corrected one-tailed serial t-tests across time to find the intervals where the deviant or incongruent condition is significantly more negative (eg, MMN or N400 component) or positive (eg, P300 component) than the standard or congruent condition.

The process described above will be used on all recorded paradigms and neurophysiological measures to create a hierarchical view on a patient's cognitive ability during coma. The outlined analysis will be used to provide preliminary labels for training ML models. Comparative data from healthy controls will be analysed to (1) establish a common baseline response to each paradigm and (2) investigate ERP within-subject differences across time (different blocks). Similarly, RS EEG will be analysed across different blocks and groups (controls vs coma vs emerged) to statistical differences in the candidate features (described below).

The supervised ML process involves two fundamental phases: training and testing. In the training phase, we assume the availability of a large quantity of training data. The principle behind ML methods is to use mathematically structured processes to compare and thus classify patterns from previously unseen samples with patterns that occur in the training set, whose respective classes are known. A supervised learning process first involves extracting various candidate features from the measured EEG data. Candidate features include power spectral density and fractal dimension,[44 45] along with other features relating to brain connectivity. These include phase-lag index, symbolic transfer entropy, directed phase-lag index and Granger causality.[42 46–49] Note that candidate features are computed from uniformly sized blocks of RS EEG or from averages of stimulus-locked trials within a block of ERP data. Block size will be optimised in the cross-validation process. The feature extraction procedure is followed by a feature selection process,[50] which identifies only the most relevant of the candidate features. The features are selected by a (supervised) feature selection algorithm to optimally discriminate between the classes, as in ref [50]. The next stage is classification, which may be described as a mathematical

function that maps distinct regions of the feature space into the respective classes. The last procedure in the training process is validation, which assesses performance of the resulting structure, often on an unseen subset of the training data. Validation is often used to empirically optimise hyperparameters used in the utilised model and/or feature extraction procedures (eg, block size). Accepted methods of validation include *leave one subject out* and *bootstrap validation*.[51] When executed properly, validation reduces the likelihood of potential pitfalls in the ML framework. Lastly, the test, or operational phase, uses the trained ML model to assign a class to new, previously unseen data points.

The study plan is to determine the features that are most relevant to prognosis for emergence for both RS and EEG/ERP data, guided by statistical results and pre-established work. These most relevant features can be used to provide clues in understanding the neurological mechanisms of coma and emergence from coma. Features extracted from single blocks, assessments and/or patients will be used to train a model to differentiate between patient outcomes (emergence vs no emergence). Identifying a formal correspondence between the frequency of detected ERP elicitations, if any, and patient outcome is critical to extend the finding that ERPs wax and wane during coma.[18 35] For model training, different classifiers will be used; notably, the local feature selection algorithm, which performs feature selection and classification, has been shown to yield promising results in a project precursor to the present study.[34] Other classifiers with demonstrated utility in EEG and other imaging techniques will also be examined for the present application (eg, random forests,[52] Gaussian processes[53] and support vector machines[32 54]).

In summary, our analysis plan for proceeding towards a prototype clinical model includes undertaking the following tasks: (1) statistical analysis and visual inspection of patient and control ERP data to establish labels for ML; (2) leverage results from (1) to extract and select an optimised feature-set for ML; 3) training, validation and testing iterations towards an optimised classification model for coma outcome prediction; and (4) implementation of an integrated prototype system using the model from (3) for use in the clinical setting.

ETHICS AND DISSEMINATION

Consent

The substitute decision maker of the eligible patient will first be informed about the study by the healthcare providers within the immediate circle of care. If the substitute decision maker is interested, the trained study staff will explain the study in more detail and will obtain written consent from the substitute decision maker (or verbal consent if the consent is to be done by phone). The healthy control participants will provide their own consent.

Confidentiality

The patient's data will not be shared with anyone except with the consent from the substitute decision maker or as required by law. All will be de-identified and assigned a patient number. A list linking the number with personal information will be kept in a secure place.

Author affiliations
[1]School of Biomedical Engineering, McMaster University, Hamilton, Ontario, Canada
[2]Vector Institute, MaRS Discovery District, Ontario, Canada
[3]Department of Psychology, Neuroscience & Behaviour, McMaster University, Hamilton, Ontario, Canada
[4]ARiEAL Research Centre, McMaster University, Hamilton, Ontario, Canada
[5]Department of Linguistics and Languages, McMaster University, Hamilton, Ontario, Canada
[6]Neuroscience Graduate Program, McMaster University, Hamilton, Ontario, Canada
[7]Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, Canada
[8]Department of Medicine, McMaster University, Hamilton, Ontario, Canada
[9]Critical Care Medicine, Hamilton Health Sciences, Ontario, Canada
[10]Brain Vision Solutions, Montreal, Quebec, Canada
[11]School of Physical and Occupational Therapy, McGill University, Montreal, Quebec, Canada
[12]Department of Surgery, McMaster University, Hamilton, Ontario, Canada

Acknowledgements The authors would like to acknowledge the efforts of Ms Chia-Yu Lin for her help in the preparation of the ethics submission and of the CHRP application. Additionally, the authors extend their gratitude to Dr Richard Kolesar for his invaluable input on refining the study parameters related to sedation.

Contributors JFC, JPR and AF-R conceived of the study. JFC, JPR, AH-D and RB designed the study protocol. RB and AH-D wrote the first draft of the paper. All authors revised the manuscript for relevant scientific content. JPR and RB specifically revised the machine learning design and validation sections of the paper. PB, CH, SB-M and RS provided an advisory role on various aspects of the paper. All authors approved the final version of the manuscript.

Funding This study is funded by the Collaborative Health Research Projects (CHRP) opportunity made possible by the Canadian Institutes of Health Research (CIHR; grant number CPG158287) and the Natural Sciences and Engineering Research Council of Canada (NSERC; grant number CHRP 523461-18) and NSERC Discovery Grant (RGPIN-2016-06633). Inkind support towards equipment, software licences and scientific consultation is provided by Brain Vision Solutions. Inkind support for patient identification and research assessment is provided by Hamilton Health Sciences (HHS). The design, management, analysis and reporting of the study are entirely independent of the funders.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval This protocol has been reviewed and approved by the Hamilton Integrated Research Ethics Board (HiREB) under project number 4840.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. Young GB, Ropper AH, Bolton CE. *Coma and impaired consciousness: a clinical perspective*. New York: McGraw Hill, 1998.
2. Giacino JT, Fins JJ, Laureys S, *et al*. Disorders of consciousness after acquired brain injury: the state of the science. *Nat Rev Neurol* 2014;10:99–114.
3. Bates D. The prognosis of medical coma. *J Neurol Neurosurg Psychiatry* 2001;71 Suppl 1:i20–i23.

4. Daltrozzo J, Wioland N, Mutschler V, *et al*. Predicting coma and other low responsive patients outcome using event-related brain potentials: a meta-analysis. *Clin Neurophysiol* 2007;118:606–14.
5. Jones C. Glasgow coma scale. *Am J Nurs* 1979;79:1551–3.
6. Brennan PM, Murray GD, Teasdale GM. Simplifying the use of prognostic information in traumatic brain injury. Part 1: The GCS-Pupils score: an extended index of clinical severity. *J Neurosurg* 2018;128:1612–20.
7. Chiappa KH, Hill RA. Evaluation and prognostication in coma. *Electroencephalogr Clin Neurophysiol* 1998;106:149–55.
8. de Sousa LC, Colli BO, Piza MR, *et al*. Auditory brainstem response: prognostic value in patients with a score of 3 on the Glasgow Coma Scale. *Otol Neurotol* 2007;28:426–8.
9. Logi F, Fischer C, Murri L, *et al*. The prognostic value of evoked responses from primary somatosensory and auditory cortex in comatose patients. *Clinical Neurophysiology* 2003;114:1615–27.
10. Vanhaudenhuyse A, Laureys S, Perrin F. Cognitive event-related potentials in comatose and post-comatose states. *Neurocrit Care* 2008;8:262–70.
11. Lew HL, Poole JH, Castaneda A, *et al*. Prognostic value of evoked and event-related potentials in moderate to severe brain injury. *J Head Trauma Rehabil* 2006;21:350–60.
12. Kane NM, Butler SR, Simpson T. Coma outcome prediction using event-related potentials: P(3) and mismatch negativity. *Audiol Neurootol* 2000;5:186–91.
13. Morlet D, Fischer C. MMN and novelty P3 in coma and other altered states of consciousness: a review. *Brain Topogr* 2014;27:467–79.
14. Fischer C, Morlet D, Bouchet P, *et al*. Mismatch negativity and late auditory evoked potentials in comatose patients. *Clin Neurophysiol* 1999;110:1601–10.
15. Garrido MI, Kilner JM, Stephan KE, *et al*. The mismatch negativity: a review of underlying mechanisms. *Clin Neurophysiol* 2009;120:453–63.
16. Holeckova I, Fischer C, Morlet D, *et al*. Subject's own name as a novel in a MMN design: a combined ERP and PET study. *Brain Res* 2008;1189:152–65.
17. Sonnadara RR, Alain C, Trainor LJ. Occasional changes in sound location enhance middle latency evoked responses. *Brain Res* 2006;1076:187–92.
18. Armanfard N, Komeili M, Reilly JP, *et al*. A Machine Learning Framework for Automatic and Continuous MMN Detection with Preliminary Results for Coma Outcome Prediction. *IEEE J Biomed Health Inform* 2018:1.
19. Dykstra AR, Cariani PA, Gutschalk A. A roadmap for the study of conscious audition and its neural basis. *Philos Trans R Soc Lond B Biol Sci* 2017;372:20160103.
20. Duncan CC, Barry RJ, Connolly JF, *et al*. Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clin Neurophysiol* 2009;120:1883–908.
21. Fischer C, Dailler F, Morlet D. Novelty P3 elicited by the subject's own name in comatose patients. *Clin Neurophysiol* 2008;119:2224–30.
22. Holeckova I, Fischer C, Giard MH, *et al*. Brain responses to a subject's own name uttered by a familiar voice. *Brain Res* 2006;1082:142–52.
23. Daltrozzo J, Wioland N, Mutschler V, *et al*. Cortical information processing in coma. *Cogn Behav Neurol* 2009;22:53–62.
24. Kotchoubey B, Daltrozzo J. *Semantic processing in a coma patient Classical conditioning effects in EEG View project Sleep in Disorders of Consciousness View project*, 2005.
25. Rämä P, Relander-Syrjänen K, Ohman J, *et al*. Semantic processing in comatose patients with intact temporal lobes as reflected by the N400 event-related potential. *Neurosci Lett* 2010;474:88–92.
26. Güler I, Ubeyli ED, Deryä Ubeyli E. Multiclass support vector machines for EEG-signals classification. *IEEE Trans Inf Technol Biomed* 2007;11:117–26.
27. Boshra R, Dhindsa K, Boursalie O, *et al*. From Group-Level Statistics to Single-Subject Prediction: Machine Learning Detection of Concussion in Retired Athletes. *IEEE Trans Neural Syst Rehabil Eng* 2019;27:1492–501.
28. Cao C, Tutwiler RL, Slobounov S. Automatic classification of athletes with residual functional deficits following concussion by means of EEG signal using support vector machine. *IEEE Trans Neural Syst Rehabil Eng* 2008;16:327–35.
29. Ravan M, Hasey G, Reilly JP, *et al*. A machine learning approach using auditory odd-ball responses to investigate the effect of Clozapine therapy. *Clin Neurophysiol* 2015;126:721–30.
30. Khodayari-Rostamabad A, Reilly JP, Hasey GM, *et al*. A machine learning approach using EEG data to predict response to SSRI treatment for major depressive disorder. *Clin Neurophysiol* 2013;124:1975–85.
31. Engemann D, Raimondo F, King J-R, *et al*. Automated measurement and prediction of consciousness in vegetative and minimally conscious patients. https://hal.inria.fr/hal-01225254 (Accessed 19 Apr 2019).
32. Sitt JD, King JR, El Karoui I, *et al*. Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state. *Brain* 2014;137:2258–70.
33. King JR, Faugeras F, Gramfort A, *et al*. Single-trial decoding of auditory novelty responses facilitates the detection of residual consciousness. *Neuroimage* 2013;83:726–38.
34. Armanfard N, Reilly JP, Komeili M. Local Feature Selection for Data Classification. *IEEE Trans Pattern Anal Mach Intell* 2016;38:1217–27.
35. Armanfard N, Komeili M, Reilly JP, *et al*. Automatic and continuous assessment of ERPs for mismatch negativity detection. *Conf Proc IEEE Eng Med Biol Soc* 2016;2016:969–72.
36. Kanich W, Brady WJ, Huff JS, *et al*. Altered mental status: evaluation and etiology in the ED. *Am J Emerg Med* 2002;20:613–7.
37. Elberling C, Callø J, Don M. Evaluating auditory brainstem responses to different chirp stimuli at three levels of stimulation. *J Acoust Soc Am* 2010;128:215–23.
38. Elberling C, Don M. A direct approach for the design of chirp stimuli used for the recording of auditory brainstem responses. *J Acoust Soc Am* 2010;128:2955–64.
39. Stone JL, Calderon-Arnulphi M, Watson KS, *et al*. Brainstem auditory evoked potentials--a review and modified studies in healthy subjects. *J Clin Neurophysiol* 2009;26:167–75.
40. Todd J, Michie PT, Schall U, *et al*. Deviant matters: duration, frequency, and intensity deviants reveal different patterns of mismatch negativity reduction in early and late schizophrenia. *Biol Psychiatry* 2008;63:58–64.
41. Connolly JF, Phillips NA. Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *J Cogn Neurosci* 1994;6:256–66.
42. Blain-Moraes S, Boshra R, Ma HK, *et al*. Normal brain response to propofol in advance of recovery from unresponsive wakefulness syndrome. *Front Hum Neurosci* 2016;10:1–6.
43. Jennett B, Bond M. Assessment of outcome after severe brain damage a practical scale. *Lancet* 1975;305:480–4.
44. Shaw NA. The neurophysiology of concussion. *Concussion* 2012;67:28–37.
45. Siamaknejad H, Loo CK, Liew WS. Fractal dimension methods to determine optimum EEG electrode placement for concentration estimation. In: 2014 Joint 7th International Conference on Soft Computing and Intelligent Systems (SCIS) and 15th International Symposium on Advanced Intelligent Systems (ISIS). *IEEE* 2014:952–5.
46. Stam CJ, Nolte G, Daffertshofer A. Phase lag index: Assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. *Hum Brain Mapp* 2007;28:1178–93.
47. Lee U, Blain-Moraes S, Mashour GA. Assessing levels of consciousness with symbolic analysis. *Philos Trans A Math Phys Eng Sci* 2015;373:20140117.
48. Blain-Moraes S, Lee U, Ku S, *et al*. Electroencephalographic effects of ketamine on power, cross-frequency coupling, and connectivity in the alpha bandwidth. *Front Syst Neurosci* 2014;8:114.
49. Barrett AB, Murphy M, Bruno MA, *et al*. Granger causality analysis of steady-state electroencephalographic signals during propofol-induced anaesthesia. *PLoS One* 2012;7:e29072.
50. Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Anal Mach Intell* 2005;27:1226–38.
51. Hastie T, Robert T, Friedman J. *The elements of statistical learning: data mining, inference and prediction*. 2nd edn. New York: Springer Series in Statistics.
52. Fraiwan L, Lweesy K, Khasawneh N, *et al*. Automated sleep stage identification system based on time-frequency analysis of a single EEG channel and random forest classifier. *Comput Methods Programs Biomed* 2012;108:10–19.
53. Tzovara A, Murray MM, Plomp G, *et al*. Decoding stimulus-related information from single-trial EEG responses based on voltage topographies. *Pattern Recognit* 2012;45:2109–22.
54. Demertzi A, Antonopoulos G, Heine L, *et al*. Intrinsic functional connectivity differentiates minimally conscious from unresponsive patients. *Brain* 2015;138:2619–31.