A multimodal biomarker for concussion identification, prognosis and management

Arnaud Jacquin, Saloni Kanakia, Doug Oberly, Leslie S. Prichep

Algorithm Development, BrainScope Co., Inc, Bethesda, MD, United States
Clinical Affairs, BrainScope Co, Inc, Bethesda, MD, United States
Department of Psychiatry, NYU School of Medicine, New York, NY, United States

ABSTRACT

Background: Prompt, accurate, objective assessment of concussion is crucial, particularly for children/adolescents and young adults. While there is currently no gold standard for the diagnosis of concussion, the importance of multidimensional/multimodal assessments has recently been emphasized.

Methods: Concussed subjects (N = 177), matched controls (N = 187) and healthy volunteers (N = 204) represented a convenience sample of male and female subjects between the ages of 13 and 25 years, enrolled at 29 Colleges and 19 High Schools in the US. Subjects were tested at time of injury and at multiple time points during recovery. Assessments included EEG, neurocognitive tests and standard concussion assessment tools.

Multimodal classifiers to maximally separate controls from concussed subjects with prolonged recovery (≥ 14 days) were derived using quantitative EEG, neurocognitive and vestibular measures, informed feature reduction and a Genetic Algorithm methodology for classifier derivation. The methodology protected against overtraining using an internal cross-validation framework. An enhanced multimodal Brain Function Index (eBFI) was derived from the classifier output and mapped to a percentile scale which expressed the index relative to non-injured controls.

Results: At time of injury eBFIs were significantly different between controls and concussed subjects with prolonged recovery, showing return to non-concussed levels at return-to-play plus 45 days. For the combined concussed population, and for the short recovery subjects, a more rapid recovery was seen.

Conclusions: This multivariate, multimodal, objective index of brain function impairment can potentially be used, along with other tools, to aid in diagnosis, assessment, and tracking of recovery from concussion.

1. Introduction

Traumatic brain injury (TBI) accounts for over 1.5 million emergency department (ED) visits annually within the United States and the vast majority of these visits are for mild injury (mTBI/concussion) [1]. In addition, the Center for Disease Control (CDC) estimates that 1.6 to 3.8 million concussions occur in sports and recreational activities annually. These figures are likely to significantly underestimate the TBI burden since many patients with mild head injuries do not seek medical care [2]. Results from a recent (2017) large scale (N = 14,765 students), voluntary Youth Risk Behavior Survey (YRBS) performed by the CDC report that approximately 15.1% of US high school students (approximately 2.5 million) reported having had at least one concussion in the past 12 months, 6% reported having two or more, and this number rises to as high as 30.3% in those who play on three or more sports teams per year [3].

Prompt, accurate, objective assessment of concussion is crucial, particularly for children/adolescents and young adults, as delayed diagnosis has been shown to lead to prolonged recovery [4,5] (reported to be as much as twice as long [4]) and may affect a child’s academic/cognitive and emotional functioning [6–8]). Further, with important brain development continuing into early adulthood, sustaining a TBI injury before age 25 has been shown to be associated with impaired adult functioning, psychiatric disorders, low education level, welfare, and disability [9].

While a number of tools are currently used by team doctors and athletic trainers to determine whether a player has suffered a sport-related concussion (SRC), none of them is considered in itself to be a gold standard for the diagnosis of concussion or adequate to assess recovery [10]. Furthermore, existing tools are largely subjective in
nature and/or suffer from poor test-retest reliability (replicability) and learning effects when used longitudinally [11].

Changes in brain electrical activity (EEG) that occur in TBI have been reported and use of such measures for the classification and identification of head injured patients has been demonstrated to be sensitive to both structural and functional brain injury [12–22]. In a study comparing diffusion tensor imaging (DTI) and electroencephalography (EEG) in blast-concussed soldiers, Sponheim and colleagues [23] reported a significant correlation between changes in mean fractional anisotropy (FA) of four major white matter tracts related to frontal interhemispheric communication and changes in phase synchrony of the EEG between frontal and frontotemporal regions, both reflecting disruption in neural transmission between brain regions. Other EEG measures reflect changes related to the physiology of mTBI/concussion including measures of changes in entropy, a measure of “complexity” reflecting disorganization of neural networks with concussive injury [24]. Changes in the frequency spectra of the EEG, power relationships, and connectivity (e.g., coherence) between regions and between neural networks have also been associated with concussive injury [14,25].

The Brain Function Index (BFI) described by Hanley and colleagues (2018) is a multivariate weighted measure that is derived from quantitative EEG features that reflect changes in brain electrical activity related to the physiology of concussive injury. In a large prospective independent FDA validation trial, the BFI was demonstrated to scale significantly with severity of clinical functional impairment in mTBI/ concussed patients within the first 3 days of head injury [19]. This objective biomarker of functional brain impairment, expressed as a percentile of the non-injured normal population is based on EEG features only. The study by Brooks et al. [21] demonstrates the potential clinical utility of the BFI in a population of athletes with a concussion studied longitudinally.

The importance of multimodal assessments of concussion, which evaluate several domains, including: somatic, cognitive, behavioral/emotional symptoms, and physical signs (e.g. vestibular-ocular deficits, LOC) has been emphasized in the recent literature [26]. Guidelines for concussion diagnosis have advanced from reliance on subjective symptom checklists and single assessment modalities to current support for multimodal assessments, as highlighted in the consensus statement from the 5th International Conference on Concussion in Sport held in Berlin, October 2016 [27]. Likewise, numerous governing bodies (e.g. the American Academy of Neurology (AAN), the National Athletic Trainers’ Association) have advocated a multidimensional approach to sport concussion management, consisting of computerized neurocognitive testing (CNT), assessment of postural stability (balance) and self-reported symptoms [28].

In light of the recently growing literature supporting multimodal evaluation for the evaluation of mTBI/concussion [26–28], this paper will describe the derivation and performance of an enhanced BFI (eBFI) which includes EEG and neurocognitive performance measures as well as selected symptoms in a multivariate index. This index will be shown to be sensitive to the presence of concussive brain injury at time of injury, trackable longitudinally, and to eventually return to baseline. Embedded in a hand-held device, this index could be used to rapidly, objectively, and reliably evaluate head injured patients.

2. Methods

2.1. Patient population

Data was collected using prototypes of the BrainScope1 Ahead Concussion Assessment System (CAS) at 42 sites (23 Colleges, 19 High Schools) across the US,2 with approval from local Institutional Review

---

1 BrainScope Company, Inc. Bethesda, Maryland.

2 College sites included: Texas Southern University, Texas A&M Prairie (Rice University); Le Moyne College, Onondaga Community College (SUNY; Syracuse University); Eastern Connecticut State University, University of Hartford, Southern CT State University (University of Connecticut); Columbia College (University of South Carolina); University of Texas – Austin; John Brown University (University of Arkansas); Michigan State University; St John Fisher College, Robert Wesleyan College, Nazareth College, Rochester Institute of Technology, The College at Brockport, SUNY (University of Rochester); University of South Florida. High School sites included: Walt Whitman High School; Rogers HS, Rogers-Heritage HS, Farmington HS, Siloam Springs HS, Gentry HS (University of Arkansas); Mason HS, Grand Ledge HS, St. John HS, Sexton HS, Everett/Perry HS, Portland HS, East Lansing HS, Waverly HS, Bath HS, Ionia HS, Williamson HS, Everett HS, Haslett HS, Holt HS (Michigan State University).
differentiate between groups are candidates for selection.

2.2. RTP information

The determination of RTP information for each subject (reported as a number of days after injury) was made as per the local return to play protocol used at each site. The criteria for RTP were not part of the protocol of the study as each site was obligated to follow the guidelines of their own institution. As such, the data is more representative of standard clinical practice. However, all the sites used a staged/graduated return to play protocol.

2.3. Clinical assessments

All study subjects were evaluated with the following symptom based scales or assessment tools: 1) Sports Concussion Assessment Tool – 3rd Edition (SCAT-3) [29,30]: Symptom Evaluation (Section 3 of that document, which assesses the presence and severity of 22 common concussion symptoms, using a Likert scale to assess symptom severity (range 0–6 per item, with a total score range of 0–122); and 2) Standard Assessment of Concussion (SAC, Section 4 and 8): a brief cognitive screening tool which includes brief subtests of orientation, immediate memory, concentration, and delayed recall (total score of 0–30). In addition, all measures which were components of the New Orleans Criteria (NOC) for referral of head injured patients for CT scans were collected by trained research assistants. These included: headache, vomiting, age > 60 years, drug or alcohol intoxication, persistent anterograde amnesia, visible trauma above the clavicle, or seizure.

Using this information in consultation with emergency medicine and sports medicine physicians, and in conjunction with published guidelines [31,32], the subjects were divided into three clinical categories for the purpose of assessing performance of the classifiers by severity of clinical functional impairment. The first subject category (referred to as “Controls” (Ctl)) corresponds to subjects who were functionally normal controls, category 2 (referred to as Cx2) subjects had mild or moderate functional/concussion symptoms with RTP under 14 days, and category 3 (referred to as Cx3) subjects had mild or moderate concussion symptoms with RTP of 14 days or more ("prolonged/protracted" recovery group). Since this study was focused on building a multimodal brain function (mTBI/concussion) index which maximally separates controls from concussed subjects, the second category was only used for testing but was not used for training, i.e. the classifier was developed for Cx3 ("concussed subjects with prolonged recovery") versus Ctl (not head injured). The age, gender and race distributions of the sample were determined by the representation of each in the populations served by the participating sites involved.

2.4. Neurocognitive data acquisition

A rapid cognitive performance assessment module developed by Vista Life Sciences was used to perform a small battery of neurocognitive tests performed on Surface Pro 3 Tablet. This rapid cognitive performance assessment module was used to assess cognitive performance. The assessment included the following tests (subtests of the ANAM battery [33]), which can be affected by TBI: Simple Reaction Time (SRT; evaluates simple motor speed, information processing speed and attention), Code Substitution Learning (CDS; evaluates associative learning and memory), Procedural Reaction Time (PRO; evaluates higher-order rapid responding including visuomotor reaction time and simple decision making), and Go/No-Go (GNG; evaluates sustained attention and impulsivity/inhibition) [33–35]. Among the raw features output by the assessment module (speed, percent accuracy, throughput), only the “throughputs” were used, which are composite scores which jointly reflect both speed and accuracy. The raw neurocognitive features were age-regressed and z-transformed to standard scores using a large normative database of subjects aged 13–35.

2.5. EEG data acquisition

Using handheld BrainScope investigational devices [18,19], five to ten minutes of EEG data was acquired under “eyes closed resting” conditions. A technician observed the subject throughout data acquisition and alerted him/her if necessary. The EEG data was recorded from the limited frontal electrode montage consisting of the Fp1, Fp2, F7, F8, AFz, A1 and A2 locations of the expanded International 10–20 Electrode Placement System, and was re-referenced to linked ears. The EEG data was acquired at a sampling rate of 1 kHz and all electrode impedances were below 10 kΩ. Amplifiers had a band pass filter from 0.3 to 250 Hz (3 dB points).

2.6. EEG data processing

EEG recordings downsampled to 100 Hz were processed using BrainScope’s algorithms for artifact detection [16] which identify for removal any physiologic and non-physiologic (i.e. from external sources) contamination, including lateral and horizontal eye movement, muscle activity (EMG), impulse artifacts, extremely low amplitude EEG activity, and atypical electrical activity. In prior studies this artifact detection showed excellent agreement (87.6%) between an experienced EEG technologist selecting artifact-free EEG segments and the automatic artifact algorithms performing this task. Previous experience has shown that sufficient artifact-free data (50–120 s; representing 20–48 epochs of length 2.56 s) can be obtained from these five to ten-minute long EEG recordings in such a population. Only artifact-free data was submitted to all further analyses of the EEG data.

2.7. Quantitative EEG (QEEG) feature extraction

The power spectrum of the artifact-free EEG data was computed using the Fast Fourier Transform (FFT) in order to extract QEEG features of absolute and relative (%) power, mean frequency, inter- and intra-hemispheric coherence and asymmetry, computed in the frequency bands: delta (1.5–3.5 Hz), theta (3.5–7.5 Hz), alpha (7.5–12.5 Hz), beta (12.5–25 Hz) and gamma (30–45 Hz). The computation of these traditional features is described in Ref. [16]. In addition, several additional features were computed, including: (a) chaotic/fractal measures (fractal dimension and scale-free activity) [36], which evaluate the global signal complexity of the brain electrical activity at each electrode location across the total spectrum; (b) measures based on information theory (entropy and wavelet entropy), which evaluate the degree of order/disorder of the brain electrical activity at each electrode location; and (c) connectivity measures (phase lag, phase synchrony, and various ratios of spectral power and coherence computed across regions) which evaluate relationships between and among cortical regions [37,38]. All these features were transformed to obtain a Gaussian distribution and z-transformed relative to age expected normal values following the QEEG methodology described in Refs. [39,40]. The importance of these steps in enhancing the sensitivity and specificity of brain electrical activity, along with its test-retest reliability and replicability has been described in detail in Ref. [41]. The resulting database of QEEG feature z-scores is referred to in this paper as the “algorithm development database”. At this stage, a final data quality check algorithm was applied to identify non-EEG data contamination based on distribution characteristics of features outside the
frequency range of interest that were not identified by the artifact algorithms. Recordings which were deemed to be affected by such contamination were eliminated from the database (approximately 3% of the total number of recordings).

2.8. QEEG feature reduction

Following the step of feature extraction, the algorithm development database contained several thousand features (\(M = 10,308\)). The problem which can arise from the availability of a very large number of features and the consequential need for data reduction are common in quantitative electrophysiology as well as in machine learning problems in general. In these cases, the large datasets involved have a number of features which is much greater than the number of subjects/recordings. In such cases, an exhaustive search of the large feature space for an optimal subset of features becomes both computationally prohibitive and statistically limited and can lead to overtraining, which in turn typically leads to poor performance on independent populations. “Informed feature reduction,” described in detail in Ref. [16], was therefore performed in order to retain only those features that are stable, replicable, physiologically meaningful, and show good separation between the two subject classes (controls and concussed subjects). The reduced feature set following this feature reduction step became the candidate feature pool for classifier building.

2.9. Multimodal classifier development

Classifier development involved searching the reduced dimensionality multimodal feature space using advanced machine learning-based techniques in order to obtain classifier candidates that can optimally separate the two categories (controls and concussed) while avoiding “overtraining”. The multimodality was represented by three broad types of quantitative features: QEEG, a small subset of neurocognitive throughput features, and a clinical sign/symptom feature reflecting vestibular/balance issues.

The dataset was first randomly split into an “All In” training set (AI; 80%) and holdout set not used in training (HO; 20%), but later used as an independent test group. The AI training dataset was further divided into 10 splits (each with train (80%) and test set (20%)) to permit internal cross-validation [42–44]. A Genetic Algorithm (GA) classifier builder (a type of machine-learning algorithm) was used to design

---

### Table 1

Descriptive statistics of subjects enrolled in the study at Day 0 (Algorithm Development “All In” (AI) + “Hold Out” (HO), Controls, Injured/Cx).

<table>
<thead>
<tr>
<th></th>
<th>Controls</th>
<th>Injured (all)</th>
<th>Injured (RTP &lt; 14)</th>
<th>Injured RTP ≥ 14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size (N; day 0)</td>
<td>187</td>
<td>177</td>
<td>75</td>
<td>102</td>
</tr>
<tr>
<td>Gender (#Female, #Male (% Female))</td>
<td>65, 122 (34.8%)</td>
<td>63, 114 (35.6%)</td>
<td>22, 53 (29.3%)</td>
<td>41, 61 (40.2%)</td>
</tr>
</tbody>
</table>

---

### Table 2

Prevalence of typical concussion signs/symptoms (Sx) at Day 0 in the different subject groups (Controls/Ctl, Injured/Cx). Cx2 are injured subjects with shorter recovery (RTP < 14); Cx3 are the injured subjects with prolonged recovery (RTP ≥ 14). “Visual Disturbances” specifically refers to blurred vision or double vision (diplopia).

<table>
<thead>
<tr>
<th></th>
<th>Loss of Consciousness (LOC)</th>
<th>Retrograde Amnesia (RGA)</th>
<th>Altered Mental Status (AMS)</th>
<th>Severe Headache (SHA)</th>
<th>Balance Problems</th>
<th>Dizziness</th>
<th>Visual Disturbances</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Ctls with symptom (Sx)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>3.2</td>
<td>3.2</td>
<td>0</td>
</tr>
<tr>
<td>% Cx2 with Sx</td>
<td>12</td>
<td>5.3</td>
<td>100</td>
<td>14.7</td>
<td>38.7</td>
<td>42.7</td>
<td>10.7</td>
</tr>
<tr>
<td>% Cx3 with Sx</td>
<td>8.8</td>
<td>2</td>
<td>100</td>
<td>27.5</td>
<td>46.1</td>
<td>58.8</td>
<td>12.7</td>
</tr>
<tr>
<td>% Cx (%) with Sx present</td>
<td>18 (10.2%)</td>
<td>6 (3.4%)</td>
<td>177 (100%)</td>
<td>39 (22%)</td>
<td>76 (42.9%)</td>
<td>92 (52%)</td>
<td>21 (11.9%)</td>
</tr>
<tr>
<td>Median RTP with Sx present</td>
<td>13.5</td>
<td>10.5</td>
<td>15</td>
<td>19</td>
<td>16.5</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>Median RTP with Sx absent</td>
<td>15</td>
<td>15</td>
<td>NA</td>
<td>14</td>
<td>15</td>
<td>13</td>
<td>15</td>
</tr>
</tbody>
</table>

**Fig. 1.** Receiver Operating Characteristic (ROC) curve for the GA classifier generated using the “All In” algorithm development database. The operating point shown on the ROC curve is for a target LOO specificity of 75%.
classifiers (Linear Discriminant Functions) based on the Training cross
correlation analysis. GA is an Evolutionary Algorithm which performs a
stochastic search involving randomness from one iteration to the next
and evaluates a series of candidate solutions, where each new candidate
is informed by high-performing predecessors, similar to genetic evolu-
tion [45–51]. The candidates’ performance was computed on the 10
splits of the training dataset in order to get an estimate of average
performance and variability across splits. Leave-One-Out (LOO) cross-
validation performance was used for feature selection using the reduced
feature pool as described above as well as for selecting the classification
threshold. This cross-validation method has been described in Statis-
tical Classification texts such as [52]. A target cross-validation specific-
ity was set to constrain the solution to acceptable ranges of sensitivity
and specificity in accordance with our tolerance for stratification of
risk. Because of the larger number of subjects on the negative side of the
classification task, a decision was made to constrain the specificity in-
stead of the sensitivity in order to reduce the chances of selecting an
operating point based on a spurious peak on the ROC curve [53].

The metrics used for evaluating classifier performance were the area
under the curve (AUC) of the ROC as well as sensitivity and specificity of
the classifiers at their operating point. ROC curves are useful for
visualizing classifier performance and the AUC is a simple to compute
scalar measure of classifier performance which is commonly used in
medical classification problems. At the end of this procedure, an “All-
In” classifier building run was conducted on the full algorithm devel-
oped database and was tested on the Hold Out (HO) dataset not used
in training the algorithm and performance reported on this set.

3. Results

Injured and matched control subjects (N = 364 at Day 0, 236 (65%)
males and 128 females (35%)) were enrolled at 42 sites (23 Colleges,
High Schools) across the US, with approval from local Institutional
Review Boards (IRBs). In addition, non-injured healthy volunteers (HV;
N = 204; 46.6% female, 53.4% male), i.e., an additional group of
males and 128 females (35%) were enrolled at 42 sites (23 Colleges, 19
High Schools) across the US, with approval from local Institutional
Review Boards (IRBs). In addition, non-injured healthy volunteers (HV;
N = 204; 46.6% female, 53.4% male), i.e., an additional group of
control subjects playing contact or non-contact sports meeting inclu-
sion/exclusion criteria, were assessed at pre-season (Baseline) and post-
season. The data was divided into two groups based on the presence or
absence of concussive brain injury: Injured/Concussed (Cx) versus not
injured (controls/Ctl). RTP in the Cx group ranged from 4 days to 115
days (70 days when we remove this clear outlier). The Cx group in-
cluded the two subtypes of Cx2 and Cx3 based on persistence of
symptoms. Table 1 shows the descriptive statistics for the control group
(187 subjects at Day 0) and the Cx groups (Cx2 and Cx3, N = 177). The
two groups Ctl and Cx3 showed significant differences at intake for
several clinical characteristics, with Cx subjects more often having typ-
ical signs and symptoms of concussive injury (especially AMS, balance
and dizziness problems related to injury; see Table 2). Out of the seven
symptoms listed in Table 2, severe headache and dizziness were the
only two for which symptom presence was associated with a return to
play longer by 5 days. Overall, no clear relationship is seen between
RTP and presence or absence of symptoms. For example, presence of
LOC and amnesia are both associated with shorter RTPs, whereas severe
headache and dizziness are associated with longer RTP. The symptoms
most reported by injured subjects at Day 0 were AMS, followed by
Dizziness, Balance Problems and Severe Headache.

3.1. Classifier performance at day of injury (Day 0)

The Receiver Operating Characteristic (ROC) curve [53] for the GA-
derived multimodal classifier generated using the “All In” algorithm
development database is shown in Fig. 1, for the separation of Ctl from
Cx3 subjects. The Leave One Out (LOO) Cross Validation sensitivity/
specificity at the selected operating point on the ROC curve (defined by
selecting a classification threshold) of the classifier were: 92.7%/75.1%.
The Negative Predictive Value (NPV) and Positive Predictive value (PPV) of the classifier were: 97.5%/49.4%. The summary per-
formance table for the classifiers developed on the 10 Train/Test splits
and for the “All In” classifier is shown in Table 3. Note from this table
that the average sensitivity (Cx3)/specificity on the 10 splits, with 95%
confi dence intervals was: sensitivity (Cx3) = 79.2%; 95%CI:

Table 3
Summary performance table for the classifiers developed on the 10 Train/Test splits and for the “All In” classifier. Cx2 are injured subjects with shorter recovery (RTP < 14); Cx3 are injured subjects with prolonged recovery (RTP ≥ 14). Area Under the ROC Curve (AUCs) are for the Cx3 vs. Ctl classification task, that is, for prediction of prolonged recovery at time of injury.

<table>
<thead>
<tr>
<th>Split#</th>
<th># Selected</th>
<th>LOOCV (Ctl vs. Cx3)</th>
<th>Test/HO Group Results</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sens (Cx3)</td>
<td>Spec</td>
<td>AUC</td>
<td>Sens (Cx)</td>
</tr>
<tr>
<td>1</td>
<td>22</td>
<td>95.5</td>
<td>74.8</td>
<td>0.928</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>90.8</td>
<td>75.3</td>
<td>0.919</td>
</tr>
<tr>
<td>3</td>
<td>23</td>
<td>93.8</td>
<td>74.9</td>
<td>0.926</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>92.4</td>
<td>74.8</td>
<td>0.931</td>
</tr>
<tr>
<td>5</td>
<td>21</td>
<td>93.9</td>
<td>74.8</td>
<td>0.92</td>
</tr>
<tr>
<td>6</td>
<td>15</td>
<td>93.8</td>
<td>74.9</td>
<td>0.918</td>
</tr>
<tr>
<td>7</td>
<td>16</td>
<td>89.2</td>
<td>75.2</td>
<td>0.925</td>
</tr>
<tr>
<td>8</td>
<td>13</td>
<td>93.8</td>
<td>74.8</td>
<td>0.923</td>
</tr>
<tr>
<td>9</td>
<td>17</td>
<td>92.4</td>
<td>74.8</td>
<td>0.912</td>
</tr>
<tr>
<td>10</td>
<td>20</td>
<td>97</td>
<td>75.2</td>
<td>0.932</td>
</tr>
<tr>
<td>Mean</td>
<td>17</td>
<td>93.3</td>
<td>75</td>
<td>0.923</td>
</tr>
<tr>
<td>Std Dev</td>
<td>4</td>
<td>2.2</td>
<td>0.2</td>
<td>0.006</td>
</tr>
<tr>
<td>95% CI</td>
<td></td>
<td>1.4</td>
<td>0.1</td>
<td>3.7</td>
</tr>
<tr>
<td>All</td>
<td>16</td>
<td>92.7</td>
<td>75.1</td>
<td>0.925</td>
</tr>
</tbody>
</table>

HO
75.3%–83.1%/specificity = 69%; 95%CI: 65.2%–72.8%, and that the sensitivity (Cx3)/spec of the “All In” classifier on the Hold Out group was 85%/73.4%. The sensitivity of this classifier on the (independent) test group of Cx2 subjects (N = 60) was 63.3%.

The top-five contributors to the classification included a clinical feature reflecting balance/dizziness (vestibular) problems, a neurocognitive throughput reaction time feature, and a set of EEG features. These EEG features contributing most to classification were specifically those which characterize connectivity between regions, including measures of phase synchrony and power asymmetry. Other important EEG feature contributors included bipolar coherence, shifts in mean frequency, within hemisphere disturbances in power gradients, and a measure of signal complexity (non-linear scale-free).

3.2. Mapping of classifier output to the enhanced multimodal Brain Function Index (eBFI)

In order to give statistical interpretability to the output of the classifier, a mapping to percentile scores (range 1–100) was created using the baseline assessments of the healthy volunteer (HV) group as a normal population. The percentile thresholds were derived so that the classifier outputs of an equal number of controls fell into each bin. A graphical representation of these thresholds is shown in Fig. 2. Note that each pair of such thresholds corresponds to a percentile bin of the eBFI. There are 100 such bins in total. This procedure results in the eBFI of the controls, expressed as a percentile, to be uniform. It also ensures that a subject with percentile score of x has x% of the normal controls with a “worse” classification output score (defined as further toward the more abnormal/Cx3 class). Note that this means that the scale is “one tailed”, i.e. “poor/abnormal” eBFI scores are expected to be only on the one side below the median eBFI score (which is by definition 50 for a normal group). When mapped to eBFI scale, the discriminant classification threshold (T = 12.5 indicated on ROC curve in Fig. 1) was equal to 23.

3.3. Analysis of longitudinal change in the eBFI

Box plots showing the distribution of the eBFI at baseline/pre-season (when available), Day 0 and at subsequent days (Day 5, RTP, RTP + 45, Post-Season (when available)) for the controls and the injured subjects are shown in Fig. 3. At baseline and all subsequent measurements, the controls have a median eBFI which close to 50, as would be expected from a percentile scale derived from a baseline control group. At Day 0, the injured Cx groups (Cx2 + Cx3 combined) have very low median eBFIs which gradually rise to a level which is on par with that of the control group at the same time point. Group comparisons of eBFI were performed at each time point, using the non-parametric Wilcoxon rank test for a median, to test for significance of the differences between the eBFI scores of the controls and the injured, reflecting return toward expected pre-season state. The significance levels are reported in Table 4. At Day 0 (day of injury) the eBFI of the injured was significantly lower than that of the controls (p = 8.1E-25) and this significance was present when considering the Cx2 and Cx3 groups separately. At Day 5, the eBFI of the injured was still significantly lower than that of the controls (p = 2.2E-4) but this significance was driven by the scores of the prolonged recovery group (Cx3). At RTP, the eBFI of the concussed group was no longer significantly lower than that of the controls (p = 0.13) but separate analyses of the short and long recovery groups reflects significant differences in the prolonged recovery group at RTP (p = 0.008), which is no longer there at RTP + 45 (p = 0.36), suggesting that for some subjects, return to play may have occurred prior to maximum recovery as seen at the later time point (RTP + 45).

For the control groups (healthy volunteers and matched controls), there were no significant differences between eBFI at (pre-season) baseline and post-season (p = 0.71). In addition, as can be seen in Table 5, there were also no significant differences between eBFI of the matched controls at Day 0 and at any of the subsequent time points (Day 5, RTP, RTP + 45).3 Importantly, the lack of any significant differences in the matched controls at Day 0 (which can be treated as their baseline measurement) and RTP + 45, in conjunction with the lack of significant differences between controls and injured at RTP + 45, supports the hypothesis the eBFI scores of the injured subjects at RTP + 45 can be considered as their “surrogate baseline” measurement.

3 The Day 0 and RTP comparison for the matched controls yielded the lowest p-value (0.09) and is an outlier within the controls.
3.4. Odds Ratio analyses

Odds Ratio (OR) comparisons were performed, involving matched controls and injured groups (Cx3; RTP \( \geq 14 \)) at Day 0, RTP and RTP + 45. It is of note that the OR of Injured subjects being called injured at RTP+45 by the classifier (i.e. with eBFI \( \leq 23 \)) versus Controls being called injured was very close to one (OR = 1.01), while the OR of Injured subjects being called injured by the classifier at Day 0 vs. Controls being called injured at Day 0 was close to 4 (OR = 3.91). Thus these ORs further support the use of RTP+45 as a surrogate baseline for the injured group.

4. Discussion

In the absence of a biomarker as a gold standard for concussion, there is reliance on self-report and brief sideline evaluations, which are largely subjective, can be learned and are inconsistent in use, both in defining the injury and in making important decisions related to recovery (e.g., RTP). Presence and severity of most concussion symptoms at time of injury are weak predictors of injury severity and of the time it takes for the athlete to return to play [55]. As noted above, unrecognized and untreated concussions can contribute to increased risk of repeat concussions, risk of mortality, slowed recovery and contribute to the presence of both short and long-term consequences [5,56–58].

Although advanced neuroimaging can be used to distinguish between groups with concussive brain injury and controls in an experimental setting [59], such technologies are not readily available at the...
sidelines, in Urgent Care Centers, concussion clinics or in the emergency department to aid in the assessment of concussive injury. Studies have suggested that EEG can act as a surrogate for other neuroimaging tools and can provide many advantages in sideline testing (in locker room or nearby venue) at the time of injury [60]. Using devices in development, studies have shown that an EEG index reflects persistence of alterations in brain function, beyond the window of abnormal findings demonstrated through clinical measures, focusing on assessment of reported symptoms, cognitive functioning, and sideline measures of vestibular functioning [61,63–65]. The BrainScope One EEG Brain Function Index [19] was also demonstrated to have potential to reflect recovery from sport-related concussion [21].

The importance of multimodal assessments of concussion, which evaluate several functional/clinical domains has been emphasized in recovery from sport-related concussion [21]. The Function Index [19] was also demonstrated to have potential to reflect recovery trajectory, potentially represents added value as an adjunct to protracted recovery. The ability to obtain such an index rapidly, in a ball athletes [55], while no other on-field symptom was associated with protracted recovery. The result also supports the hypothesis that some of the more severely injured subjects (those with prolonged RTP) may have been allowed to return to play too soon. It is of note that the injured subjects that recovered within 14 days, and which can therefore be considered to be only very mildly injured, were largely identified as “not prolonged recovery” by the classifier and that their eBFIs were no longer significantly different from those of controls at Day 5. The demonstration that the injured group shows no significant differences from the control subjects at RTP + 45 suggests that the eBFIs in the injured subjects at RTP + 45 have returned to the point where they fall within the baseline range of a non-injured population.

This study demonstrated the potential clinical utility of a multimodal biomarker based on brain electrical activity, neurocognitive and vestibular features to provide a means for personalized quantitative tracking from time of injury throughout recovery of brain function after concussive injury. It is of note that others have reported the importance of a vestibular clinical feature (dizziness at time of injury) as being associated with a protracted recovery from a sports-related concussion (defined as RTP ≥ 21 days) in a sample of 107 male high school football athletes [55], while no other on-field symptom was associated with protracted recovery. The ability to obtain such an index rapidly, in a hand held device, at the point of care and at any point along the recovery trajectory, potentially represents added value as an adjunct to current clinical practice to assist in more objective, timely, confident and optimal determination of, and tracking of recovery from, mTBI/concussion. An independent prospective validation trial of such a multimodal Brain Function Index is under way.

Gioia states in Ref. [66] that “Ultimately, the practice of concussion assessment and management will benefit from an evidence-based medicine approach […] whereby clinicians have access to statistical bases for evaluating a patient’s scores relative to those seen in subjects with known concussions versus non-injured subjects.” The eBFI, a multivariate, multimodal, objective and quantitative index of brain function impairment directly addresses this need.

Conflicts of interest

All authors of this manuscript are employed by BrainScope Company, Inc. Dr. Prichep is an inventor on patents licensed by BrainScope from NYU School of Medicine.

Sources of support

The clinical study was funded in part by a contract to BrainScope Company Inc. from the U.S. Navy (Naval Health Research Center), contract #W911QY-14-C-0098. Disclaimer: The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Navy position, policy or decision unless so designated by other documentation. Also supported in part by BrainScope Company, Inc.

Acknowledgements

The authors acknowledge the contributions of those who made this research possible, including the primary investigators and research staff at all the clinical sites and the athletes who volunteered for the study.

References
