Abstract **Background:** Hypotension is recognized as a potentially damaging secondary insult after traumatic brain injury. Systems to give clinical teams some early warning of likely hypotensive instability could be added to the range of existing techniques used in the management of this group of patients. By using the Edinburgh University Secondary Insult Grades (EUSIG) definitions for hypotension (systolic arterial pressure <90 mmHg OR mean arterial pressure <70 mmHg) we collected a group of ~2,000 events by analyzing the Brain-IT database. We then constructed a Bayesian Artificial Neural Network (an advanced statistical modeling technique) that is able to provide some early warning when trained on this previously collected demographic and physiological data.

**Materials and Methods:** Using EUSIG defined event data from the Brain-IT database, we identified a Bayesian artificial neural network (BANN) topology and constructed a series of datasets using a group of clinically guided input variables. This allowed us to train a BANN, which was then tested on an unseen set of patients from the Brain-IT database. The initial tests used a particularly harsh assessment criterion whereby a true positive prediction was only allowed if the BANN predicted an upcoming event to the exact minute. We have now developed the system to the point where it is about to be used in a two-stage Phase II clinical trial and we are also researching a more realistic assessment technique.

**Key Results:** We have constructed a BANN that is able to provide early warning to the clinicians based on a model that uses information from the physiological inputs; systolic and mean arterial pressure and heart rate; and demographic variables age and gender. We use 15-min SubWindows starting at 15 and 30 min before an event and process mean, slope and standard deviations. Based on 10 simulation runs, our current sensitivity is 36.25% (SE 1.31) with a specificity of 90.82% (SE 0.85). Initial results from a Phase I clinical study shows a model sensitivity of 40.95% (SE 6%) and specificity of 86.46% (SE 3%) Although this figure is low it is considered clinically useful for this dangerous condition, provided the false positive rate can be kept sufficiently low as to be practical in an intensive care environment.

**Conclusion:** We have shown that using advanced statistical modeling techniques can provide clinical teams with useful information that will assist clinical care.

**Keywords** Traumatic Brain Injury • Prediction • Hypotension • Bayesian Artificial Neural Network
Introduction

Patients being treated in an intensive care unit (ICU) for traumatic brain injury (TBI) are at risk of secondary insults [8]. Hypotension, critically low blood pressure, is one type of secondary insult and has been shown to be related to poor outcome [3, 7].

As part of a current European Union (EU) funded research project AvertIT [1], we have been researching the use of a statistical technique called a Bayesian artificial neural network (BANN) to assess whether this technique can be used to give clinicians early warning of this condition.

This report details our approach to using the BANN technique and provides some early indications of the results of using this technique in a Phase I clinical trial in six neuro-intensive care facilities throughout Europe: Uppsala, Sweden; Glasgow, Scotland; Vilnius, Lithuania; Heidelberg, Germany; Monza, Italy; and Barcelona, Spain.

Materials and Methods

This section of the report provides details on the data source used for initial investigations; the data preparation required; and the use of this prepared material to train the BANN.

Data Source

The data for our research comes from the BrainIT consortium [2]. The database comes from a multi-center study across 22 hospitals in Europe and it contains three classes of information: demographic data; physiological data; and episodic data.

The BrainIT database contains 199 patients although only 119 patients met our criteria of both BPs and BPm available with a sufficiently low number of short data gaps. The average age of the 119 patients was 39.04 years (range = 15−83, median = 34.05), with 85% being male. Nineteen patients had no events but these patients are included in the training data to ensure that vectors from a patient in the training group are not included in the test group.

Data Preparation

After statistical classification using definitions from clinical practice, we decided to use the published work on secondary insults by Jones and colleagues at Edinburgh University [7]. We use the definition for hypotension which is BPs ≤ 90 mmHg or BPm ≤ 70 mmHg for a hold-down of 5 min. Using the EUSIG definition we built a list of 2,081 events from 100 patients with project-specific software [4].

Our next task was to produce a series of datasets that could be used to decide the topology of the BANN and then calculate the various coefficients that form the heart of the BANN software, which runs in the clinical environment. We chose to use an approach called SubWindows in which each relevant physiological signal is divided into sections of time, and each SubWindow is summarized by multiple statistical calculations. Each statistical calculation within a SubWindow becomes a single input into the BANN. With reference to Fig. 1, we can see that for BPm, the slope of SubWindow 1 forms an input into the BANN. An input definition syntax was developed and a software application was produced that scanned the BrainIT database on a minute by minute basis for each patient to build up a series of vectors (rows of data). Each vector of data contains a timestamp, a group of demographic and physiological signals as specified by the input definition file, and finally a flag that indicates whether or not a hypotensive event started or not at the appropriate prediction window.

BANN Topology and Training

This process used software based on a core of routines from Radford Neal at the University of Toronto [9] to perform the Bayesian calculations. We were guided by the clinical experience of the AvertIT neurosurgical teams in selecting likely inputs and time ranges. We carried out 33 runs over a 2-month period from September to October 2008. The top five runs showed a sensitivity of >40% and these are summarized in Table 1. Full details can be found in Howells and Donald [6]. We chose to further investigate the 30_15 BasicWithSlope model as its constituent measurements of Age, Sex, BPs, BPm, and heart rate were available in all centers.

A single BANN run is constructed by randomly assigning half of the patients (~50) to the training group and the other half to the test group. The positive vectors for each group are balanced by a proportion of negative vectors to form the training and test sets for the training run. It is important to ensure that vectors from a patient in the training group are not included in the test group.

We then constructed a series of BANN training runs. After each run an assessment was made on model performance as described in the section Test Set Model Assessment. The performance of the top five models is shown in the Results section in Table 1.

BANN Test Runs

A training run produces a series of coefficients for the BANN and the performance of this set of model parameters is tested by running the calculations over a set of unseen inputs provided from the test set. This produces a prediction list with
the probability of a hypotensive event occurring in the given prediction window. We then use the standard ROC procedure \cite{5} for assessing a classifier.

**Clinical Tests**

Our testing has moved beyond the research phase and we are currently (October 2010) about to conclude a Phase I clinical study designed to quantify measures of the sensitivity and specificity of the system when operating under clinical conditions. The Phase I study software has been installed on a dedicated server in each one of the six participating neurointensive care units. A research nurse from the project works with ward staff to admit a patient to the study and once relative assent has been obtained, the system transfers the minute by minute physiological data on an hourly basis to a central data store at the national eScience center at the University of Glasgow (NeSC) \cite{11}.

**Test Set Model Assessment**

The initial assessment method used for feature selection was particularly harsh. During this early phase of the research we took the more mathematical view that if the model had been trained on data at 15 min ahead of an event, we would only score a correct prediction if the model predicted this event to

---

**Table 1** Top five performing models

<table>
<thead>
<tr>
<th>Date</th>
<th>Group</th>
<th>Name</th>
<th>Duration (H:M:S)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008-10-09</td>
<td>30_15</td>
<td>BasicWithETCO2</td>
<td>01:14:11</td>
<td>43.9</td>
<td>85.1</td>
</tr>
<tr>
<td>2008-10-16</td>
<td>30_15</td>
<td>BasicSlopeVar550Nets</td>
<td>15:16:45</td>
<td>43.3</td>
<td>88.3</td>
</tr>
<tr>
<td>2008-10-09</td>
<td>30_15</td>
<td>BasicWithSlope</td>
<td>05:27:44</td>
<td>42.1</td>
<td>85.0</td>
</tr>
<tr>
<td>2008-10-07</td>
<td>25_10</td>
<td>Basic</td>
<td>03:01:08</td>
<td>41.4</td>
<td>91.1</td>
</tr>
<tr>
<td>2008-10-13</td>
<td>30_15</td>
<td>BasicWithSlope16HL</td>
<td>02:43:39</td>
<td>41.2</td>
<td>85.4</td>
</tr>
</tbody>
</table>
the minute. We applied this method to all 33 BANN training runs as described in the section BANN Topology and Training.

We felt that a more realistic approach was required because we have noticed that frequently groups of false positive events occur just before the 15-min actual event time. This means that we were getting more than 15 min warning, a better performance, but our assessment criterion was penalizing the model. We have therefore recently been using a technique that we feel provides a more clinically practical assessment of model performance.

**Phase I Model Assessment**

We present a proposal for a scoring system that we believe makes better clinical sense and gives a more realistic assessment of model performance. This section describes how the events are grouped into blocks of time to allow the calculation of values for true positives, true negatives, false positives and false negatives.

With reference to Fig. 2, consider a system where the overall time in the ICU is split into 30-min blocks. For this example we have 4 days in the ICU, which equates to 192 possible blocks. We have, from the HypoPredict system, the times and durations of the actual hypotensive events that occurred given the EUSIG definitions. These can be assigned to the appropriate block as seen in the “EUSIG Events” line.

We then perform a further grouping of these events into “Episodes” of hypotension as it has been observed that the events sometimes occur in groups one after the other. The important characteristic of an early warning system from a clinician’s point of view is a system that gives a warning of the start of an episode. We will consider an inter-event time of 15 min or less to define the existence of a group. This allows us to fill in the line “Episodes.” In this line it can be seen that the two events in block 4 have become a single episode. The events
that started in blocks 15 and 17 have also been combined into a single episode running from block 15 to block 19.

The gated output from the model is used where “gated” means that the model probabilities are not considered during a hypotensive event or for a recovery period after the end of an event. This recovery period is required so that the model is using non-hypotensive data in its calculations. The recovery period is model-specific; the current model being tested has a 30-min recovery time. Using these definitions, we obtain the “Predictions” line in Fig. 2. We now consider each type of prediction block in turn using a prediction threshold of 0.3. The value of 0.3 was chosen from an ROC analysis carried out during the model selection procedure.

**True Positive**
A single true positive is generated if the model produces an output above 0.3 for 3 consecutive minutes and an event occurs within 30 min (the block size) of the start of the warning. Note that it is possible for a warning to start just at the end of one block and the episode occurs, for example, 20 min into the following block. This situation is handled by the protocol described in the section Episode Prediction in the Following Block.

**True Negative**
A true negative means that the model output was below the 0.3 threshold for the entire 30-min block and indeed there were no episodes during that time.

**False Positive**
A single false positive is generated if the model output stays above 0.3 for 3 consecutive minutes and no episode starts within 30 min (i.e. a block size time) of the start of the warning. Again, this has the potential to spill over into the next block and if this occurs it is handled by the rules described in the section Episode Prediction in the Following Block.

**False Negative**
A false negative is generated when an episode occurs and there is no warning output from the model.

**Episode Prediction in the Following Block**
When the model produces a warning output – in the examples above this has been defined as a probability greater than 0.3 – the actual event may occur in the following block. This is shown in the expanded view in Fig. 2. The scoring rule is defined such that the following block (block 15 in this case) is given the true positive score rather than the block where the prediction first occurred.

The rule to determine a false positive score is defined in the opposite manner. In the case of a false positive prediction that occurs just at the end of one block, if no event occurs within 30 min of the start of the prediction, which requires checking in the following block, then the block where the false prediction occurred is marked as false positive (see Fig. 2).

**False Positive Suppression**
We recognized from an early stage that the BANN classifier would only form one part of an eventual HypoPredict Engine, which could possibly be used within a clinical environment. This is shown diagrammatically in Fig. 3.

As a consequence of the initial indications from the Phase I clinical trial, we have recently begun to investigate using suppression techniques based on the statistical properties of the distribution curve produced by the BANN for each minute’s prediction along with, crucially, clinical experience used when assessing the raw blood pressure and heart rate traces. These clinical heuristics have given us a promising tool in our efforts to reduce the false positive rate of the system.

**Results**

**Model Selection**
The model selection is shown in Tables 1.
Phase I Trial Early Indications

Early indications from the Phase I clinical study, based on 20 patients collected from the 6 participating neurointensive care units, indicate that we are achieving a sensitivity of 43.85% (SE 7%) with a specificity of 83.21% (SE 3%). By applying our first pass suppression techniques these values change to a sensitivity of 40.95% (SE 6%) with a specificity of 86.46% (SE 3%). The average warning time is 21 min (SE 0.74).

Discussion

We have shown that it is possible to construct a system that provides useful early warning of the dangerous traumatic brain injury secondary insult condition of hypotension. This system is based on an advanced statistical modeling technique called a Bayesian Artificial Neural Network (BANN) and we have described the not inconsiderable data preparation required to use this technique. We feel that the topic of model assessment has been addressed in a manner that includes the requirements from clinical teams to produce systems that not only provide accurate warning of an event of interest, but which also recognize the need to maintain a low number of false positives. Finally, we have shown that this research has translated from an experimental result to a system that works under real clinical conditions.

Conclusion

Areas of future research include assessing the influence of different probability cutoff thresholds on calculated false positive rate. Although higher sensitivities are desirable and the addition of other more complex signals such as EEG may improve accuracy, the lead emphasis clinically is on a technology that has more ubiquitous application in ITUs with only basic physiological monitoring, and of course that it be clinically practical by generating as few false warnings as possible. A technology that could detect even 1 in 3 (33% sensitivity) hypotensive events could be clinically significant, but only if the clinicians alter their management of hypotension accordingly. To show this would ultimately require a further clinical trial.

Conflict of interest statement  The authors declare that they have no conflict of interest.

References