DECISION TREE ANALYSIS OF DATA FROM A NEUROLOGICAL INTENSIVE CARE UNIT

Forecasting recovery after traumatic brain injury using intelligent data analysis of admission variables and time series physiological data- a comparison with logistic regression.

The data collection for this study was supported by the Medical Research Council, UK. G9508752. Grant holder, P Andrews.

Authors: Dr. Peter JD Andrews MD, Professor Derek Sleeman PhD, FRS(Ed), Mr. Andrew McQuatt MSc, Dr. Vincent Corruble PhD, Mrs PA Jones MSc, Dr. Tim Howells PhD, Dr. Carol S A Macmillan MB ChB.

Institutions:
Department of Anaesthetics, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU. Tele No. +44 131 537 1666, Fax No. +44 131 537 1021, email Pandrews@ed.ac.uk

Department of Computing Science, Meston Building, King's College, University of Aberdeen, Aberdeen, AB24 3UE. Tele No. +44 1224 272295/6, Fax +44 1224 273422, email Sleeman@csd.abdn.ac.uk

Corresponding Author:
Peter Andrews. Department of Anaesthetics, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU.
Tele No. +44 131 537 1666, Fax No. +44 131 537 1021, email Pandrews@ed.ac.uk

Keywords: **Head Injury, Adults, Secondary Cerebral Insults, Intensive Care, Outcome, Decision Tree Analysis**

This work was presented in part to the Artificial Intelligence in Medicine Conference (AIMDM'99), Aalborg, Denmark, June 1999.

ABSTRACT

Decision tree analysis highlights patient sub-groups and critical values in variables assessed. Importantly the results are visually informative and often have clear clinical interpretation about risk factors faced by these subgroups.

The aim of this prospective study was to compare logistic regression with decision tree analysis of an observational, head injury data set, which includes a wide range of secondary insults and 12 month outcome.

Methods: 124 adult head injured patients were studied during intensive care, using a computerized data collection system. Verified values falling outside threshold limits were analyzed by insult grade and duration using logistic regression. A Decision Tree was automatically induced from root node to target classes (Glasgow Outcome Score, GOS).

Results: In 69 patients, in whom 8 insult categories could be assessed, outcome at 12 months was analyzed by logistic regression, to determine the relative influence of age, admission Glasgow Coma Sumscore, Injury Severity Score (ISS), pupil response on admission and insult duration. The most significant predictors of mortality in this patient set were duration of hypotensive, pyrexic and hypoxemic insults. When good vs. poor outcome was considered, hypotensive insults and pupil responses on admission were significant.

Decision tree results: Hypotension and low cerebral perfusion pressure (CPP) are the best predictors of death with a 9.2% improvement in predictive accuracy (PA) over that which could be achieved by simply predicting the largest outcome category as the outcome for each patient. Hypotension was a significant predictor of poor (GOS 1-3) outcome. Low CPP, age, hypocarbia and pupil response were also good predictors of outcome (good/poor) with a 5.1% improvement in PA. In certain subgroups of patients pyrexia predicted good outcome.

Conclusion: Decision tree analysis confirmed some of the results of logistic regression and challenged others. This investigation shows there is knowledge to be gained from analysing observational data using this methodology.

Keywords: **Head Injury, Adults, Secondary Cerebral Insults, Intensive Care, Outcome, Decision Tree Analysis**

INTRODUCTION

Patients who sustain head trauma suffer not only from the effects of the primary injury, but also from additional secondary, largely ischemic, brain damage. In 1978 Graham and colleagues [17] reported a 91% incidence of ischemic brain damage in a series of 151 fatal cases of severe head injury subjected to autopsy. The same group [18] reported on a further 112 fatalities from severe head injury treated in the 1980s where, despite more intensive treatment, the incidence of ischemic brain damage remained higher than 80%. Other authors[8;9;14] have considered those patients who "talk and die" or "talk and deteriorate". In these cases the primary injury is judged not to have been severe; it is subsequent events that result in fatal or disabling cerebral injury, thus introducing the concept of "avoidable brain damage".

Secondary insults that are likely to be responsible for ischemic and other forms of secondary brain damage may be intracranial or systemic in origin and may arise during initial management or later in the Intensive Care Unit (ICU)[1;3;12]. Between 1993 -1998 a number of authors reported that in comatose head injured patients arriving at hospital, hypoxemia was found in 30% and arterial hypotension in 15% of patients upon arrival in the Emergency Room (ER)[10;15;31;36;37]. More recently, probably because of better at-the-scene resuscitation and transport arrangements, there has been a reduction in the frequency of these early insults[15].

Forecasting recovery following traumatic brain injury is a complex but important task and any predictive model is likely to be better calibrated when information on secondary insults are included. These data are of value when discussing prognosis with relatives and important when considering the wider issues of resource allocation, comparison of outcomes between Intensive Care Units[5;6] and perhaps measuring the effect of novel pharmaceutical compounds on recovery. Many previous attempts have predicted outcome after traumatic brain injury, however, statistical analysis, including complex modelling, treats all the patients as one homogeneous pool assuming that each variable is likely to be as important for each patient within this diagnostic category of “head injury”. The advantage of decision tree analysis is that it creates subgroups of patients and identifies factors that are important to the individual or subgroups. These analyses are visually informative and will often challenge current knowledge and suggest new hypotheses for future testing. Therefore, there is considerable theoretical advantage in this type of analysis when exploring large, clinical, observational datasets.

We have induced a number of decision trees using an historic traumatic brain injury dataset that has previously been analysed using statistical techniques to allow comparison between the two methodologies.

Aims of the Study

The aims of this study were; (i) to identify areas where the decision tree analysis agreed with statistical analysis, (ii) identify areas where there were differences between the two analyses and (iii) look at unexpected or unexplained results that challenge our understanding of the process that occurs after traumatic brain injury.

PATIENTS AND METHODS

Patients

This prospective study group consisted of 124 patients aged 16 years and over, admitted to the regional Head Injury Unit, Edinburgh between 1st January, 1989 and 16th June, 1991; a statistical analysis has been reported[20].

Admission details

Details were retrieved from the referring hospital (68 patients), or from the Emergency Room for patients brought direct from the scene of the accident (56 patients). Data included cause and nature of injury, age, condition including Glasgow Coma Score (GCS) of the patient on arrival at hospital and after resuscitation, pupil response, computerized tomography (CT) results, operative treatment and medical therapy.

Inclusion criteria

Patients were eligible for this study if the following inclusion criteria were met:

1. The post resuscitation GCS was <13 or GCS >12 if the Injury Severity Score [4;35] was >15.
2. There were clinical indications for monitoring the patients in ICU
3. A computer data collection system was available for use within 24 hours of admission.

Patients were classified according to the post-resuscitation Glasgow Coma Sumscore (GCSs) into severe head injuries with a GCSs of 8 or less, with no eye opening; moderate head injuries whose post-resuscitation GCSs was 9 - 12 (or GCS $\leq$8 with eye opening) and minor head injuries (GCSs 13-15) that were associated with other injuries sufficient to achieve an ISS of at least 16[4]. The management strategy for head injury was carried out by experienced neurosurgical, anesthesiology and nursing staff and remained unchanged during the study period.

Mechanical ventilation and ICP monitoring were employed and additional invasive monitoring where clinically indicated, thus patients had data collected only for those physiological parameters which were appropriate to them as judged by the clinical staff. Once connected to the computerized data collection (CDC) system, patients remained on it until clinical monitoring ceased. In some cases this was as long as 2-3 weeks. As only 2 systems were available, inevitably there were patients who fulfilled the criteria but could not be connected to the CDC system as they were already in use. From a total population of 2010 head injured patients admitted to the ward during this time period, 429 fitted study criteria. Of these 51 were children, and of the remaining adults 33% (124 patients) had computer monitored data collected. Overall, 45% of the severe head injuries were computer-monitored (Table 1) and were representative of the study group, but only 18% of moderate and 20% of minor head injuries could be connected to the computer system (and were the more severely injured of these groups).

Computerized data collection (CDC) – Time Series Physiological Data

A continuous data collection system, based on a personal computer, has been developed with the capability of recording, minute-by-minute, at least 14 physiological variables[20]. Monitored variables included heart rate (measured from the ECG), systolic and mean arterial blood pressure (MAP) measured invasively, intracranial pressure (ICP) recorded by a Camino catheter, arterial oxygen saturation measured from a pulse oximeter (SaO_2), jugular venous oxygen saturation ($SjvO_2$) using the Oximetrix 3 system (Abbott Critical Care Systems, USA)[2], end tidal CO_2 ($ETCO_2$), peripheral and core temperature. From these variables, cerebral perfusion pressure (CPP = MAP - ICP) and cerebral arterio-jugular venous oxygen content difference ($Ca\text{-}jvO_2$) were calculated on line: $Ca\text{-}jvO_2 = (SaO_2 - SjvO_2) \times Hb \times 1.39$ mls O_2/100 mls blood.

Definitions and Scoring of secondary insults

At the outset of this study, threshold limits were set for abnormal values to define insult grades for each variable, based on earlier research[3;20;30] and clinical experience. Values falling outside these limits were displayed in colored font to help in the identification of insults. Three ranges of increasingly abnormal values were defined (Grade 1 (yellow), Grade 2 (orange), and Grade 3 (red)) for each insult variable, except for Ca-jvO_2 where one grade of insult was recorded for increased and decreased oxygen extraction. The ranges of abnormal values used, described as Edinburgh University Secondary Insult Grades (EUSIG) are shown in Table 2 and were analyzed off-line by Edinburgh Browser® software.

To be considered a secondary insult in this study, abnormal values had to persist for 5 or more consecutive minutes. The insult was deemed to have ended only when values returned to normal for 5 consecutive minutes. Pyrexia was considered to be an insult when it lasted for 1 hour, and finished when normal for 1 complete hour to reflect the slower rate of change of this variable. The date, grade, start and stop time for each insult type was recorded. A change in grade of insult was recorded only if the new grade lasted for at least 5 minutes or 1 hour in the case of pyrexia. Temperature was managed according to a standard protocol that aimed for normothermia (or $<37^0$C). Surface cooling and nasogastric or rectal paracetamol 1gm were the principle interventions to facilitate this.

The Edinburgh Monitor® software automatically detects when a channel is not working by testing the ICU monitor's internal error flags as well as comparing the 'valid' data against upper and lower rejection limits specific for each channel. Values falling outside these limits are deemed non-physiological and rejected. The system also allows the insertion of comments by nurses and doctors with the time of day recorded so that additional events which may be classified as an artifact, can be identified e.g. nursing maneuvers or blood gas sampling.

Several channels of data had to be verified against other results before being accepted as valid. $ETCO_2$ insults were verified from arterial blood gas results. $SjvO_2$ values were validated against the Oximetrix print-out to ensure adequate light intensity readings and twice daily calibration against an IL482 co-oximeter[2]; acceptable simultaneous recordings from ICP and arterial blood pressure were prerequisites for CPP calculation. $Ca\text{-}jvO_2$ computation necessitated correct data for SaO_2, hemoglobin and $SjvO_2$. When there was doubt as to the validity of abnormal values shown on the Edinburgh Browser file, they were rejected. Final identification and verification of all insults by grade and duration was done manually by 4 research personnel. Although time consuming and labor intensive, we judged this process to be essential to ensure that the data were reliable and accurate. The intention was therefore to *under estimate* insults, rather than include dubious data.

ICP data recorded in this study included readings taken during treatment of intracranial hypertension. Our practice was to correct, where possible, remediable causes of raised ICP and then treat raised ICP when this exceeded 25 mm Hg in the first 48 hours, or 30 mm Hg thereafter unless signs of brain herniation were observed at lower levels of ICP.

Intervals of data loss occur periodically: while computer files are backed-up, when patients are removed from the ICU for CT scan or operation, during some nursing and physiotherapy procedures or due to computer failure.

Outcome Scoring

The 5-point Glasgow Outcome Scale (GOS) was chosen as the measure of outcome, being widely used and reliable[19]. Before assigning an outcome score, information was collected from several sources: a patient interview in conjunction with a battery of seven neuropsychometric tests; a 60 item questionnaire sent to relatives; a letter to the patient's General Practitioner and scrutiny of all medical notes and outpatient clinic letters. At the time of assigning a GOS, the scorer was unaware of the secondary insult data. GOS scoring was done at 12 months (strictly between 11 and 13 months) after injury. Only one patient was completely lost to follow-up at 12 months. Two others were known to be alive, although a reliable GOS could not be ascertained.

Statistical Analysis

Start and stop time and grade of insult were entered onto a computer database (Dbase IV; Ashton Tate). SPSS was used for data compilation and analysis. This included tabulation of insult grades and duration for each variable. This report presents analyses based on insult duration and frequency in all 124 patients.

Examination of relationships between insult variables and other prognostic indicators such as severity of injury and age was made primarily using Logistic Regression. Other statistical tests (Kruskal-Wallis, Mann-Whitney U Tests, Chi Square) supplemented the findings. Significance was assumed at the 5% level.

DECISION TREE METHODOLOGY

Building decision trees

A definition of decision trees and their creation is as follows: "The traditional approach to constructing a decision tree from a training set of cases described in terms of a collection of attributes is based on successive refinement. Tests on the attributes are constructed to partition the training set into smaller and smaller subsets until each subset contains cases belonging to a single class. These tests form the interior nodes of the decision tree and each subset is associated with one of its leaves. An unseen case is classified by tracing a path from the root of the tree to the appropriate leaf and asserting that the case belongs to the same class as the set of training cases associated with that leaf" [27].

The decision trees used in this report were generated using the See5 environment [7;29]. This is a PC application that uses the C5.0 algorithm, that is an upgraded version of the C4.5 algorithm. As the new facilities of C5.0 have not been utilised in this analysis, the resulting decision trees can be considered as being the same as produced by the C4.5 algorithm. There has been much work done on decision trees since the 1970s, both by Quinlan [27;28] and others in the machine learning and statistics communities[16;22].

The power of this method lies in its ability to choose the value of the attribute at which the cases can be divided into the lowest entropy categories. Pruning of decision trees is useful for simplifying models that have "grown" too much and therefore "overfit" the data. Generally, when generating a decision tree there will be some misclassifications, i.e. some cases that are assigned to one class by the tree though they belong to another one. These errors are represented in the leaf nodes of the decision trees in the following way: (**n/m**), where **n** is the total number of cases in the leaf node and **m** is the number of misclassified cases, and includes decimal places because of pruning.

Assessing tree accuracy

This has been presented using the following measurements:

The *training accuracy* is a measure of how accurately the tree represents all known cases. However, these figures give little indication of how well these decision trees would predict the outcome of a new patient. Predictive accuracy calculated using ten fold cross-validation gives a better measure of predictive accuracy (see below). However, it could be misleading to consider that, say an 80% accuracy is very good, in the situation where one of the classes represents 75% of the population. Therefore all the results reported in this paper include the size of the biggest category (as a percentage of population) as a comparison.

Predictive accuracy is a measure of how well the tree classifies new cases. The standard method of testing the predictive accuracy of a decision tree is to use a new set of as-yet-unseen cases from the same population. The problem with this is that the predictive accuracy is very much dependent upon which cases are in the test and in the training sets. To avoid this problem, the set of cases is randomly split into a number of equally sized subsets (say, 10). Each subset in turn becomes the test set, with the remaining 9 subsets forming the training set. This, therefore, requires ten different decision trees to be generated and tested. The overall predictive accuracy is then the average of all the predictive accuracy. This method is called *crossvalidation*, and as it uses 10 subsets, it is referred to as *10 fold cross-validation*. However, when the cross-validation is repeated, it will randomly choose 10 different subsets and so produce a different overall predictive accuracy. For this reason, the average of ten 10-fold cross validation tests is used for an overall predictive accuracy measurement.

RESULTS

Severity of head injury, age and sex distribution:

Of the 124 patients in this study 68 patients had sustained severe head injuries, 36 moderate, and 20 minor head injuries. Data sets include; (**1**) all 124 patients, (**2**) 121 patients for whom GOS is known at 12 months, (**3**) a subset of 71 patients who had a minimum of 8 channels of physiological monitoring in ICU with secondary insult processing and (**4**) 69 patients of this subgroup, with outcomes at 12 months (two of the 71 were known to be alive but GOS not known). Diagnosis and outcome data are shown in Table 1.

Injury Severity Score:

Median Injury Severity Scores for the groups of severe, moderate and minor head injured patients were 25.0, 16.0 and 20.5 respectively, and the overall median ISS was 25.0. Sixty-five of the patients had an ISS of 25 or more. Multiple injuries were defined as those patients with a head injury in association with another injury with an Abbreviated Injury Score of >1. The numbers of patients with isolated head injuries and those with associated multiple injuries were 53 and 71 respectively.

Secondary Insults: Distribution

All the insults identified in this study were found in the subset of valid data collection time remaining after deducting lost CDC time. Those channels using pressure devices (ICP, BP and CPP) had *valid* CDC time of *the total CDC time* of 91%, 86% and 88% respectively. This apparent inconsistency in the percentages of BP and CPP is due to the considerably longer total duration of BP monitoring. The channels for heart rate and pyrexia were found to have a similar usability: hypoxemia and $ETCO_2$ were valid for 80% and 74% respectively of the total CDC time. The $SjvO_2$ channel and its derivatives had the least valid CDC time at 56% of the total recording time.

One hundred and thirteen (91%) of the patients studied had secondary insults detected during their period of computerized data collection in the Intensive Care Unit, and nearly 1 million minutes were identified where values were at insult level, of which 83.5% (804,276 minutes) were at the least severe EUSIG grade 1 level. Hypertensive and hypotensive insults were found in 89% and 73% of CDC patients respectively. However nearly 90% of the total duration of these blood pressure insults were at grade 1 level. Seventy-seven patients had ICP monitored, of whom 65 (84%) had at least one ICP insult, and 61 (81%) had one or more CPP insults. Insults were found in patients of all grades of severity of injury, and in all parameters monitored.

Insults of short duration were common. The median insult duration for ICP grade 1 and grade 3 was 42 and 21 minutes respectively, while the median time for hypoxemia grade 1 was only 12 minutes. (Table 3).

Secondary Insults: Relationship to mortality and morbidity

The subset of 71 patients, who had computerized data collection for the 8 variables of ICP, arterial hypo- and hypertension, CPP, pyrexia, hypoxemia, brady- and tachycardia, were analyzed using logistic regression of mortality and morbidity at 12 months. Because there is a mathematical relationship between ICP, hypotension and CPP. Only two of these three should be included in the analysis; CPP was excluded. Eleven factors (age, GCSs and pupil response on admission, ISS, and insult duration of 7 insults) were therefore entered into the regression algorithm.

When mortality (death or survival) at 12 months was considered, (Table 4) only three of the variables - duration of hypotensive, pyrexic and hypoxemic insults were significant predictors of outcome (p=0.0064, p=0.0137 and p=0.0244 respectively). However the age variable was found to be very close to significance in this group of patients (p=.0652).

When good vs poor outcome was used as the outcome measure (n=69), i.e. good recovery and moderate disability vs severe disability, vegetative survival or death, only

2 variables were significant. Total duration of hypotensive insults (p=0.0118) and the reaction of the pupils at the time of admission to the neurosurgical unit (p=0.0226) were significant predictors of quality of outcome at 12 months. Table 4.

Secondary Insults: Relationship to severity of head injury.

Although the logistic regression data (n=71) showed that hypotensive insults were important predictors of outcome, independent of the coma score on admission, we wanted to verify across the whole data set that insult frequency was not simply a measure of severity of injury. Comparing the patient groups of those in coma on admission (i.e. severe head injuries, n=68) and those not in coma (i.e. moderate and minor head injuries, n=56) by the insult variables, no significant differences between the group medians were found for duration of ICP, hypertensive, CPP, hypoxemic, hypocarbic, hypercarbic, pyrexic or bradycardic insults. However in the whole group, the duration of hypotensive and tachycardic insults were significantly longer in patients who were comatose on admission (Mann-Whitney U test, p = 0.0021 and p =0.0013 respectively.)

Decision Tree Analyses

Decision trees were generated for nine different combinations. They were obtained from 3 selections of the data; namely (i) demographic, (ii) insult (time series physiological), (iii) demographic and insult, against 3 types of predictions (death/survival, good/poor outcome, the Glasgow Outcome Score [19]).

Decision Tree Results

Accuracy of the models

Data sets are the same as those used for the statistical analysis (data sets 1-4). Tables 5a&b give the results for the nine combinations discussed above. The first column is the largest category and is given as an item of comparison. Decision trees were induced using the incomplete data from the **121** patients (as decision trees can handle missing data) and also from the **69** patient subset.

Improvements in accuracy are greatest where the data are complete.

All decision trees described subsequently are based on the *69 patient dataset* as this allows comparison with the statistical analysis (figures 1,2 &3).

Decision trees

As there were a high number of different decision trees generated by this analysis, we only show a selection here. The rest are presented in full elsewhere[24]. Figure 1 shows a decision tree for predicting good or poor outcome that is based only on demographic data for the n=69 dataset; Figure 2 predicts good/poor outcome based upon on insult data for the subset (n=69) and Figure 3 predicts good/poor outcome based upon on demographic and insult data for the subset (n=69).

Predictive Accuracy

Using only demographic data. Demographic data, when analysed on its own, consistently produces poor accuracy improvements or a slight reduction in accuracy, therefore the decision trees produced are no more accurate than simply predicting the largest outcome category. However, the sub-groups of patients they identified are of interest to the medical community. Given a larger sample of data, or perhaps additional demographic data, these accuracies would probably improve.

Using only insult data. This is the dataset which generally produced the best test accuracy and improvements in predictive accuracy of 2.3 to 9.2% (table 5b).

Using demographic and insult data together. These data produced some improvements in accuracy. However, the decision trees produced as a result of this particular analysis are probably of most interest to the medical community as they show which physiological abnormalities are most useful in summarising cases. For example, age of a patient appearing to affect recovery from a particularly low blood pressure.

Taking all the decision trees generated[24], there are often *good predictors of outcome* in each of the data categories (demographic, insult). The best predictors (i.e. those at "high" positions in the decision trees, or those which appear often in decision trees) are as follows: In the demographic data, severity of injury (minor, moderate, severe), GCSs, age, cause of injury, pupil response on admission and ISS. In the insult (time series physiological) data: hypotension, cerebral perfusion pressure (CPP), bradycardia and ICP.

The decision trees were discussed with three clinicians experienced in head injury. The findings are summarised below under three broad headings: those which confirm current medical views, those which challenge them, and those which raise issues to be resolved.

DISCUSSION

As noted below, the results were discussed with several clinicians experienced in this area. The discussions covered a general review of the decision trees produced, and more particularly we discussed [with clinicians] in general the several anomalous cases which occurred in the analyses. For example, the left most node of figure 4 predicts a good outcome when patients "both pupils react to light" and when "their age <= 52". This node covers 65.7 patients (decimals because of tree pruning); however 10.7 patients do not have good outcomes. These patients represent anomalies and these are the sorts of patients that we discussed with the clinicians - posing them the question “why do you think their outcome was different from the rest of the group".

Views Confirmed

The decision trees confirm current thinking about hypotension and *pupil response being major predictors of poor outcome*. The absence of any pupillary response to light is associated with poor outcome (see figures 1-4). The management of blood pressure was also confirmed (see figures 2&3) as being of importance in the treatment of head injury patients[34].

CPP insults were found to be more important than ICP insults. This also mirrors current clinical thinking and reflects the CPP oriented protocol by which these patients were managed[23]. Recently emphasis has moved back again to ICP as episodes of “neuro-worsening” have been shown to be associated with ICP increases and not CPP changes[21;38]. In the future, Decision Tree analysis might be able to identify sub-groups in whom ICP orientated therapy is more appropriate than CPP therapy. It is unlikely that one therapy will be best suited to all patients, as the process of head injury is complex.

Views Challenged

The way that decision trees represent the cases in *sub-groups was of great interest* to the clinicians. The idea of *hypothesis generation* (suggestion) rather than hypothesis confirmation was also appealing to the clinicians associated with this study (Peter Andrews, Carol Macmillan, Patrick Statham (Consultant Neurosurgeon), David Wright (Consultant in Intensive Care)).

It was suggested that some of the insults (hypertension, tachycardia and bradycardia) are consequences of clinical actions. For example, tachycardia could be the result of a drug being administered to regulate blood pressure.

The *effect of alcohol and morphine* on the Glasgow Coma Score and pupil response on admission was discussed. If a patient is very drunk when the accident happens, their low coma score or unreactive pupils may be due to the alcohol rather than being a symptom of brain damage. Similarly, if there has been other bodily injuries to the patient, it is quite common for morphine to be given as an analgesic. This also affects GCS and pupil response. Such medication may make the patient appear to be in a worse state than they actually are and could explain some of the "better than expected" results.

The scale used for measuring *motor response* is also of interest. Point 1 on the scale means no response. This could be due to severe brain damage or possibly too external factors such as alcohol or morphine. However, point 2 on the scale shows an extension response of the limbs that is always associated with brain damage. This means that patients with a motor score of 2 may be more severely brain damaged than those with a score of 1, meaning that the scale is not a continuous one. As this score is a constituent part of the GCS score, it could explain why a patient with a score of 3 (the lowest possible score) may have better outcome that a patient with a score of 4.

Issues to be Resolved

When considering the different grades of insult in an analysis, it is more realistic to think about higher grades of each insult also embodying insults at a lower threshold. For example, when a patient is suffering grade 2 or grade 3 of an insult, then consider them to be suffering grade 1 of that insult as well (similarly, grade 3 is also grade 2). This would make insult duration and number of occurrences more realistic.

The clinicians were *interested in "mis-classified" cases*. Generally, the clinicians would agree with the sub-groups and expected outcomes given by the decision trees. However, they were interested in those patients who died/had poor outcome when their expected outcome was survival/ good outcome. Perhaps if additional parameters (such as alcohol intake of the patient, whether they are sedated or not) were taken into account, the algorithm could better classify these patients. During the final few hours before death, a patient's heart rate and blood pressure often wildly fluctuates. Usually these parameters will rise to a peak and then fall. This could explain many of the hypotension grade 3 insults and some of the hypertension, bradycardia and tachycardia insults. This "pre-morbid" data is not really useful in predicting outcome of new cases, and so should be removed.

Currently grade 3 of one insult type is not necessarily as bad as grade 3 of a different insult type. For example, grade 3 hypotension is nearly always fatal, however patients do survive small periods of grade 3 ICP insult. *Perhaps the current thresholds that* *define the grades of insult need to be adjusted.*

The decision trees for *prediction of Glasgow Outcome Score* were generally thought to be too big to provide useful predictions; some of the outcome classes contained too few patients to be representative. To improve these decision trees, *more training cases* would need to be provided.

Pyrexia was identified by statistical analysis as an important insult associated with outcome. The authors of the statistical paper considered that the likely causal explanation was that pyrexia increased cerebral metabolic requirements for oxygen, increased excitatory neurotransmitter traffic and exacerbated the injury process. The considerable experimental evidence that hypothermia may be neuro-protective supported this view. However, the decision tree analysis has highlighted pyrexia as an important predictor of favorable outcome. This finding requires further evaluation[36].

Comparison between the Edinburgh statistical analysis and decision tree analysis show a number of similarities, particularly that hypotension is a strong indicator of poor outcome. Important differences include a creation of smaller patient groups by the decision tree, each of which was characterised. The statistical analysis suggests factors that discriminate for the set of patients as a whole. Within one of the smaller patient groups, pyrexia was seen as a predictor of good outcome but was assumed to be a predictor of poor outcome for the logistic regression and discriminate analysis. The statistical analysis and decision tree analysis showed age as an important co-variant in prediction of outcome and the decision tree gave threshold values of 50 years for the prediction of death and 30 years for prediction of good or poor outcome. Although less accurate at the prediction of the individual outcome classes of the GOS, it is useful to note that the discrimination between the better outcome categories

(Good/Moderate/Poor) are more usually at grade 1 insult threshold (for blood pressure, hypoxia and episodes of hypocarbia etc). These results support our hypothesis that even minor physiological derangement may adversely impact upon cerebral oxygen delivery in a brain that has impaired autoregulation. We can hypothesise that grade III physiological insults are associated with death and severe disability and in patients with no grade III insults, the presence or absence of grade I&II insults determines independent or dependent survival. This is important as most grade I insults are easily treated and early detection and prevention might reduce morbidity.

A shortcoming of this analysis is the small numbers in certain sub-groups. The results for these sub-groups can therefore be difficult to explain.

Decision tree classification suggests that refinement of data processing is required.

For example, insult thresholds are classified according to three increasing grades but it is shown from the decision tree analysis that patients never recover from some Grade III insults but they do from others such as pyrexia. Also as a Grade II insult is still a Grade I insult on the analysis done on the duration of grades of insult, processing is required to accommodate this. Furthermore the powerful predictive value of low arterial pressure and low CPP mean that pre-death physiology should not be included in such analysis in future as they bias the data for the predictions including death. It is noteworthy that this principle was not followed in the Acute Physiology and Chronic Health Evaluation modelling program[32;33]. However, we conclude from the decision tree analysis that handling of blood pressure and intracranial pressure is extremely important in the analysis of such head injury datasets.

Previous Literature on Decision Tree Analysis of Head Injured Patients

The accuracy obtained in this Decision Tree analysis study is better than these reported by Choi et al.[11] or Pilah et al.[26]. As the focus of these papers are a technical discussion of the accuracy of their approach, and not a discussion of the medical relevance of their results, these papers are not reviewed further here: interested readers are referred to references 25&26.

CONCLUSION

This study has shown that improved monitoring techniques will give a more realistic estimate of the nature, frequency and duration of secondary pathophysiological insults. These data emerge at a time when there is a major interest in therapies that act on basic mechanisms underlying primary and secondary brain damage such as anti-oxidants, ion-channel blockers, membrane stabilizers and inflammation[25]. What is clear from this study is that secondary insults play a significant role in determining the outcome and occur commonly with current management[13]. In any trial of therapy, their occurrence must be recorded to ascertain whether therapies reduce incidence or improve cerebral tolerance to secondary insults.

Reference List

1. Andrews PD: Head injury: Complications and management. **Current Opinion in Anaesthesiology 11**:473-477, 1998

2. Andrews PD, Dearden NM, Miller JD: Jugular bulb cannulation: Description of a cannulation technique and validation of a new continuous monitor. **British Journal of Anaesthesia 67**:553-558, 1991

3. Andrews PD, Piper IR, Dearden NM, et al: Secondary insults during intrahospital transport of head-injured patients. **Lancet 335**:327-330, 1990

4. Baker SP: Injury classification and the international classification of diseases codes. **Accident Analysis & Prevention 14**:1982

5. Bekar A, Ipekoglu Z, Tureyen K, et al: Secondary insults during intrahospital transport of neurosurgical intensive care patients. **Neurosurgical Review 21**:98-101, 1998

6. Block EJ: Diagnostic modalities in acute trauma. **New Horizons 7**:10-25, 1999

7. Blumer, A. ID3 Decision tree learning algorithm. 1991. Computer Program

8. Chesnut RM: Medical management of severe head injury: Present and future. **New Horizons 3**:581-593, 1995a

9. Chesnut RM: Secondary brain insults after head injury: Clinical perspectives. **New Horizons 3**:366-375, 1995b

10. Chesnut RM: The management of severe traumatic brain injury. **Emergency Medicine Clinics of North America 15**:581-604, 1997

11. Choi SC, Muizelaar JP, Barnes TY, et al: Prediction tree for severely head-injured patients. **Journal of Neurosurgery 75**:251-255, 1991

12. Davies G, Deakin C, Wilson A: The effect of a rigid collar on intracranial pressure. **Injury 27**:647-649, 1996

13. Dearden NM: Mechanisms and prevention of secondary brain damage during intensive care. **Clinical Neuropathology 17**:221-228, 1998

14. Doberstein CE, Hovda DA, Becker DP: Clinical considerations in the reduction of secondary brain injury. **Annals of Emergency Medicine 22**:993-997, 1993

15. Dunn LT: Secondary insults during the interhospital transfer of head-injured patients: An audit of transfers in the Mersey Region. **Injury 28**:427-431, 1997

16. Dzeroski S, Lavrac N: Rule induction and instance-based learning applied in medical diagnosis. **Technology & Health Care 4**:203-221, 1996

17. Graham DI, Adams JH, Nicoll JA, et al: The nature, distribution and causes of traumatic brain injury. **Brain Pathology. 5**:397-406, 1995

18. Graham DI, Ford I, Adams JH, et al: Ischaemic brain damage is still common in fatal non-missile head injury. **Journal.of.Neurology, Neurosurgery &**

Psychiatry 52:346-350, 1989

19. Jennet, B and Bond.M. Assessment of outcome after severe brain damage. A practical scale. Lancet 1, 480-484. 1975.

20. Jones PA, Andrews PD, Midgley S, et al: Measuring the burden of secondary insults in head-injured patients during intensive care. **Journal of Neurosurgical Anesthesiology 6**:4-14, 1994

21. Juul N, Morris GF, Marshall SB, et al: Intracranial hypertension and cerebral perfusion pressure: Influence on neurological deterioration and outcome in severe head injury. **Journal of Neurosurgery 92**:1-6, 2000

22. Lavrac N: Selected techniques for data mining in medicine. **Artificial Intelligence in Medicine 16**:3-23, 1999

23. Maas AR, Dearden M, Teasdale GM, et al: EBIC-Guidelines for management of severe head injury in adults. **Acta Neurochirurgica 139**:286-294, 1997

24. McQuatt, A. Using machine learning techniques to predict clinical outcome. 1998. Computing Science Department, University of Aberdeen. Thesis

25. Morganti-Kossmann MC, Hans VJ, Lenzlinger PM, et al: TGF-beta is elevated in the CSF of patients with severe traumatic brain injuries and parallels blood-brain barrier function. **Journal of Neurotrauma 16**:617-628, 1999

26. Pilah IA, Mladenic D, Lavrac N, et al: Data analysis of patients with severe head injury, in Boston, Kluwer Academic Publishers:1997

27. Quinlan, J. R. Induction of decision trees. Machine Learning 1(1), 81-106. 1986.

28. Quinlan, J. R. Simplifying decision trees. International Journal of Man-Machine Studies 27, 221-234. 1987.

29. Quinlan JR: Data mining tools C5.0 and See5. **Rulequest Research** 1997

30. Robertson CS, Gopinath SP, Goodman JC, et al: SjvO2 monitoring in head-injured patients. **Journal of Neurotrauma 12**:891-896, 1995

31. Sarnaik AP, Lieh-Lai MW: Transporting the neurologically compromised child. **Pediatric Clinics of North America 40**:337-354, 1993

32. Seneff M, Knaus WA: APACHE: A prognostic scoring system. **Problems in Critical Care 3**:563-577, 1989

33. Seneff M, Knaus WA: Predicting patient outcome from intensive care: A guide to APACHE, MPM, SAPS, PRISM, and other prognostic scoring systems. **Journal of Intensive Care Medicine 5**:33-52, 1990

34. Signorini DF, Andrews PD, Jones PA, et al: Adding insult to injury: The prognostic value of early secondary insults for survival after traumatic brain injury. **Journal of Neurology, Neurosurgery & Psychiatry 66**:26-31, 1999

35. Trunkey DD, Siegel J, Baker SP, et al: Panel: Current status of trauma severity indices. **Journal of Trauma-Injury Infection & Critical Care 23**:1983

36. Verweij BH, Muizelaar JP: Avoiding secondary brain injury after severe head trauma: Monitoring and management. **Journal of Cranio-Maxillofacial**

Trauma, Vol 2(3) (pp 8-17), 1996

37. Wald SL, Shackford SR, Fenwick J: The effect of secondary insults on mortality and long-term disability after severe head injury in a rural region without a trauma system. **Journal of Trauma-Injury Infection & Critical Care 34**:377-382, 1993

38. Youn JS, Marshall LF: Cerebral perfusion pressure or intracranial pressure?. **Journal of Neurosurgery 92**:191-192, 2000

Table 1. Demographic features of the 124 computer monitored head injured patients.

	SEVERE	**MODERATE**	**MINOR**
NUMBER	68	36	20
SEX			
male	58	30	16
female	10	6	4
MEAN AGE (years)	34	37	43
DIAGNOSIS			
Evacuated Hematoma	29	13	6
Non Evacuated **Hematoma**	**13**	**14**	**9**
Diffuse Injury	26	7	0
-Other Diagnosis*	0	2	5
INJURY SEVERITY SCORE			
Median ISS	25.0	16.0	20.5
12 MONTH OUTCOME (GOS)			
Dead	21	4	1
Vegetative	0	0	0
Severe Disability	10	6	2
Moderate Recovery	18	4	3
Good Recovery	19	20	13
Lost to Follow up	0	2	1

Other Diagnoses* = Moderate or Minor Head Injury with no intracranial abnormality apparent on CT scan
ISS = Injury Severity Score
GOS = Glasgow Outcome Scale

Table 2. Edinburgh University secondary insult grades (EUSIG)

	GRADE 1	GRADE 2	GRADE 3
Raised ICP mmHg	≥20	≥30	≥40
Hypotension mmHg Systolic	≤90	≤70	≤50
or Mean	≤70	≤55	≤40
Hypertension mmHg Systolic	≥160	≥190	≥220
or Mean	≥110	≥130	≥150
CPP mmHg	≤60	≤50	≤40
Hypoxemia SaO_2 %	≤90	≤85	≤80
or PaO_2 kPa	≤8.0	≤7.0	≤6.0
Increased oxygen extraction $SjvO_2$ %	≤54	≤49	≤45
Decreased oxygen extraction $SjvO_2$%	≥75	≥85	≥95
Hypercarbia kPa	≥6.0	≥8.0	≥10.0
Hypocarbia kPa	≤3.0	≤2.5	≤2.0
Pyrexia ^{o}C	≥38	≥39	≥40
Tachycardia bpm	≥120	≥135	≥150
Bradycardia bpm	≤50	≤40	≤30
Increased oxygen extraction grade only)	Ca-jvO_2 mlO_2/100ml blood	≥9	(one
Decreased oxygen extraction grade only)	Ca-jvO_2 mlO_2/100ml blood	≤4	(one

Table 3. Median duration in minutes of each insult type by grade for sub-group n=69.

Variable	Grade 1	Grade 2	Grade 3
ICP	42	24	21
HYPOTENSION	29	22.5	32
HYPERTENSION	28	21	14
CPP	27	21.5	25.5
HYPOXEMIA	11.5	19	20
Increased oxygen extraction $SjvO_2$ %	28	18.5	60
Decreased oxygen extraction $SjvO_2$%	32	17	80.5
HYPOCARBIA	58	36	13
HYPERCARBIA	39	144	-
PYREXIA	254	195	104
BRADYCARDIA	23	13.5	-
TACHYCARDIA	30.5	25	22.5
Increased oxygen extraction	80	n/a	n/a
Decreased oxygen extraction	54	n/a	n/a

Table 4. Logistic regression at 12 months

Survival vs Death **Good vs Poor**

Outcome

VARIABLE	SIGNIF.		VARIABLE
DURATION OF HYPOTENSION	.0064		DURATION OF HYPOTENSION
DURATION OF PYREXIA	.0137		PUPIL RESPONSE O/A
DURATION OF HYPOXEMIA	.0244		DURATION OF PYREXIA
AGE	.0652		AGE
DURATION OF RAISED ICP	.1162		DURATION OF HYPOXEMIA
DURATION OF HYPERTENSION	.3689		DURATION OF RAISED ICP
ISS	.3855		DURATION OF BRADYCARDIA
GCSs Post-Resuscitation	.3858		ISS
DURATION OF TACHYCARDIA	.4001		DURATION OF HYPERTENSION
PUPIL RESPONSE O/A	.4857		DURATION OF TACHYCARDIA
DURATION OF BRADYCARDIA	.8733		GCSs Post-Resuscitation
GOODNESS OF FIT	90.00%		GOODNESS OF FIT

Table 5a. Decision Tree accuracy (both training and cross-validated).

Accuracy for the analyses of 121 patients data, where x/y/z gives Training Accuracy/Test Accuracy/Predictive Accuracy improvement over biggest class as a percentage (n=121).

Outcome	Biggest Category	Demographic Data	Insult data	Demographic & insult data
Dead/ Alive	**78.5**	**90.9/ 77.6/ -0.9**	**95.0/ 82.7/ +4.2**	**98.3/ 84.2/ +5.7**
Good/ Poor	**63.6**	**82.6/ 64.2/+0.55**	**78.5/ 64.4/ +0.7**	**84.3/ 60.9/ -2.7**
GOS	**43.0**	**81.8/ 44.3/ +1.3**	**62.0/ 47.0/ +4.0**	**80.2/ 44.3/ +1.3**

Table 5b. Accuracy for the Subset of Patients (CDC) where x/y/z gives Training Accuracy/Test Accuracy/Predictive Accuracy improvement over biggest class as a percentage (n = 69).

Outcome	Biggest Category	Demographic Data	Insult data	Demographic & Insult data
Dead/Alive	**78.3**	**85.5/ 73.5/ -4.8**	**94.2/ 87.5/ +9.2**	**94.2/ 87.9/ +9.6**
Good/Poor	**62.3**	**95.7/ 60.4/ -1.9**	**87.0/ 67.4/ +5.1**	**89.9/ 64.0/ +1.7**
GOS	**39.1**	**78.3/ 39.2/ +0.1**	**88.4/41.4/ +2.3**	**85.5/ 37.5/ -1.6**

Figure 1. Predicting good (GOS 4&5) or poor (*bad*, GOS 1-3) outcome using only demographic data for the set of patients, n=69. The numbers in the outcome boxes reflect the total number of cases for that outcome (n) and the number mis-classified cases (m); decimals are due to pruning of the tree (n/m). Diagnosis 1,2&3 = Evacuated extradural hematoma, acute sudural hematoma, intracerebral hemorrhage respectively. Diagnosis 4 = non-evacuated hematoma, Diagnosis 5=diffuse injury and Diagnosis 6&7 = other.

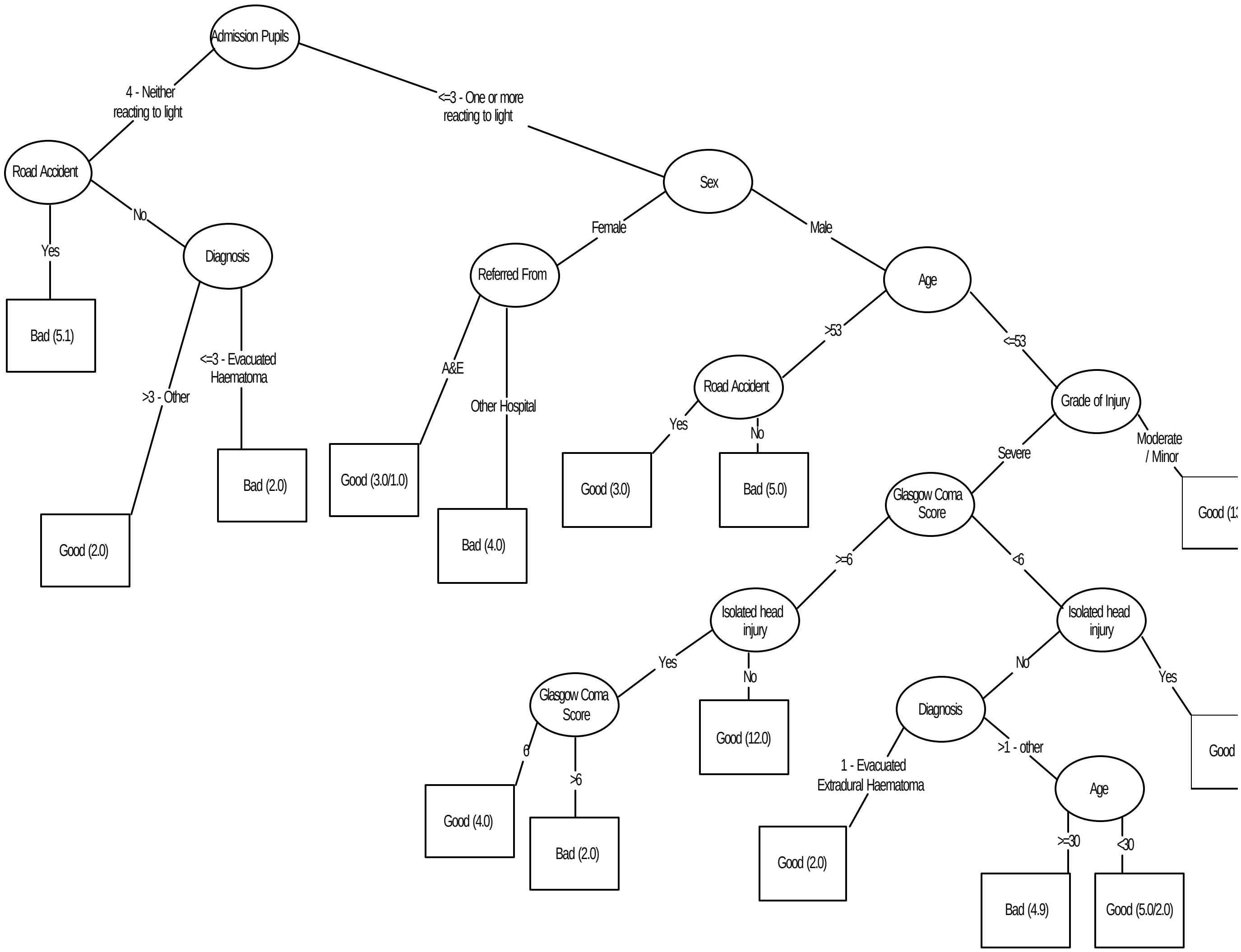

Figure 2. Predicting good (GOS 4&5) or poor (*bad*, GOS 1-3) outcome using only insult data
for the set of patients, n=69. The numbers in the outcome boxes reflect the total number of cases for that outcome (n) and the number mis-classified cases (m); decimals are due to pruning of the tree (n/m).

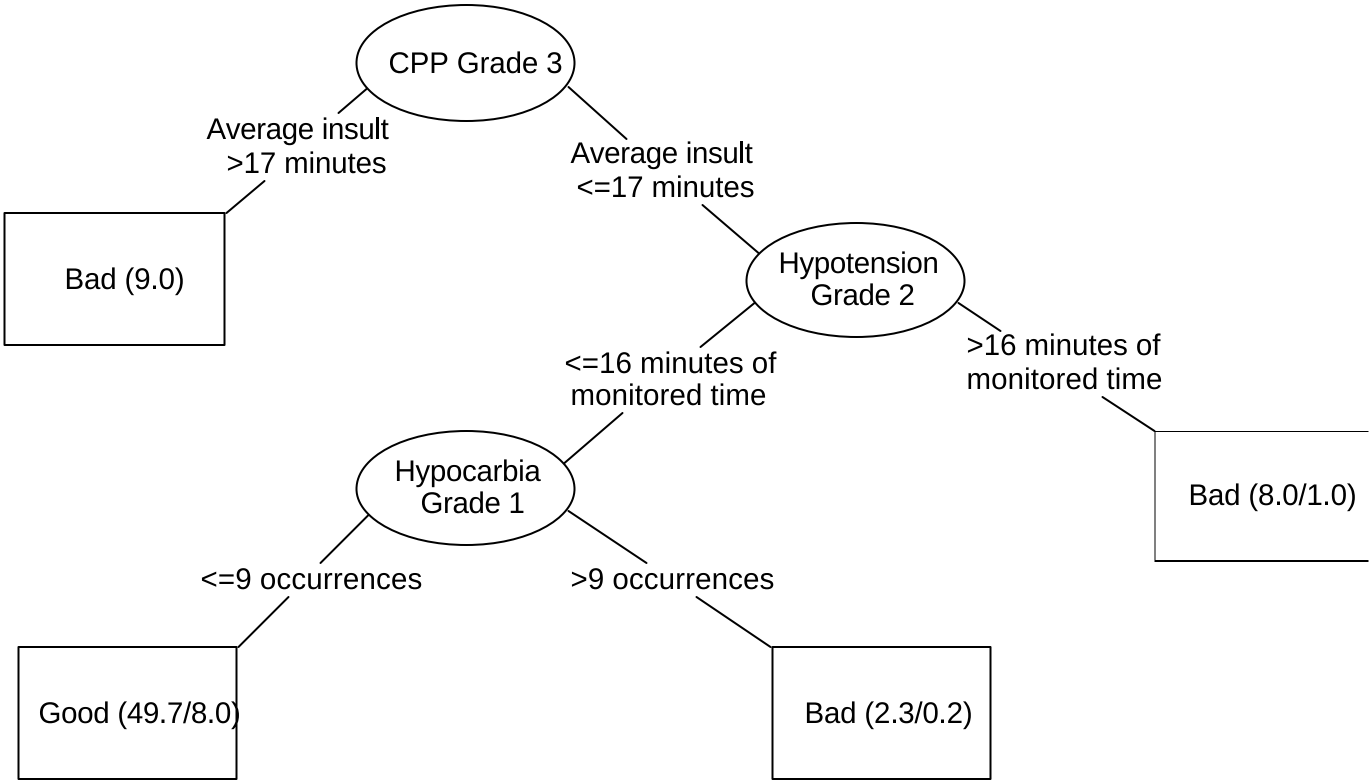

Figure 3. Predicting good (GOS 4&5) or poor (*bad*, GOS 1-3) outcome using demographic and insult data for the set of patients, n=69. The numbers in the outcome boxes reflect the total number of cases for that outcome (n) and the number mis-classified cases (m); decimals are due to pruning of the tree (n/m).

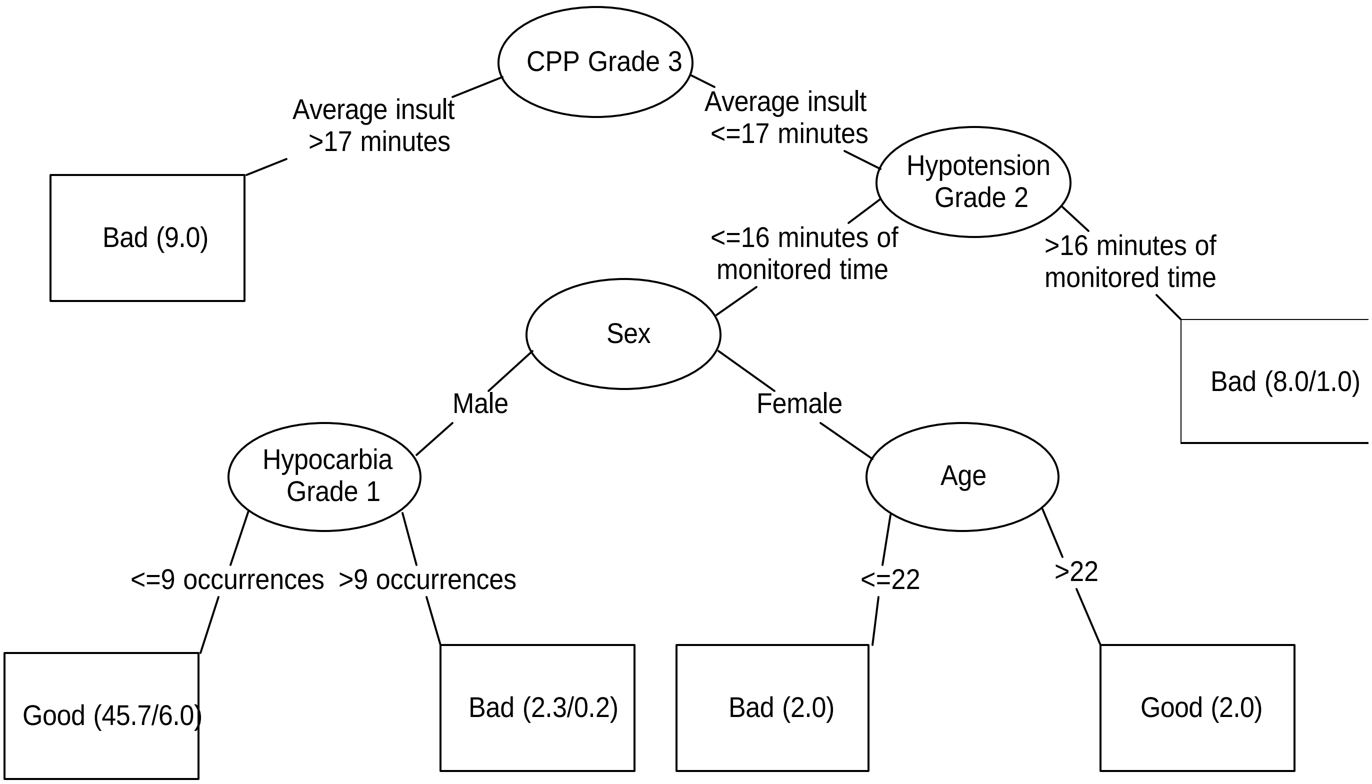

Figure 4. Predicting good (GOS 4&5) or poor (*bad*, GOS 1-3) outcome using only demographic data for the set of patients, n=121. The numbers in the outcome boxes reflect the total number of cases for that outcome (n) and the number mis-classified cases (m); decimals are due to pruning of the tree (n/m).