J Neurosurg 97:326–336, 2002

Predicting recovery in patients suffering from traumatic brain injury by using admission variables and physiological data: a comparison between decision tree analysis and logistic regression

PETER J. D. ANDREWS, M.D., F.R.C.A., DEREK H. SLEEMAN, PH.D., F.R.S.(ED), PATRICK F. X. STATHAM, F.R.C.S.(SN), ANDREW MCQUATT, M.SC., VINCENT CORRUBLE, PH.D., PATRICIA A. JONES, M.APP.SCI., TIMOTHY P. HOWELLS, PH.D., AND CAROL S. A. MACMILLAN, M.R.C.P., F.R.C.A.

Departments of Anaesthetics and Clinical Neuroscience (Neurosurgery), University of Edinburgh, Western General Hospital, Edinburgh; Department of Computing Science, King's College, University of Aberdeen, United Kingdom

Object. Decision tree analysis highlights patient subgroups and critical values in variables assessed. Importantly, the results are visually informative and often present clear clinical interpretation about risk factors faced by patients in these subgroups. The aim of this prospective study was to compare results of logistic regression with those of decision tree analysis of an observational, head-injury data set, including a wide range of secondary insults and 12-month outcomes.

Methods. One hundred twenty-four adult head-injured patients were studied during their stay in an intensive care unit by using a computerized data collection system. Verified values falling outside threshold limits were analyzed according to insult grade and duration with the aid of logistic regression. A decision tree was automatically produced from root node to target classes (Glasgow Outcome Scale [GOS] score).

Among 69 patients, in whom eight insult categories could be assessed, outcome at 12 months was analyzed using logistic regression to determine the relative influence of patient age, admission Glasgow Coma Scale score, Injury Severity Score (ISS), pupillary 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 and poor outcomes were compared, hypotensive insults and pupillary response on admission were significant.

Using decision tree analysis, the authors found that hypotension and low cerebral perfusion pressure (CPP) are the best predictors of death, with a 9.2% improvement in predictive accuracy (PA) over that obtained by simply predicting the largest outcome category as the outcome for each patient. Hypotension was a significant predictor of poor outcome (GOS Score 1–3). Low CPP, patient age, hypocarbia, and pupillary response were also good predictors of outcome (good/poor), with a 5.1% improvement in PA. In certain subgroups of patients pyrexia was a predictor of good outcome.

Conclusions. Decision tree analysis confirmed some of the results of logistic regression and challenged others. This investigation shows that there is knowledge to be gained from analyzing observational data with the aid of decision tree analysis.

KEY WORDS • head injury • secondary cerebral insult • decision tree analysis • intensive care unit • outcome

PATIENTS who sustain head trauma suffer not only from the effects of the primary injury, but also from secondary, mainly ischemic, brain damage. In 1978 Graham and colleagues[19] reported a 91% incidence of ischemic brain damage in a series of 151 fatal cases of severe head injury that had been subjected to autopsy. The same group of authors[20] reported on another 112 fatalities resulting from severe head injury in the 1980s; despite even more intensive treatment, the incidence of ischemic brain damage among these patients remained higher than 80%. Other authors[10,11,15] have considered these patients as those who "talk and die" or "talk and deteriorate." In such cases the primary injury is judged to have been not severe; rather it is subsequent events that have resulted 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 while the patient is in the ICU.[2,4,13] Between 1993 and 1998 a number of authors re-

Abbreviations used in this paper: BP = blood pressure; Ca-jvO_2 = cerebral arteriojugular venous oxygen content difference; CoDC = computerized data collection; CPP = cerebral perfusion pressure; ETCO_2 = end-tidal CO_2; EUSIG = Edinburgh University Secondary Insult Grades; GCS = Glasgow Coma Scale; GOS = Glasgow Outcome Scale; ICP = intracranial pressure; ICU = intensive care unit; ISS = Injury Severity Score; MABP = mean arterial BP; SaO_2 = arterial oxygen saturation; SjvO_2 = jugular venous oxygen saturation.

ported that in comatose head-injured patients arriving in the emergency room, hypoxemia was demonstrated in 30% and arterial hypotension in 15%.[9,16,34,36,37] More recently, probably because of better at-the-scene resuscitation efforts and transport arrangements, there has been a reduction in the frequency of these early insults.[16]

Predicting recovery of patients suffering from traumatic brain injury is a complex but important task, and any predictive model is likely to be better calibrated when information on secondary insults is included. These data are valuable in discussing prognosis with relatives and in considering the greater issues of resource allocation, comparison of patient outcomes among ICUs,[6,7] and the effect of novel pharmaceutical compounds on recovery. There have been many previous attempts to predict outcome after traumatic brain injury. Researchers who have used statistical analysis, including complex modeling, however, have considered all patients as one homogeneous pool and have assumed that each variable will similarly affect each patient within this diagnostic category of head injury. The advantage of using decision tree analysis is that with it one can create subgroups of patients and identify factors that are important to the individual or subgroups. These analyses are visually informative, often challenge current knowledge, and suggest new hypotheses for future testing. Therefore, there is considerable theoretical advantage in using this type of analysis when exploring large, clinical, observational data sets.

We have produced a number of decision trees to study a historical traumatic brain injury data set that was previously analyzed with the aid of statistical techniques to compare the two methodologies.

The aims of this study were the following: to identify areas in which results from the decision tree analysis agreed with those from statistical analysis; to identify areas in which there were differences between the two analyses; and to look at unexpected or unexplained results that challenged our understanding of the process following traumatic brain injury.

Clinical Material and Methods

Patient Population

This prospective study group consisted of 124 patients aged 16 years and older, who had been admitted to the regional Head Injury Unit located in Edinburgh between January 1, 1989 and June 16, 1991. A statistical analysis of the data has been previously reported.[22]

Admission details were retrieved from the referring hospital (68 patients) or from the emergency room for patients brought directly from the scene of the accident (56 patients). Data included cause and nature of injury, patient age, GCS score of the patient on arrival at the hospital and after resuscitation, pupillary response, results of computerized tomography scans, surgical treatment, and medical therapy.

Patients were eligible for this study if the following inclusion criteria were met: 1) a postresuscitation GCS score less than 13 or GCS score greater than 12 if the ISS[5,18] was greater than 15; 2) clinical indications for monitoring the patient in the ICU; and 3) a CoDC available for use within 24 hours of hospital admission. Patients were classified according to the following postresuscitation GCS scores: severe head injury, a GCS score of 8 or less, with no eye opening; moderate head injury, a postresuscitation GCS score from 9 to 12 (or a GCS score $\leq$ 8, with eye opening); and minor head injury, GCS score from 13 to 15 that was associated with other injuries sufficient to warrant an ISS of at least 16.[5] The management strategy for patients with head injury was performed by experienced neurosurgical, anesthesiological, and nursing staff and remained unchanged during the study period.

Mechanical ventilation and ICP monitoring as well as additional invasive monitoring when clinically indicated were used; thus, data were collected for only those physiological parameters appropriate for each patient. Patients remained connected to the CoDC unit until clinical monitoring ceased, and in some cases this period extended to as long as 2 to 3 weeks. Only two CoDC systems were available, and inevitably there were patients who fulfilled the inclusion criteria but could not be connected to a system because both were already in use. From a total population of 2010 head-injured patients admitted to the ward during this time period, 429 (51 children and 378 adults) fit the study criteria. Computer-monitored data were collected from 124 adults (33%). Overall, 45% of the patients with severe head injuries were monitored with the aid of a computer (Table 1) and were representative of the study group, but only 18% of patients with moderate and 20% of those with minor head injuries could be connected to the computer system (and were the more severely injured in these groups).

TABLE 1

Demographic features of 124 head-injured patients who underwent computer monitoring

Variable	Type of Head Injury		
	Severe	Moderate	Minor
no. of patients	68	36	20
sex			
M/F	58:10	30:6	16:4
mean age (yrs)	34	37	43
diagnosis			
evacuated hematoma	29	13	6
nonevacuated hematoma	13	14	9
diffuse injury	26	7	0
other*	0	2	5
median ISS	25	16	20.5
12-mo GOS score			
death	21	4	1
vegetative state	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

* Moderate or minor head injury with no intracranial abnormality apparent on computerized tomography scan.

Computerized Collection of Physiological Data

A continuous data collection system, based on a personal computer, was developed with the capability of recording, minute-by-minute, at least 14 physiological variables.[22] Monitored variables included heart rate (measured by an electrocardiograph), systolic pressure and MABP (mea-

TABLE 2
*Secondary insult grades according to the EUSIG**

Insult Type	Grades of Secondary Insult 1	2	3
raised ICP (mm Hg)	≥20	≥30	≥40
hypotension (mm Hg)			
systolic BP	≤90	≤70	≤50
MABP	≤70	≤55	≤40
hypertension (mm Hg)			
systolic BP	≥160	≥190	≥220
MABP	≥110	≥130	≥150
low CPP (mm Hg)	≤60	≤50	≤40
hypoxemia			
SaO_2 (%)	≤90	≤85	≤80
PaO_2 (kPa)	≤8	≤7	≤6
increased OEF (% $SjvO_2$)	≤54	≤49	≤45
decreased OEF (% $SjvO_2$)	≥75	≥85	≥95
hypercarbia (kPa)	≥6	≥8	≥10
hypocarbia (kPa)	≤3	≤2.5	≤2
pyrexia (˚C)	≥38	≥39	≥40
tachycardia (bpm)	≥120	≥135	≥150
bradycardia (bpm)	≤50	≤40	≤30
increased OEF†	≥9		
decreased OEF†	≤4		

* bpm = beats per minute; OEF = oxygen extraction fraction.
† One grade only.

sured invasively), ICP (recorded using a Camino catheter), SaO_2 (measured using a pulse oximeter), $SjvO_2$,[3] $ETCO_2$, and peripheral and core temperatures. From these variables, CPP (MABP − ICP) and Ca-jvO_2 were calculated online: Ca-jvO_2 = (SaO_2−$SjvO_2$) × hemoglobin × 1.39 ml O_2/100 ml blood.

Definitions and Scoring of Secondary Insults

At the outset of this study, threshold limits were set at abnormal values to define insult grades for each variable, which were based on earlier research[4,22,33] and clinical experience. Values falling outside these limits were displayed in a 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 for which only one grade of insult was recorded for increased or decreased oxygen extraction. The ranges of abnormal values used, according to the EUSIG, are listed in Table 2 and were analyzed offline by using appropriate software.

To be considered a secondary insult, abnormal values had to persist for 5 consecutive minutes or longer. 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; when body temperature returned to normal for 1 complete hour, the insult was considered to be ended to reflect the slower rate of change of this variable. The date, insult grade, and start and stop time for each insult type were recorded. A change in the 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 (< 37˚C). Surface cooling and nasogastric or rectal administration of paracetamol (1 g) were the principal interventions used to facilitate this. Software automatically detects whether a monitor channel is working by testing its internal error flags, as well as by comparing valid data against upper and lower rejection limits specific for each channel. Values falling outside these limits are deemed nonphysiological and rejected. The system also allows for the insertion of comments by nurses and doctors, with the time of day recorded so that additional events that may be classified as an artifact can be identified (for example, nursing maneuvers or blood gas sampling).

Several channels of data had to be verified against other results before being accepted as valid. The $ETCO_2$ insults were verified from arterial blood gas results. The $SjvO_2$ values were validated against those on the printout obtained with the oximeter to ensure adequate light intensity readings and twice daily calibration against a cooximeter.[3] Acceptable simultaneous recordings of ICP and MABP were prerequisites for CPP calculation. Computation of Ca-jvO_2 necessitated correct data for SaO_2, hemoglobin, and $SjvO_2$. If there was doubt as to the validity of abnormal values shown, the values were rejected. Final identification and verification of all insults by grade and duration were manually performed by four research personnel. Although time consuming and labor intensive, this process was judged to be essential to ensure that data were reliable and accurate. The intention was therefore to underestimate insults, rather than to include dubious data.

Intracranial pressure data recorded in this study included readings obtained during treatment of intracranial hypertension. We corrected, when possible, remediable causes of raised ICP and then treated raised ICP when it 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 occurred periodically for the following reasons: while computer files were being backed up, when patients were removed from the ICU for computerized tomography scan or surgery, during some nursing and physiotherapy procedures, or due to computer failure.

Outcome Scoring

The five-point GOS[21] was used to measure patient outcome, because it is widely used and reliable. Before assigning an outcome score, information was collected from several sources: a patient interview, results from a battery of seven neuropsychometric tests, a 60-item questionnaire answered by relatives, a letter from the patient's general practitioner, medical notes, and outpatient clinic letters. When assigning a GOS score, the scorer was unaware of the secondary insult data. Scoring was performed 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 score could not be ascertained.

Statistical Analysis

Start and stop times and grade of insult were entered into a computer database. Data compilation and analysis, including tabulation of insult grades and duration of each insult, were performed using appropriate software. This report presents results of 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 patient age was made primarily by using logistic regression. Results of other statistical tests (Kruskal–Wallis test, Mann–Whitney U-test, and chi-square test) supplemented the findings. Significance was set at the level of 0.05.

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.[31]

There has been much work done on decision trees since the 1970s, both by Quinlan[32,33] and others in the machine learning and statistics communities.[17,25] The decision trees in this report were generated with the aid of a personal computer application that uses the C5.0 algorithm, which is an upgraded version of the C4.5 algorithm.[8,30] Note that because the new upgrades for C5.0 have not be utilized in our analysis, the decision trees in this report can be considered as being the same as those produced with the C4.5 algorithm.

The power of decision tree analysis lies in its ability to indicate 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, that is, some cases will be assigned to one class although they belong to another. These errors are represented in the leaf nodes of the decision tree 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. The training accuracy is a measure of how accurately the tree represents all known cases. These figures give little indication of how well the decision trees can predict the outcome of a new patient, however. Predictive accuracy calculated using 10-fold crossvalidation gives a better measure of predicted outcome (see later). Nonetheless, it could be misleading to consider, for example, that an 80% rate of accuracy is very good if 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 the population) as a comparison.

The standard method of testing the predictive accuracy of a decision tree, that is, how well the tree classifies new cases, is to use new, as-yet-unseen cases from the same population. The problem with this approach is that the predictive accuracy is very dependent on which cases are included in the test and training sets. To avoid any problems, the new set of cases is randomly split into a number of equally sized subsets (for example, 10). Each subset in turn becomes the test set, with the remaining nine subsets forming the training set; thus, 10 different decision trees must be generated and tested. The overall predictive accuracy then is the average of all the predictive accuracies. This method is called crossvalidation and, because it uses 10 subsets, it is referred to more specifically as a 10-fold crossvalidation. When the crossvalidation is repeated, however, it will randomly choose 10 different subsets and so produce a different overall predictive accuracy. For this reason, the average of 10 10-fold crossvalidation tests is used for an overall predictive accuracy measurement.

Sources of Supplies and Equipment

Jugular venous oxygen saturation was measured using the Oximeter 3 system provided by Abbott Critical Care Systems (Morgan Hill, CA). The Edinburgh browser and monitor software were written by Timothy Howells. The cooximeter (IL482) was obtained from Instrument Laboratory Co. (Lexington, MA). Data were compiled and analyzed with the aid of software purchased from SPSS, Inc. (Chicago, IL). The Dbase IV computer database used during statistical analysis was manufactured by Ashton Tate/Borland (Scotts Valley, CA). The ID3 decision tree learning algorithm and data mining tools C5.0 and See5 were used during decision tree analysis.[30,31]

Results

Severity of Head Injury and Data Sets Studied

Of the 124 patients in this study, 68 had sustained severe head injuries, 36 moderate, and 20 minor. Data sets studied included the following groups: 1) all 124 patients; 2) 121 patients whose GOS score was known at 12 months; 3) a subset of 71 patients in whom a minimum of eight channels of physiological data were monitored in the ICU, with secondary insult processing; and 4) 69 patients from the preceding subgroup who had known outcomes at 12 months (two of the 71 were known to be alive but their GOS scores were unknown). Diagnosis and outcome data are listed in Table 1.

Injury Severity Score

Median ISSs for the groups of patients with severe, moderate, and minor head injuries were 25, 16, and 20.5, respectively, and the overall median ISS was 25. Sixty-five of the patients had an ISS of 25 or higher. Patients with multiple injuries were defined as those patients with a head injury in association with another injury assigned an Abbreviated Injury Score greater than 1. The numbers of patients with isolated head injuries and those with associated multiple injuries were 53 and 71, respectively.

Secondary Insults

Distribution. All insults identified in this study were found in the subset of valid data collection time remaining after deducting lost CoDC time. Those channels using pressure devices (ICP, MABP, and CPP) had valid CoDC times of the total CoDC time of 91%, 86%, and 88%, respectively. This apparent inconsistency in the percentages of MABP

TABLE 3

*Median duration of insult type by grade for 69 patients**

Variable	Insult Grade (min) 1	2	3
raised ICP	42	24	21
hypotension	29	22.5	32
hypertension	28	21	14
low CPP	27	21.5	25.5
hypoxemia	11.5	19	20
increased OEF (% $SjvO_2$)	28	18.5	60
decreased OEF (% $SjvO_2$)	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 OEF	80	NA	NA
decreased OEF	54	NA	NA

* NA = not applicable; — = unknown.

and CPP is due to the considerably longer total duration of MABP monitoring. The channels for heart rate and pyrexia were found to have a similar validation: hypoxemia and $ETCO_2$ were valid for 80% and 74%, respectively, of the total CoDC time. The $SjvO_2$ channel and its derivatives had the least valid CoDC time at 56% of the total recording time.

Secondary insults were detected in 113 (91%) of the patients studied during computer monitoring in the ICU. Values were at insult level for nearly 10^6 minutes, of which 83.5% (804,276 minutes) were at the least severe EUSIG Grade 1 level. Hypertensive and hypotensive insults were demonstrated in 89% and 73% of patients during CoDC, respectively; however, almost 90% of the total duration of these BP insults remained at Grade 1 level. A total of 77 patients underwent monitoring of ICP, and 65 (84%) had at least one ICP insult and 61 (81%) had one or more CPP insults. Insults were experienced by patients in all injury grades and in all parameters monitored.

Insults of short duration were common. The median insult durations for ICP Grades 1 and 3 were 42 and 21 minutes, respectively, whereas the median time for hypoxemia Grade 1 was only 11.5 minutes (Table 3).

TABLE 4

Logistic regression analysis at 12-month outcome in 69 patients

Variable	Outcome (probability): Survival Compared W/ Death	Good Compared W/ Poor
patient age	0.0652	0.0964
duration of hypotension	0.0064	0.0118
pupillary response on admission	0.4857	0.0226
duration of pyrexia	0.0137	0.0772
duration of hypoxemia	0.0244	0.1217
duration of raised ICP	0.1162	0.1941
duration of hypertension	0.3689	0.6133
ISS	0.3855	0.5701
postresuscitation GCS score	0.3858	0.9051
duration of bradycardia	0.8733	0.3737
duration of tachycardia	0.4001	0.6327
goodness of fit	90.00%	83.32%

TABLE 5

*Accuracy of analyses of data obtained in 121 patients**

Outcome	Largest Category	Analysis of Data (%): Demographic	Insult	Demographic & Insult
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 score	43.0	81.8/44.3/+1.3	62.0/47.0/+4.0	80.2/44.3/+1.3

* Values indicate percentage of improvement over largest class in training accuracy/test accuracy/predictive accuracy.

Relationship to Morbidity and Mortality. The subset of 71 patients in whom computerized data collection was obtained for the eight variables of ICP, arterial hypo- and hypertension, CPP, pyrexia, hypoxemia, bradycardia, 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 variables should be included in the analysis and therefore CPP was excluded. Eleven factors (age, GCS score, and pupil response on admission, ISS, and insult duration of seven insults) were subsequently entered into the regression algorithm.

When patient mortality (death compared with 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). The age variable was demonstrated to be very close to significance in this group of patients ($p = 0.0652$).

When good and poor outcomes were compared, that is, good recovery and moderate disability compared with severe disability, vegetative state, or death (69 patients), only two 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).

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

Decision Tree Analysis

Decision trees were generated for nine different com-

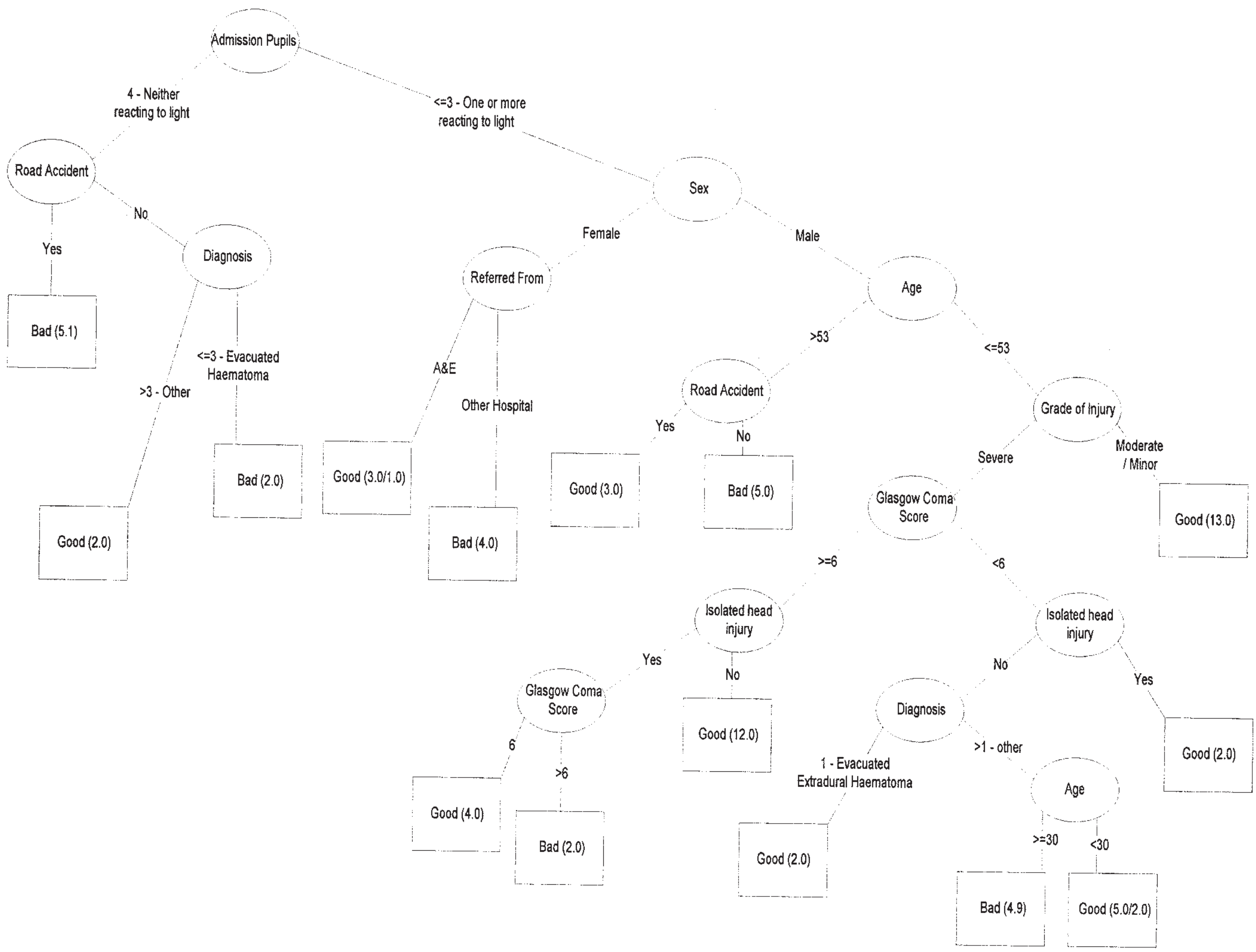

FIG 1. Decision tree analysis set up to predict good (GOS Scores 4 and 5) or bad (GOS Scores 1–3) outcome by using only demographic data from 69 patients. Numbers in the outcome boxes reflect the total number of cases for that outcome (n) and the number of misclassified cases (m); decimals appear due to pruning of the tree (n/m). Diagnoses: 1, evacuated extradural hematoma; 2, acute subdural hematoma; 3, intracerebral hemorrhage; 4, nonevacuated hematoma; 5, diffuse injury; 6 and 7, other.

binations of data. Three sets of data (demographic, insult [based on physiological data over time], and demographic together with insult), were combined with three types of predictions (death/survival, good/poor outcome, and the GOS score[21]).

Data sets are the same as those used in the statistical analysis (data sets one–four). Tables 5 and 6 give the results for the nine combinations of data discussed earlier. The first column is the largest category and is provided as an item of comparison. Decision trees were generated using the incomplete data from the 121 patients (decision trees can handle missing data) and also from the 69-patient subset. Note that improvements in accuracy are greatest when data are complete.

All decision trees described subsequently are based on the 69-patient data set to allow comparison with the results of statistical analysis (Figs. 1–3).

Although there were a number of different decision trees generated by this analysis, we show only a limited selection here; the rest are presented in full elsewhere.[27] Figure 1 illustrates a decision tree for predicting good or poor (bad) outcome, which is based only on demographic data for the 69-patient data set; Fig. 2 predicts good or poor (bad) out-

TABLE 6

*Accuracy of analyses of data obtained in 69 patients**

		Analysis of Data (%)		
Outcome	Largest Category	Demographic	Insult	Demographic & Insult
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 score	39.1	78.3/39.2/+0.1	88.4/41.4/+2.3	85.5/37.5/−1.6

* Values indicate percentage of improvement over largest class in training accuracy/test accuracy/predictive accuracy.

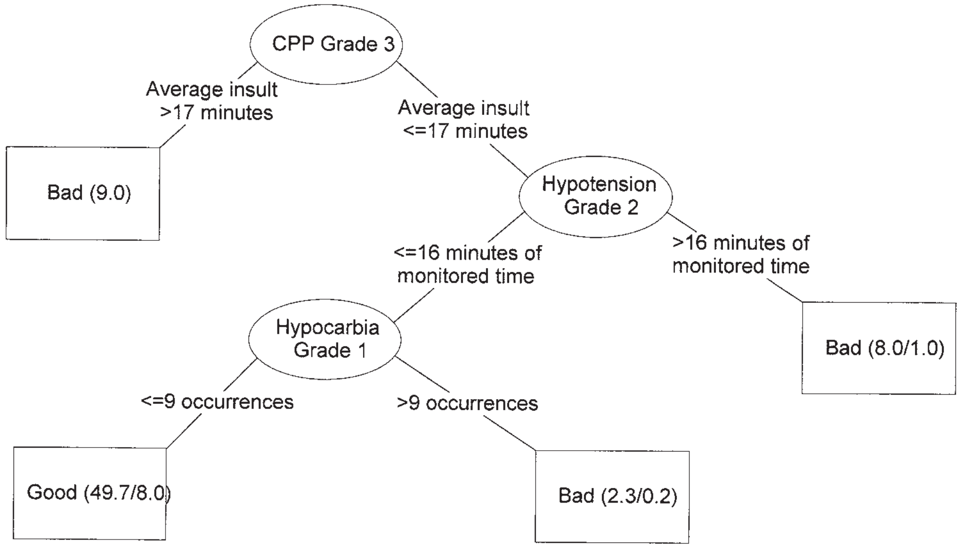

FIG. 2. Decision tree analysis set up to predict good (GOS Scores 4 and 5) or bad (GOS Scores 1–3) outcome by using only insult data from 69 patients. See Fig. 1 legend for explanatory notes.

come based on insult data for the subset of 69 patients; and Fig. 3 predicts good or poor outcome based on combined demographic and insult data for the subset of 69 patients.

Predictive Accuracy of Decision Trees

Using Only Demographic Data. Demographic data, when analyzed separately, consistently produces poor accuracy improvements or a slight reduction in accuracy; therefore, decision trees produced using this data are no more accurate than simply predicting the largest outcome category. Nonetheless, the subgroups of patients identified are of interest to the medical community. Given a larger sample of data or perhaps additional demographic data, predictive accuracies of decision trees based on demographic data alone would probably improve.

Using Only Insult Data. Decision trees based on insult data generally produced the best test accuracy and improvements in predictive accuracy of 2.3 to 9.2% (Table 6).

Using Demographic and Insult Data. These combined data produced some improvements in accuracy. Nonetheless, the decision trees produced with these data are probably of most interest to the medical community because they show which physiological abnormalities are most useful in summarizing cases (for example, the age of a patient appearing to recover from particularly low BP).

Considering all the decision trees generated,[27] we note that there are often good predictors of outcome in each of

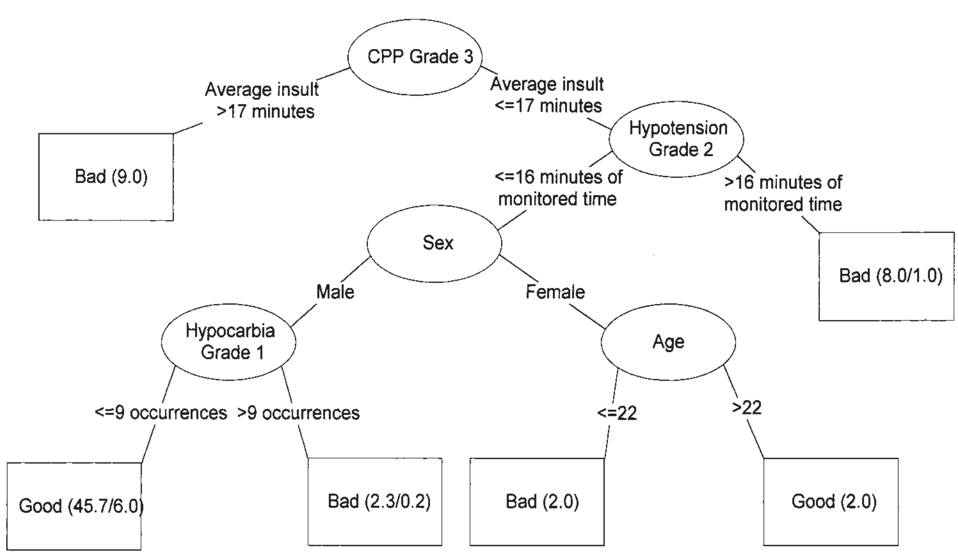

FIG. 3. Decision tree analysis used to predict good (GOS Scores 4 and 5) or bad (GOS Scores 1–3) outcome using both demographic and insult data from 69 patients. See Fig. 1 legend for explanatory notes.

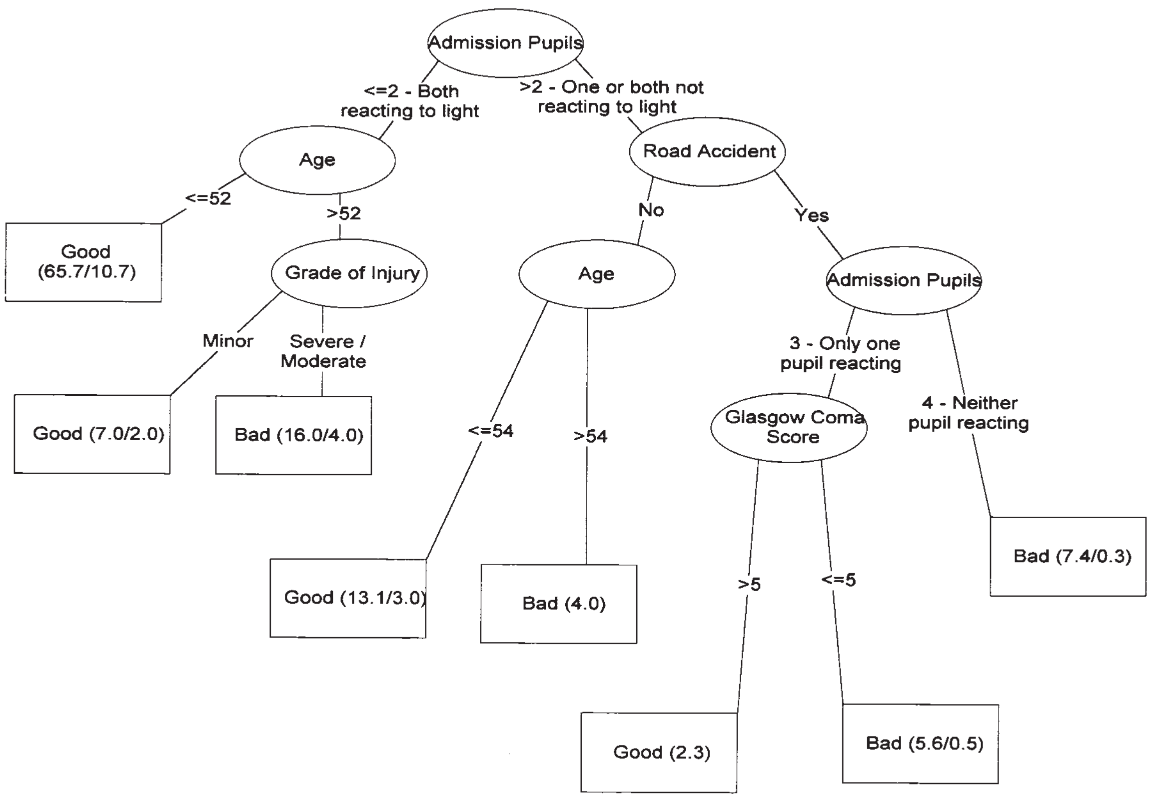

FIG. 4. Decision tree analysis used to predict good (GOS Scores 4 and 5) or bad (GOS Scores 1–3) outcome by using only demographic data from 121 patients. See Fig. 1 legend for explanatory notes.

the data categories (demographic and insult). The best predictors, that is, those at high positions in the decision trees or those that appear often, in the demographic and insult data sets, respectively, are as follows: severity of injury (minor, moderate, severe), GCS score, patient age, cause of injury, pupillary response on admission, and ISS; and hypotension, CPP, bradycardia, and ICP.

The decision trees were discussed with three clinicians experienced in head injury. The findings are summarized in the *Discussion*.

Discussion

The results of the statistical and decision tree analyses were discussed with several clinicians experienced in the fields of neurosurgery and intensive care management of head-injured patients. These discussions covered a general review of the decision trees produced and, more particularly, the several anomalous cases that appeared during the analyses. For example, the node located farthest left on Fig. 4 predicts a good outcome for patients aged 52 years or younger and whose pupils react to light. This node applies to 65.7 patients (decimals appear because of tree pruning); however, 10.7 patients have poor outcomes. Such examples of patient-outcome anomalies were discussed with clinicians by posing the question, "why do you think their outcome was different from the rest of the group." The conclusions of these discussions are summarized in the following sections.

Views Confirmed

Data obtained using decision tree analysis confirmed current thinking about hypotension and pupillary response as being major predictors of poor outcome. The absence of any pupillary response to light was associated with poor outcome (Figs. 1–4). Also, management of BP was proved to be important in the treatment of patients after head injury (Figs. 2 and 3).[35]

Cerebral perfusion pressure insults were found to be more important than ICP insults. This finding mirrors current clinical thinking and reflects the CPP-oriented protocol by which these patients were managed.[26] Recently, emphasis has moved again to ICP because episodes of "neuroworsening" have been shown to be associated with ICP increases and not changes in CPP.[23,38] In the future, decision tree analysis might be used to identify subgroups of patients in whom ICP-orientated therapy is more appropriate than CPP-oriented therapy. It is unlikely that a single therapy will be best suited to all patients, especially given that the pathological processes after head injury are complex.

Views Challenged

The graphic representation of clinical cases in subgroups by using the decision tree analysis was of great interest to the clinicians involved in this study. This approach more closely follows clinical decision making than statistical analysis and was found to be visually informative. The possibility of actually generating hypotheses rather than merely confirming existing ones was also appealing to the clinicians (P.J.D.A., C.S.A.M., P.F.X., and a consultant David Wright). A number of possible avenues of clinical research have become evident from our analysis.

Certain insults (hypertension, tachycardia, and bradycardia) are potential consequences of clinical or pharmacological intervention. For example, tachycardia and hypertension could result from the administration of inotropic or vasoactive medication to regulate BP. The current analysis did not include medical and nursing maneuvers, but future analysis and data interpretation should take into account medication given and respiratory therapy and nursing interventions. Such day-to-day patient care may relate to clinical outcomes either independently or through secondary insults (for example, tracheal suction leading to hypertension).

The effect of alcohol and morphine on the GCS score and pupillary response on admission was also discussed. If intoxicated when the trauma-inducing accident happens, a patient's low GCS score or unreactive pupils may be due to alcohol or drugs (or both) rather than symptoms of brain damage. Similarly, if the patient has incurred other injuries, it is quite common for morphine to be administered or, in other cases of severe injury, sedation and paralysis may be induced and intubation initiated. Such interventions also affect the GCS score and pupillary response, and may make the patients appear to be more seriously injured neurologically than they actually are, which could explain some better-than-expected results.

The GCS score used for measuring the best motor response is also of interest. Point 1 on the scale denotes no response, which could be due to severe brain damage or possible external factors such as alcohol or drugs. Point 2 on the scale, an extension response of the limbs, is always associated with brain damage. This means that a patient with a motor score of 2 may be more severely brain damaged than a patient with a score of 1. The significance of this observation is that the motor score is not a continuous scale. Because this score is a constituent part of the GCS, it could explain why a patient with a total score of 3 (the lowest possible score on the scale) may have a better outcome than a patient with a score of 4.

Issues to be Resolved

When considering different grades of insult in an analysis, it is more realistic to think about the higher grades of each insult as including those at a lower grade. For example, if a patient is suffering from Grade 2 or Grade 3, then consider them to be suffering from a Grade 1 insult of the same type (similarly, Grade 3 includes Grade 2). From this perspective, insult duration and number of occurrences become more realistic.

Misclassified cases were of considerable interest. Generally, the clinicians agreed with the subgroups and expected outcomes in the decision trees. Nevertheless, they were interested in those patients who died or had a poor outcome when the expected outcome was survival or a good outcome. Perhaps if additional parameters (such as alcohol intake of the patient or whether the patient was sedated) were taken into account, the algorithm could better classify these patients. During the final few hours before death, a patient's heart rate and BP often wildly fluctuate. Usually these parameters will rise to a peak and then fall; this phenomenon could explain some of the hypotension Grade 3 insults and some of the hypertension, bradycardia, and tachycardia results. Of the 15 patients who died in this cohort, data collection stopped in nine when it became obvious that additional interventions were futile before they died (at least 4 hours before death). There were five patients in whom data collection continued until the time of death, and one in whom monitoring was stopped 45 minutes before death. A review of the nursing charts and the minute-by-minute physiology data indicate that death was sudden and not inevitable in these cases. These premorbid data are not useful in predicting outcome of new cases and should be excluded from analysis in the future.

Currently Grade 3 of one insult type is not necessarily as adverse as Grade 3 of another insult type. For example, Grade 3 hypotension is nearly always fatal; however, patients do survive short periods of Grade 3 ICP insult. Perhaps the current thresholds that define grades of insult must be adjusted.

Decision trees for the prediction of GCS scores were generally thought to be too large to provide useful predictions, and some of the outcome classes contained too few patients to be representative. To improve these decision trees, more training cases must be provided.

During statistical analysis pyrexia was identified as an important insult associated with outcome. Authors of the statistical paper asserted 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 neuroprotective supported their view. Within one of the smaller patient groups, pyrexia was seen as a predictor of good outcome but a predictor of poor outcome for the logistic regression and discriminate analyses. This finding requires further evaluation.[36]

Comparison between the Edinburgh statistical analysis and our decision tree analysis shows a number of similarities, particularly that hypotension is a strong indicator of poor outcome. An important difference includes the creation of smaller patient groups by using the decision tree, each of which was characterized. With the statistical analysis, one uncovers factors that discriminate for the set of patients as a whole. Both the statistical analysis and decision tree analysis revealed age as an important covariant in the prediction of outcome, and the decision tree gave threshold values of 50 years of age for the prediction of death and 30 years of age for the prediction of good or poor outcome. Although less accurate in the prediction of individual outcome classes of the GOS, it is interesting to note that the distinction between better outcome categories (good/moderate/poor) is usually more clear at Grade 1 insult thresholds (for BP, hypoxia, and episodes of hypocarbia, and so forth). These results support our hypothesis that even minor physiological disruption may adversely affect cerebral oxygen delivery in a brain that has impaired autoregulation. We can hypothesize that Grade 3 physiological insults are associated with death and severe disability and that in patients with no Grade 3 insults, the presence or absence of Grade 1 and 2 insults determines independent or dependent survival. This is important because most Grade 1 insults are easily treated, and early detection and prevention might reduce morbidity.

A shortcoming of our analysis is the small numbers in certain subgroups, thus making results for these subgroups difficult to explain.

Decision tree classification indicates that refinement of data processing is required. For example, insult thresholds were classified according to three increasing grades; however, it has been shown with the decision tree analysis that although patients never recover from some Grade 3 insults, they do recover from others such as pyrexia. Also, given that a Grade 2 insult incorporates a Grade 1 insult in the analysis performed 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 predeath physiology should not be included in such analysis in the future because these factors bias the data for predictions, including death. It is noteworthy that this principle was not followed in the Acute Physiology and Chronic Health Evaluation modeling program.[1,24] We conclude from the decision tree analysis, however, that handling of BP and ICP is extremely important in the analysis of such head injury data sets.

Previous Literature on Decision Tree Analysis of Head-Injured Patients

The accuracy achieved in this decision tree analysis study is better than those reported by Choi, et al.,[12] or Pilah, et al.[29] Because the focus of these papers is 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 (if interested, refer to references 28 and 29).

Conclusions

In this study, we have shown that improved monitoring techniques offer 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 antioxidants, ion channel blockers, membrane stabilizers, and inflammation.[28] What is clear from this study is that secondary insults play a significant role in determining patient outcome and occur commonly with current management.[14] In any trial of therapy, their occurrence must be recorded to ascertain whether therapies reduce the incidence of or improve cerebral tolerance to secondary insults.

References

1. Alvarez M, Nava JM, Rue M, et al: Mortality prediction in head trauma patients: performance of Glasgow Coma Score and general severity systems. **Crit Care Med 26:**142–148, 1998
2. Andrews PJ: Head injury: complications and management. **Curr Opin Anaesthesiol 11:**473–477, 1998
3. Andrews PJ, Dearden NM, Miller JD: Jugular bulb cannulation: description of a cannulation technique and validation of a new continuous monitor. **Br J Anaesth 67:**553–558, 1991
4. Andrews PJ, Piper IR, Dearden NM, et al: Secondary insults during intrahospital transport of head-injured patients. **Lancet 335:** 327–330, 1990
5. Baker SP, O'Neill B, Haddon W Jr, et al: The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. **J Trauma 14:**187–196, 1974
6. Bekar A, Ipekoglu Z, Tureyen K, et al: Secondary insults during intrahospital transport of neurosurgical intensive care patients. **Neurosurg Rev 21:**98–101, 1998
7. Block EF: Diagnostic modalities in acute trauma. **New Horiz 7:** 10–25, 1999
8. Blumer AC: ID3 Decision tree learning algorithm. **Electrical Engineering and Computer Science at Tufts University.** (http://www.eecs.tufts.edu/g/150ML/slides/id3.html) [Accessed 5 June 2002]
9. Chesnut RM: The management of severe traumatic brain injury. **Emerg Med Clin North Am 15:**581–604, 1997
10. Chesnut RM: Medical management of severe head injury: present and future. **New Horiz 3:**581–593, 1995
11. Chesnut RM: Secondary brain insults after head injury: clinical perspectives. **New Horiz 3:**366–375, 1995
12. Choi SC, Muizelaar JP, Barnes TY, et al: Prediction tree for severely head-injured patients. **J Neurosurg 75:**251–255, 1991
13. Davies G, Deakin C, Wilson A: The effect of a rigid collar on intracranial pressure. **Injury 27:**647–649, 1996
14. Dearden NM: Mechanisms and prevention of secondary brain damage during intensive care. **Clin Neuropathol 17:**221–228, 1998
15. Doberstein CE, Hovda DA, Becker DP: Clinical considerations in the reduction of secondary brain injury. **Ann Emerg Med 22:** 993–997, 1993
16. 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
17. Dzeroski S, Lavrac N: Rule induction and instance-based learning applied in medical diagnosis. **Technol Health Care 4:**203–221, 1996
18. Gennarelli TA, Champion HR, Copes WS, et al: Comparison of mortality, morbidity, and severity of 59,713 head injured patients with 114,447 patients with extracranial injuries. **J Trauma 37:** 962–968, 1994
19. Graham DI, Adams JH, Nicoll JA, et al: The nature, distribution and causes of traumatic brain injury. **Brain Pathol 5:**397–406, 1995
20. Graham DI, Ford I, Adams JH, et al: Ischaemic brain damage is still common in fatal non-missile head injury. **J Neurol Neurosurg Psychiatry 52:**346–350, 1989
21. Jennett B, Bond M: Assessment of outcome after severe brain damage. A practical scale. **Lancet 1:**480–484, 1975
22. Jones PA, Andrews PJ, Midgley S, et al: Measuring the burden of secondary insults in head-injured patients during intensive care. **J Neurosurg Anesthesiol 6:**4–14, 1994
23. Juul N, Morris GF, Marshall SB, et al: Intracranial hypertension and cerebral perfusion pressure: influence on neurological deterioration and outcome in severe head injury. The Executive Committee of the International Selfotel Trial. **J Neurosurg 92:**1–6, 2000
24. Knaus WA: APACHE 1978–2001: the development of a quality assurance system based on prognosis: milestones and personal reflections. **Arch Surg 137:**37–41, 2002
25. Lavrac N: Selected techniques for data mining in medicine. **Artif Intell Med 16:**3–23, 1999
26. Maas AI, Dearden M, Teasdale GM, et al: EBIC-guidelines for management of severe head injury in adults. European Brain Injury Consortium. **Acta Neurochir 139:**286–294, 1997
27. McQuatt A: **Using machine learning techniques to predict clinical outcome. Thesis.** Aberdeen: University of Aberdeen, 1998
28. Morganti-Kossmann MC, Hans VH, Lenzlinger PM, et al: TGF-beta is elevated in the CSF of patients with severe traumatic brain injuries and parallels blood-brain barrier function. **J Neurotrauma 16:**617–628, 1999
29. Pilah IA, Mladenic D, Lavrac N, et al: Data analysis of patients with severe head injury, in Lavrac N, Keravnou ET, Zupan B (eds): **Intelligent Data Analysis in Medicine and Pharmacology.** Boston: Kluwer Academic, 1997

30. Quinlan JR: Data mining tools C5.0 and See5. **Rulequest Research.** (http://www.rulequest.com/see5-info.html) [Accessed 5 June 2002]
31. Quinlan JR: Induction of decision trees. **Mach Learn 1:**81–106, 1986
32. Quinlan JR: Simplifying decision trees. **Int J Hum-Comput St 51:**497–510, 1999
33. Robertson CS, Gopinath SP, Goodman JC, et al: SjvO2 monitoring in head-injured patients. **J Neurotrauma 12:**891–896, 1995
34. Sarnaik AP, Lieh-Lai MW: Transporting the neurologically compromised child. **Pediatr Clin North Am 40:**337–354, 1993
35. Signorini DF, Andrews PJ, Jones PA, et al: Adding insult to injury: the prognostic value of early secondary insults for survival after traumatic brain injury. **J Neurol Neurosurg Psychiatry 66:** 26–31, 1999
36. Verweij BH, Muizelaar JP: Avoiding secondary brain injury after severe head trauma: monitoring and management. **J Craniomaxillofac Trauma 2:**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. **J Trauma 34:** 377–382, 1993
38. Young JS: Cerebral perfusion pressure or intracranial pressure? **J Neurosurg 92:**191–192, 2000 (Letter)

Manuscript received August 3, 2000.

Accepted in final form April 29, 2002.

Data collection for this study was supported by Grant No. G9508752 from the Medical Research Council, United Kingdom, awarded to Dr. Andrews.

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

Address reprint requests to: Peter J. D. Andrews, M.D., Department of Anaesthetics, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, United Kingdom. email: Pandrews@ed.ac.uk.