A STATISTICAL INVESTIGATION INTO THE MANAGEMENT STRATEGY OF HEAD INJURED PATIENTS

by

Elaine Margaret Anderson

Submitted for the Degree of Master of Science in the Department of Statistics, University of Glasgow, January 1988.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. G.D. Murray, for his guidance throughout this period of research.

I am also grateful to Professor W.B. Jennett at the Department of Neurosurgery, University of Glasgow, for providing useful comments and the data for the study.

Thanks to Alan and my family for their support and encouragement.

C O N T E N T S

SUMMARY

Radiologists and neurosurgeons debate the need to refer all head injured patients for radiography. Whilst radiologists have constructed a management strategy for referring recent head injured patients for computerised tomography scanning and X-ray, neurosurgeons have devised guidelines for the management of patients with a head injury for which the presence/absence of a skull fracture is an important feature for admission to hospital.

To examine whether individuals with a high risk of a skull fracture could be identified, a study based on 3424 patients from the Accident and Emergency Department at the Monklands District General Hospital was carried out. Twelve variables for each attender were considered. By employing different discrimination procedures, it was hoped that a classification rule with low error rates could be identified to determine the presence of a skull fracture.

The second section of this thesis deals with estimating the risk of a head injured patient developing an intracranial haematoma in an attempt to reduce the number of unnecessary admissions. Data were available on 8504 head injured patients from Accident and Emergency Departments in Great Britain plus 988 head injured patients from the Haematoma Study at the Southern General Hospital.

The medical background to the two questions posed in this study are described in more detail in Chapter 1.

In Chapter 2, a comprehensive examination of the twelve variables recorded at Monklands District General Hospital is carried out. Combining categories of some variables and the construction of the new variable Glasgow Coma Sum from the three

variables Eyes Open, Motor and Verbal are discussed.

The application of the linear logistic regression model to the two class discrimination problem (absence/presence of a skull fracture) is considered in Chapter 3. The performance of this method is assessed using error rates - sensitivity and specificity, the Youden Index and the area under the receiver operating characteristic curve. The linear logistic regression model employing the three variables - Glasgow Coma Sum, Headache/Vomiting and Facial Injury - seem to perform well in terms of the three aforementioned methods of assessment. At the end of this chapter, the linear discriminant is contrasted with the linear logistic regression model utilising the subsets {COMASUM, VOM, FAC} and {COMASUM, VOM, FAC, SCALP}. Both models appeared to perform equally well.

In Chapter 4, an alternative procedure to the linear logistic regression model for the discrimination problem based on classification trees is described. A comparison of the two methods is made using the Brier Score. Although only small differences existed, the classification tree approach is preferable on the basis of being a simpler method in practice for allocating future patients and being able to handle missing data.

To answer the second question, the absolute and relative risks of developing an intracranial haematoma are considered in Chapter 5. The calculation of confidence intervals for both types of risks are described. Two approaches for calculating the confidence intervals for relative risks are considered.

In Chapter 6, four features of the data set - Cause of Injury, Glasgow Coma Sum, Sex and Skull Fracture - are employed to estimate the risk of a head injured patient developing an intracranial haematoma. Risks are also extended to include

children. Only Skull Fracture and Glasgow Coma Sum were useful for considering future management of a head injured patient. An adult head injured patient with no skull fracture and Glasgow Coma Sum of 15 has a low risk of a haematoma. Patients who have no skull fracture and Glasgow Coma Sum of 9-14 or 3-8 or have a skull fracture present with Glasgow Coma Sum of 15 have intermediate levels of risk, while patients with a skull fracture and Glasgow Coma Sum of 9-14 or 3-8 have a high risk of developing an intracranial haematoma.

Children suffering from a head injury with no skull fracture and Glasgow Coma Sum of 15 have a low risk of developing an intracranial haematoma, while children with a skull fracture and Glasgow Coma Sum of 3-8 have a high risk of developing an intracranial haematoma.

Ammendments to the existing guidelines for admission or transferral to a Neurosurgical Unit plus appropriate management strategy for referring head injury patients for radiography are discussed in Chapter 7.

CHAPTER 1

INTRODUCTION

Head injuries account for a notable part of the work of Accident and Emergency (A and E) Departments, and because of the risk of serious complications, are a source of concern to the clinicians who staff them. A recent survey in Scotland indicated that head injuries accounted for 11% of all attenders at Accident and Emergency Departments (Strang et al., 1978). This represents in the United Kingdom, about one million new head injury attenders every year; one for every 65 of the population (Jennett and MacMillan, 1981). Most attenders are only mildly injured and require no special treatment. In 1983, Guidelines for the management of patients with a recent head injury were drawn up by a group of neurosurgeons (Appendix I) to define criteria aimed at reducing unnecessary admissions and allocating more effective care to those patients whose brain injury demands close attention.

The aim of the clinician is to prevent or reverse any preventable or reversible complication as a result of a recent head injury, and in particular to detect and remove intracranial haematomas - accumulations of blood within the tissues of the brain that clots to form solid swellings - as soon as possible.

Whilst infection can be prevented by appropriate head injury management, the recovery of a head injured patient with an intracranial haematoma will depend on early recognition of the haematoma and rapid surgical intervention.

The traditional approach in the management of patients with a recent head injury was to wait until a patient showed a deterioration in the level of consciousness and then to transfer

the patient for treatment. This approach is too late to prevent or reverse any serious complications.

Intracranial haematomas are now often detected by computed tomography but only for patients with a high risk of an intracranial haematoma. Therefore there is a need to estimate the risk of an intracranial haematoma to decide which patients are at such a low risk that they can be sent home, which patients have a medium risk and need to be admitted to hospital simply for observation and which patients have such a high risk that referral for immediate scanning is justified without a period of observation.

Developing a management strategy using risk levels may result in haematomas being detected earlier and by reducing the number of admissions and the number of CT scans, resources may be better employed.

Earlier work by Mendelow et al. (1983) indicated that the presence of a skull fracture is a powerful indicator of the risk that an adult (aged 15 or over) will develop a surgically significant haematoma after a recent head injury. Only a small percentage of radiographs indicate the presence of a skull fracture. A survey in 1974 of all A and E Departments in Scotland showed that of 2865 patients who attended after a head injury, 58% had a skull X-ray and of these, only 2.7% had a fracture (Strang et al.,1978). The incidence of a skull fracture then falls to 1.5% if it is assumed that patients not X-rayed do not have a skull fracture.

Radiologists dispute the need to X-ray so many head injuries even although the presence of a skull fracture may indicate intracranial damage (Brocklehurst et al., 1987). A Multidisciplinary Panel of Radiologists (Masters et al., 1987)

identified two main groups of head injured patients - those at high risk of intracranial injury and those at low risk of such injury - and developed a management strategy for managing the two groups. The patients in the high risk group were designated as candidates for emergency CT scanning, neurosurgical consultation, or both. Patients identified as belonging to the low risk group - patients who are asymptomatic or who have one or more of: headache, dizziness, scalp haematoma, laceration, contusion or abrasion - should not be recommended for radiographic imaging. In a study to validate this management strategy, data on 7035 patients in 31 hospitals were collected. The results showed that on the basis of the panel's criteria no intracranial injuries would have been detected in any of the patients assigned to the low risk group. Employing such a management strategy would result in a large decrease in the use of skull radiography, reduce unnecessary radiation exposure and savings of resources.

In order to use this management strategy in practice, certain results should be investigated. This study attempts to answer the two main questions:

(i) Can individuals with a high risk of skull fracture be identified (or at least can individuals with negligible risk of a skull fracture be eliminated) as needing an X-ray?

(ii) Can the risk of developing a haematoma be estimated to improve the existing guidelines for admission or transfer to a neurosurgical unit? Alternatively, can any other easily elicited clinical features replace skull fracture to assess the presence of an intracranial haematoma?

CHAPTER 2

DESCRIPTION OF THE DATA SET

2.1 Explanation of Variables Used in the Study

Over 48000 new attenders are treated at the Accident and Emergency Department at Monklands District General Hospital per annum. Approximately 4200 of these are diagnosed as having a head injury.

The basic data for each attender is directly entered onto the computerised records system in the Accident and Emergency Department and for the duration of the head injury study, 1st April to 31st December 1984, additional data relating to the head injury were recorded.

Over this 9 month period, data were collected on 3971 head injury patients. During this time, head injuries accounted for 10.5% of all attenders.

Before formal analysis was carried out, the variables in the data set, received from Monklands hospital, were examined and screened for consistency.

16 variables were recorded and bearing in mind that the goal is the identification of individuals with a high risk of a skull fracture, the variable "Skull Fracture" is of initial importance. This variable is divided into six exclusive categories:

1. Unknown
2. No skull X-ray
3. X-ray no fracture
4. Clinical fracture of base (CSF/blood in nose or ear)
5. Linear fracture

6. Depressed fracture

As this variable is of primary importance to future analysis, categories 1-4 are grouped as "no skull fracture" and categories 5 and 6 are grouped to give the "skull fracture" category, creating the "new" variable:

SKULL FRACTURE:

1. No fracture (n=3905)
2. Fracture (n=66)

Of the total study population, 98.3% fall into category 1 and 1.7% into category 2. However, treating the category "no skull X-ray" as "no fracture" can lead to serious complications. The absence of a fracture does not exclude serious brain damage. A patient may therefore present at the Accident and Emergency Department as a mild head injury and is sent home rather than being admitted. The patient is re-admitted for developing complications and by then it is too late for intervention by the doctor (Jennett and Miller,1972).

The remaining variables in the study were examined to assess their usefulness as predictors. The percentage of patients with a skull fracture for different features are shown in Table 2.1.

An additional variable which identified where the patients went on their discharge from hospital was also recorded. However, as the problem posed was to discover what variables aid in the prediction of a skull fracture, it was clear that the aforementioned variable was irrelevant in this context and it was thus omitted from the analysis.

To use the variables as shown in Table 2.1 would produce tedious calculations and take up excessive computing time in any further analysis to be carried out. Also, it is noticed that 13 of the 14 variables contain less than 6 patients with a skull

Table 2.1 Percentage of Skull Fracture for Variables in the Study

VARIABLE	n	no. of fracture	%age with Skull Fracture
AGE			
1.14 and under	1965	21	1.1
2.15-64	1692	33	2.0
3.Over 65	205	5	2.5
Not recorded	109	7	6.4
SEX			
1.Male	2704	55	2.0
2.Female	1249	11	0.9
Not recorded	18	-	-
SCALP INJURY			
0.Unknown	13	-	-
1.No external injury	1446	12	0.8
2.Swelling only	653	17	2.6
3.Abrasion/contusion only	598	15	2.5
4.Superficial laceration <5cm long	1088	13	1.2
5.Superficial laceration >5cm long	104	2	1.9
6.Laceration through galea <5cm long	38	2	5.4
7.Laceration through galea >5cm long	31	5	16.2
FACIAL INJURY			
0.Unknown	30	3	10
1.None	2457	36	1.5
2.Facial abrasion/contusion	590	7	1.2
3.Facial laceration	686	5	0.7
4.Periorbital haematoma	100	10	10
5.Fractured nose	66	1	1.5
6.Fractured other facial bones	36	4	11.1
7.Fractured mandible	6	-	-
EYES OPEN			
1.None	22	8	36.3
2.To pain	24	4	16.7
3.To speech	25	3	12.0
4.Spontaneously	3886	51	1.3
Not recorded	14	-	-
MOTOR			
1.No motor response	15	5	33.4
2.Extension	4	2	50
3.Spastic flexion	4	1	25
4.Normal flexion	13	4	30.8
5.Localises	56	4	7.1
6.Obeys commands	3854	50	1.3
Not recorded	25	-	-

Table 2.1 (con)

VARIABLE	n	no. of fracture	%age with Skull Fracture
VERBAL			
1.None	21	8	38.1
2.Incomprehensible sounds	63	3	4.8
3.Inappropriate words	10	2	20
4.Confused	83	6	7.2
5.Orientated	3768	46	1.2
Not recorded	26	1	3.8
HISTORY OF UNCONSCIOUSNESS/AMNESIA			
0.Unknown	167	8	4.8
1.None	3497	41	1.2
2.Less than 5 mins with full recovery of consciousness (i.e. orientated)	195	4	2.0
3.5-30 mins with full recovery of consciousness (i.e. orientated)	68	1	1.5
4.30-60 mins with full recovery (i.e. orientated)	7	1	15
5.>1 hour and/or still disorientated or worse	37	11	30
DETERIORATION			
0.Unknown	46	3	6.5
1.No	3897	57	1.5
2.Yes	28	6	21.4
EPILEPSY			
0.Unknown	129	5	3.9
1.No	3799	59	1.6
2.Focal	6	-	-
3.Generalised	15	2	13.0
4.Chronic	21	-	-
HEADACHE/VOMITING			
0.Unknown	52	5	9.6
1.None	3224	34	1.1
2.Headache	422	13	3.1
3.Vomiting	177	7	3.9
4.Headache and vomiting	96	7	7.3
PUPILS			
0.Unknown	43	3	7.0
1.Both reacting equal	3876	55	1.4
2.Both reacting unequal	28	2	7.1
3.One reacting	4	-	-
4.Neither reacting	11	5	45.5
5.Local factors affecting one or both pupils	9	1	11.1

Table 2.1(con)

VARIABLE	n	no. of fracture	%age with Skull Fracture
ALCOHOL			
0.Unknown	179	10	5.6
1.No	3213	41	1.3
2.Suspected	98	4	4.1
3.Definite (not measured)	405	7	1.7
4.Definite (measured <199)	24	1	4.2
5.Definite (measured 200-399)	50	3	6.0
6.Definite (measured >400)	2	-	-
FOCAL SIGNS			
0.Unknown	140	2	1.4
1.None	3810	56	1.5
2.Hemiparesis	10	2	20.0
3.Hemiplegia	2	2	100.0
4.Dysphasia/Aphasia	5	2	40.0
5.2 or 3+4	4	2	50.0

fracture in one or some of their categories. As the "skull fracture" variable is of primary importance to the study, to have numbers as small as these would produce results from which very little information could be obtained. Thus, after discussion with clinical colleagues, the variables were regrouped as presented in Table 2.2.

The Glasgow Coma Scale (Teasdale and Jennett, 1974) consists of the three variables Eyes Open (Eye opening in response to stimulation), Motor (Motor response of best limb in response to stimulation) and Verbal (Verbal response to stimulation). In this study, the "not recorded" category within each of the variables was omitted. These three variables when added together produce the variable COMASUM (Table 2.2) which measures, in this case, the depth of coma for a patient entering the Accident and Emergency Department. This term is known worldwide in medical literature as the Glasgow Coma Sum (Teasdale et al., 1979a).

After the initial screening of the data, it was decided to use a more stringent definition of a head injury to be consistent with the Scottish Head Injury Management Study (SHIMS) (Jennett et al., 1977). To conform with this and other studies, patients with only (1) Simple facial abrasion/contusion **or**

(2) Simple facial lacerations

were excluded from the analysis. This did not alter the regrouping of the variables. This leaves a total of 3424 patients with a head injury, representing 9.1% of all new attenders at Monklands Accident and Emergency Department.

2.2 Missing Values

This sample consisted of 66 attenders with a skull fracture and 3358 without. Values were missing from a total of 90 data

Table 2.2 Regrouping of Variables

Variable	Description
AGE	Age, grouped as 14 and under, 15 and over
ALC	Level of alcohol on admission, recoded as 0 (unknown), 1 (none) or 2-6 (else)
COMASUM	The sum of the raw Eyes Open, Motor and Verbal scores, in the range of 3 to 15 but recoded as 3-14 (abnormal) or 15 (normal)
DETERN	Deterioration, recoded 0 (unknown), 1 (no) or 2 (yes)
EP	Epilepsy, recoded 1 (none) or 0,2-5 (some)
FAC	Facial injury, recoded as 0-2 (none) and 3-7 (some)
FOC	Focal signs on admission, recoded as 0,1 (none) or 2-5 (some)
PUP	Pupil reaction to light, recoded 0,1 (both) or 2-5 (one or neither)
SCALP	Scalp injury, recoded as 1 (none) or 2-7 (some)
SEX	Sex, recorded 1 (male) or 2 (female)
UNCON	History of unconsciousness/amnesia, recoded as 1 (none) or 0,2-5 (unconsciousness)
VOM	Headache/vomiting after head injury, recoded 1 (none) or 0,2-4 (some)

vectors.

2 data vectors which were classified as fracture (3%) and 88 vectors classified as no fracture (3%) were incomplete.

From Table 2.2, it is easily seen that within certain variables, the "unknown" category has been grouped along with the recorded categories. This was decided by the clinicians involved in the study as the relevant "unknown" category contained a relatively large percentage of skull fractures in comparison to the remaining classes within the variable.

Several variables, however, have the number of missing values recorded within each category of SKULL FRACTURE as shown in Table 2.3.

Table 2.3 Missing values within SKULL FRACTURE

Variable	No. of missing values in fracture category	No. of missing values in no fracture category
AGE	-	53 (1.6%)
SEX	-	17 (0.5%)
SCALP	1 (1.5%)	9 (0.3%)
COMASUM	1 (1.5%)	26 (0.8%)

2.3 Identification of Possible Predictors

As a preliminary examination to determine potential predictors to be considered in the linear logistic regression model and discrimination models, Chi-squared statistics were evaluated.

The tabulated values in Table 2.4 indicate that all but one of the variables (SCALP) show a significant marginal association with SKULL FRACTURE at the 5% significance level. However, "marginal" significance can alter in a multivariate approach. Although one variable may be non-significant "marginally", when used in multiple regression, the variable along with the other

variables may produce useful information. It was therefore decided to construct discrimination functions based on all or subsets of the above variables.

Table 2.4 Chi-squared statistics of explanatory variables in the study

Variable	X^2	degrees of freedom	Tail probability
AGE	5.79	1	0.02
ALC	21.59	2	<0.001
COMASUM	92.00	1	<0.001
DETERN	64.10	2	<0.001
EP	5.84	1	0.02
FAC	10.12	1	0.001
FOC	146.20	1	<0.001
PUP	50.58	1	<0.001
SCALP	3.07	1	0.08
SEX	5.38	1	0.02
UNCON	34.10	1	<0.001
VOM	31.30	1	<0.001

CHAPTER 3

STATISTICAL DISCRIMINATION TECHNIQUES

3.1 General Introduction to Discrimination

When a population is known to consist of two totally exclusive classes π_1 and π_2, the two class predictive discrimination problem may be thought of as determining which class a future individual belongs given a vector of k observations. To develop a discrimination procedure, a training sample of size N is randomly selected from the population, for which the class of origin is known and a k-vector of observations is available; say n_1 come from class π_1 and n_2 come from class π_2 ($n_1+n_2=N$). In general, let the k dimensional observation vector be denoted by $\underline{X} = (X_1,\ldots,X_k)^T$ and let the sample data-vector of the j^{th} individual in class π_i be denoted by $\underline{X}_{ij}$ ($i=1,2$ and $j=1,\ldots,n_i$). For both classes, $\underline{X}$ is assumed to be a random vector with random variables $X_1,\ldots,X_k$. Then a decision rule for discriminating between future individuals, constructed on the training sample of N observations, assigns an individual with data vector $\underline{X}$ to population π_1 if some real-valued function of $\underline{X}$ is less than a given real number, otherwise the individual is classified as belonging to π_2. Clearly the task of identifying whether a future patient belongs to the skull fracture or no skull fracture class is a discrimination problem. In this instance, there are 12 potential discriminators, with $n_1=66$ and $n_2=3358$.

3.2 The Linear Logistic Regression Model

3.2.1 Introduction

The linear logistic regression model is commonly used in medical statistics to model outcomes which are binary in nature. In this particular application the outcome variable, Y, is SKULL FRACTURE. As Y can take only two possible values of interest (skull fracture or no skull fracture), it is unrealistic to assume the linear regression model

$$E(Y)= \alpha + \beta_1 X_1 + \ldots\ldots + \beta_k X_k = \alpha + \sum_{i=1}^{k} \beta_i X_i$$

where α is a constant,

$\underline{\beta}=(\beta_1,\ldots\ldots,\beta_k)^T$ is a vector of known parameters,

$\underline{X}=(X_1,\ldots\ldots,X_k)^T$ is a vector of known constants.

As $\alpha + \sum_{i=1}^{k} \beta_i X_i$ is assumed to be the expected value of a normal distribution and if the vector $\underline{\beta}$ is unrestricted, $\alpha + \sum_{i=1}^{k} \beta_i X_i$ will lie in the interval $(-\infty,\infty)$.

Considering the expected value of Y, which is equivalent to the probability that Y=1 (denoted by p), a more sensible and realistic approach would be to model $\log(p/1-p)$ which belongs to the interval $(-\infty,\infty)$. Therefore representing the probability of a fracture, Y=1, by $Pr(Y=1|\underline{X})$, for an individual with covariate values $\underline{X}=\underline{x}$, the linear logistic regression model is specified as:

$$\log\left[\frac{Pr(Y=1|\underline{x})}{1-Pr(Y=1|\underline{x})}\right] = \log\left[\frac{Pr(Y=1|\underline{x})}{Pr(Y=0|\underline{x})}\right]$$

$$= \alpha + \sum_{i=1}^{k} \beta_i x_i$$

or equivalently,

$$Pr(Y=1|\underline{x}) = \frac{\exp(\alpha + \sum_{i=1}^{k} \beta_i x_i)}{1 + \exp(\alpha + \sum_{i=1}^{k} \beta_i x_i)}$$

Having the model in this form, several methods of analysis can be carried out. It is worth noting that non-linear functions of the explanatory variables (e.g. interactions) can be included in models of this type. Fitting such models by maximum likelihood is computationally much more complicated than for standard normal theory based models, but there is a very close analogy with such methods.

3.2.2 Variable Selection

Many variables may be initially considered as suitable predictors. However, it is important to eliminate those variables which are irrelevant to the analysis. Using a smaller set of variables also reduces the computing time for setting up the discriminant procedure and the effort involved in allocating future individuals of interest. Thus some method of finding an "optimal" set of predictors based on a criterion, f say, for comparing the discriminative power of two sets of variables is required. Two such criteria are suggested and are described below. In this particular application, a stepwise procedure is used to select the subset of variables to be employed in future analysis.

Using a criteria, this procedure selects the best one dimensional predictor: denote the best one dimensional subset by $S_1=\{X_{(1)}\}$. If the realisation of the criterion function, $f(S_1)$,

exceeds a certain threshold value, d_1, then the procedure continues to a second stage. Otherwise, it is assumed that there are no worthwhile predictors in the full set of variables.

If the second stage is reached, the (p-1) two-variable subsets containing $X_{(1)}$ are considered, and that with the maximum f, say $S_2=\{X_{(1)},X_{(2)}\}$, is selected. If $f(S_2)-f(S_1) > d_1$, then the procedure continues; otherwise S_1 is chosen for use in the linear logistic regression model.

Future steps using this procedure are similar. However, in some procedures at stage i ($\geq$3) once S_i has been selected, each of the subsets S_{ij} of S_i produced by deleting $X_{(j)}$ is considered. If, for any j, $f(S_i)-f(S_{ij}) < d_2$ (another threshold) then that $X_{(j)}$ for which $f(S_i)-f(S_{ij})$ is minimal, is deleted from S_i. S_i is then redefined to be S_{ij} and the next step continues as before. The procedure carries on until no variables can be added or removed. However, the subset obtained by such a procedure may not maximise f over all possible subsets.

When k is large, to consider all 2^{k-1} variable-subsets directly requires a vast amount of computation and hence a "suboptimal" approach such as a stepwise procedure is commonly considered.

This stepwise strategy is implemented with the BMDP program PLR which can fit models either by full maximum likelihood (method MLR) or by an approximation to this which is far more efficient computationally (method ACE - asymptotic covariance). These are described in more detail below.

(1)Maximum Likelihood Ratio

At each stage, the maximum likelihood estimate of $\underline{\beta}$, say $\underline{\beta}^*$, is calculated, some of whose components are held at zero. A

revised estimate of $\underline{\beta}$ is computed for each variable, X_i, that may be entered or removed. Denote this estimate by $\tilde{\beta}_{(i)}$. From these estimates the maximum likelihood:

$$L(\hat{\beta}) = \prod_{j=1}^{n} \frac{\exp(y_j(z_j.\hat{\beta}))}{1 + \exp(z_j.\hat{\beta})}$$

where $j=1,\ldots,n$ (n = sample size)

y_j=response variable

and $\underline{z}=(z_1,\ldots,z_n)$ designates the vector of design variables generated from the set of categorical variables. For example, a categorical variable with 3 categories - 2 degrees of freedom - would be translated into 2 design variables.

Variables may then be entered or removed from the model on the basis of the significance of the approximate Chi-squared test:

$$x_i^2 = 2 \mid \log(L(\hat{\beta})/L(\tilde{\beta}_{(i)})) \mid$$

(2) Asymptotic Covariance

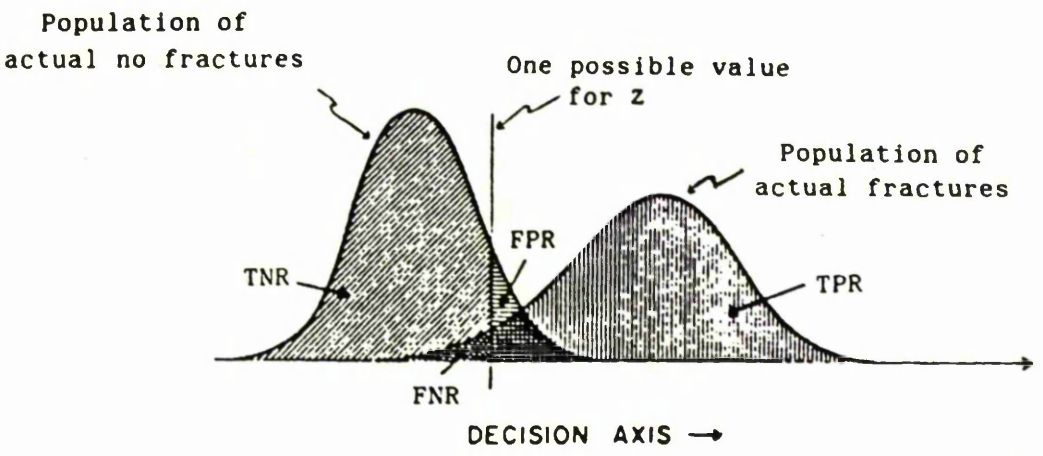
In this method, the asymptotic covariance matrix of the parameter estimates is evaluated at each stage. From this, the significance of the change in the residual sum of squares due to entering or removing a particular variable can be computed by means of an F-test.

Although computationally more quicker, the Asymptotic Covariance method may yield less reliable estimates of the significance of entering or removing a variable. Therefore, it is possible that the ACE method may select a slightly different subset of predictors.

3.3 Error Rates

Let π_1, π_2 denote a set of populations to which a patient may be allocated; it will be assumed that a patient belongs to one and only one of these populations.(In this application, π_2 may be assumed to be the population of patients with a skull fracture). The performance of a discrimination rule is assessed by how well individuals are allocated to their true population. In the linear logistic regression context, a rule is a value, Z, for the probability of belonging to π_2, such that if a case has probability (p) greater than Z, then action is taken as if the patient belongs to π_2, and if p is less than Z, then no skull fracture is assumed. Clearly, as shown in figure 3.1, an arbitrary choice of Z results in two possible errors.

Figure 3.1



The false-positive error rate (FPR) is the proportion of patients without a skull fracture who obtain a probability greater than Z, while the false-negative error rate (FNR) is the proportion of patients with a skull fracture who obtain a probability less than Z. No overlapping between the two populations in figure 3.1 above would result in a perfect test.

The above diagram assumes that there are an equal number of patients in the two populations. In this study however, the number of patients in population π_1 vastly exceeds that of π_2. The area under the curve corresponding to population π_1 would therefore increase with the area under the curve corresponding to population π_2 decreasing accordingly. Consequently, this results in population π_1 having more patients with $p>z$ than population π_2.

Expressed in the terms used in figure 3.1, **sensitivity** is defined as $\frac{TPR}{TPR + FNR}$ and **specificity** as $\frac{TNR}{TNR + FPR}$. The values yielded by each of these measures may vary from 0 to 1, with 1 representing perfection.

3.4 Youden Index

Various proposals for a combined single score have been recommended. One such recommendation is the Youden Index which has been defined as: **sensitivity + specificity - 1** (Youden,1950)

This index has many desirable features:

(1) The values of the Index lie between 0 and 1 inclusively if the test indicates a greater proportion of positive results for population π_2 than for population π_1: it takes the value 0 whenever a test gives the same proportion of positives for both populations. This class of test is clearly worthless. The Index takes the value 1 when no errors are present (i.e. a perfect test).

(2) The Index is not dependent of the relative and absolute sizes of the study populations.

(3) Tests having the same Index make the same total number of misclassifications per hundred patients.

(4) A standard error for the Index can be calculated which allows

two classification rules to be compared.

One drawback of the Index, however, is that false positive errors and false negative errors are assumed to be equally undesirable, which is clearly not the case in this application.

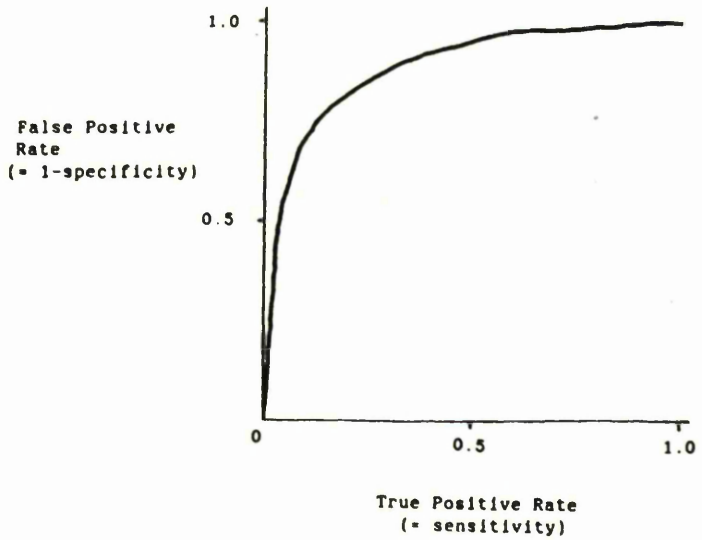
3.5 The Receiver Operating Characteristic (ROC) Curve

Sensitivity and specificity do not provide a unique description of the discrimination performance or accuracy of a test as they depend on the arbitrary selection of a cut-off point, z. By lowering z, both the True Positive (TPR) and False Positive (FPR) rates will increase, both rates being independent of the disease prevalence. By plotting the pairs
(True Positive , False Positive) for a range of values of the cut-off point z, a receiver operating characteristic curve is obtained (see figure 3.2). This curve must pass through the origin, as all tests can be called negative, and similarly must pass through the point (1,1) as all tests can be called positive.

The curve also lies above the line y=x, as a positive decision is more probable when a case is actually positive than when a case is actually negative.

Essentially, z must be chosen to yield an appropriate compromise between the two types of error. When the disease prevalence is low, the false positive rate has to be kept small otherwise all positive predictions will be false positive decisions. This may lead to unnecessary expensive follow up treatment. Alternatively, if finding positive cases is of overriding importance, then selection of a low cut-off point is essential.

Figure 3.2 An Example of an ROC Curve



Metz (1978) describes various methods to compare the discriminative ability of two tests by means of ROC curves. By plotting the curves on the same diagram, the performance of the two tests concerned can be compared. In general, better discrimination performance is indicated by the ROC curve which is further toward the top left hand corner in the diagram. Alternatively, if the two ROC curves cross, the situation to which the discrimination procedure has been employed will have to be reexamined.

Another technique to compare the discriminative power of two tests using ROC curves has been suggested by Hanley and M^{C}Neill (1982). This method reduces the entire ROC curve to a single quantitative index, namely, the area under the ROC curve. Hanley and M^{C}Neill interpret this area as "the probability that a randomly chosen diseased subject is (correctly) rated or ranked with greater suspicion than a randomly chosen non-diseased subject". This probability is equivalent to the non-parametric Wilcoxon Statistic, W, where

$$W = \frac{1}{n_1 n_2} \sum_{1}^{n_1} \sum_{1}^{n_2} S(x_1, x_2)$$

where n_1 = total number of patients with skull fracture

n_2 = total number of patients with no skull fracture

$$S(x_1,x_2) = \begin{cases} 1 & \text{if } x_1 > x_2 \\ 1/2 & \text{if } x_1 = x_2 \\ 0 & \text{if } x_1 < x_2 \end{cases}$$

and x is some quantative variable. (In this particular case, x is the levels of the probabilities from linear logistic regression).

An advantage of this procedure is that no assumptions have to be made about the underlying population distributions of π_1 and π_2. The value of W is equivalent to the area under the curve calculated using the trapezoidal rule. The statistic, W, is therefore an approximation to the true area under the curve.

3.6 Results

Data from 3424 cases with known outcome were recorded. This data set was split randomly into two sets, Set A and Set B, using the random number generator within BMDP. (Each case in the data set was assigned a value from the Un(0,1) generator: cases with values greater than or equal to 0.5 were allocated to Set A; cases with values less than 0.5 were allocated to Set B). Using the particular value of 3394173 for the pseudo-random generator resulted in 1788 cases being allocated to Set A - 43 (2.4%) skull fractures and 1745 no skull fractures - and 1636 cases allocated to Set B - 23 (1.4%) skull fractures and 1613 no skull fractures.

3.6.1 Performance of Linear Logistic Regression

To identify the number of variables to be used in a linear logistic regression equation for predicting future outcomes, equations with 1 to 10 variables were established. It was expected that using a maximum of 10 predictor variables would be more than adequate to identify patients with a skull fracture. The maximum number of probabilities for each logistic equation can be explained by the expression $\prod_{i=1}^{j} c_i$ where c_i = number of categories in i^{th} variable entered and j = number of variables entered into the equation.

Method 1

Using Set A as the training data set to generate the equations mentioned above, resulted in the variables indicated in Table 3.1(a) being entered. The goodness of fit for each equation was calculated using the Chi-squared statistic - $2 \Sigma \text{ observed}.\log\left[\frac{\text{observed}}{\text{expected}}\right]$ - as generated by BMDP. This statistic is approximated by the Chi-squared distribution when the sample size is much larger than the number of categories. However, if the observed number of individuals in some categories is small, usually taken to be < 5, this approximation may break down. Modelling with many variables, some categories may contain less than 5 patients and so caution has to be taken when interpreting the Chi-squared statistic. Set B, the test data set, was then run through the 10 equations, and by adjusting the cut-off point (or probability of a fracture, say Z), "true" sensitivity and "true" specificity could be assessed.

Method 2

The analysis as described above was carried out with the roles of the data sets A and B reversed. Results are shown in

TABLE 3.1 (a)

Variables entering linear logistic equation.

Method 1

Variables	Number of variables in equation. 1	2	3	4	5	6	7	8	9	10
AGE										
SEX									*	*
SCALP				*	*	*	*	*	*	*
FAC			*	*	*	*	*	*	*	*
COMASUM	*	*	*	*	*	*	*	*	*	*
UNCON							*	*	*	*
DETERN										
EP						*	*	*	*	*
VOM		*	*	*	*	*	*	*	*	*
PUP										*
ALC								*	*	*
FOC					*	*	*	*	*	*
GOODNESS OF FIT STATISTIC	133.2	122.2	115.2	100.0	97.6	96.3	95.7	93.9	93.4	93.3

TABLE 3.1 (b)

Variables entering linear logistic equation.

Method 2

Variables	Number of variables in equation.									
	1	2	3	4	5	6	7	8	9	10
AGE										*
SEX					*	*	*	*	*	*
SCALP										
FAC										
COMASUM		*	*	*	*	*	*	*	*	*
UNCON							*	*	*	*
DETERN			*	*	*	*	*	*	*	*
EP								*	*	*
VOM									*	*
PUP				*	*	*	*	*	*	*
ALC	*	*	*	*	*	*	*	*	*	*
FOC						*	*	*	*	*
GOODNESS OF FIT STATISTIC	106.9	93.3	85.2	81.9	79.4	77.7	76.9	76.5	76.2	76.1

Table 3.1(b).

Tables 3.1 (a) and (b) indicate that the Chi-squared goodness of fit criteria decreases when a further variable is added. However, after 3 or 4 variables have been added in both methods, smaller decreases are obtained at the cost of adding a further variable. To identify the "best" variable subset, it is necessary to assess the relative merits of each equation by the performance of the sample error and true error rates. In this Head Injury Study, as the ratio of patients with no skull fracture to those with a skull fracture is approximately 50:1, the best sample error and true error rate are obtained by taking a cut-off point such that all patients are allocated to the no skull fracture category. Clearly, unless the cost of misclassifying a patient with a skull fracture is greatly increased, using such error rates to select the best variable subset are invalid. Alternatively, the best variable subset may be chosen by considering the sensitivity and specificity of the discrimination rules for each equation.

3.6.2 Assessment of Sensitivity and Specificity

From the complete data cases, sensitivity, specificity, "true" sensitivity and "true" specificity ("true" sensitivity and "true" specificity were assessed using the test data set) were evaluated for each equation generated in Method 1 and Method 2. Values for the equations containing 1-8 variables from both methods are shown in Tables 3.2(a) and (b) respectively. To compensate for the small number of patients with a skull fracture, small values of the probability of a skull fracture in the range 0.01-0.2 were selected to obtain interpretable results. The actual cut-off points chosen were identified to reflect the

TABLE 3.2 (a)

Values of sensitivity, specificity, 'true' sensitivity and 'true' specificity from Method 1

1 Variable

		Sensitivity	Specificity	'True' Sensitivity	'True' Specificity
cut-off point	0.02	0.33	0.94	0.41	0.94

2 Variables

		Sensitivity	Specificity	'True' Sensitivity	'True' Specificity
cut-off point	0.02	0.58	0.77	0.59	0.76
	0.05	0.33	0.94	0.41	0.94
	0.1	0.21	0.98	0.36	0.98

3 Variables

		Sensitivity	Specificity	'True' Sensitivity	'True' Specificity
cut-off point	0.02	0.7	0.65	0.68	0.63
	0.03	0.58	0.77	0.59	0.76
	0.06	0.44	0.91	0.41	0.92
	0.1	0.26	0.97	0.41	0.97

4 Variables

		Sensitivity	Specificity	'True' Sensitivity	'True' Specificity
cut-off point	0.01	0.98	0.14	0.86	0.14
	0.03	0.6	0.78	0.43	0.78
	0.06	0.44	0.95	0.38	0.95
	0.1	0.35	0.97	0.38	0.98
	0.2	0.23	0.98	0.38	0.99

TABLE 3.2 (a) (cont'd)

5 Variables

	Sensitivity	Specificity	'True' Sensitivity	'True' Specificity
0.01	0.98	0.14	0.86	0.14
0.03	0.6	0.78	0.43	0.77
0.06	0.46	0.91	0.38	0.91
0.1	0.35	0.97	0.38	0.97
0.2	0.14	0.997	0.19	0.996

cut-off point

6 Variables

	Sensitivity	Specificity	'True' Sensitivity	'True' Specificity
0.01	0.98	0.16	0.86	0.16
0.03	0.58	0.79	0.43	0.78
0.06	0.47	0.91	0.38	0.91
0.1	0.35	0.98	0.38	0.97
0.2	0.23	0.99	0.33	0.99

cut-off point

7 Variables

	Sensitivity	Specificity	'True' Sensitivity	'True' Specificity
0.01	0.98	0.16	0.86	0.16
0.03	0.58	0.8	0.43	0.78
0.06	0.44	0.94	0.38	0.95
0.1	0.37	0.97	0.38	0.97
0.2	0.21	0.99	0.33	0.99

cut-off point

8 Variables

	Sensitivity	Specificity	'True' Sensitivity	'True' Specificity
0.01	0.93	0.21	0.67	0.23
0.03	0.58	0.82	0.43	0.81
0.06	0.47	0.93	0.38	0.93
0.1	0.35	0.97	0.38	0.98
0.2	0.23	0.99	0.24	0.992

cut-off point

TABLE 3.2 (b)

Values of sensitivity, specificity, 'true' sensitivity and 'true' specificity from Method 2

1 Variable

		Sensitivity	Specificity	'True' Sensitivity	'True' Specificity
cut-off point	0.02	0.65	0.8	0.23	0.81

2 Variables

		Sensitivity	Specificity	'True' Sensitivity	'True' Specificity
cut-off point	0.02	0.73	0.77	0.44	0.77
	0.1	0.32	0.98	0.11	0.98

3 Variables

		Sensitivity	Specificity	'True' Sensitivity	'True' Specificity
cut-off point	0.02	0.73	0.8	0.26	0.81
	0.03	0.45	0.94	0.16	0.93
	0.06	0.36	0.98	0.14	0.98
	0.1	0.32	0.99	0.09	0.99

4 Variables

		Sensitivity	Specificity	'True' Sensitivity	'True' Specificity
cut-off point	0.02	0.77	0.76	0.44	0.76
	0.03	0.55	0.94	0.19	0.93
	0.06	0.45	0.97	0.16	0.97
	0.1	0.32	0.99	0.09	0.99
	0.2	0.18	0.995	0.07	0.997

TABLE 3.2 (b) (cont'd)

5 Variables

cut-off point	Sensitivity	Specificity	'True' Sensitivity	'True' Specificity
0.01	0.77	0.8	0.39	0.81
0.03	0.55	0.95	0.19	0.95
0.06	0.45	0.98	0.16	0.97
0.1	0.32	0.99	0.09	0.99
0.2	0.18	0.996	0.05	0.998

6 Variables

cut-off point	Sensitivity	Specificity	'True' Sensitivity	'True' Specificity
0.01	0.77	0.8	0.39	0.8
0.03	0.55	0.95	0.19	0.95
0.06	0.45	0.97	0.16	0.97
0.1	0.27	0.99	0.09	0.99
0.2	0.27	0.995	0.09	0.998

7 Variables

cut-off point	Sensitivity	Specificity	'True' Sensitivity	'True' Specificity
0.01	0.77	0.79	0.4	0.8
0.03	0.59	0.92	0.21	0.92
0.06	0.45	0.98	0.14	0.98
0.1	0.36	0.99	0.12	0.99
0.2	0.27	0.995	0.09	0.997

8 Variables

cut-off point	Sensitivity	Specificity	'True' Sensitivity	'True' Specificity
0.01	0.77	0.79	0.4	0.8
0.03	0.59	0.93	0.21	0.92
0.06	0.45	0.98	0.14	0.98
0.1	0.32	0.99	0.12	0.99
0.2	0.27	0.994	0.09	0.997

probabilities of a skull fracture obtained from the linear logistic regression models.

As mentioned previously, sensitivity and specificity are inversely related: increasing the cut-off point z reduces sensitivity and increases specificity. Consequently, it is important to consider the values of "true" sensitivity and "true" specificity at each cut-off point. These values are considered to assess the relative merits of each cut-off point, using data which has not been used to generate the equations.

It appears that in both methods, after the third or fourth variable has been entered, there is no improvement in the best pair of values for "true" sensitivity and "true" specificity.

For the ideal test, the best pair of values for "true" sensitivity and "true" specificity would be (1,1). Inevitably, a perfect test very seldom exists and hence it is necessary to select a pair of values which will have the best overall consequence for the whole population.

Although four variables may be the maximum considered in Method 1, it would seem difficult from this crude method to identify the best subset. The use of three variables with cut-off point 0.1, may not appear to be any different from selecting one variable with the cut-off point of 0.02. However, in the practical context, the equation based on three variables would assign 44 less patients for skull fracture treatment.

It would appear that there is no clear superior variable subset in Method 2. To obtain a "true" sensitivity value greater than 0.4 results in "true" specificity of approximately 0.8. Similarly, to achieve a "true" specificity value of greater than 0.95, reduces the "true" sensitivity value to approximately 0.15.

Clearly, in this method the relative importance of the two types of errors has to be considered: is it more important to not diagnose a skull fracture than to refer a patient for unnecessary skull fracture treatment.

3.6.3 Results of the Youden Index

Calculating the values of the Youden Index, for all ten equations in both methods (see Tables 3.3(a) and (b) for the equations containing 1-8 variables), as expected, yield similar results to those obtained by considering "true" sensitivity and "true" specificity. An advantage of the Youden Index is that one figure characterises the performance of the rule. It is noticed that in Method 1, the 8 variable subset achieved negative values of the Youden Index at the lowest cut-off point. This has been caused by a test showing a greater proportion of positive results for the no skull fracture population than for the skull fracture population. Such a test is clearly worthless.

Although this Index appears to identify the variable subset selected using the "true" sensitivity and "true" specificity pair, the identification of the best cut-off point may be more difficult. As discussed previously, identifying the best cut-off point requires some consideration of the relative importance of the two types of errors. The Youden Index, in this example, may therefore be impractable as it assumes that both errors have the same weight.

3.6.4 Assessment of the ROC Curve

As with sensitivity, specificity and the Youden Index, the receiver operating characteristic curve requires knowledge of the true outcome of each patient. The entire curve represents the

TABLE 3.3 (a) Youden Index

Method 1

Cut-off point	Number of Variables								
	1	2	3	4	5	6	7	8	
0.01				0.12	0.12	0.14	0.14	0.14	*
				0	0	0.02	0.02	-0.1	+
0.02	0.27	0.35	0.35						
	0.35	0.35	0.31						
0.03			0.35	0.38	0.38	0.37	0.38	0.4	
			0.35	0.21	0.2	0.21	0.21	0.24	
0.05		0.27							
		0.35							
0.06			0.35	0.39	0.37	0.38	0.38	0.4	
			0.33	0.33	0.29	0.29	0.33	0.31	
0.1		0.19	0.23	0.32	0.32	0.33	0.34	0.32	
		0.34	0.38	0.36	0.35	0.35	0.35	0.36	
0.2				0.21	0.137	0.14	0.2	0.22	
				0.37	0.186	0.32	0.32	0.23	

* - Youden Index using sensitivity and specificity.

+ - Youden Index using 'true' sensitivity and 'true' specificity.

TABLE 3.3 (b) Youden Index

Method 2

Cut-off point	Number of Variables: 1	2	3	4	5	6	7	8	
0.01					0.57	0.57	0.56	0.56	*
					0.2	0.19	0.2	0.2	+
0.02	0.45	0.5	0.53	0.53					
	0.04	0.21	0.07	0.2					
0.03			0.39	0.49	0.5	0.5	0.51	0.52	
			0.09	0.12	0.14	0.14	0.13	0.13	
0.06			0.34	0.42	0.43	0.42	0.43	0.43	
			0.12	0.13	0.13	0.13	0.12	0.12	
0.1		0.3	0.31	0.31	0.31	0.26	0.35	0.31	
		0.09	0.08	0.08	0.08	0.08	0.11	0.11	
0.2				0.18	0.18	0.26	0.26	0.26	
				0.067	0.048	0.09	0.09	0.09	

* - Youden Index using sensitivity and specificity.

+ - Youden Index using 'true' sensitivity and 'true' specificity.

performance of each equation over all possible cut-off points. This curve can then be used to compare the equations with, in general, the better discriminator having a curve closer to the upper left hand corner.

A smooth curve fitted subjectively by eye for each ROC curve, in equations containing 1-8 variables for both methods, (Figures 3.3(a)-(h),3.4(a)-(h)), will often provide an adequate estimate of the full ROC curve. Subjective impressions initially were that in Method 1, a discrimination procedure using the 3-variable subset would produce better results. As would be expected, the shape of the ROC curves for all 8 equations based on Set A is similar. The relative merits of each equation should be assessed using the completely new data set, Set B; hence the choice of the 3 variable subset .

Considering Method 2, the 4 variable subset would appear to give a better discrimination of skull fracture but a more objective procedure may be required. The area under the curve (W), was evaluated for all eight equations (see Tables 3.4(a) and (b)). A detailed calculation of W and its standard error for 4 variables in Method 2 (Set B) is detailed on the following page. It would seem that the area under the curve, W, derived in this way agrees with the subjective impressions. However, the statistic W identifies a measure of the difference between the equations. As Metz explains, no fully satisfactory procedure has been constructed to test the significance of the apparent differences between the area under the ROC curves.

To assess how well the 3 and 4 variable subsets perform, the subsets {COMASUM,VOM,FAC} , {COMASUM,VOM,FAC,SCALP}, {COMASUM,ALC,DETERN} and {COMASUM,ALC,DETERN,PUP} were used to generate equations in both methods. The data set not used in the

Row	Contents	Column (Rating) : x						Total	Remarks
		≤0.01	0.01-0.03	0.03-0.06	0.06-0.1	0.1-0.2	≥0.2		
1	Number of no fractures rated x	1219	285	56	23	12	9	n_{NF} = 1604	Obtained from linear logistic output.
2	Number of fractures rated > x	17	12	10	7	4	0		Obtained from (3) by successive subtraction from n_F = 22
3	Number of fractures rated x	5	5	2	3	3	4	n_F = 22	Obtained from linear logistic output
4	Number of no fractures rated < x	0	1219	1504	1560	1583	1595		Obtained from (1) by successive add^{n}. to 0
5	(1)x(2)+1/2x(1)x(3)	23770.5	4132.5	616	195.5	66	18	28798.5	W = Total (5) ÷ (n_F.n_{NF}) = 0.816
6	(3)x[(4)^2+(4)x(1)+1/3x(1)^2]	2476601.7	9302255	4694570.7	7408969	7574799	10233601	41690796.4	Q_2 = Total (6) ÷ (n_F.n_{NF}^2) = 0.736
7	(1)x[(2)^2+(2)x(3)+1/3x(3)^2]	466064.3	60515	6794.7	1679	372	48	535473	Q_1 = Total (7) ÷ (n_{NF}.n_F^2) = 0.69

$$W = \hat{\theta} = \text{Total (5)} \div (n_{NF} \cdot n_F) = 28798.5 \div (1604 \cdot 22) = 0.816 = 81.6\%$$

$$SE(\hat{\theta}) = \sqrt{\frac{\hat{\theta}(1-\hat{\theta}) + (n_F - 1)(Q_1 - \hat{\theta}^2) + (n_{NF} - 1)(Q_2 - \hat{\theta}^2)}{n_F \cdot n_{NF}}} = \sqrt{\frac{0.1501 + 0.504 + 112.21}{22.1604}} = 0.0566 = 5.66\%$$

Method 1

Figure 3.3 (a)

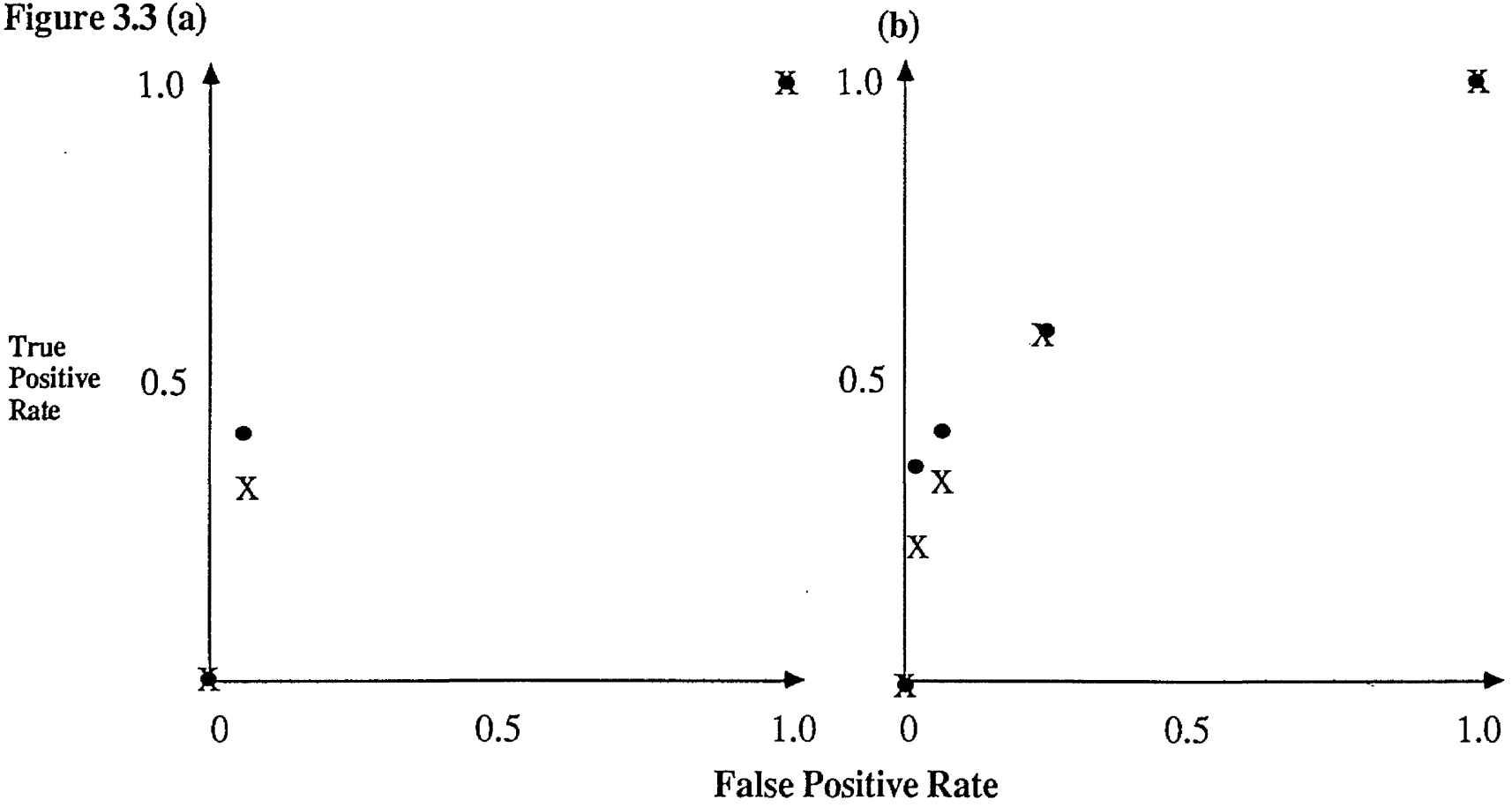


Figure 3.3 (c)

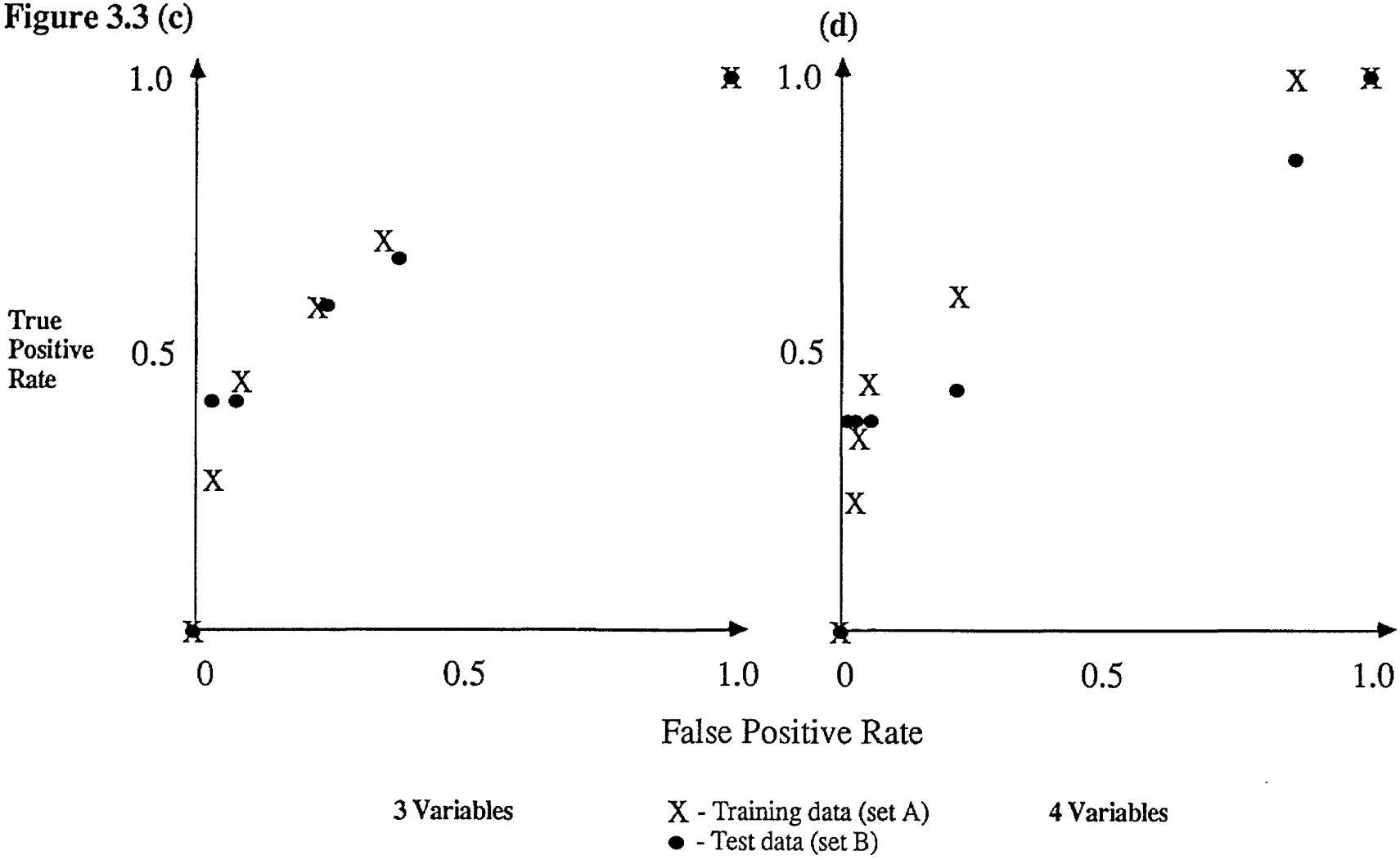

(d)
1.0
0.5
0
0.5
1.0
True Positive Rate
False Positive Rate
3 Variables
4 Variables
X - Training data (set A)
● - Test data (set B)

Figure 3.3 (e)

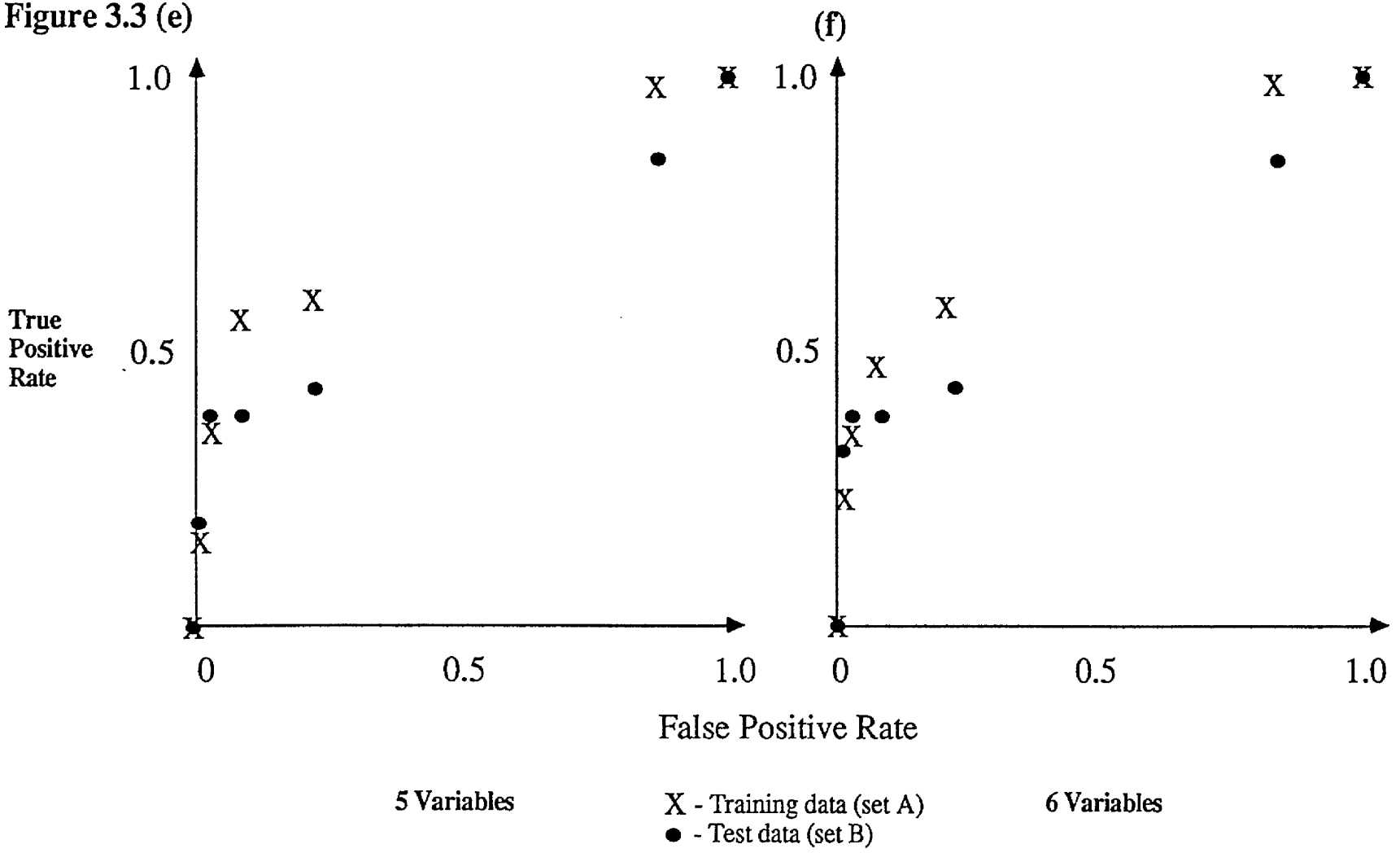
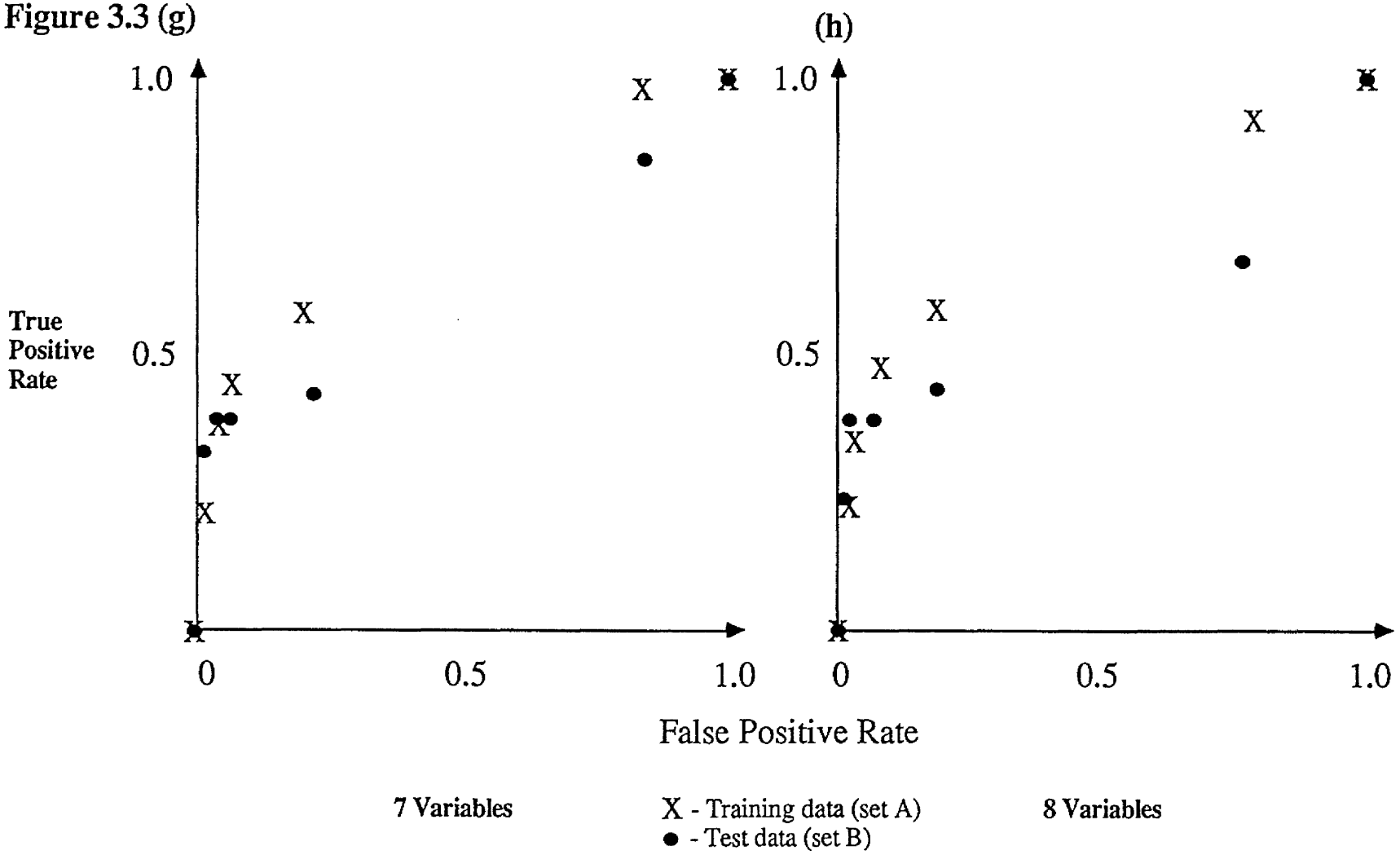

(f)
1.0
0.5
True Positive Rate
0
0.5
1.0
False Positive Rate
5 Variables
6 Variables
X - Training data (set A)
● - Test data (set B)

Figure 3.3 (g)


(h)
1.0
0.5
0
0.5
1.0
True Positive Rate
False Positive Rate
7 Variables
8 Variables
X - Training data (set A)
● - Test data (set B)

Method 2

Figure 3.4 (a)

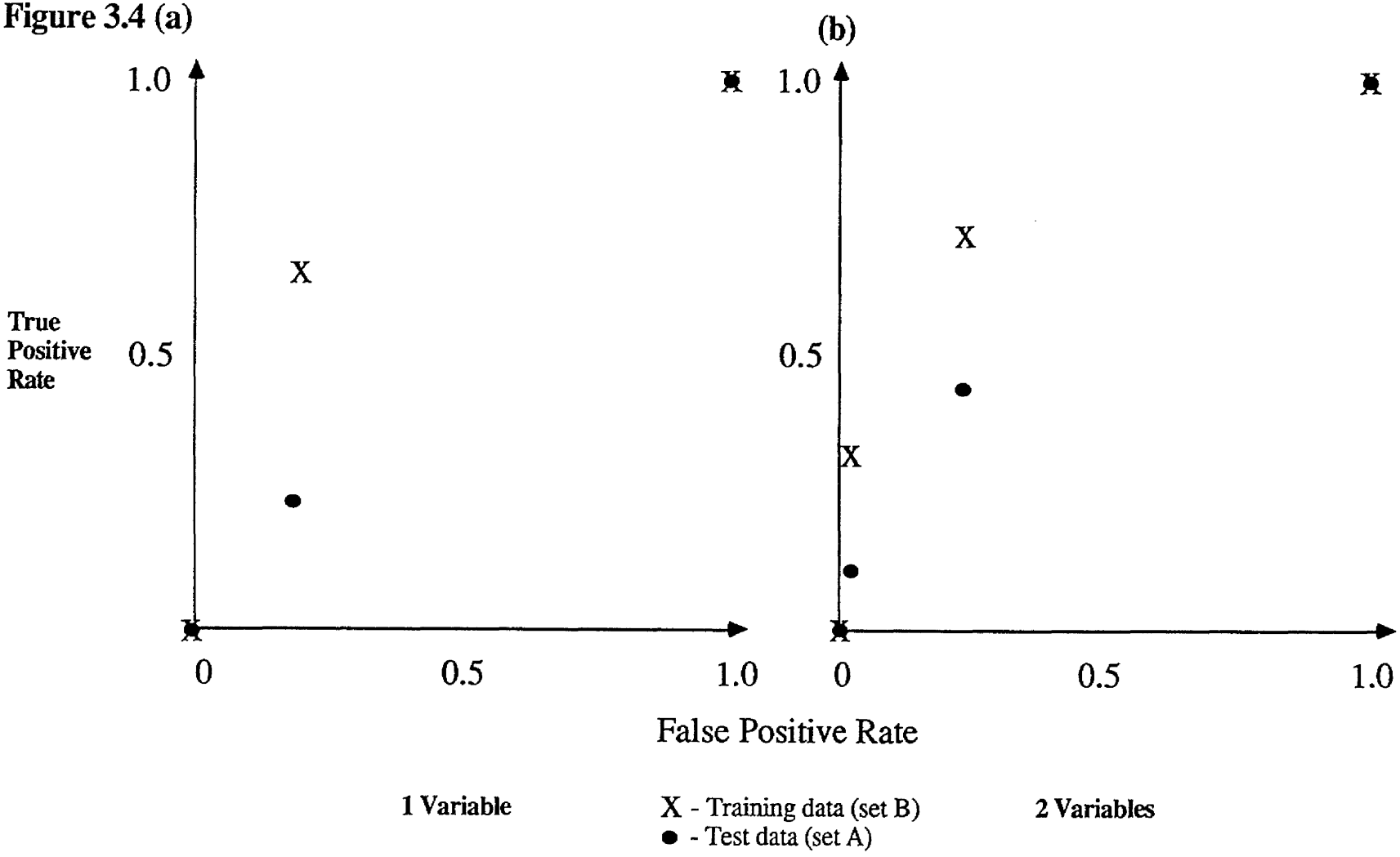


Figure 3.4 (c)

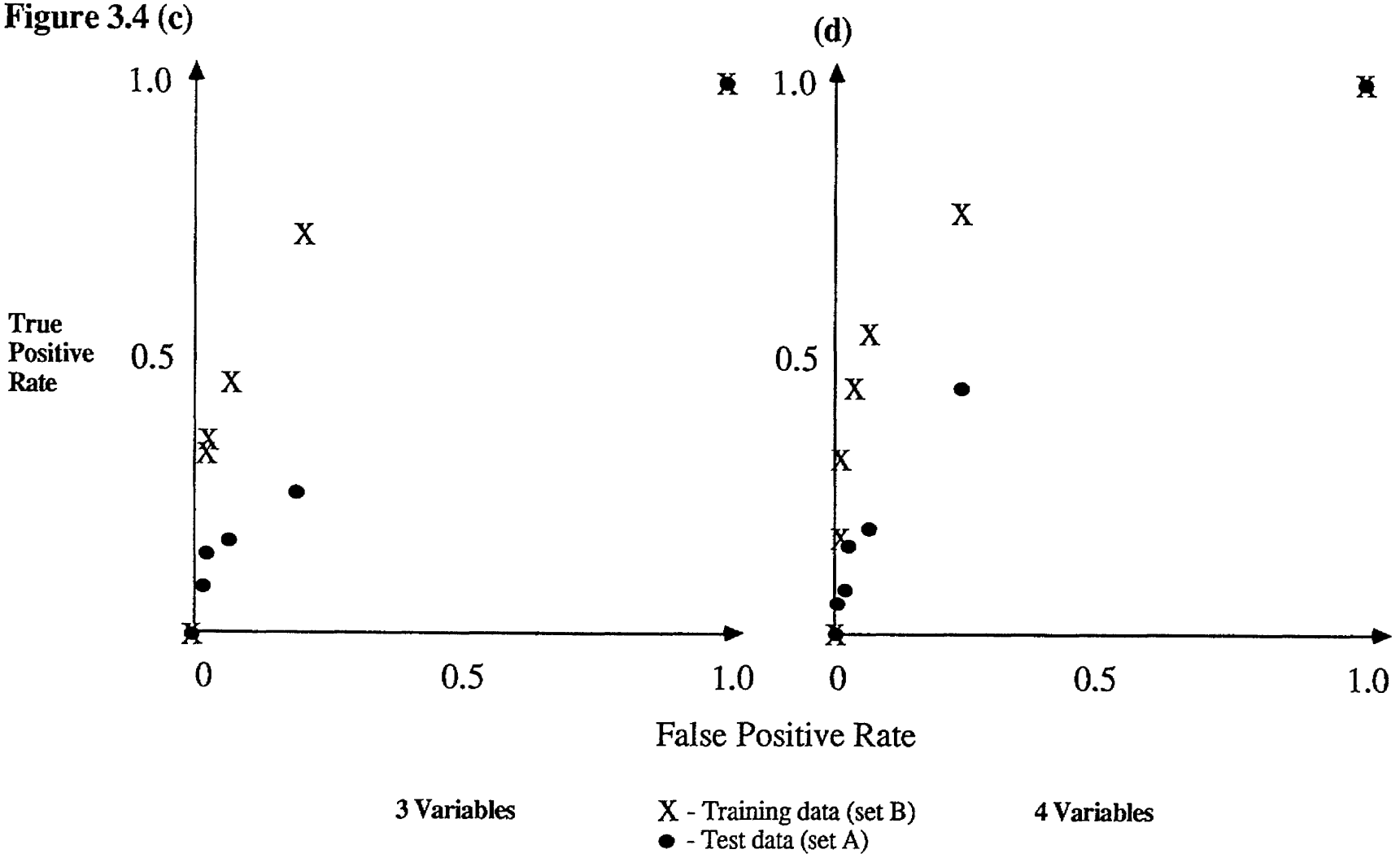

(d)
1.0
0.5
0
0.5
1.0
True Positive Rate
False Positive Rate
3 Variables
4 Variables
X - Training data (set B)
● - Test data (set A)

Figure 3.4 (e)

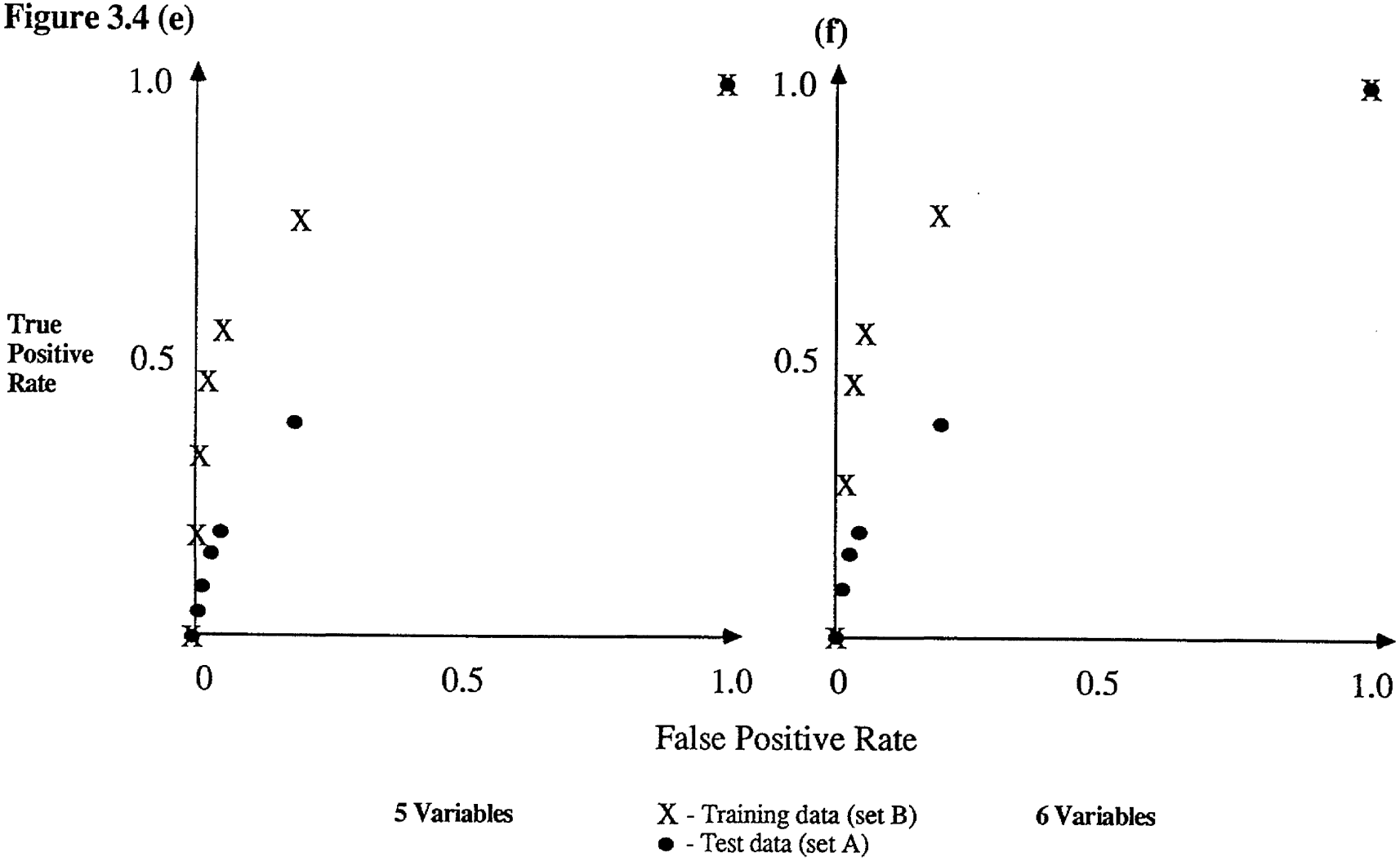

(f)
1.0
0.5
0
0.5
1.0
True Positive Rate
False Positive Rate
5 Variables
6 Variables
X - Training data (set B)
● - Test data (set A)

Figure 3.4 (g)

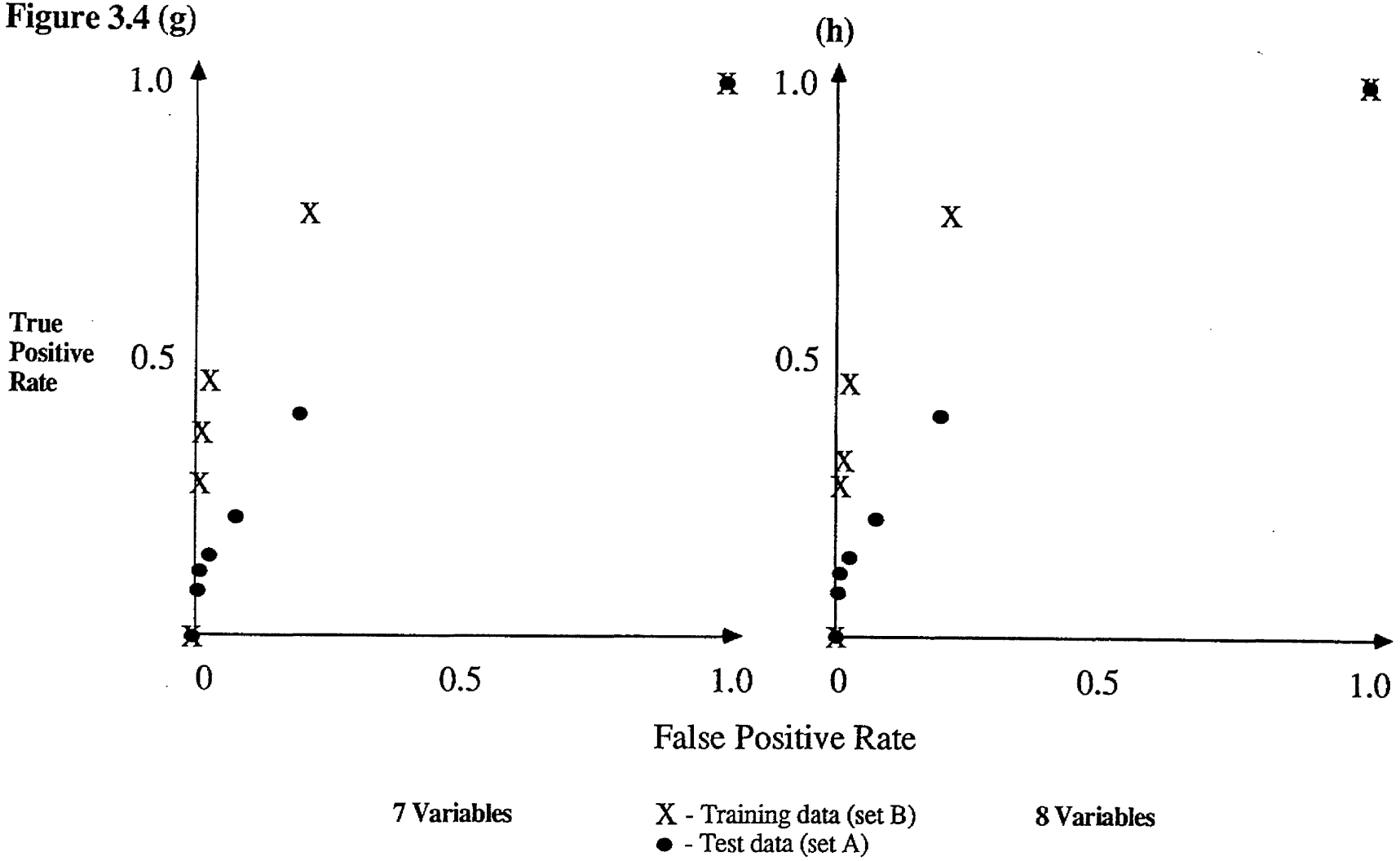

(h)
1.0
0.5
0
0.5
1.0
True Positive Rate
False Positive Rate
7 Variables
8 Variables
X - Training data (set B)
● - Test data (set A)

TABLE 3.4 Area under the ROC curve

(a) Method 1

		Training data (Set A)		Test data (Set B)	
		Area (W)	S.E.(W)	Area (W)	S.E.(W)
Number of variables	1.	63.2	0.052	67.6	0.069
	2.	69.8	0.047	71.2	0.067
	3.	72.6	0.045	71.8	0.066
	4.	74.8	0.044	62.7	0.077
	5.	74.6	0.044	62.3	0.077
	6.	74.4	0.044	62.9	0.076
	7.	74.9	0.043	63.2	0.076
	8.	74.5	0.045	58.0	0.085

(b) Method 2

		Training data (Set B)		Test data (Set A)	
		Area (W)	S.E.(W)	Area (W)	S.E.(W)
Number of variables	1.	72.7	0.058	52.0	0.045
	2.	77.7	0.058	61.4	0.048
	3.	79.4	0.059	54.1	0.048
	4.	81.6	0.057	61.1	0.048
	5.	82.8	0.056	61.0	0.048
	6.	82.8	0.055	61.0	0.048
	7.	82.6	0.056	60.6	0.048
	8.	82.5	0.056	60.8	0.048

generation of the equations was then used to assess the performance of the discrimination. Areas under the ROC curve were calculated as a measure of performance (Table 3.5)

Table 3.5 Areas Under the ROC Curve

	Subsets of Variables			
	COMASUM VOM FAC	COMASUM ALC DETERN	COMASUM VOM FAC SCALP	COMASUM ALC DETERN PUP
Method 1				
Training data (Set A)	72.6	62.8	74.8	63.2
Test data (Set B)	71.8	71.0	62.7	69.7
Method 2				
Training data (Set B)	67.6	79.4	69.0	81.6
Test data (Set A)	68.1	54.1	71.3	61.1

From this table, the best 3 and 4 variable subsets are {COMASUM,VOM,FAC} and {COMASUM,VOM,FAC,SCALP}. The relative performance of these two equations may be assessed by the unbiased (or less biased) area achieved by running Set B and Set A through Method 1 and Method 2 respectively. Although both subsets have their own particular merits, the 3 variable subset {COMASUM,VOM,FAC} should be recommended, as the unbiased area appears to be larger on the whole, and the inclusion of another variable does not alter the performance of the discrimination.

3.7 The Linear Discriminant

It has been established on a similar set of data (Titterington et al., 1981) that the performance of the linear logistic regression model can be similar to that of the linear discriminant.

In population π_i, i=1,2, if $\underline{X}$ is a multivariate normal random vector with mean vector $\underline{\mu}_i$ and common covariance matrix, Σ, then

in linear discrimination, an individual is assigned to π_1 if and only if:

$$(\underline{\mu}_1-\underline{\mu}_2)^T\Sigma^{-1}\underline{X} > \log\left[\frac{p(\pi_2)}{p(\pi_1)}\right] + 1/2(\underline{\mu}_1-\underline{\mu}_2)^T\Sigma^{-1}(\underline{\mu}_2+\underline{\mu}_1) \qquad (3.1)$$

i.e $\quad \underline{X}^T\underline{\lambda} > c_1$

where $\quad \underline{\lambda} = [(\underline{\mu}_1-\underline{\mu}_2)^T\Sigma^{-1}]^T = \Sigma^{-1}(\underline{\mu}_1-\underline{\mu}_2)$

$$c_1 = \log\left[\frac{p(\pi_2)}{p(\pi_1)}\right] + 1/2(\underline{\mu}_1-\underline{\mu}_2)^T\Sigma^{-1}(\underline{\mu}_2+\underline{\mu}_1)$$

and $p(\pi_i)$ is the probability of an individual belonging to group i before $\underline{X}$ is observed.

Using the maximum likelihood estimates of $\underline{\mu}_1$, $\underline{\mu}_2$ and Σ obtained from the training sample:

$$\hat{\underline{\mu}}_i = \bar{\underline{X}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\underline{X}_{ij} \qquad (i=1,2)$$

and $$\hat{\Sigma} = S = \frac{1}{n_1+n_2}(n_1\hat{\Sigma}_1 + n_2\hat{\Sigma}_2)$$

where $$\hat{\Sigma}_i = \frac{1}{n_i}\sum_{n=1}^{n_i}(\underline{X}_{ij} - \bar{\underline{X}}_i)(\underline{X}_{ij} - \bar{\underline{X}}_i)^T \qquad (i=1,2)$$

the sample based equivalent of (3.1) is assign a future individual to population π_1 if and only if:

$$(\bar{\underline{X}}_1-\bar{\underline{X}}_2)^TS^{-1}\underline{X} > \log\left[\frac{p(\pi_2)}{p(\pi_1)}\right] + 1/2(\bar{\underline{X}}_1-\bar{\underline{X}}_2)^TS^{-1}(\bar{\underline{X}}_2+\bar{\underline{X}}_1) \qquad (3.2)$$

However in this application, all the variables are discrete and clearly non-normal. Fisher (1936) derived exactly the same rule as (3.2) using a different approach to the discrimination problem, not based on any particular parametric form but by merely looking for some sensible rule based on a linear function

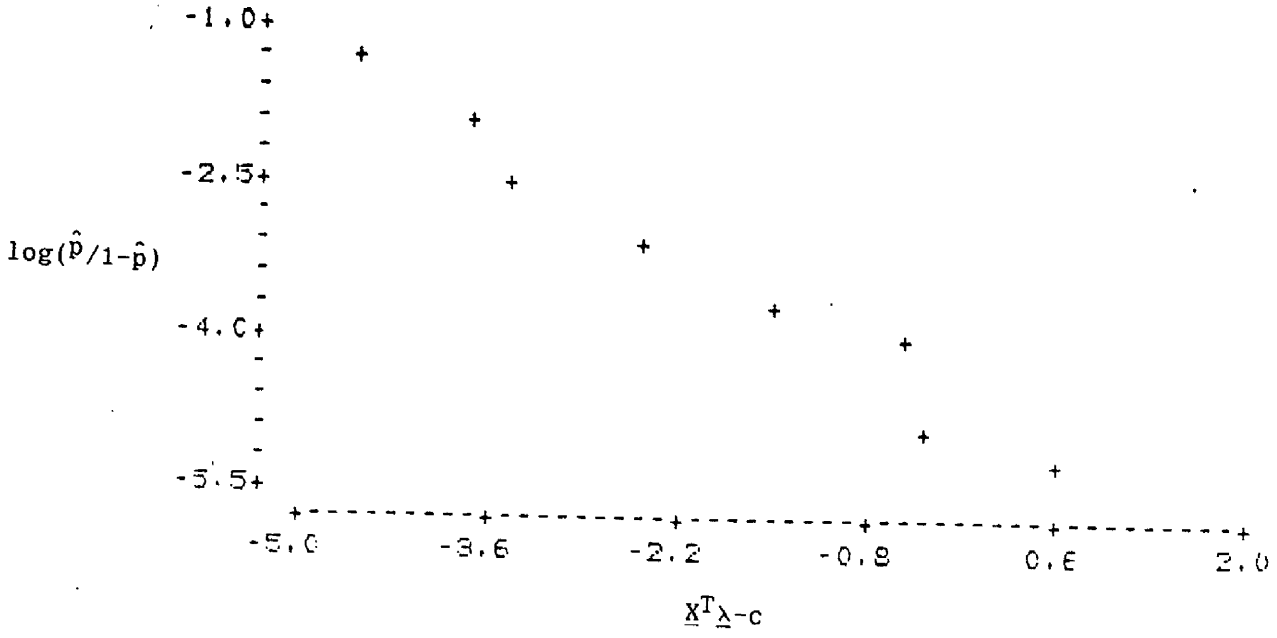
of the $\underline{X}$'s.

Using sample estimates to ascertain whether the linear logistic regression model performed as well as linear discrimination using this particular data set, the values from the linear logistic regression equation (for the 3 and 4 variable subset) were plotted against the corresponding linear combination of the variables ($\underline{X}^T\underline{\lambda}-c$) obtained from the BMDP program P7M (see Figures 3.5(i) and (ii)).

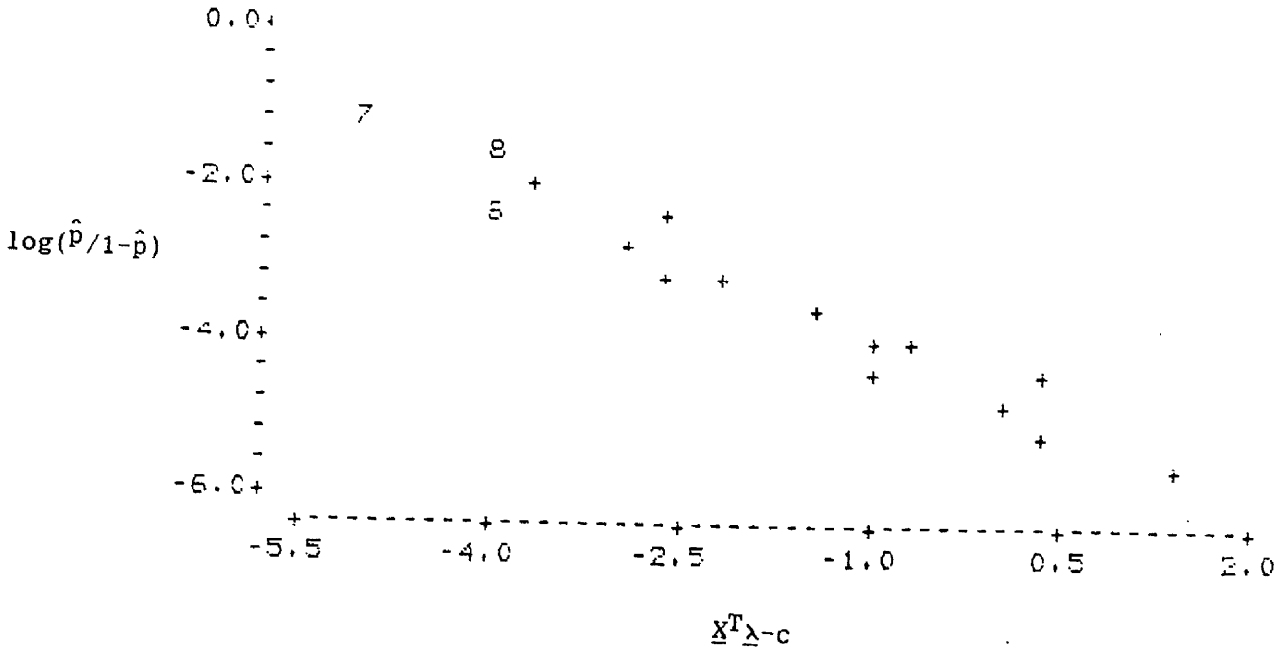
The values from the linear logistic regression are negatively correlated with the corresponding values obtained from linear discrimination, since the large values of $\log\left[\frac{\hat{p}}{1-\hat{p}}\right]$ and values of $\underline{X}^T\underline{\lambda}-c < 0$ correspond to predicting patients with a skull fracture. Thus the linearity of both these graphs indicate that the two methods have similar discriminative power.

Figure 3.5 Plots of the linear discriminant and linear logistic regression

(i) Three Variables



(ii) Four Variables



CHAPTER 4

CLASSIFICATION AND REGRESSION TREES (CART)

4.1 Explanation of the Method

A classification tree, in this study, is a tree structured classification rule which assigns an incoming head injured patient to the Accident and Emergency Department to one of the mutually exclusive groups, skull fracture or no skull fracture. In more general terms, given J mutually exclusive classes, a classification tree or classification rule is a systematic way of predicting what class a case is in given a measurement or data vector $\underline{x} = (x_1, x_2, \ldots, x_n)$, say. That is, given any $\underline{x} \in X$, where X denotes the measurement space containing all possible measurement vectors, a classification rule, $d(\underline{x})$, assigns one of the classes to $\underline{x}$.

Ideally, a classification tree should be constructed using past knowledge or as in this particular situation, using a training set of data which would provide insight and understanding into the predictive nature of the data (different relationships will exist in different subsets of the measurement space X). The binary tree structured classification trees are constructed by repeated splits of subsets of the measurement space X into two descendant subsets beginning with X itself. These subsets are termed nodes in tree theory, with the root node $t_1 = X$. The entire construction of a tree, then, is determined by three characteristics:

(1) Selection of the splits.

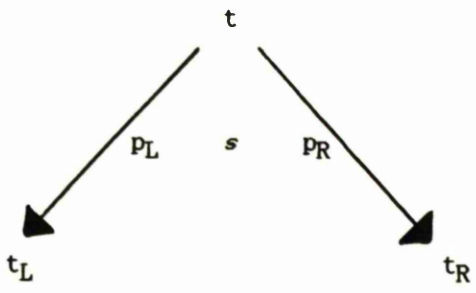
(2) Determining whether a node is terminal or non-terminal.

(3) The assignment of each terminal node to a class, or more

generally, probabilities of class membership.

For any node t, suppose that there is a split s of the node which divides it into t_L and t_R such that a proportion p_L of the cases in t go into t_L and a proportion p_R go into t_R (Figure 4.1).

Figure 4.1



The first problem in the construction of a classification tree is to determine which binary splits separate out the different classes in the measurement space. Each split of a subset should be selected such that the descendant subsets are "purer".

The goodness of split of a node is defined to be the decrease in the impurity measure

$$\Delta i(s,t) = i(t) - p_L i(t_L) - p_R i(t_R)$$

which is derived from an impurity function. As defined by Breiman et al.(1984), an impurity function Φ defined on the set of all J-tuples of numbers $(p_1,.....,p_j)$, where p_i denotes the proportion of class i profiles in any node, satisfying $p_i \geq 0$, $j=1,....J$, $\sum_j p_j = 1$ with the properties:

(1) Φ is a maximum only at the point $(1/j,1/j,.....,1/j)$,

(2) Φ achieves its minimum only at the points $(1,0,.....,0)$,

$(0,1,0,\ldots,0)$, $\ldots\ldots,(0,0,\ldots,0,1)$

(3) Φ is a symmetric function of $p_1,\ldots\ldots,p_j$.

The impurity measure of any node is then defined as:

$$i(t) = \Phi(p(1|t),p(2|t),\ldots\ldots,p(j|t))$$

where $p(j|t)$ is the proportion of cases $\underline{x}_n \in t$ which belong to class j.

The selection of the next split, s^*, is then chosen at the split which gave the largest decrease in impurity. By successive splits, a large binary tree, T_{max}, is developed which has all nodes declared terminal.(A node is declared terminal when the node cases are all in the same class or the nodes are small - contain less than 5 cases).

A tree constructed in this manner will generally be much too large and will require "pruning" back to obtain the right sized tree. Selecting the "best" sized tree requires estimating the "true misclassification rate", $R^*(T)$,(see Appendix II), calculated from a test set for example, at each stage in the pruning process. The pruning process begins with the largest tree T_{max}, computing the misclassification rate R(T) for each node $t \in T_{max}$, and progressively pruning T_{max} upward to the root node such that at each stage of pruning, R(T) is as small as possible. Thus in this process, the sequence of smaller and smaller trees $T_{max},T_1,T_2,\ldots\ldots,t_1$ (the root node) is constructed.

The "best" sized tree is then identified as the simplest tree whose accuracy or estimated true misclassification rate is comparable to the minimum $R^*(T)$ (within one standard error).

In mathematical terms, the right sized tree selected, T_{k1}, can be defined as the maximum k satisfying

$$\hat{R}(T_{k1}) \leq \hat{R}(T_{k0}) + SE(\hat{R}(T_{k0}))$$

where $\hat{R}(T_{k0}) = \min_k \hat{R}(T_k)$

and $SE(\hat{R}(T_{k0})) = [\hat{R}(T_{k0})(1-\hat{R}(T_{k0}))/N]^{\frac{1}{2}}$

where N = number of cases in the test sample or cross validation technique.

Finally, having identified this tree with a set of terminal nodes denoted by $\tilde{T}$, each $t \in \tilde{T}$ has to be assigned a class $j \in \{1,\ldots,J\}$ according to a class assignment rule. Essentially a class assignment rule allocates class j to node t, if

$$p(j|t) = \max_i p(i|t)$$

(In the case of ties, the assignment rule arbitrarily assigns one of the maximising classes to node t).

4.2 Splitting Rules within CART

Two splitting rules available within CART to construct classification trees are the Gini Index of diversity and the twoing rule.

The Gini Index of diversity assigns the measure of node impurity to be

$$i(t) = \sum_{i \neq j} p(i|j)p(j|t)$$

At a node t, with s splitting t into t_L and t_R, the twoing rule choses the split s that maximises

$$\frac{p_L p_R}{4} \left[\sum_j |p(j|t_L) - p(j|t_R)|\right]^2$$

It has been suggested (Breiman et al.) that properties of the final tree are insensitive to the choice of splitting rule and it has been proposed that the criterion used to prune or recombine upward is more important.

4.3 Missing Values

Unlike most discrimination techniques, CART uses an algorithm to deal with missing values. At a node t, the algorithm identifies the best split s of t using the variable x_s and then selects the next best split, s', on the variables other than x_s. s' is defined as the best surrogate for s. The algorithm continues to identify a second best surrogate, third best, and so on. Therefore, if a case has x_s missing, it goes to t_L or t_R using the best surrogate split, or if $x_{s'}$ is missing, use the second best surrogate split and so on.

4.4 Performance of CART

The resubstitution estimate of the misclassification cost of a tree T, as described in Appendix 2, tends to be less accurate than the other two estimates. Using an independent test sample is computationally more efficient and is preferred when the learning sample contains a large number of cases resulting in a relatively unbiased estimate of the misclassification costs.

Although computationally more expensive, the cross-validation estimate makes use of all the cases and gives more information regarding the stability of the tree structure.

In the particular application to the head injury study, the method using cross-validation estimates was selected to construct the tree, T, with the training set (Set A) mentioned in the previous chapter.

Using all the cases in the training set, no useful classification tree could be constructed.(A tree with only 2

terminal nodes). However, a table of variable importance (Table 4.1) was listed. Evidence from this table suggests that the method of linear logistic regression (LLR) selects nearly the same variable subset as classification trees - LLR selected {COMASUM,VOM,FAC,SCALP} as the best variable subset. These variables are in the top 5 in order of variable importance.

Thus employing only the four variables selected by LLR and by varying the cost of misclassifying a class j object as a class i object, say C(i|j), and the number of cases in the randomly selected subset of no fractures, a more useful tree was constructed. Subset sizes of 43, 86, 172 (ratios of skull fracture patients to no skull fracture patients of 1:1, 1:2, 1:4) and a larger subset of 500 no fracture cases were chosen. Using the CART package, only the subset of 172 no skull fracture cases and 43 skull fractures with the cost of misclassifying a no skull fracture as a skull fracture equal to 1 and misclassifying a skull fracture as a no skull fracture equal to 5, gave a useful tree - Figure 4.1 (all other trees had 4 or less terminal nodes).

Having obtained the best tree, all training cases (1788 patients) and test cases (1636 patients) were run down the tree. The number of cases misclassified and the probability of misclassifying a patient at each node were calculated. (see Tables 4.2(i) - (iii)).

It would appear that the probability of misclassification at each node differs - at terminal nodes assigned the class no skull fracture, the performance is extremely good with the performance at terminal nodes assigned the class skull fracture being extremely poor. As in the linear logistic regression model, the poor performance of the method results from the small proportion of cases in the data set having a skull fracture. However, there

Table 4.1 Variable Importance

Variable	Relative Importance*	Number of Categories
COMASUM	100	2
FAC	37	2
VOM	36	2
SEX	33	2
SCALP	31	2
FOC	28	2
DETERN	11	3
AGEGROUP	9	2
UNCON	8	2
EP	4	2
ALC	3	3
PUP	0	2

* Defining the measure of importance of variable x_m, using the Gini splitting rule, as

$$M(x_m) = \sum_{t \in T} \Delta i(\tilde{s}_m, t)p(t)$$

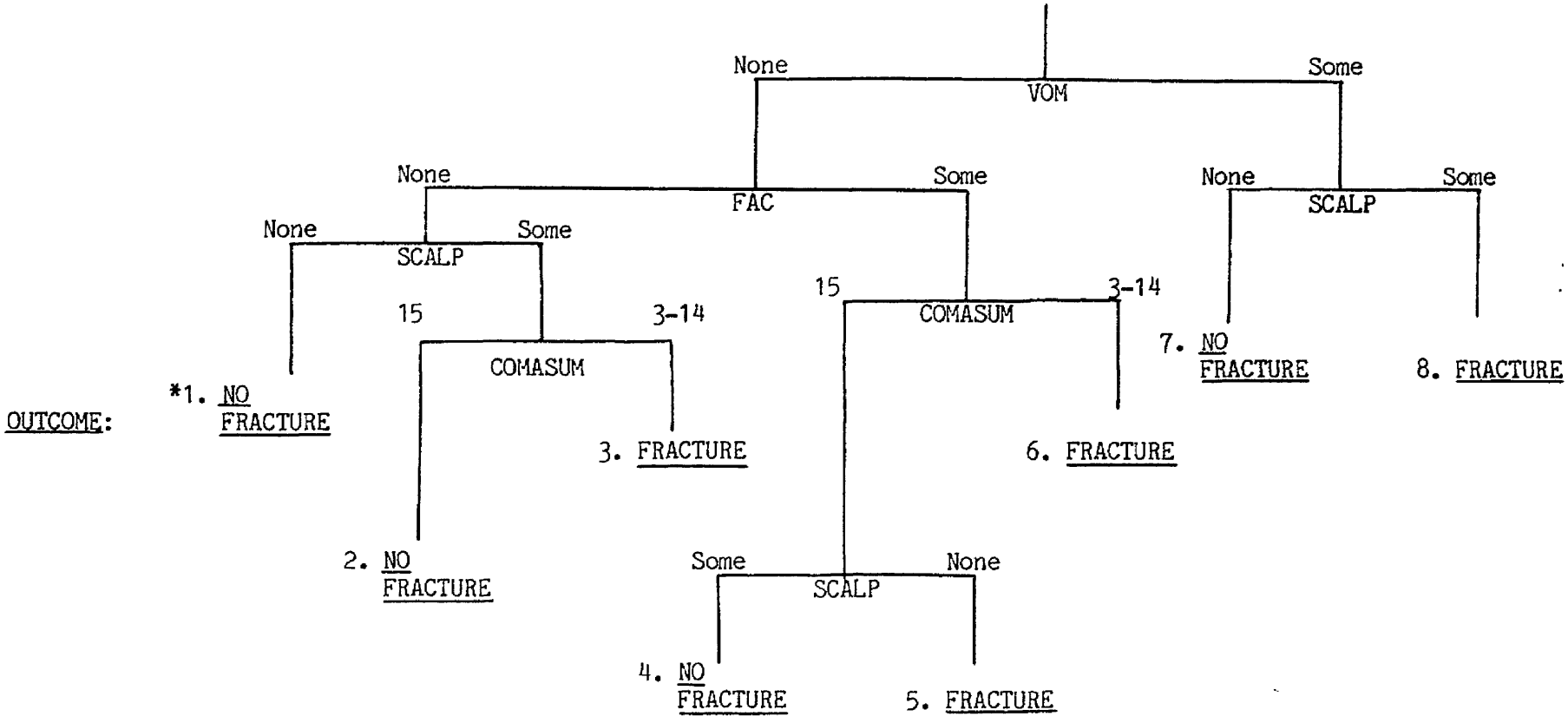
where p(t) = the probability that a case is in node t and $\tilde{s}_m$ is the surrogate split on x_m.

Then relative importance is defined as:

$$100\ M(x_m)/\max_M M(x_m)$$

FIGURE 4.1

'Best' Classification Tree



* denotes node number

Table 4.2 The proportion of cases misclassified for the three populations

(i) 215 cases - 43 fractures and 172 no fractures

Node	Classification	No.of cases	No. of cases misclassified	Probability of misclassification
1	No fracture	27	1	0.04
2	No fracture	100	12	0.12
3	Fracture	11	8	0.73
4	No fracture	10	1	0.10
5	Fracture	16	12	0.75
6	Fracture	3	1	0.33
7	No fracture	7	-	-
8	Fracture	41	21	0.51
Total	Tree	215	56	0.26

(ii) 1788 cases - Training Set (Set A)

Node	Classification	No.of cases	No. of cases misclassified	Probability of misclassification
1	No fracture	197	1	0.01
2	No fracture	966	12	0.01
3	Fracture	45	42	0.93
4	No fracture	64	1	0.02
5	Fracture	146	142	0.97
6	Fracture	13	11	0.85
7	No fracture	113	-	-
8	Fracture	244	224	0.92
Total	tree	1788	433	0.24

(iii) 1636 cases - Test data (Set B)

Node	Classification	No.of cases	No. of cases misclassified	Probability of misclassification
1	No fracture	173	-	-
2	No fracture	857	7	0.01
3	Fracture	28	28	1.00
4	No fracture	59	1	0.02
5	Fracture	152	150	0.99
6	Fracture	12	11	0.92
7	No fracture	104	3	0.03
8	Fracture	251	242	0.96
Total	tree	1636	442	0.27

may be subtle differences between these two types of analysis.

4.5 Comparison of CART with Linear Logistic Regression

The difference between the classification tree analysis and linear logistic regression was assessed for skull fracture and no skull fracture separately using the Brier Score (Brier,1950). In this particular application, the Brier Score can be thought of as the average "distance" between the estimated probabilities of fracture and no fracture and a perfect prediction which assigns probability 1 to the correct classification. For a fracture case, the contribution to the Brier Score is:

$$[\hat{p}(\#) - 1]^2 + [\hat{p}(\text{no } \#) - 0]^2$$

which reduces to $2[1 - \hat{p}(\#)]^2$, and for a no fracture case the contribution is:

$$[\hat{p}(\#) - 0]^2 + [\hat{p}(\text{no } \#) - 1]^2$$

which reduces to $2[\hat{p}(\#)]^2$.

($\hat{p}(\#)$ denotes the estimated probability of a skull fracture).

The Brier Score is then obtained by averaging these contributions over all cases in the test data set. The score may take values between 0 and 2, with small values indicating good performance. Table 4.3 shows the contributions separately for the skull fracture and no skull fracture cases, and separately for each cell. The overall scores were 0.026 for CART and 0.025 for LLR which again emphasise the similarity of performance of the two approaches.

Although only small differences exist, the classification

TABLE 4.3 Calculations for the Brier Score.

Cell (SCALP, VOM, COMASUM, FAC)				Test Data		CART	LLR	No Fracture		Fracture	
				No #	#	$\hat{p}$ (#/cell)	$\hat{p}$ (#/cell)	$2\hat{p}$ (#/cell)^{2}x no# CART	$2\hat{p}$ (#/cell)^{2}x no# LLR	$2(1-\hat{p}$(#/cell)^{2}x# CART	$2(1-\hat{p}$(#/cell)^{2}x# LLR
1	1	1	1	16	-	0.0051	0.0172	0.0008	0.0096	0.0000	0.0000
1	1	1	2	10	-	0.1540	0.0806	0.4740	0.1300	0.0000	0.0000
1	1	2	1	157	-	0.0051	0.0023	0.0078	0.0016	0.0000	0.0000
1	1	2	2	149	2	0.0274	0.0113	0.2235	0.0373	3.7840	3.9200
1	2	1	1	11	-	0.0000	0.0497	0.0000	0.0550	0.0000	0.0000
1	2	1	2	2	-	0.0000	0.2079	0.0000	0.1720	0.0000	0.0000
1	2	2	1	68	3	0.0000	0.0067	0.0000	0.0061	6.0000	5.9580
1	2	2	2	18	-	0.0000	0.0330	0.0000	0.0360	0.0000	0.0000
2	1	1	1	28	-	0.0670	0.0876	0.2520	0.4200	0.0000	0.0000
2	1	1	2	1	1	0.1540	0.3251	0.0474	0.2110	1.4310	0.9100
2	1	2	1	845	7	0.0124	0.0123	0.2535	0.2535	13.6570	13.6570
2	1	2	2	58	-	0.0156	0.0589	0.0070	0.4060	0.0000	0.0000
2	2	1	1	20	5	0.0820	0.2233	0.2680	1.9940	8.4250	6.0300
2	2	1	2	3	2	0.0820	0.5906	0.0402	2.0931	3.3700	0.6700
2	2	2	1	195	1	0.0820	0.0360	2.6130	0.5070	1.6850	1.858
2	2	2	2	19	-	0.0320	0.1579	0.2546	0.9462	0.0000	0.0000
								P_1=0.0028	P_1=0.0045	P_2=1.8260	P_2=1.5700

# denotes fracture

tree approach results in a simpler method for allocating future patients. Medical staff could easily follow the pathway in the tree for a particular patient in the prediction of a skull fracture without having to carry out the complex mathematical procedure of calculating the probability of a skull fracture from the linear logistic regression approach.

Another advantage of a tree structure is that once a classification tree has been constructed, there is no need to interpret the probability of a skull fracture or what probability to select for deciding whether a patient has a skull fracture.

CHAPTER 5

CALCULATION OF CONFIDENCE INTERVALS FOR RISKS OF HAEMATOMA

Recently in the medical literature, Mendelow et al.(1983) estimated the risk of an adult (age 15 or over) developing a surgically significant intracranial haematoma after a head injury using two easily measured features - presence/absence of a skull fracture and determination of the conscious level. Mendelow et al. calculated their risks of a haematoma based on a sample of 545 patients with haematomas at the Southern General Hospital, 2773 head injured patients at A and E Departments, and 2783 head injured patients at Primary Surgical Wards. The main aim of this study is to extend the risk factors of haematoma to children and to include more features to estimate risks such that patients could be identified as either having a low or high risk of developing an intracranial haematoma

5.1 The Data Set

For the purpose of this study, the A and E Department data discussed in Mendelow et al. was used with the addition of the head injury data from the A and E department at Monklands District General Hospital employed in the previous chapters. The head injury data from the Haematoma Study conducted at the Southern General Hospital was extended to cover the years 1974-1984 inclusively. The number of head injured patients, of all ages, in each study involved in the A and E data is shown on Table 5.1.

Table 5.1 No. of patients in the different studies in the A and E data

Study	No. of patients
Monklands	3424
Glasgow Royal Infirmary	797
SHIMS	3567
Teesside	716
Total	8504

Patient A and E data from the SHIMS and Teesside studies were recorded on the same type of form which differed from both forms used at the Glasgow Royal Infirmary (GRI) and Monklands District General Hospital. Despite this, 5 features common to all studies could be identified. Each of the 5 features identified - Age, Cause of Injury, Glasgow Coma Sum, Sex and Skull Fracture - had a good proportion of patients in each category and a low proportion of missing observations as compared to other variables recorded (Table 5.2).

For consistency over all studies, the Cause of Injury variable had categories reduced to Road Traffic Accident (RTA) or Non-Road Traffic Accident (Non-RTA).

The Glasgow Coma Sum was recorded in both Monklands and GRI studies but had to be "manufactured" for the SHIMS and Teesside studies in the following manner:

If a patient was talking sensibly and obeying commands he was scored as having a Coma Sum of 15 ;

a) If a patient was obeying commands but was not orientated **or**

b) If a patient was not obeying but was talking sensibly or confused, he was allocated to the 9-14 category ;

Table 5.2 A and E Characteristics (n = 8504)

VARIABLE	n	%
AGE		
1.<15	3614	42.5
2.15-64	4258	50.1
3.65 or over	534	6.3
Not recorded	98	1.2
CAUSE OF INJURY		
1.RTA	1068	12.6
2.Non-RTA	7045	82.8
Not recorded	391	4.6
GLASGOW COMA SUM		
1.3-8	69	0.8
2.9-14	421	5.0
3.15	7973	93.8
Not recorded	41	0.5
SEX		
1.Male	5941	69.9
2.Female	2543	29.9
Not recorded	20	0.2
SKULL FRACTURE		
1.No	8333	98.0
2.Yes (Vault and/or Base)	171	2.0

If a patient was not obeying and was not talking or talking unspecified, he was allocated to the category 3-8.

The remaining 3 variables were consistent throughout all studies.

The "not recorded" category for each feature was omitted and initially calculations were conducted on adults only (i.e. patients ≥15 years of age). This left a total of 4574 (95%) adults with complete data out of 4792 head injured patients from the A and E data set.

The corresponding number from the Haematoma Study carried out at the Southern General Hospital resulted in 844 (85%) adults with complete data from a total sample of 988 haematoma cases.

Employing these data sets, the frequency of features in the study were calculated for the A and E and Haematoma data (Tables 5.3 and 5.4 respectively) and the subsequent absolute and relative risks evaluated. Using so many risk factors leads to a small number of patients in some categories (e.g. Table 5.3 has 13 of the 24 categories with <10 patients). Due to this characteristic, an alternative methodology based on log linear modelling was used to produce more reliable estimates of the relative risk.

To demonstrate the statistical methodology, it is easier to consider 3 dimensional tables. Results from such tables can be readily extended to 4 and higher dimensional contingency tables. To apply the procedure to a 3-way table, the Skull Fracture variable was collapsed leaving the three variables; Cause of Injury, Glasgow Coma Sum and Sex to be employed in the methodology.

Table 5.3 A and E Data Set (n = 4574)

CAUSE	COMASUM	SEX	FRACTURE	
			NO	YES
RTA	15	Male	481	3
		Female	251	5
	9-14	Male	36	2
		Female	9	3
	3-8	Male	8	8
		Female	4	2
Non-RTA	15	Male	2547	37
		Female	920	14
	9-14	Male	147	15
		Female	58	0
	3-8	Male	13	8
		Female	3	0

Table 5.4 Haematoma Data Set (n = 844)

CAUSE	COMASUM	SEX	FRACTURE	
			NO	YES
RTA	15	Male	6	12
		Female	2	5
	9-14	Male	9	50
		Female	4	20
	3-8	Male	11	83
		Female	7	14
Non-RTA	15	Male	22	53
		Female	5	15
	9-14	Male	53	160
		Female	21	14
	3-8	Male	37	198
		Female	15	28

5.2 Risks of Intracranial Haematoma

Although the estimated probability ($\hat{\theta}_{ijk}$) of falling into the (i,j,k)th cell is useful, more information may be gained by estimating the risk, or calculating approximate 95% confidence intervals for the risk, that a head injured patient will develop a surgically significant intracranial haematoma.

5.2.1 Calculating Absolute Risks from the Raw Data

The absolute risk is expressed as the frequency of a traumatic haematoma in the total number of patients with a given set of features in the referral population - that is, Accident and Emergency departments in the West of Scotland. The total number of head injured patients who attend A and E departments in the West of Scotland during the eleven years over which the haematoma data had been collected had to be estimated. These yearly estimates were based both on the Scottish Head Injury Management Study and on the Scottish Mortality records. Over the eleven years, the number of adult attenders with a head injury, allowing for an increase in such patients in the A and E departments each year, was estimated to be 344000. Estimation of the risk of an intracranial haematoma in absolute terms, requires data about the total number of head injured patients in the different groups seen at A and E departments. Complete data were available in only 844 of the 988 patients with a haematoma. Therefore the total A and E estimates for this period were reduced by a corresponding factor, assuming that the missing cases were randomly distributed.

Using the corrected figure of 293711 as a base, the values of the 12 different combinations, of the features in the A and E patients during the period of analysis, were estimated from their

frequencies in the samples in Table 5.5(i). These estimates, A and E "scaled up" are shown in Table 5.5(ii). The absolute risks for each of the feature combinations were evaluated by dividing the A and E "scaled up" value by the corresponding values from the Haematoma sample (Table 5.5(ii)).

5.2.2 Confidence Intervals for the Absolute Risks

In this section, interest is in estimating the ratio, Ψ, of the proportions of head injured patients in the A and E and Haematoma studies who fall into a particular category. This ratio, when multiplied by the overall risk, is often called the Absolute Risk. The overall risk is simply:

$$\frac{\text{total no. of patients in A and E "scaled up" sample}}{\text{total no. of patients in Haematoma sample}}$$

Katz et al.(1978) produced a method for calculating confidence intervals for this ratio when each of the proportions were relatively small. Recently however, Bailey (1987) has produced an alternative method allied to that recommended by Katz et al. but which is more stable and simpler to use.

5.2.3 Calculating Intervals

Let X be the number of patients in the A and E sample in cell (i,j,k) say, with associated probability λ. Then $X \sim Bi(m,\lambda)$ where m is the total number of patients in the A and E sample.

Similarly, for the Haematoma sample, let $Y \sim Bi(n,\mu)$ where n is the total number of patients in the Haematoma sample and μ is the probability of a head injured patient with a haematoma falling into cell (i,j,k).

Table 5.5 Data Sets Employed in the Construction of Absolute Risks

(i)

CAUSE	COMASUM	SEX	Haematoma Sample	A and E Sample
RTA	15	Male	18	484
		Female	7	256
	9-14	Male	59	38
		Female	24	12
	3-8	Male	94	16
		Female	21	6
Non-RTA	15	Male	75	2584
		Female	20	934
	9-14	Male	213	162
		Female	35	58
	3-8	Male	235	21
		Female	43	3
			844	4574

(ii)

CAUSE	COMASUM	SEX	A and E sample Scaled Up	Absolute Risk 1:
RTA	15	Male	31079	1700
		Female	16439	2300
	9-14	Male	2440	41
		Female	771	32
	3-8	Male	1027	11
		Female	385	19
Non-RTA	15	Male	165927	2200
		Female	59975	3000
	9-14	Male	10403	49
		Female	3724	110
	3-8	Male	1348	6
		Female	193	4
			293711	348

Denote $\theta = \lambda/\mu$ and let $p_X = X/m$ and $p_Y = Y/n$ be the observed proportions of patients falling into cell (i,j,k) in the A and E and Haematoma study respectively. Provided m and n are not too small, it can be assumed that $U = p_X^t - \theta^t p_Y^t$ is approximately normally distributed for constant t.

Bailey states that U has zero mean and

$$\mathrm{Var}(U) = t^2[\lambda^{2t-1}(1-\lambda)/m + \theta^{2t}\mu^{2t-1}(1-\mu)/n]$$

Replacing λ and μ with the estimators p_X and P_Y respectively, results in the pivotal random variable

$$Z = (p_X^t - \theta^t p_Y^t)/(t(p_X^{2t-1}q_X/m + \theta^{2t}p_Y^{2t-1}q_Y/n)^{1/2}) \qquad (5.1)$$

where $q_X = 1-p_X$ and $q_Y = 1-p_Y$. Approximate confidence intervals for θ can be obtained by setting Z equal to an appropriate deviate of the standard normal distribution. Solving (5.1) for θ, the general form of the two limits for the confidence interval are given by:

$$\theta^-, \theta^+ = \frac{p_X}{p_Y}\left[\frac{1 \pm zt\sqrt{\frac{q_X}{x} + \frac{q_Y}{y} - \frac{z^2t^2q_Yq_X}{xy}}}{1 - z^2t^2\frac{1}{y}q_Y}\right]^{1/t}$$

Bailey suggests that a suitable choice of t is the value that minimises the skewness of Z. After rearranging, the first order term in the third central moment of Z is 0 when

$$t = \frac{1}{3} + \frac{2}{3}\,[m^2\lambda^2\mu(1-\mu) - n^2\mu^2\lambda(1-\lambda)] \,/\, [m^2\lambda^2(1-\mu)^2 - n^2\mu^2(1-\lambda)^2]$$

5.2.4 Numerical Results

Employing the data from the Haematoma and A and E studies, 95% confidence intervals for Ψ were obtained for all 12 feature vectors. By multiplying these intervals by the overall risk (estimated to be 350), the corresponding confidence intervals for the absolute risk of an intracranial haematoma were induced - Table 5.6. These intervals may however, be too narrow due to the uncertainty in the overall risk

5.3 Relative Risks

For the IJK cells 1,2,....,IJK (IJK = 12 in this case) defined by the three variables Cause of Injury, Glasgow Coma Sum and Sex, the A and E sample of head injured patients over all IJK cells will have a multinomial distribution with expected proportions $\theta^{(A)}_{111}, \theta^{(A)}_{211}, \ldots, \theta^{(A)}_{IJK}$ where $\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \theta^{(A)}_{ijk} = 1$. Denote the respective observed proportions by $\hat{\theta}^{(A)}_{111}, \hat{\theta}^{(A)}_{211}, \ldots, \hat{\theta}^{(A)}_{IJK}$. Similarly, the Haematoma sample of head injured patients over the IJK cells will have expected proportions $\theta^{(H)}_{111}, \theta^{(H)}_{211}, \ldots, \theta^{(H)}_{IJK}$, with $\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \theta_{ijk} = 1$ and respective observed proportions $\hat{\theta}^{(H)}_{111}, \hat{\theta}^{(H)}_{211}, \ldots, \hat{\theta}^{(H)}_{IJK}$. The relative risk is then defined as the risk that a patient with feature vector (i,j,k) will develop an intracranial haematoma as compared with a patient in a reference cell. If, say, cell (I,J,K) is the reference cell, denoted by θ_{REF}, the relative risk of a haematoma for cell (i,j,k) is defined as

$$\frac{\theta^{(H)}_{ijk}}{\theta^{(A)}_{ijk}} \Big/ \frac{\theta^{(H)}_{REF}}{\theta^{(A)}_{REF}}$$

Table 5.6 95% Confidence Intervals for Absolute Risks

CAUSE OF INJURY	COMASUM	SEX	ABSOLUTE RISK 1:
RTA	15	Male	[1100,2900]
		Female	[1200,5600]
	9-14	Male	[28 , 62]
		Female	[16 , 63]
	3-8	Male	[6 , 18]
		Female	[6.7, 43]
Non-RTA	15	Male	[1800,2800]
		Female	[2000,4800]
	9-14	Male	[41 , 59]
		Female	[71 , 160]
	3-8	Male	[3.6, 8.8]
		Female	[1.2, 13]

and is estimated by

$$\frac{\hat{\theta}^{(H)}_{ijk}}{\hat{\theta}^{(A)}_{ijk}} \Big/ \frac{\hat{\theta}^{(H)}_{REF}}{\hat{\theta}^{(A)}_{REF}} \qquad (5.2)$$

In this study, two methods of calculating approximate 95% confidence intervals for relative risks were proposed. Although relative risk may not be as relevant as the absolute risk, it does not need an estimate of the A and E population and hence the properties of the model can be modelled.

5.3.1 Confidence Intervals for Relative Risks: Method A

In general for the (i,j,k)th cell, considering only the four cells with expected proportions $\theta^{(H)}_{ijk}$, $\theta^{(A)}_{ijk}$, $\theta^{(H)}_{REF}$, $\theta^{(A)}_{REF}$ in a 2x2 table formed from two unrelated binomial distributions with observed proportions $\hat{\theta}^{(H)}_{ijk}$, $\hat{\theta}^{(A)}_{ijk}$, $\hat{\theta}^{(H)}_{REF}$, $\hat{\theta}^{(A)}_{REF}$, the relative risk may be written as:

$$\frac{\theta^{(H)}_{REF}}{1 - \theta^{(H)}_{REF}} \Big/ \frac{\theta^{(A)}_{REF}}{1 - \theta^{(A)}_{REF}} \qquad (5.3)$$

In the 2x2 table, $\theta^{(H)}_{ijk} = 1 - \theta^{(H)}_{REF}$ and similarly $\theta^{(A)}_{ijk} = 1 - \theta^{(A)}_{REF}$.

The maximised likelihood estimate of (5.3) is obtained by replacing the expected proportions by their respective observed proportions. Taking logarithms, the estimated relative frequency is:

$$\log\left[\frac{\hat{\theta}^{(H)}_{REF}}{1 - \hat{\theta}^{(H)}_{REF}}\right] - \log\left[\frac{\hat{\theta}^{(A)}_{REF}}{1 - \hat{\theta}^{(A)}_{REF}}\right]$$

i.e.

$$[\log \hat{\theta}^{(H)}_{REF} - \log(1 - \hat{\theta}^{(H)}_{REF})] - [\log \hat{\theta}^{(A)}_{REF} - \log(1 - \hat{\theta}^{(A)}_{REF})]$$

which is a linear combination of the observed cell proportions with asymptotic variance

$$\frac{1}{n^{(H)}_{REF}} + \frac{1}{n^{(H)}_{ijk}} + \frac{1}{n^{(A)}_{REF}} + \frac{1}{n^{(A)}_{ijk}}$$ (see Bishop et al.,1975)

where $n^{(H)}_{REF}$, $n^{(H)}_{ijk}$, $n^{(A)}_{REF}$, $n^{(A)}_{ijk}$ are the observed entries in the 2x2 table. The reference cell for this method was identified as the cell corresponding to $\min_{ijk} \left[\frac{1}{n^{(A)}_{ijk}} + \frac{1}{n^{(H)}_{ijk}}\right]$.

Approximate 95% confidence intervals calculated using this method are shown in Table 5.7.

In general, these intervals, which are easily computed, may be too wide in practice. An alternative method of calculating 95% confidence intervals for the relative risk of an intracranial haematoma based on the log linear model is discussed in the following section. Firstly, the statistical methodology of the log linear model is described.

5.4 The Log Linear Model

Looking at several categorical variables simultaneously presents particular problems of analysis and interpretation. Such multidimensional contingency tables where each variable corresponds to one dimension of the table have become easier to handle by the wide range of statistical computing packages now available. In this particular application, three categorical

Table 5.7 95% Confidence Intervals for Relative Risks

CAUSE OF INJURY	COMASUM	SEX	RELATIVE RISK
RTA	15	Male	[0.9, 3.4]
		Female	[0.5, 3.1]
	9-14	Male	[39 , 130]
		Female	[41 , 220]
	3-8	Male	[140, 550]
		Female	[58 , 460]
Non-RTA	15	Male	[0.8, 2.2]
		Female	1
	9-14	Male	[37 , 100]
		Female	[15 , 52]
	3-8	Male	[280, 990]
		Female	[190,2400]

variables are dealt with.

In general, for a three dimensional IxJxK contingency table with a total sample size of N, refer to the number of individuals in the (i,j,k)th cell as x_{ijk}. Denote θ_{ijk} to be the probability that an individual falls into the (i,j,k)th cell. The simplest model for a three dimensional table corresponds to complete independence between all the three variables. For this model, the natural logarithm of the cell probabilities can be written in the form:

$$\log \theta_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k$$

subject to the restrictions:

$$\sum_{i=1}^{I} \alpha_i = \sum_{j=1}^{J} \beta_j = \sum_{k=1}^{K} \gamma_k = 0$$

where μ denotes the grand mean

and the α_i, β_j, γ_k represent main effects.

In any modelling exercise, it is very unlikely that the three variables for the data are indeed independent and hence a more complex log linear model containing two factor and higher order interaction terms will be required to adequately explain the data, particularly if N is large. In this study, only hierarchical log linear models will be considered. A hierarchical log linear model is one in which whenever an interaction term is included, all lower order interactions involving variables in the higher order term must be involved in the final equation. For example, in the full model

$$\log \theta_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk}$$

if the term $(\alpha\beta\gamma)_{ijk}$ is represented in the model, the terms $(\alpha\beta)_{ij}$, $(\alpha\gamma)_{ik}$ and $(\beta\gamma)_{jk}$ must also be included. Conversely, if $(\alpha\beta)_{ij} = 0$ for all values of i and j, then $(\alpha\beta\gamma)_{ijk} = 0$ for all i,j and k.

To select a working model, it is necessary to perform goodness of fit tests to identify the particular model which achieves a balance between simplicity and adequately fitting the data. The likelihood ratio test statistic is one general criterion for comparing expected values for two different log linear models, where one model is a special case of the second. Denoting the two fitted values for the observed frequency x_{ijk}, for model 1 and model 2 as $N\hat{\theta}_{ijk}(1)$ and $N\hat{\theta}_{ijk}(2)$ respectively, where model 2 is a special case of model 1, the likelihood ratio test statistic is defined as:

$$2 \sum_{i,j,k} x_{ijk} \log\left[\frac{N\hat{\theta}_{ijk}(1)}{N\hat{\theta}_{ijk}(2)}\right] \qquad (5.3)$$

This statistic tests whether the difference between the expected values for the two models is simply due to random variation, given that the true expected values satisfy model 1. This conditional test statistic has an asymptotic Chi-squared distribution under the null hypothesis (i.e. the extra parameter in model 2 equals 0), with degrees of freedom equal to the difference in the degrees of freedom of models 1 and 2. Expression (5.3) is simply the difference in the values of the likelihood ratio goodness of fit statistics for the two models. Thus by formulating a nested hierarchy of models of interest and by calculating the respective likelihood ratio goodness of fit statistics, it is possible to identify a working model which

adequately fits the data (Fienberg, 1977).

In multidimensional tables in general, before formulating a nested hierarchy of models of interest and simplifying the task ahead, it is often useful to identify the highest-order interactions in the full model which are definitely not zero. Parameters in the full model having approximate 95% confidence intervals not containing zero can be identified as those parameters whose standardised estimates (i.e. parameter estimate ÷ standard error of estimate) are greater than 1.96. Having identified the highest-order interactions, the fullest model in the sequence of the nested hierarchy should include these parameters plus all lower order interactions including the appropriate variables to obtain a hierarchical model.

5.4.1 Examples on Selection of a Model

Whilst 4 and higher dimensional contingency tables may require a method to identify the highest order interactions in the full model which are definitely not zero, to simplify the task ahead, the sequence in a nested hierarchy for a 3-way table, because of their relative simplicity, can begin with the full model or the model containing all two-way interactions (if the three way interaction can be assumed to equal zero). If a different nested hierarchy of models was chosen, by adding the two-factor effect terms in a different order, the method described previously may yield alternative "best" models.

The following notation is used in all further analysis:

Denote μ as the grand mean

s_i as the i^{th} level of the variable Sex (i=1,2)

m_j as the j^{th} level of the variable Coma Sum (j=1,2,3)

c_k as the k^{th} level of the variable Cause of Injury (k=1,2)

$(sm)_{ij}$ as the $(i,j)^{th}$ level of the Sex and Coma Sum interaction

$(sc)_{ik}$ as the $(i,k)^{th}$ level of the Sex and Cause of Injury interaction

$(mc)_{jk}$ as the $(j,k)^{th}$ level of the Coma Sum and Cause of Injury interaction

5.4.2 The A and E Study

Assuming the 3-way interaction to be zero, employing the model containing all 2-way interactions, the sequence of 4 models forming the nested hierarchy in Table 5.8(i) was constructed. Evaluation of the respective likelihood ratio goodness of fit statistics, $G^2(1)$-$G^2(4)$ say, and the degrees of freedom df_1-df_4 was performed using the BMDP program P4F (shown in Table 5.8(i)). As mentioned earlier in the chapter, the differences between the likelihood ratio goodness of fit statistics are of the form (5.3), and so $G^2(3)$-$G^2(4)$, $G^2(2)$-$G^2(3)$, $G^2(1)$-$G^2(2)$ can be used to test whether the difference between the expected values of models (3) and (4), (2) and (3), and (1) and (2) respectively, might simply be due to random variation.

From Table 5.8(i), the value of $G^2(3)$-$G^2(4)$ (=3.19) when referred to a Chi-squared distribution with df_3-df_4 (=2) degrees of freedom is not significant at the 0.05 level. This indicates that model 3 is preffered to model 4 and hence continue up the hierarchy. Proceeding to $G^2(2)$-$G^2(3)$ (=17.19) which exceeds the upper 5% level of the Chi-squared distribution with 1 df by a considerable amount. It thus makes sense to continue no further and to employ the model

$$\log \theta_{ijk} = \mu + s_i + m_j + c_k + (sc)_{ik} + (mc)_{jk}$$

to describe the data. Using this model, the estimate of the

A and E DATA

Table 5.8(i) Selection of Model

Model	G^2	d.f.
1. $\mu+s_i+m_j+c_k$	45.19	7
2. $\mu+s_i+m_j+c_k+(mc)_{jk}$	22.92	5
3. $\mu+s_i+m_j+c_k+(sc)_{ik}+(mc)_{jk}$	5.73	4
4. $\mu+s_i+m_j+c_k+(sm)_{ij}+(sc)_{ik}+(mc)_{jk}$	2.54	2

$G^2(3)-G^2(4)= 3.19$ referred to $X^2(4-2;0.95) = 5.99$

$G^2(2)-G^2(3)=17.19$ referred to $X^2(5-4;0.95) = 3.84$

$G^2(1)-G^2(2)=22.27$ referred to $X^2(7-5;0.95) = 5.99$

Table 5.8(ii) Probabilities Calculated from "best" Model

		SEX	
CAUSE OF INJURY	COMASUM	MALE	FEMALE
RTA	15	0.1072	0.0546
	9-14	0.0072	0.0037
	3-8	0.0032	0.0016
NON-RTA	15	0.5657	0.2034
	9-14	0.0354	0.0127
	3-8	0.0039	0.0014

probability of falling into each of the 12 cells is given in Table 5.8(ii).

5.4.3 Haematoma Study

Using exactly the same methodology as in 5.4.2 on the Haematoma data, the model

$$\log \theta_{ijk} = \mu + s_i + m_j + c_k + (sc)_{ik}$$

was selected to explain the data, when employing the sequence of four models in Table 5.9(i). The estimated probabilities of a head injured patient falling into each of the 12 cells, using the above model, are given in Table 5.9(ii)

5.4.4 Confidence Intervals for Relative Risks: Method B

Approximate 95% confidence intervals for relative risk of a haematoma taking as a reference the cell corresponding to female, Coma Sum 15 and Non-RTA, (θ_{232}), were calculated. θ_{232} was identified as the cell with the lowest absolute risk of haematoma in Table 5.5(ii).

Taking logarithms and replacing $\theta^{(A)}_{REF}$ and $\theta^{(H)}_{REF}$ with $\theta^{(A)}_{232}$ and $\theta^{(H)}_{232}$ respectively, 5.1 reduces to:

$$[\log \theta^{(A)}_{232} - \log \theta^{(A)}_{ijk}] - [\log \theta^{(H)}_{232} - \log \theta^{(H)}_{ijk}]$$

with corresponding approximate 95% Confidence Interval:

$$[(\log \theta^{(A)}_{232} - \log \theta^{(A)}_{ijk}) - (\log \theta^{(H)}_{232} - \log \theta^{(H)}_{ijk})] \pm 1.96\sqrt{\mathrm{Var}(\log \theta^{(A)}_{232} - \log \theta^{(A)}_{ijk}) + \mathrm{Var}(\log \theta^{(H)}_{232} - \log \theta^{(H)}_{ijk})}$$

Substituting $\log \theta^{(A)}_{ijk}$ and $\log \theta^{(H)}_{ijk}$ by the parameters in the log linear model obtained from section 5.4.2 and 5.4.3

HAEMATOMA DATA

Table 5.9(i) Selection of Model

Model	G^2	d.f.
1. $\mu+s_i+m_j+c_k$	15.84	7
2. $\mu+s_i+m_j+c_k+(sc)_{ik}$	9.74	6
3. $\mu+s_i+m_j+c_k+(sm)_{ij}+(sc)_{ik}$	7.41	4
4. $\mu+s_i+m_j+c_k+(sm)_{ij}+(sc)_{ik}+(mc)_{jk}$	2.90	2

$G^2(3)-G^2(4)=4.51$ referred to $X^2(4-2;0.95) = 5.99$

$G^2(2)-G^2(3)=2.33$ referred to $X^2(6-4;0.95) = 5.99$

$G^2(1)-G^2(2)=6.10$ referred to $X^2(7-1;0.95) = 3.84$

Table 5.9(ii) Probabilities Calculated from "best" Model

		SEX	
CAUSE OF INJURY	COMASUM	MALE	FEMALE
RTA	15	0.0288	0.0088
	9-14	0.0795	0.0242
	3-8	0.0943	0.0287
NON-RTA	15	0.0882	0.0165
	9-14	0.2430	0.0455
	3-8	0.2885	0.0540

respectively, formulae for the interval estimates for the logarithm of the relative risk of haematoma were obtained. For example, using the log linear model obtained from section 5.4.2 $[\log \theta^{(A)}_{232} - \log \theta^{(A)}_{111}]$ reduces to:

$$[\mu+s_2+m_3+c_2+(sc)_{22}+(mc)_{32}] + [\mu+s_1+m_1+c_1+(mc)_{11}+(sc)_{11}]$$

$$=-2s_1-2m_1-2c_1-m_2+(mc)_{21} \qquad (5.4)$$

with $\sum_{i=1}^{2} s_i = \sum_{j=1}^{3} m_j = \sum_{k=1}^{2} c_k = \sum_{j=1}^{3} (mc)_{jk} = \sum_{k=1}^{2} (mc)_{jk} = 0$

The variance term associated with (5.4) is :

$$4var(s_1) + 4var(m_1) + 4var(c_1) + var(m_2) + var((mc)_{21})$$
$$+ 8cov(s_1,m_1) + 8cov(s_1,c_1) + 4cov(s_1,m_2) - 4cov(s_1,(mc)_{21})$$
$$+ 8cov(m_1,c_1) + 4cov(m_1,m_2) - 4cov(m_1,(mc)_{21}) + 4cov(m_2,c_1)$$
$$- 4cov(c_1,(mc)_{21}) - 2cov(m_2,(mc)_{21})$$

Similarly, formulae for $\log \theta^{(H)}_{232} - \log \theta^{(H)}_{ijk}$ was obtained. Using the maximum likelihood estimates for the parameters in the log linear models in sections 5.4.2 and 5.4.3 (obtained from BMDP, P4F package) interval estimates of the relative risk of haematoma were induced from the corresponding interval estimates of the logarithm of the relative risk of haematoma (Table 5.10). Although more difficult to computate, and requiring more distributional assumptions, these intervals are generally narrower than the corresponding intervals evaluated using Method A.

For simplicity, the methodology used in this chapter was evaluated on the 3 categorical variables; Cause of Injury, Glasgow Coma Sum, and Sex. However, as mentioned previously, the main interest of the clinicians involved in the study was to

Table 5.10 95% Confidence Intervals for Relative Risks

CAUSE OF INJURY	COMASUM	SEX	RELATIVE RISK
RTA	15	Male	[2.5, 4.4]
		Female	[1.4, 2.9]
	9-14	Male	[87 , 210]
		Female	[49 , 130]
	3-8	Male	[210, 630]
		Female	[120, 400]
Non-RTA	15	Male	[1.5, 2.4]
		Female	1
	9-14	Male	[60 , 120]
		Female	[34 , 57]
	3-8	Male	[550,1500]
		Female	[300, 760]

calculate the risks of intracranial haematoma based on the four categorical variables and to evaluate the risks for children. These points are dealt with in the following chapter.

CHAPTER 6

FURTHER ANALYSIS

The two objectives in this particular study were identified in Chapter 5. The first objective was the construction of 95% confidence intervals for Relative and Absolute risks of an intracranial haematoma employing four potential features of the data. The second was to widen the risk of intracranial haematoma to include children and will be discussed later in this chapter. The 95% confidence intervals for the relative risk of an intracranial haematoma were calculated employing Method B as the computation was reduced substantially using the widely available statistical package BMDP, P4F. Also, the small number of observations in some of the cells would evoke large confidence intervals.

Considering the four variables Cause of Injury, Glasgow Coma Sum, Skull Fracture and Sex recorded in the A and E study and the Haematoma Study, and employing the statistical methodology discussed in the previous chapter, log linear models were identified which adequately described the A and E and Haematoma data (Tables 6.1 and 6.2). In Table 6.1, when $G^2(3)$ (=20.72) is referred to the 95% level of the Chi-squared distribution with 13 degrees of freedom (=22.36) it can be concluded that this particular model fits the A and E data reasonably well. Similarly from Table 6.2, the model identified as being the most adequate had a likelihood goodness of fit statistic equal to 22.74 which, when compared to the 95% level of the Chi-squared distribution with 13 degrees of freedom (=22.36) indicates that model 3 fits the data very well. It is worth mentioning that in

Table 6.1 Selection of a Model - A and E Data

	Model	G^2	d.f.
	1. $\mu+s_i+m_j+c_k+f_l+(mf)_{jl}$	62.88	16
	2. $\mu+s_i+m_j+c_k+f_l+(sc)_{ik}+(mf)_{jl}$	45.22	15
*	3. $\mu+s_i+m_j+c_k+f_l+(sc)_{ik}+(mc)_{jk}+(mf)_{jl}$	20.72	13
	4. $\mu+s_i+m_j+c_k+f_l+(sm)_{ij}+(sc)_{ik}+(mc)_{jk}+(mf)_{jl}$	18.62	11
	5. $\mu+s_i+m_j+c_k+f_l+(sm)_{ij}+(sc)_{ik}+(mc)_{jk}+(mf)_{jl}+(cf)_{kl}$	18.45	10
	6. $\mu+s_i+m_j+c_k+f_l+(sm)_{ij}+(sc)_{ik}+(sf)_{il}+(mc)_{jk}+(mf)_{jl}+(cf)_{kl}$	18.45	9
	7. $\mu+s_i+m_j+c_k+f_l+(sm)_{ij}+(sc)_{ik}+(sf)_{il}+(mc)_{jk}+(mf)_{jl}+(cf)_{kl}+(scf)_{ikl}$	14.97	8
	8. $\mu+s_i+m_j+c_k+f_l+(sm)_{ij}+(sc)_{ik}+(sf)_{il}+(mc)_{jk}+(mf)_{jl}+(cf)_{kl}+(scf)_{ikl}+(mcf)_{jkl}$	12.86	6

$G^2(7)-G^2(8) = 2.11$ referred to $X^2(8-6;0.95) = 5.99$

$G^2(6)-G^2(7) = 3.48$ referred to $X^2(9-8;0.95) = 3.84$

$G^2(5)-G^2(6) = 0$ referred to $X^2(10-9;0.95) = 3.84$

$G^2(4)-G^2(5) = 0.17$ referred to $X^2(11-10;0.95) = 3.84$

$G^2(3)-G^2(4) = 2.10$ referred to $X^2(13-11;0.95) = 5.99$

$G^2(2)-G^2(3) = 24.5$ referred to $X^2(15-13;0.95) = 5.99$

$G^2(1)-G^2(2) = 17.66$ referred to $X^2(16-15;0.95) = 3.84$

* denotes "best" model

(Note that s_i, m_j, c_k are as described in Chapter 5 with f_l denoting the l^{th} level of the variable SKULL FRACTURE)

Table 6.2 Selection of a Model - Haematoma Data

	Model	G^2	d.f.
	1. $\mu+s_i+m_j+c_k+f_l+(sc)_{ik}+(sf)_{il}$	39.62	16
	2. $\mu+s_i+m_j+c_k+f_l+(sc)_{ik}+(sf)_{il}+(cf)_{kl}$	33.47	15
*	3. $\mu+s_i+m_j+c_k+f_l+(sc)_{ik}+(sf)_{il}+(mf)_{jl}+(cf)_{kl}$	22.47	13
	4. $\mu+s_i+m_j+c_k+f_l+(sc)_{ik}+(sf)_{il}+(mc)_{jk}+(mf)_{jl}+(cf)_{kl}$	19.76	11
	5. $\mu+s_i+m_j+c_k+f_l+(sm)_{ij}+(sc)_{ik}+(sf)_{il}+(mc)_{jk}+(mf)_{jl}+(cf)_{kl}$	17.62	9
	6. $\mu+s_i+m_j+c_k+f_l+(sm)_{ij}+(sc)_{ik}+(sf)_{il}+(mc)_{jk}+(mf)_{jl}+(cf)_{kl}+(smf)_{ijl}$	12.01	7
	7. $\mu+s_i+m_j+c_k+f_l+(sm)_{ij}+(sc)_{ik}+(sf)_{il}+(mc)_{jk}+(mf)_{jl}+(cf)_{kl}+(smf)_{ijl}+(mcf)_{jkl}$	7.89	5

$G^2(6)-G^2(7) = 4.12$ referred to $X^2(7-5;0.95) = 5.99$

$G^2(5)-G^2(6) = 5.61$ referred to $X^2(9-7;0.95) = 5.99$

$G^2(4)-G^2(5) = 2.14$ referred to $X^2(11-9;0.95) = 5.99$

$G^2(3)-G^2(4) = 2.71$ referred to $X^2(13-11;0.95) = 5.99$

$G^2(2)-G^2(3) = 11$ referred to $X^2(15-13;0.95) = 5.99$

$G^2(1)-G^2(2) = 6.15$ referred to $X^2(16-15;0.95) = 3.84$

* denotes "best" model

Tables 6.1 and 6.2 only one sequence of nested hierarchical models was considered. Although no other nested hierarchies were deliberated, it is noted that by employing a different sequence of nested hierarchies from the one chosen may produce an alternative "best" model.

After identifying the log linear models to explain the A and E and Haematoma data, 95% confidence intervals for the relative risk of an intracranial haematoma were produced (Table 6.3) using Method B in 5.4.4. The relative risk here, is defined as the risk that a patient with feature vector (i,j,k,l) will develop an intracranial haematoma as compared with a patient who is female, has Glasgow Coma Sum equal to 15, had not been involved in a road traffic accident and had no skull fracture.

Interpreting Table 6.3, it is noticed that the relative risks for females are similar to that of males in the no skull fracture group - with the exception of Non-RTA and Glasgow Coma Sum 15. Considering the skull fracture group, the relative risks for females are approximately double that of males. However, as these risks tend to be low, a clinician would not be interested in recording the sex of a patient.

In addition, the relative risks of intracranial haematoma for Non-RTA are similar to RTA in the no skull fracture group and approximately double in the skull fracture group. These two results indicate that there is little point in using either of the features Sex or Cause of Injury in any of the models to predict risk.

It would therefore seem reasonable to omit these features from the model and recalculate relative and absolute risks of intracranial haematoma employing the two variables; Glasgow Coma Sum and Skull Fracture.

Table 6.3 95% Confidence Intervals for Relative Risk of a Haematoma, Taking as a Reference Non-RTA, Glasgow Coma Sum 15, Female, No Fracture:

CAUSE	COMASUM	SEX	FRACTURE	
			NO	YES
RTA	15	Male	[0.70, 2.0]	[200, 640]
		Female	[0.88, 2.2]	[93 , 350]
	9-14	Male	[14 , 58]	[1400,5800]
		Female	[28 , 110]	[700,2900]
	3-8	Male	[56 , 290]	[790,3400]
		Female	[69 , 330]	[360,1900]
Non-RTA	15	Male	[0.74, 1.5]	[110, 330]
		Female	1	[43 , 140]
	9-14	Male	[26 , 76]	[860,3100]
		Female	[28 , 65]	[340,1300]
	3-8	Male	[240,1000]	[1900,7600]
		Female	[250, 890]	[730,3100]

Although interest was initially focussed on extending the number of risk factors to that used by Mendelow et al., it is noted that in this study a different pair of risk factors are suggested to identify those patients at high risk of developing an intracranial haematoma. The variable SKULL FRACTURE is common to both studies.

Based on these two variables, the A and E and Haematoma data sets were reexamined with the risks and 95% confidence intervals for the risks being extended to children. 4767 adults with complete data out of 4792 and 3599 children with complete data out of 3614 were available from the A and E data. The Haematoma data produced 861 adults with complete data out of a total sample of 988 and 99 children with complete data out of 119.

These patients were allocated to one of the six categories - produced from the two variables Glasgow Coma Sum and Skull Fracture - and their Absolute risk of developing an intracranial haematoma within each category was calculated. Tables 6.4 and 6.5 indicate these features for Adults and Children respectively.

Using Method A in 5.3.2 and the method described in 5.2.3, 95% confidence intervals for the relative and absolute risks of an intracranial haematoma were then produced for both adults and children (Tables 6.6 and 6.7 respectively). (The relative risk in these tables is defined as the risk that a patient with feature vector (x,y) will develop an intracranial haematoma as compared with a patient who has a Glasgow Coma Sum of 15 and no skull fracture present. For children, the overall risk required to induce the confidence intervals for absolute risk was calculated to be 2100).

In these tables, the intervals for the absolute risks are probably too narrow due to the uncertainty of the overall risk,

Various Features for Adults and Children

Table 6.4

ADULTS

		Haematoma sample	A and E sample	A and E scaled up	Absolute risk 1:
No Fracture					
GCS	15	35	4378	275318	7900
	9-14	90	258	16225	180
	3-8	72	31	1949	27
Fracture					
GCS	15	86	61	3836	45
	9-14	248	20	1258	5
	3-8	330	19	1195	4
Total		861	4767	299781	348

Table 6.5

CHILDREN

		Haematoma sample	A and E sample	A and E scaled up	Absolute risk 1:
No Fracture					
GCS	15	16	3409	200943	13000
	9-14	12	118	6956	580
	3-8	10	11	648	65
Fracture					
GCS	15	18	48	2829	160
	9-14	19	8	472	25
	3-8	24	5	295	12
Total		99	3599	212143	2100

95% Confidence Intervals for the Relative and Absolute Risks of Intracranial Haematoma for Adults and Children

Table 6.6

ADULTS

		Relative Risk	Absolute Risk 1:
No Fracture			
GCS	15	1	[5800,11000]
	9-14	[29 , 66]	[140, 230]
	3-8	[170, 500]	[19 , 44]
Fracture			
GCS	15	[110, 280]	[33 , 63]
	9-14	[870,2800]	[3.0, 7.4]
	3-8	[1200,3900]	[2.2, 5.6]

Table 6.7

CHILDREN

		Relative Risk	Absolute Risk 1:
No Fracture			
GCS	15	1	[8200,20000]
	9-14	[10 , 47]	[350, 1100]
	3-8	[71 , 530]	[25 , 160]
Fracture			
GCS	15	[38 , 170]	[92 , 260]
	9-14	[190,1300]	[9.2, 47]
	3-8	[340,3100]	[2.9, 20]

thus giving the impression that the estimates of risk are more precise than they actually are. Nevertheless, from Tables 6.6 and 6.7, there is clearly a consistent rank order of risks of a patient developing a haematoma with the presence of a skull fracture or Glasgow Coma Sum of 3-8 or both. In both adults and children, the presence of a skull fracture and having a Glasgow Coma Sum of 3-8 indicates the highest risk of developing an intracranial haematoma (Diagram 6.1). When both these features are present, the risk of developing an intracranial haematoma is several orders of magnitude greater than if one/neither was present.

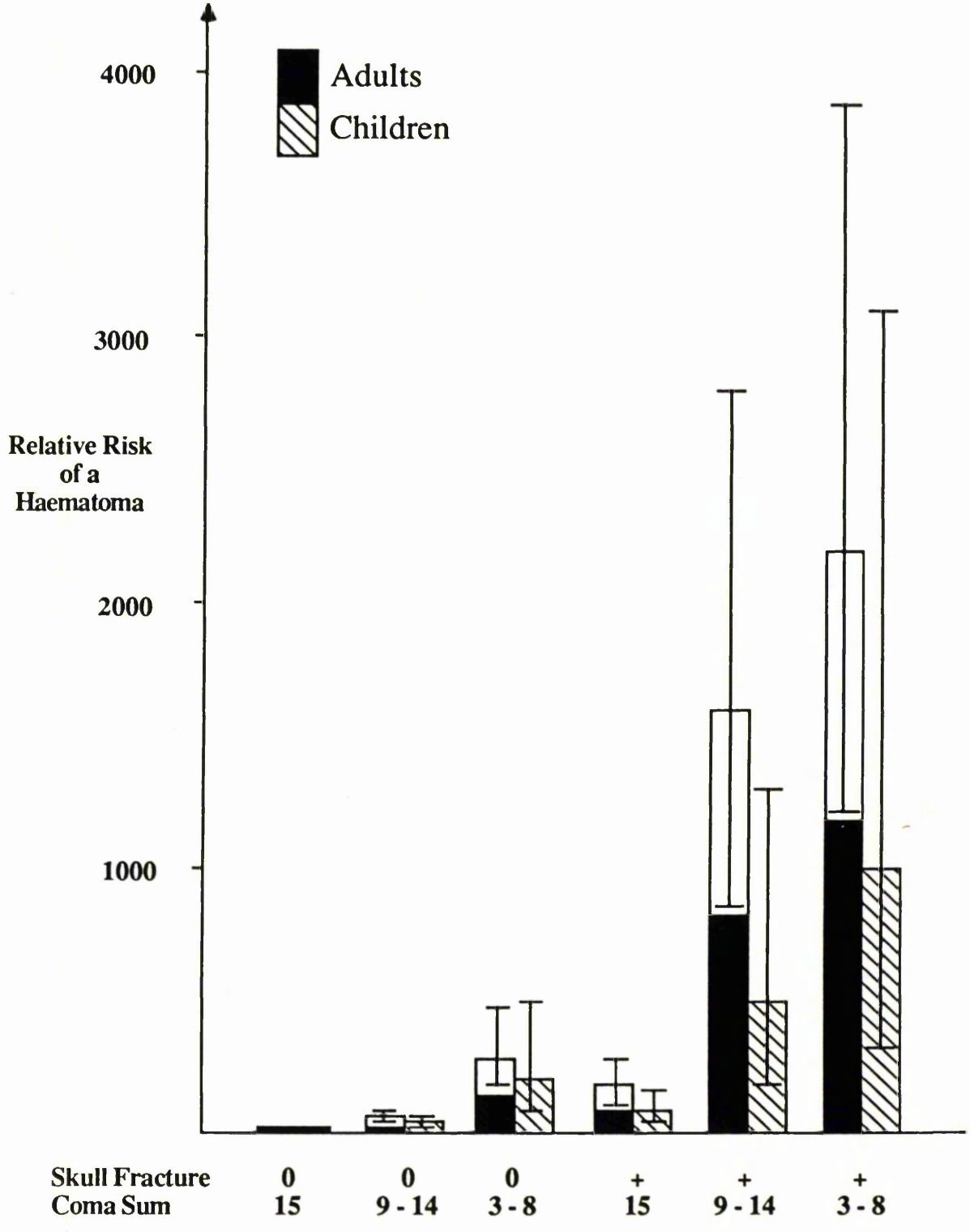
Finally, it was suggested by the clinicians involved in this study to include the variable "disruption of consciousness" with the two variables Glasgow Coma Sum and Skull Fracture in the construction of 95% confidence intervals for the relative and absolute risks of developing an intracranial haematoma. This variable is defined as a patient who has any post traumatic amnesia or any history of unconsciousness.

Using these three variables in further analysis involves 4767 adults with complete cases out of 4972 from the A and E sample and 357 adults with complete cases out of 399, from the pre-1978 Haematoma Study together with the 1984 cases from the Haematoma Study, to produce the Haematoma data. Calculations were restricted to Adults as only data from 45 children were available from the aforementioned sources. (This number is too small to obtain any meaningful estimates of risks).

An initial look at the A and E and Haematoma data for the three variables indicated that the "disruption of consciousness" variable should only be recorded for those patients with Glasgow Coma Sum 15. The number of patients in the four cells determined

Diagram 6.1

Relative Risks of Haematoma for Adults and Children with Approximate 95% Confidence Intervals



by Glasgow Coma Sum 9-14 and 3-8 with the "disruption of consciousness" categories were too small. The information about the Haematoma, A and E and A and E "scaled up" data are summarised in Tables 6.8(i)-(iii) respectively. The relative risk in this analysis is where each feature vector is compared to a patient who has Glasgow Coma Sum 9-14 and has no skull fracture. 95% confidence intervals for the relative risk of a patient developing an intracranial haematoma are shown in Table 6.9. The absolute risks calculated from Tables 6.8(i) and (iii) are shown in Table 6.10 with the corresponding 95% confidence intervals.

It is noticed (Table 6.10) that the inclusion of the variable "disruption of consciousness" provides the clinician with more information as to the risk of a patient developing a haematoma in the Glasgow Coma Sum 15 category. For example, taking a patient with no skull fracture present, the risk of developing an intracranial haematoma with no signs of disruption is 1 in 31000 whereas if the patient does have signs of disruption, the risk is vastly increased to 1 in 6700 - although both these risks are very low. Similarly for the skull fracture group i.e. if a patient has no disruption, the risk of developing a haematoma is 1 in 81 whereas if the patient shows signs of disruption, the risk is increased to 1 in 29. The category of no skull fracture and Glasgow Coma Sum of 9-14 produces a slightly lower risk of haematoma in Table 6.10 to that of Table 6.4 and similarly for no skull fracture and Glasgow Coma Sum of 3-8. However, when comparing the skull fracture category along these two Glasgow Coma Sum categories to Table 6.4, it is noticed that although the risk in the 9-14 category remains the same, the risk of developing an intracranial haematoma rises from 1 in 4 in Table

Table 6.8(i)

Haematoma Sample

	COMASUM 15		COMASUM 9-14	COMASUM 3-8
	No disruption	Disruption		
No Fracture	3	3	30	26
Fracture	16	10	111	158

Table 6.8(ii)

A and E Sample

	COMASUM 15		COMASUM 9-14	COMASUM 3-8
	No Disruption	Disruption		
No Fracture	3611	767	258	31
Fracture	50	11	20	19

Table 6.8(iii)

A and E Scaled Up

	COMASUM 15		COMASUM 9-14	COMASUM 3-8
	No Disruption	Disruption		
No Fracture	94109	19989	6724	808
Fracture	1303	287	521	495

Table 6.9 95% Confidence Intervals for Relative Risks

	COMASUM 15		COMASUM 9-14	COMASUM 3-8
	No Disruption	Disruption		
No Fracture	[0.002,0.024]	[0.010,0.11]	1	[3.7,14]
Fracture	[1.4,5.5]	[3.0,20]	[26,89]	[38,130]

Table 6.10 Absolute Risks and 95% Confidence Intervals for the Risks

	COMASUM 15		COMASUM 9-14	COMASUM 3-8
	No Disruption	Disruption		
No Fracture	31000 [13000,140000]	6700 [2700,29000]	220 [170,350]	31 [16,52]
Fracture	81 [46,140]	29 [11,59]	5 [2.7,6.9]	3 [1.9,4.9]

6.4 to 1 in 3 in Table 6.10 for the 3-8 category.

Thus the "disruption of consciousness" variable has considerable use in the detection of a haematoma and had this variable been omitted, necessary and useful information regarding the management of head injured patients may have been lost.

CHAPTER 7

DISCUSSION

Radiologists and Neurosurgeons debate the correct management strategy for head injured patients, particularly the necessity to perform radiography on all patients. To identify the individuals with a high risk of a skull fracture, two discrimination models were employed, namely linear logistic regression and Classification and Regression Trees (CART). Both of these procedures perform reasonably well using the identical subset of four indicator variables {COMASUM, VOM, FAC, SCALP}.

Initially using test and training data sets, the best subset was identified using true error rates (obtained by testing the model on a large number of new cases) and receiver operating curves. Having obtained the "best" subset, the resulting linear logistic regression equation was compared to the corresponding linear discrimination model employing the same subset. The results from this comparison indicated that both models would perform equally well.

Using the same test and training data sets, the classification and regression tree procedure was executed by the CART package. Again the same four indicator variables were selected to construct the best tree. A direct comparison between the linear logistic regression equation and the tree obtained from CART was made using the Brier Score. Results of this test indicated a slight improvement in the performance of the discrimination using the classification method.

Conveniently, all the methods discussed above essentially discriminate between no skull fracture and skull fracture equally well. It is important to distinguish the method which is easily

communicated to clinicians and whose application would lead to few difficulties. Clearly, to apply linear logistic regression and the linear discriminant would require access to either a programmable calculator or computer. Also, difficulties arise if any components of the measurement vector are missing. Of the three discrimination procedures mentioned, only the classification tree analysis can deal with missing data and moreover it is easily communicated to clinicians.

In conclusion, the procedure based on classification trees using the 4 variable subset {COMASUM, VOM, FAC, SCALP} would classify future patients most easily. In practice, clinicians should have no inhibitions from employing this method, and no interpretation of probabilities is required.

On the more theoretical side, the linear logistic regression model used in this study would almost certainly have more discriminative power if interaction terms were included. However, to achieve a simple system which, in practice, has to be used and understood by clinicians, interaction terms in this method were omitted. CART, on the other hand, will include some interactions in the optimal tree without complicating the discrimination.

Due to the large number of no skull fracture patients in the data set, both discrimination procedures could be used to identify those individuals at a low risk of having a skull fracture. Using CART, nodes at which there is a very low probability of misclassifying a patient with no skull fracture are easily identifiable. Therefore, eliminating those patients at a low risk of having a skull fracture for radiography may be the best course to take and may be acceptable to both radiologists and neurosurgeons.

In calculating the risk of developing an intracranial haematoma, the data from several sources were combined, perhaps causing opportunities for inaccuracies. Nevertheless, the data from each study were recorded on specifically designed forms thus enhancing the validity of the results produced.

The actual levels of risk calculated assumed that the data included all the surgically significant haematomas to have occurred in the West of Scotland over the eleven years under study. This study provides a basis for determining the management of head injured patients which will ensure that the maximum available resources are allocated to minimising avoidable mortality and morbidity. This can be achieved by reducing the total number of head injured patients admitted to hospital, but at the same time providing adequate facilities for the urgent scanning of patients at highest risk.

From the analysis, it is determined that the adult head injured patients with no skull fracture and Glasgow Coma Sum of 15 at the time of examination have an extremely low risk of a haematoma, even if there are signs of disruption. If such adults were sent home from A and E departments, there would be major savings with minimal risk.

Patients who have no skull fracture and Glasgow Coma Sum of 9-14 or 3-8, or have a skull fracture present with Glasgow Coma Sum of 15, have intermmediate levels of risk. Clearly these patients should be admitted to hospital for observation.

Patients with a skull fracture and Glasgow Coma Sum of 9-14 or 3-8 have a very high risk of developing an intracranial clot. After any necessary initial resuscitation, all such patients should be referred without further delay for CT scanning. This should not unduly overload existing facilities, as it is

estimated that there are only nine such patients per 100000 population a year (Mendelow et al.).

In this thesis, the risks of an intracranial haematoma were also evaluated for children, employing only the two variables skull fracture and Glasgow Coma Sum. It was determined that children with no skull fracture and Glasgow Coma Sum of 15 have a very low risk of developing an intracranial haematoma. Children with a skull fracture and Glasgow Coma Sum 3-8 have a relatively high risk of developing an intracranial haematoma and should be referred immediately for CT scanning.

The existing guidelines for admission or transfer to a Neurosurgical Unit could be altered accordingly. It appears that no other easily elicited clinical features investigated in this study could replace the presence or absence of a skull fracture for determining the presence or absence of an intracranial haematoma.

APPENDIX I

Guidelines for Skull X-ray after Head Injury

CRITERIA FOR NEUROSURGICAL CONSULTATION ABOUT PATIENTS WITH RECENT HEAD INJURY

Neurosurgical Department

Institute of Neurological Sciences, Glasgow

1. Fractured skull

   with confusion or worse impairment of consciousness. or

   with focal neurological signs, or

   with fits, or

   with any other neurological symptoms or signs.

2. Coma continuing after resuscitation - even if no skull fracture.
3. Deterioration in level of consciousness or other neurological signs.
4. Confusion or other neurological disturbances persisting for more than 6-8 hours, even if there is no skull fracture.
5. Compound depressed fracture of the vault of the skull.
6. Suspected fracture of base of skull (CSF rhinorrhoea or otorrhoea, bilateral orbital haematoma, mastoid haematoma) or other penetrating injury (gunshot etc.).

Patients in categories 1-3 should be referred urgently.

Note: The diagnosis and initial treatment of serious extracranial injuries should always take priority over transfer to the neurosurgical unit.

Treatment of Head Injured Patients in Coma

or with Possible Multiple Injuries

1. Assess for respiratory difficulty, for shock, and for internal injuries especially after a high velocity injury, e.g. a road traffic accident.
2. Perform: a) chest x-ray; b) blood gas estimation; c) cervical spine x-ray; d) other investigations as relevant.
3. Appropriate treatment may include:

   Intubate (e.g. if airway obstructed or threatened)

   Ventilate (e.g. cyanosis, PaO_2<60mmHg, $PaCO_2$>45mmHg

   Mannitol, only after consultation with neurosurgeon

   Application of cervical collar or cervical traction

   Immobilisation of fractures, treatment of internal injuries.
4. If accepted for transfer the patient should be accompanied by medical or nursing staff who are able to insert or to re-position endotracheal tube, and to initiate or to maintain ventilation.

GUIDELINES FOR THE MANAGEMENT OF PATIENTS WITH RECENT HEAD INJURY

Criteria for Skull X-ray after recent Head Injury

Clinical judgement is necessary but the following criteria are helpful: (any of the following)

1. Loss of consciousness or amnesia at any time.
2. Neurological symptoms or signs.
3. Cerebrospinal fluid or blood from the nose or ear.
4. Suspected penetrating injury.
5. Scalp bruising or swelling.
6. Difficulty in assessing the patient (i.e. Alcohol intoxication, epilepsy, children)

Criteria for Admission of Adults to Hospital

1. Confusion or any other depression of the level of consciousness at the time of examination.
2. Skull fracture.
3. Neurological symptoms or signs.
4. Difficulty in assessing the patient e.g. alcohol, epilepsy.
5. Other medical conditions - e.g. haemophilia.
6. The patient's social conditions or lack of a responsible adult/relative.

Post-traumatic amnesia or unconsciousness with full recovery is not necessarily an indication for admission.

Patients sent home should receive advice to return immediately if there is any deterioration.

Adapted from Harrogate Seminar Report 8 "The Management of Acute Head Injury" DHSS 1983 and "Guidelines for the Initial Management after Head Injury in Adults" British Medical Journal 1984 288 p983-985

APPENDIX II

Misclassification Rates

Every patient in the population is assumed to belong to one of J mutually exclusive classes - denote the set of classes to be $C = \{1,.....,J\}$. Partition the measurement space X into J distinct subsets $A_1,....,A_j$ such that for every $\underline{x} \in A_j$, j is the predictive class.

The misclassification cost of classification rule d, $R^*(\underline{d})$. can be estimated using the CART package in three ways, namely the resubstitution estimate, the test sample estimate and the cross-validation estimate. Denote these misclassification cost estimates for tree T. by R(T), $R^{ts}(T)$, and $R^{cv}(T)$ respectively.

Before describing misclassification costs further, it is useful to introduce further notation:

Let

(1) $\tilde{T}$ denote the set of terminal nodes of tree T.

(2) C(i|j) is defined on the cost of misclassifying a class j patient as a class i patient.

(3) p(j|t) denote the probability of a patient falling into class j given that it falls into node t.

(4) p(t) denote the probability that a case falls into node t.

(5) $Q^*(i|j)$ be the probability that a patient in j is classified into i by d.

(6) $R^*(j) = \sum_i C(i|j)Q^*(i|j)$

i.e. the expected cost of misclassification for class j patients.

1. Resubstitution Estimate

The resubstitution estimate calculated with the same data

used to construct T, is defined to be

$$R(T) = \sum_{t \in \tilde{T}} r(t)p(t)$$

where $r(t) = \min_i \sum_j C(i|j)p(t)$

2. The Test Sample Estimate

In this method, all patients are divided randomly into two sets $\mathcal{L}_1$ and $\mathcal{L}_2$ with sample sizes N_1 and N_2 respectively. The cases in $\mathcal{L}_1$ are used to construct the tree T and the cases in $\mathcal{L}_2$ are used to estimate $R^*(T)$. Using $\mathcal{L}_2$, the test sample estimate is defined as :

$$R^{ts}(T) = \frac{1}{N_2} \sum_{i,j} C(i|j)\, N_{ij}$$

where N_{ij} is the number of cases in class j whose predicted outcome is class i.

3. The V-fold Cross Validation Estimate

Define the complexity of a tree T as the number of terminal nodes, denoted by $|\tilde{T}|$. The cost - complexity measure $R_\alpha(T)$ is then defined as:

$$R_\alpha(T) = R(T) + \alpha|\tilde{T}|$$

where α (≥ 0) is called the complexity parameter and R(T) is estimated using the learning sample.

To estimate $R^{CV}(T(\alpha))$, all cases, $\mathcal{L}$, are randomly divided into V subsets, denoted by $\mathcal{L}_1, \ldots, \mathcal{L}_V$ of approximately equal sizes. For every ν, $\nu = 1, \ldots, V$, the procedure is applied to obtain the largest tree, $T_{max}^{(\nu)}$, using the learning sample $\mathcal{L}-\mathcal{L}_\nu$, and the corresponding minimal cost - complexity subtree of T_{max}, $T_{max}^{(\nu)}$. For complexity parameter α, the cross-validation estimate $R^{CV}(T(\alpha))$ is then defined as

$$R^{CV}(T(\alpha)) = \frac{1}{N} \sum_{i,j} C(i|j)N_{ij}$$

where

$$N_{ij} = \sum_{\nu} N_{ij}^{(\nu)}$$

and $N_{ij}^{(\nu)}$ is the number of class j cases in L_ν classified as i by $T^{(\nu)}(\alpha)$.

REFERENCES

1. Bailey, B.J.R.: 1987. "Confidence Limits to the Risk Ratio". Biometrics, Vol. 43 pp201-205.
2. Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W.: 1975. "Discrete Multivariate Analysis". The MIT Press, London.
3. Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J.: 1984. "Classification and Regression Trees". Wadsworth International Group, California.
4. Brier, G.W.: 1950. "Verification of Forecasts Expressed in Terms of Probability". Monthly Weather Review, Vol. 78 pp1-3.
5. Brocklehurst, G., Gooding, M. and James, G.: 1987. "Comprehensive care of patients with head injuries". British Medical Journal, Vol. 294 pp345-347.
6. Fienberg, S.E.: 1977. "The Analysis of Cross-Classified Categorical Data". The MIT Press, London.
7. Fisher, R.A.: 1936. "The use of multiple measurements in taxonomic problems". Annals of Eugenics, Vol. 7 pp111-117.
8. Hanley, J.A. and M^{c}Neill, B.J.: 1982. "The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve". Radiology, Vol. 143 pp29-36.
9. Jennett, B. and M^{ac}Millan, R.: 1981. "Epidemiology of head injury". British Medical Journal, Vol. 282 pp101-104.
10. Jennett, B. and Miller, J.D.: 1972. "Infection after depressed fracture of the skull". Journal of Neurosurgery, Vol. 36 pp333-339.

11. Jennett, B., Murray, A., MacMillan, R., MacFarlane, J., Bentley, C., Strang, I. and Hawthorne, V.: 1977. "Head Injuries in Scottish Hospitals". Lancet ii pp696-698.

12. Katz, D., Baptista, J., Azen, S.P. and Pike, M.C.: 1978. "Obtaining Confidence Intervals for the Risk Ratio in Cohort Studies". Biometrics, Vol. 34 pp469-474.

13. Masters, S.J., McClean, P.M., Arcarese, J.S., Brown,R.F., Campbell, J.A., Freed, H.A., Hess, G.H., Hoff, J.T., Kobrine, A., Koziol, D.F., Marasco, J.A., Merten, D.F., Metcalf, H., Morrison, J.L., Rachlin, J.A., Shaver, J.W. and Thornbury, J.R.: 1987. "Skull X-ray Examinations after Head Trauma". The New England Journal of Medicine, Vol. 316 pp84-91.

14. Mendelow, A.D., Teasdale, G., Jennett, B., Bryden, J., Hessett, C. and Murray, G.: 1983. "Risks of intracranial haematoma in head injured patients". British Medical Journal, Vol. 287 pp1173-1176.

15. Metz, C.E.: 1978. "Basic Principles of ROC Analysis". Seminars in Nuclear Medicine, Vol. VIII pp283-298.

16. Strang, I., MacMillan, R. and Jennett, B.: 1978. "Head injuries in accident and emergency departments at Scottish hospitals". Injury, Vol. 10 pp154-159.

17. Teasdale, G. and Jennett, B.: 1974. "Assessment of coma and impaired consciousness. A practical scale". Lancet ii pp81-84.

18. Teasdale, G., Murray, G., Parker, L. and Jennett, B.:1979a. "Adding up the Glasgow Coma Score". Acta Neurochirurgica Supplement, Vol. 28 pp13-16.

19. Titterington, D.M., Murray, G.D., Murray, L.S., Spiegelhalter, D.J., Skene, A.M., Habbema, J.D.F. and Gelpke, G.J.: 1981. "Comparison of discrimination techniques applied to a complex data set of head injured patients". Journal of the Royal Statistical Society Series 'A', Vol. 144 pp145-175.

20. Youden, W.D.: 1950. "Index for Rating Diagnostic Tests". Cancer, Vol. 3 pp32-35.