MODELLING THE RECOVERY PROCESS AFTER SEVERE HEAD INJURY

by

LILIAN S. MURRAY

A Dissertation Submitted to the

UNIVERSITY OF GLASGOW

for the degree of

DOCTOR OF PHILOSOPHY

Department of Neurosurgery

Department of Statistics

November 1988.

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

ACKNOWLEDGEMENTS

I first wish to thank my supervisors, Professor G M Teasdale and Professor D M Titterington, for their support throughout the period of my research.

I am deeply grateful to both Professor G M Teasdale and Professor W B Jennett for allowing me to carry out this research with the Head Injury Study data bank. The help, encouragement, advice and many hours of their time which they have given me since I joined the Department of Neurosurgery have been invaluable. Their expertise and enthusiasm is an inspiration to all who work with them.

I am also indebted to Mrs Myra Smith for help with the preparation of the text.

My thanks are also due to my many friends, particularly in the Departments of Statistics and Neurosurgery, who have given me continual support and encouragement.

My deepest thanks, however, must go to my husband and three children for the sacrifices they have made. Gordon, without your support this work would not have been possible.

The Head Injury Study was funded by the Medical Research Council while this work was carried out.

DECLARATION

The results of Sections 3.4 and 3.5 have been published previously. Section 3.4 gives an expansion of the results in Jennett et al. (1976) and in particular Tables 3.5 and 3.6 are reproduced. Section 3.5 summarises the work of Titterington et al. (1981) and Tables 3.11 - 3.18 are reproduced. Figure 7.1 is reproduced from Gray's Anatomy (1958) and Figure 7.2 from Jennett and Teasdale (1981).

TABLE OF CONTENTS

CHAPTER 2 HISTORICAL DEVELOPMENT OF THE HEAD INJURY STUDY DATA BANK

CHAPTER 3 THE USE OF DISCRIMINANT ANALYSIS TO PREDICT THE SIX MONTH OUTCOME OF PATIENTS IN THE HEAD INJURY STUDY

CHAPTER 4 MODELS USING TIME TRENDS

CHAPTER 5 ASSESSMENT OF PERFORMANCE

CHAPTER 6 THE INTRODUCTION OF AGE INTO THE MODELS

CHAPTER 7 PRACTICAL IMPLICATIONS AND FURTHER WORK

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CT	Computerised tomography
D/V	Death or vegetative state
E	Eye opening component of the Glasgow Coma Scale
GCS	Glasgow Coma Scale
GR	Good recovery
M	Motor response component of the Glasgow Coma Scale
MD	Moderate disability
ML	Maximum likelihood
MRI	Magnetic resonance imaging
MRP	Motor response pattern
M/G	Moderate disability or good recovery
OCS	Oculocephalic response
OVS	Oculovestibular response
PET	Positron emission tomography
PML	Pseudo maximum likelihood
SD	Severe Disability
SEM	Spontaneous eye movements
V	Verbal response component of the Glasgow Coma Scale
24H	The first twenty-four hours after the onset of coma
2-3D	Two to three days after the onset of coma
4-7D	Four to seven days after the onset of coma
8-14D	Eight to fourteen days after the onset of coma
15-28D	Fifteen to twenty-eight days after the onset of coma

SUMMARY

The problem considered in this thesis is the prediction of the quality of survival after severe head injury. A model of the recovery trend of the patient through time is derived and this model is used to predict ultimate outcome.

Chapter 1 introduces the problem of prognosis in clinical decision making, and in particular, its importance in the context of severe head injuries. It identifies the need for a new statistical approach to this problem.

Chapter 2 describes the development of the Head Injury Study data bank from the initial stages when terminology needed to be carefully defined to the present day. It gives a detailed description of the Glasgow Coma and Outcome Scales. The data collection methods are described along with the problems encountered in establishing a reliable data bank. Suggestions are given to minimise these problems.

In Chapter 3 discriminant analysis is introduced and its terminology defined. The factors involved in variable selection, the problem of missing data and the assessment of the performance of a discriminant rule are discussed in general terms. Two major studies are described where the prediction of outcome after severe head injury is made using information from the Head Injury Study data bank: first the early work using an independence model, and then a comparative study which was carried out to assess the relative merits of different discrimination techniques. Chapter 3 finishes by illustrating that, while these methods are successful in the prediction of death or survival, a new approach is required to predict the quality of survival.

Chapter 4 contains the work involved in modelling the recovery

trend of the survivors. This is done by modelling the coma score through time. The first order autoregressive model which was initially adopted is described along with the modifications required to give an adequate decription of the data. Ways of reducing the number of parameters which need to be estimated are considered, as well as the effect of using a pseudo maximum likelihood approach to reduce the computation involved in obtaining the parameter estimates. Three methods which adequately model the recovery trend are obtained.

Chapter 5 examines the performance of these methods by assessing their ability to predict the quality of survival. This assessment is based on the classification matrices and three separation measures (the error rate, average logarithmic score and average quadratic scores). How performance is affected by different priors and the 'jack-knife' technique is examined. The performance of the models incorporating trend is compared with that of other available models. Age is shown to have a substantial effect on the prediction of prognosis.

In Chapter 6, age is incorporated into the models considered in Chapter 5 and the performance is re-assessed.

Chapter 7 discusses the possible clinical reasons for the general lack of success of the methods considered in Chapter 5 and Chapter 6. The use of the verbal component of the coma scale is considered, and alternative data which may be useful to predict the quality of survival are discussed. Recommendations are made for future work, the importance of the quality of the information collected is stressed, and the vital role which simple statistical techniques have to play is emphasised.

CHAPTER 1

INTRODUCTION

1.1 The Importance of Prognosis in Clinical Decision Making

The ability to predict the course or consequence of disease is fundamental to most clinical decisions. Before an investigation is ordered, a drug prescribed or even an operation advised the clinician needs to know the likely benefit, the associated risk and the result of withholding any such measures. These decisions demand an estimate of prognosis, however crude. The extent of the problem is illustrated by Wagner et al. (1978), who, in a review of the problems in diagnosis, cite 827 references.

In a formal approach to such difficult decisions the problem can be split into four steps

(i) the strict definition of relevant terminology

(ii) careful record keeping

(iii) the identification of factors affecting prognosis

(iv) the construction of a model to estimate prognosis.

In adopting such a strategy, the clinician uses his accumulated experience and judgement to go through these steps and make an estimate of prognosis.

1.2 The Importance of Prognosis in Severe Head Injury

Predictive thinking is particularly important in the management head injured patients. In those with a minor head injury the doctor needs to decide if admission for observation or transfer to a specialist unit is justified, and this depends on the likelihood of certain complications developing. In patients with a severe

head injury who remain in coma after effective treatment has been given, prognosis about the ultimate outcome is of particular concern. Many of those patients will die or be permanently disabled no matter what treatment is given. Resources are always limited and the deployment of facilities to one patient limits their availability to others, whether this is in the Intensive Therapy Unit in the early stages after injury or later, in the rehabilitation of survivors. Even if this were not the case, needlessly prolonged intensive care can be demoralising for the patient and relatives, as well as being sometimes unnecessary or even hazardous. Similarly over optimism about recovery can lead to fruitless efforts at rehabilitation rather than realistic adjustments to cope with handicaps and minimise consequent limitations.

It is therefore important to identify as early as possible the patients who will benefit from the facilities of a specialist unit in order that their management can be appropriate and humane, rather than intuitive or defensive.

1.3 Prognosis of Severe Head Injury

The reasons that even experienced clinicians have difficulty in making firm predictions about outcome after severe head injury are not hard to discover. It would take the average consultant until he retired to look after 1000 patients with severe head injury. Even if he could remember the clinical details and ultimate outcome of all his patients, his capacity to analyse accurately how these inter-relate, and how they can be used in a new case, would be imperfect. In practice clinicians tend to remember the 'remarkable recovery' or the 'disappointing failure to respond', and their

estimations of prognosis are affected by their most recent memories.

In the late sixties the advances in computer technology and their increased availability made the storage and analysis of data on head injured patients a realistic possibility. The Head Injury Study was initiated in Glasgow in 1968 by Professor Bryan Jennett and was joined by centres in the Netherlands and the U.S.A., in 1972 and 1974 respectively.

As already mentioned, before clinical information can be stored on computer in a form suitable for analysis it is vital to develop accurate record keeping methods, and this in turn requires clear definitions of the features to be recorded. The work involved in setting up the data bank is described in detail in Chapter 2. The problems encountered were similar to those encountered later by de Dombal (1978) in his study of the computer diagnosis of acute abdominal pain and by Marshall et al. (1983) in setting up a National Traumatic Coma Data Bank in the U.S.A..

Once the data bank was established it was possible to identify features which affected prognosis. There are many reports identifying such features, both by those involved in the Head Injury Study (Avezaat et al., 1979; Braakman et al., 1980; Jennett et al., 1977b; Jennett et al., 1979; Teasdale et al., 1982b) and others (Becker et al., 1977; Marshall et al., 1979; Miller et al., 1977; Overgaard et al., 1973; Pagni, 1973; Pazzaglia et al., 1975).

The natural progression from identifying single features which affect prognosis is to use combinations of features in the hope that a more accurate prediction of prognosis can be made. Thus the aim was to use the data bank of stored information to construct a model of the recovery pattern after severe head injury which would allow a prediction of prognosis for a new case. There are many

different methods of model building, but the simple approach used initially in the Head Injury Study gave promising results (Jennett et al., 1976). Stablein et al. (1980) used a logistic regression approach and this was the basis of a criticism of the Head Injury Study methods by Becker (1979). This criticism was a factor in instigating a comparative study (Titterington et al., 1981) of the different methods available, and so the various approaches were applied to the cases in the Head Injury Study. Chapter 3 describes the evolution of the study from the identification of single features affecting prognosis through to the comparative study.

1.4 The Need for a New Approach in Modelling Prognosis

While the models studied had been largely successful in predicting which patients would make a good recovery and which would die, no method was successful in identifying which patients would remain disabled after their head injury. The identification of this group of patients is important for several reasons. First, they impose a burden, both financial and social, on the community in which they live. Secondly, it is this group who are most likely to benefit from a new or intensified treatment since most other patients will clearly die regardless of treatment or will recover with conventional intensive care.

I was disappointed by the fact that no method used in the comparative study had been successful in identifying this group of patients and therefore it seemed to be worthwhile to try to find a new approach to this problem. It is this exercise which forms the main part of this thesis.

During the course of the study, which I joined in 1975, I dealt with data from many head injured patients, and after some time I

was struck by the fact that many cases who remained disabled had shown little change over time (either improvement or deterioration) in the early stages after injury. By contrast, those with good or very poor outcomes seemed to separate out more quickly. The existing methods of prediction had used the state of the patient at a particular time point, so that an attempt to model the time trends in the data offered a new approach. In Chapter 4 the various possible models that can incorporate time trends are reviewed and a model derived which described the data adequately while using as few parameters as possible.

The logical sequel to the derivation of a model is to assess its discriminatory power. The performance of the derived model was assessed by comparison with that of other more standard models. This exercise is described in Chapter 5. The comparison of the results obtained from the derived model with those from a model incorporating the age of the patient showed that a worthwhile improvement in performance could be achieved by incorporating age into the derived model. Two different methods of incorporating age were considered and the performance of all models is re-assessed, with age included, in Chapter 6. Chapter 7 reviews the results of Chapters 5 and 6 and considers their further implications. These are relevant both to the statistical approach, and in particular, to the clinical data that seem likely to be relevant in future work.

CHAPTER 2

HISTORICAL DEVELOPMENT OF THE HEAD INJURY STUDY DATA BANK

2.1 Epidemiology of Head Injury

In Britain almost one million patients attend Accident and Emergency Departments each year with a head injury (Jennett et al., 1977a). Fortunately most are only mildly injured; only one in five of these patients is admitted to hospital and of these two thirds are discharged within 48 hours (McMillan et al., 1979; Strang et al., 1978).

In Scotland, about 15,000 head injuries each year are admitted to hospital. The majority of minor head injuries are wholly treated in primary surgical wards which are in general surgical, accident, orthopaedic and paediatric departments, while the more severe head injuries are treated in the four regional neurosurgical units in Glasgow, Aberdeen, Dundee and Edinburgh.

In Glasgow the regional unit is the Institute of Neurological Sciences, located within the Southern General Hospital, and it serves a population of 2.7 million in the West of Scotland. Patients are taken only as transfers from other hospital units, never directly from the scene of the accident; even in the Southern General Hospital all head injuries are dealt with first by primary surgeons. Jennett et al. (1977a) showed that Glasgow was similar to Dundee and Aberdeen in its head injury practice with 4-5% of admitted cases being transferred to the regional unit. However, their survey did not include hospitals in the Lothian Health Board. These are served by the Edinburgh Neurosurgical Unit which has a policy of admitting a large proportion of head injured patients, including the mildly injured, directly to a ward

supervised by neurosurgeons, although still not in the main neurosurgical department.

In 1977 guidelines (Teasdale et al., 1982a) were adopted and later formalised (Briggs et al., 1984) for the transfer of head injured patients to the Institute. The current guidelines for the management of head injured patients are shown in Figure 2.1. This change in transfer policy led to an increase in the number transferred per year, from just over 220 to around 500. Essential details of the features of the head injured patients admitted to the Institute of Neurological Sciences in 1986 are given in Table 2.1.

2.2 Definition of Terminology

2.2.1 Assessment of Conscious Level

Impairment of consciousness is an indication of dysfunction in the brain as a whole and is one of the most consistent features of head injury. In the acute stages, changes in conscious level provide the best indication of the development of complications such as an intracranial haematoma, while the depth and duration of coma indicate the degree of ultimate recovery which can be expected. Reliable assessment of the extent of impaired consciousness is therefore of prime, practical importance in the management of head injured patients.

The various levels of impaired consciousness have been described and recorded by an abundance of alternative terms. Expressions such as comatose, semi-comatose, stuporous and semi-conscious have often been used, and a range of inconsistent systems described (Frowein, 1976). This problem led Teasdale and Jennett (1974) to examine the existing systems and to develop the

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:

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.

Criteria for 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: In all cases the diagnosis and initial treatment of serious extracranial injuries takes priority over transfer to the neurosurgical unit.

Management of Head Injured Patients in Coma or with Possible Multiple Injuries

1. Assess for respiratory insufficiency, 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, $Pa0_2 < 60mmHg$, $PaCO_2 > 45mmHg$
   Commence IV infusion (1500ml/24h)
   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 personnel able to insert or to re-position endotracheal tube, and to initiate or to maintain ventilation.

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

Figure 2.1 Guidelines currently used for the management of head injured patients

Table 2.1 Details of the cases admitted to the Institute of Neurological Sciences in 1986

Number of cases	592
Under 20 years old	31%
Road traffic accident victim	38%
Admitted within 6 hours of injury	45%
In coma or intubated on admission	23%
In coma > 6 hours	19%
Skull fracture	58%
Operated haematoma	17%
Extracranial complications	57%
Non-reacting pupils an admission	7%

Glasgow Coma Scale. They found that existing systems suffered from one or more of three defects. Some depended on specific anatomical-clinical correlations, whereas studies of the brain after severe blunt injury (Mitchell and Adams, 1973) had shown that most cases had lesions widespread throughout the brain. Some described coma by a series of arbitrary steps, assuming groups of clinical features unique to each level, whereas Teasdale and Jennett observed that the reality is a continuous spectrum of responsiveness between deep coma and full consciousness. Finally, few scales had been tested for the consistency with which the signs and symptoms upon which they depended could be elicted by different observers. To find wide practical application, a system must be simple and based upon clearly defined criteria, which can be elicited reliably by a wide range of medical and nursing staff. The Glasgow Coma Scale took account of all these considerations and provided an effective method of describing the various states of impaired consciousness encountered in clinical practice.

2.2.2 The Glasgow Coma Scale (GCS) and Score

Three separate aspects of the patient's behaviour are evaluated independently of each other:-

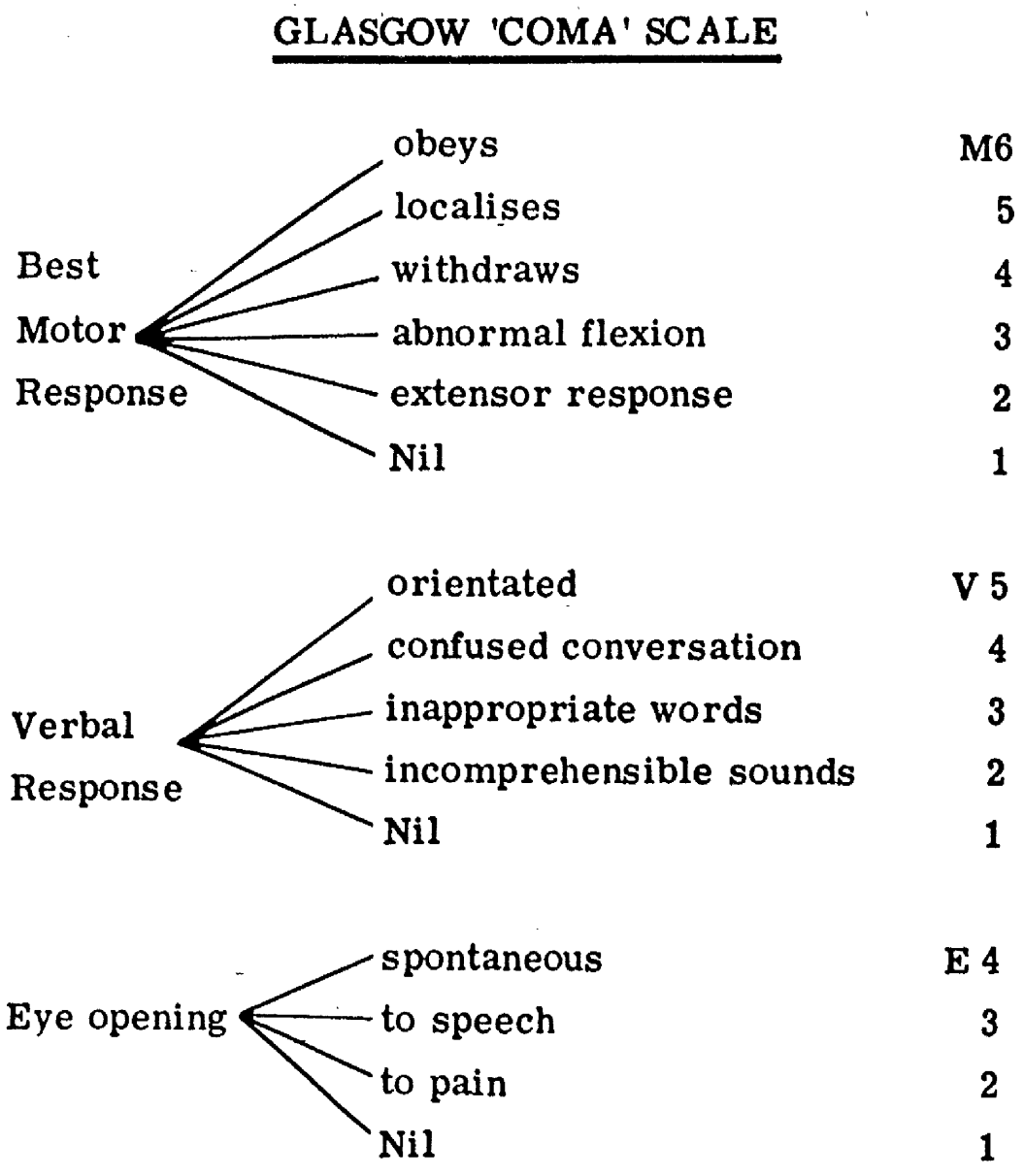
(i) the stimulus required to induce eye opening (E)

(ii) the verbal response (V)

(iii) the best motor response (M).

Each aspect of behaviour is assessed in terms of a well defined series of responses which indicate the degree of dysfunction. Each step in each component has to be allocated a notation, with a score of 1 indicating maximum dysfunction, and Figure 2.2 illustrates the scores possible for each aspect of behaviour. Summation of the scores of the three components yields the overall Glasgow Coma

Figure 2.2 The Glasgow Coma Scale



Score (Teasdale et al., 1979a).

When eliciting the eye opening score, **spontaneous** opening indicates that the arousal mechanisms in the brain stem are active. It does not necessarily imply awareness. If spontaneous opening is not present then a **spoken** command is given, usually the patient's name is called and he is requested to open his eyes. If this is unsuccessful then a **painful** stimulus is applied by exerting pressure on the finger-nail bed with a pen or pencil. **No eye opening** in response to a painful stimulus implies a marked degree of depression of the arousal system.

After the patient has been roused as fully as possible, verbal and motor performance are assessed.

With the verbal response, **orientation** requires the patient to know who he is, where he is, and the month and year. If he is unable to answer these questions but capable of producing phrases, sentences and even conversational exchange, then the patient is termed **confused.** **Inappropriate words** refer to intelligible articulation used in an exclamatory, random way while moaning and groaning constitute **incomprehensible sounds.** While the presence of speech indicates a high degree of integration in the nervous system, **no verbal response** may, of course, be the result of causes other than impaired consciousness, such as dysphasia.

When scoring the best motor response, to reflect the functional state of the brain as a whole, the best or highest response from any limb is recorded. **Obeying** commands is judged from the response to instructions such as 'lift your arms' or 'put out your tongue'. Reflex grasp responses occur in unconscious patients, and asking a patient to squeeze the examiner's fingers is not a reliable test. If the patient does not obey commands then a painful stimulus is applied. This is applied first at the finger-nail bed, but

subsequently it may be necessary to apply pressure over the supra-orbital notch. **Localising** is recorded if the patient moves a limb in such a way as to locate the painful stimulus on the head in an attempt to remove it. If the arm bends at the elbow but does not achieve a localising response then a **flexion** response is recorded. This can vary from **normal** rapid withdrawal to **abnormal** slow dystonic movements in which the limbs assume stereotyped postures. **Extension** responses of the limbs when the elbows or knees straighten are clearly abnormal. The limbs may even adopt this position without stimulation. **No response** to pain is scored when repeated and varied stimulation elicits no detectable movement or change in tone of the limbs. The GCS has been universally adopted as a bedside test, and the introduction of the scale has greatly enhanced the value of routine observations (Teasdale, 1975). A chart on which the responses are recorded provides a visual profile of the patient's progress which can be rapidly assessed (Teasdale et al., 1975). A typical chart of a patient who suddenly deteriorates but after operation gradually improves is shown in Figure 2.3.

With the help of the GCS, Teasdale and Jennett (1976) defined coma as the inability

(i) to open the eyes to any stimulus

(ii) to utter any recognisable words

and (iii) to obey commands.

In terms of the GCS, this implies an eye score (E) of 1, a verbal score (V) of 2 or less and a best motor score (M) of 5 or less. If any of (i)-(iii) above could be achieved then the patient was regarded as not in coma.

The Glasgow Coma Scale

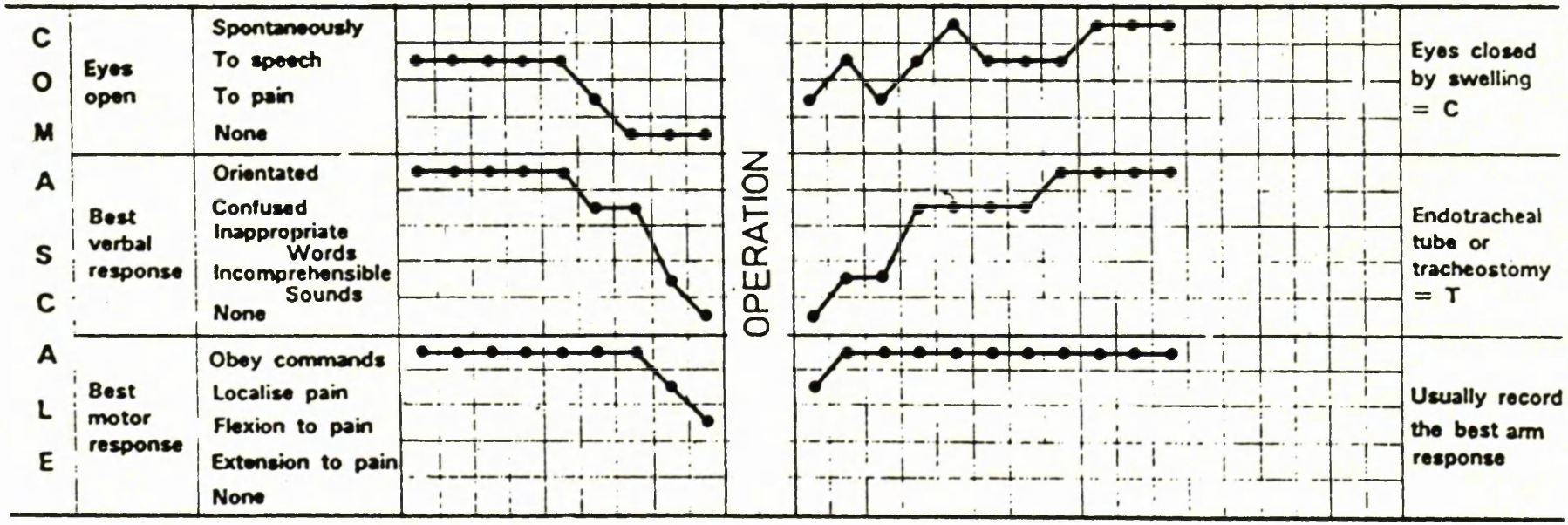


Figure 2.3 Typical coma scale chart of a patient who suddenly deteriorates but after operation gradually improves

2.2.3 The Glasgow Outcome Scale

Much of the difficulty which doctors experience when making decisions about head injured patients, both in the acute stage and during recovery, results from the uncertainty about the outcome. Barlow and Teasdale (1986) found that, in a multi-national group of 59 neurosurgeons, 56% chose 'estimated prognosis' as the most important factor in determining a difficult clinical decision. Jennett and Bond (1975) saw the definition of outcome as the first step towards making possible the prediction of outcome. They reviewed recent papers on outcome after head injury and found that a wide range of terms were used. As persisting disability after head injury usually comprises both mental and physical handicap they devised a simple scale, the Glasgow Outcome Scale, for describing overall social outcome. This scale has five categories :-

(i) death

(ii) vegetative state

(iii) severe disability

(iv) moderate disability

(v) good recovery.

Death might seem to require no further definition. However, advanced technology can now keep other major organs functioning when irreversible brain damage has occurred and strict criteria now exist to determine brain death. It is now agreed that in such cases the time of death is when brain death is confirmed and not some later time when the heart stops.

The **vegetative state** was defined by Jennett and Plum (1972) in rigorous terms which limited it to patients who showed no evidence of meaningful responsiveness. Patients who obey even simple commands or utter any words are assigned to a better category.

Vegetative patients breathe spontaneously, have periods of spontaneous eye opening, when they may follow moving objects with their eyes, show reflex responses in their limbs (to postural or painful stimuli), and they may swallow food placed in their mouths. This non-sentient state must be distinguished from other conditions of wakeful, reduced responsiveness – such as the locked-in syndrome, akinetic mutism and severe aphasia.

Severe disability indicates that a patient is conscious but dependent and needs the assistance of another individual every day for some activities of daily living. This may range from continuous total dependency to the need for assistance with only one activity such as dressing, getting out of bed, moving about the house or going outside to shop. Most often, dependency is due to a combination of physical and mental disability, but many patients who have little or no physical deficit are unable to organise their day to day lives effectively and must be classed as severely disabled. Some require the care and protection which only a mental hospital can provide: others cope at home with the support of attentive relatives but could not be left alone for a whole day because they would be unable to organise their meals, or to deal with callers, or any domestic crisis which might arise.

Moderate disability means that patients are independent but disabled. Such a patient is able to look after himself at home, to go out to shop and to travel by public transport. However, some previous activities, either at work or in their social life, are now no longer possible because of physical or mental deficit. Some patients in this category are able to return to certain kinds of work, even to their own job if this happens not to involve a high level of performance in the area of their major deficit.

Good recovery indicates the capacity to resume normal

occupational and social activities although there may be minor physical or mental deficits. The patient need not have resumed all his previous activities, and may not be working because unemployment may be due to many factors other than the degree of recovery.

The time after injury at which outcome is assessed is important. During the first year an increasing number of those initially vegetative or severely disabled die: on the other hand, some severely or moderately disabled reach a better outcome. Jennett and Bond (1975) state that a third of those still moderately disabled at 3 months after injury had made a good recovery by 12 months, and over 80% of those who improved their 3 month outcome by 12 months had already achieved the higher grade within 6 months of injury.

2.3 The Establishment of an International Data Bank

2.3.1 Introduction

Head injury is a common cause of death and disability, particularly in the young, and patients with a severe head injury put a considerable burden on acute hospital services in the early stages after injury. If they survive, the burden then falls on many aspects of the health services in the community and can last for many years (Jennett, 1975).

The value of a data bank of clinical cases collected in a standardised way as a basis for the management of new cases and for relating therapeutic efforts to outcome was pointed out by Fries (1976). The collection of such a data bank of patients with severe head injury was initiated at the Institute of Neurological Sciences in 1968. Extension of the data collection to two Dutch centres

(Rotterdam and Groningen) in 1972 and to an American centre (Los Angeles) in 1974 (Jennett et al., 1977b) made it possible to test the feasibility of standardising methods of clinical recording among several teams of clinicians. Another American centre (San Francisco) joined the study in 1980.

2.3.2 Definition of Criteria for Admission to the Data Bank

It was essential to establish at an early stage if a case was sufficiently severe to be admitted to the data bank. The most widely accepted indicator of brain damage was the extent and duration of impairment of conscious level. Because of the lack of generally agreed scales of assessment, it was during this time that the Glasgow Coma Scale and the Glasgow Outcome Scale were devised and developed. For inclusion in the data bank, the patient had to be in coma, as defined in Section 2.2.2, for at least six hours. Patients who were lucid after injury and then deteriorated so that they were in coma for six hours or more were included. Patients who died within six hours of injury were excluded.

The choice of the duration of coma as six hours was to some extent arbitrary. However, the period was chosen to allow time for the diagnosis and management of other injuries and their associated complications, such as shock and respiratory insufficiency. It is well known that these may affect several parameters of neurological function, in particular the pupil reaction and the level of responsiveness. The extent of brain damage may therefore be over-estimated on the basis of the patient's state in the first few hours after injury.

All cases who were admitted to each of the participating centres during their period of study and who were eligible were accepted into the data bank.

2.3.3. Data Collection

Data were recorded by one of a series of specified clinical trainees who had been made aware of the purposes of the study and the categorisation of the various features agreed between participating centres. The evolution of these naturally began in Glasgow, but when other centres joined the study considerable care was taken to ensure uniformity of eliciting, interpreting and recording clinical data.

Several different types of data were collected. Personal details such as the age and sex of the patient were recorded, as well as the history of the patient from injury until admission to neurosurgery. The investigations carried out, such as X-rays and intra-cranial pressure monitoring, and treatment given were also noted, as well as aspects of coma. Data from the bedside day sheets on the coma scale, pupil reaction, eye signs and several autonomic activities (respiration, heart rate, blood pressure and temperature) were collected. Some of these observations are so labile that at any one time they alone may be unreliable as a guide to the degree of brain damage, while others change less rapidly. All are dynamic, however, and an essential feature of the study was that data on the patient were noted repeatedly at the bedside. This gave rise to a massive amount of data, so it was decided to summarise them by the best and the worst state within a series of time periods after the onset of coma. The time periods chosen were:-

(i) the first twenty-four hours (24H)

(ii) two to three days (2-3D)

(iii) four to seven days (4-7D)

(iv) eight to fourteen days (8-14D)

(v) fifteen to twenty-eight days (15-28D).

It was also noted whether the best coma score came before or after the worst, by coding whether the patient was improving, not changing, deteriorating or fluctuating within the time period.

The data collected also included an assessment of the six month outcome using the Glasgow Outcome Scale and, once this was available, the data were transferred into the computer data bank. Great efforts were made to follow up every case – an extremely time-consuming occupation. However, a few cases had to be classed as out of hospital and lost to follow-up.

2.4 Problems Associated with Data Collection

2.4.1 Observer Variability

An important feature of any practical measurement scale is that it should give consistent results when used by different observers. Teasdale et al. (1978) performed a detailed study of the observer variability associated with the GCS and with some alternative terms often used to describe patients with acute brain injury. The observers used ranged from nurses to consultant neurosurgeons. Patients in a specially prepared film were observed and scored by groups in Britain, Europe and North America. Nurses and general surgeons were found to be as consistent as neurosurgeons when using the GCS and it was relatively resistant to language or cultural differences between observers. The practical reliability of the scale enhanced its value both in monitoring individual cases and for making meaningful comparisons between series of patients with acute brain injury. Indeed, such has been the success of the scale that within a few years it was used in more than half the neurosurgical units in Britain (Gentleman & Teasdale, 1981) and is now widely used in North America and throughout the world,

including countries such as Russia and Japan (Schein, 1988).

As with the GCS, there was good agreement between different observers when the Glasgow Outcome Scale was used (Maas et al., 1983). Indeed Langfitt (1978) suggested that both Glasgow scales should be adopted worldwide, at least for a period of five years, to facilitate the comparison of different studies.

2.4.2 Form Design

Even with strict definition of the terminology, substantial problems can still be encountered in data collection. The physical means by which this is achieved can vary. Nowadays it is relatively simple to collect data directly on to a microcomputer using a database program. However when the Head Injury Study was initiated such programs were not available and the data collected were transcribed onto a form by the clinicians involved.

Good design of such a form (or database program) plays an important role in careful record keeping. The larger the study the more important it becomes to have a well designed form, and effort at this stage can be rewarded when the subsequent data collection and analysis are made as simple as possible. Useful general guidelines have been produced by Gore and Altman (1982) and comments specific to head injury on this problem have been made by Miller and Teasdale (1985).

Numerous versions of the data collection form have been used throughout the duration of the study. In 1978 a great deal of effort, in which I had a leading part, was put into refining existing approaches to create a final well designed method of data collection. The form which is included in Appendix 1 was used to collect most of the data used in this thesis.

2.4.3 Data Checking

After the data were stored on computer they were checked before any analysis was carried out. At a simple level all values recorded were confirmed to be within their permitted range and important features which had not been coded were identified. More sophisticated checks were also incorporated to detect unusual or unlikely combinations. For example it is unlikely that a head injured patient whose best motor response is nil will be opening his eyes spontaneously.

2.5 Reduction of Dimensionality

2.5.1 Introduction

As many as 300 items of information can be collected on some head injured patients and not all of this information is relevant to prognosis. A reduction in the dimensionality of the data normally takes place at the modelling stage, but a rational clinical approach to this problem can also be incorporated at an earlier stage. The Glasgow Coma Score and created eye indicant are examples where the dimensionality of the data has been reduced after discussing the results of exploratory data analyses with the clinical staff involved in the study.

2.5.2 The Glasgow Coma Score

In spite of summarising aspects of coma by the best and worst in the period, the dimensionality of the data is still large. In particular, the coma scale is composed of three separate responses. These responses tend to be related to each other, particularly when responsiveness is severely depressed in the first few days after injury. For example, a patient whose best motor response is

extension will not be speaking and is unlikely to be opening his eyes at this stage. There is, therefore, some redundancy in recording all three items in this circumstance, indicating that combination of all three results into an overall measure of responsiveness might be accomplished without undue loss of information.

The simplest measure is the sum of the three component scores and indeed this total is now widely known as the Glasgow Coma Score. It ranges from three to fifteen, with scores of less than eight usually indicating coma; scores of nine or more are out of coma (Teasdale et al., 1983). However, at least in theory, the same total score could be made up in a number of different ways. Teasdale et al. (1979a) showed that, in practice, the overall score proves to result, in the majority of cases, from one characteristic combination of responses. This was particularly the case with scores in the lower half of the range (3 - 8) during the first week after injury. Even when the same overall score encompasses groups of patients with different component scores, the outcomes of the different groups prove to be similar. However, information is lost when using the sum or some subset of the components instead of the three individual results. This loss of information may be partly compensated for by the conceptual simplicity of one number versus three.

If any of its component scores is missing then the Glasgow Coma Score is also missing. This happens most frequently with ventilated patients, who have an endotracheal tube in position or a tracheostomy, in which case the verbal response cannot be elicited.

2.5.3 Created Eye Indicant

Plum and Posner (1972) pointed out the value of studying eye

movements as an indication of brain stem function. The data bank records spontaneous eye movements and reflex eye movements from the oculocephalic and oculovesticular reflex response. Patients frequently had one or more of these three observations missing because the test was not carried out. For example, it would be unlikely that the reflex tests would be carried out if the patient had orientating (normal) spontaneous eye movements. To reduce the problems of missing data, dimensionality and dependency between features, a method was devised to take account of one or other feature being absent, impaired or normal and a 'created eye indicant' was devised to combine the information contained in the three features, whether tested or not.

2.6 The Current State of the Data Bank

In spite of the effort and cost involved, data collection in Glasgow has been continuous and all head injuries up to the end of 1986, a total of 2005 cases, are now available for analysis. This brings the total in the data bank to 3078, with 305 cases from Rotterdam, 113 cases from Groningen, 225 from Los Angeles and 430 from San Francisco in addition to those from Glasgow.

Since 1985, Edinburgh, Liverpool and Southampton have joined Glasgow in a multi-centre study of the effect of providing predictions of prognosis to clinicians. By the end of this study in December 1988 it is anticipated that the data bank will have around 4000 cases of severe head injury.

This extensive carefully documented data bank provides a unique resource which is invaluable in the study of severe head injury.

CHAPTER 3

THE USE OF DISCRIMINANT ANALYSIS TO PREDICT THE SIX MONTH OUTCOME OF PATIENTS IN THE HEAD INJURY STUDY

3.1 Introduction

After scales for the assessment of conscious level and outcome had been developed and a database of the features of patients with a severe head injury set up, an attempt could be made to predict, using data collected shortly after injury, the degree of recovery which patients will attain. The first step in the procedure was to examine the relationship between the individual features and outcome. As mentioned in Chapter 1, there are numerous reports identifying such features, and the data bank of 1356 Glasgow cases was used to illustrate examples of the relationships of coma score, pupil reaction and eye indicant to the six month outcome, based on the patient's best state in the first twenty-four hours (Tables 3.1 - 3.3). Similar relationships can be shown using the patient's worst state or data from different time periods. These findings confirmed that depth and duration of coma are reliable markers of severity of brain damage and hence indicators of likely outcome. The natural way forward from this was to use combinations of features to predict outcome. This brings the problem into the framework of discriminant analysis, where the aim is to assign an observation to one of two or more distinct classes or groups, on the basis of a training set of observations whose classes of origin are known. However, here the analysis is used for prognosis rather than the more common medical application of discriminant analysis for diagnosis.

Table 3.1 Relationship between 24 hour best coma score and six month outcome from the data bank of 1356 Glasgow cases

Six Month Outcome	Coma Score 3-5	6-7	8-10	11-15
Death	226 78%	236 54%	88 27%	24 16%
Vegetative State	9 3%	8 2%	3 1%	1 1%
Severe Disability	21 17%	66 15%	44 14%	17 11%
Moderate Disability	16 6%	58 13%	70 22%	44 29%
Good Recovery	16 6%	70 16%	117 36%	66 43%
Total	288 100%	438 100%	322 100%	152 100%

Table 3.2 Relationship between 24 hour best pupil reaction and six month outcome from the data bank of 1356 Glasgow cases

Six Month Outcome	Pupil Reaction	
	Reacting	Not Reacting
Death	409 39%	222 83%
Vegetative State	13 1%	10 4%
Severe Disability	146 14%	21 8%
Moderate Disability	196 19%	9 3%
Good Recovery	276 27%	7 3%
Total	1040 100%	269 100%

Table 3.3 Relationship between 24 hour best eye indicant and six month outcome from the data bank of 1356 Glasgow cases

Six Month Outcome	Eye Indicant		
	Absent/Bad	Impaired	Good
Death	196 90%	102 59%	168 29%
Vegetative State	2 1%	6 3%	10 2%
Severe Disability	9 4%	36 19%	85 15%
Moderate Disability	8 4%	22 12%	124 22%
Good Recovery	3 1%	25 13%	185 32%
Total	218 100%	191 100%	572 100%

3.2 Terminology of Discriminant Analysis

Before proceeding, it is useful to introduce the notation and terminology of discriminant analysis (Duda and Hart, 1973; Aitchison and Dunsmore, 1975; Lachenbruch, 1975) in relation to prognosis.

(i) Individuals in the study are assumed to belong to one of a finite set of k outcome categories, $\Pi_1, \ldots, \Pi_k$.

(ii) Associated with these outcome categories there may be a set of prior probabilities, arrival rates or relative incidences, $p(\Pi_1), \ldots, p(\Pi_k)$ which sum to unity and which summarise our knowledge of the frequency of occurrence of the different categories.

(iii) Each individual has information available in the form of a finite set of feature variables or indicants. These measurements will form a feature vector for the patient.

(iv) A training data set, D, is available of n individuals whose outcome categories and feature vectors are known and represented as

$$D = \{(o_i, x_i), i=1, \ldots, n\}.$$

The outcome category of individual i is denoted by o_i and the feature vector by x_i.

(v) A discriminant rule is set up for assigning an individual to one of the outcome categories or for specifying the probability of each of the different outcome categories, given the feature vector of the individual. The discriminant rule is developed from the training set of data, D.

(vi) A test data set is provided of individuals whose outcome categories and feature vectors are also known, so that the

performance of the discriminant rule can be evaluated. Often the training and test data sets are the same, and less biased evaluation can be achieved provided cross-validatory assessment is used (Lachenbruch and Mickey, 1968).

For a new individual with feature vector y, the discriminant rule gives a means of obtaining estimates for the conditional probabilities $\{p(\Pi_i|y,D),\ i=1,\ \ldots,\ k\}$. These estimates may then be used to assign that individual to the outcome category associated with the largest probability.

There are two approaches to this problem, which Dawid (1976) calls the diagnostic and the sampling paradigms. With the diagnostic paradigm $p(\Pi_i|y,D)$, the distribution of the outcome category for a given feature vector, is modelled directly. With the sampling paradigm, Bayes' theorem is used to give

$$p(\Pi_i|y,D) \propto p(y|\Pi_i,D)\ p(\Pi_i), \qquad i=1,\ \ldots,\ k$$

and both $p(\Pi_i)$, the prior probability, and $p(y|\Pi_i,D)$, the distribution of the feature vector within a given disease category, are modelled.

The diagnostic paradigm is restricted mainly to the use of generalised logistic models, whereas by adopting the sampling paradigm the main effort is in modelling $p(y|\Pi_i,D)$ and so density estimation, either parametric or non-parametric, is of prime concern. This approach gives wide scope for the many methods of density estimation available, but certain decisions have to be made, such as which variables to include, before proceeding with the density estimation problem.

3.3 Factors Important in Comparative Discriminant Analysis Studies

3.3.1 Introduction

In a comparative study of different discriminant rules, some important choices have to be made in addition to that of the model to be used. These include:-

(i) the variables to be selected for inclusion in the rule,

(ii) the criteria to be used for the evaluation of the performance of the rule

(iii) the method to be adopted to deal with missing data, if it exists.

3.3.2 Variable Selection

In many practical discriminant analysis problems, data on a very large number of variables are collected. Indeed, in the Head Injury Study over 300 separate items can be available for some patients. In such cases a subset of the variables has to be selected which it is hoped will be almost as informative as the entire set.

There are many factors which might influence the choice of variables and this makes the problem a difficult one. If the aim was to produce a simple nomogram for diagnostic screening then perhaps only three or four variables could be chosen. However, if computing facilities were available, more variables could be incorporated in a more complex screening rule. Missing data can be important in variable selection. A variable might be a powerful discriminator but be recorded so rarely that it cannot be incorporated into the model. Similarly, the cost or time involved in measuring certain variables has to be considered, in view of the fact that the result of an expensive and time-consuming bioassay

may be no more informative than an easily obtained item of clinical information.

These factors are all concerned with the practical applicability of the final discriminant rule, but its statistical properties are also important. Lachenbruch (1975), Hand (1981) and Habbema and Gelpke (1981) discuss various methods of variable selection which are widely used, and the problem of variable selection is in itself a separate research area. The problem of variable selection for head injury prognosis was important in the early development of a prognostic model and will be discussed further in Section 3.4.3.

3.3.3 Criteria for the Evaluation of the Performance of a Discriminant Rule

To compare the performance of different discriminant rules on a particular sets of variables or of a single discriminant rule on different sets of variables, appropriate criteria must be employed. Two quite separate aspects of performance must be considered. Historically, the more important aspect is how well the groups corresponding to the various outcome categories are separated. However, more recently it has become increasingly important to know whether or not the probabilities assigned to each group are realistic. For example, if there are two outcome categories, a rule which invariably assigns a probability of 0.51 to the correct category gives perfect separation but unrealistic probabilities. At the other extreme a rule which for every case just assigns the prior probabilities does in some sense give accurate probabilities but is of no use for separation. Habbema et al. (1978, 1981), Habbema and Hilden (1981) and Hilden et al. (1978a, 1978b) give an extensive discussion of these points and present a large number of

measures of efficiency for a discrimination procedure.

The measures of separation considered here will be

(i) the error rate,

(ii) the average logarithmic score,

(iii) the average quadratic or Brier score.

(i) The error rate (the proportion of cases allocated to an incorrect category) is the most commonly used measure of separation and was the one used in the early work on variable selection for the Head Injury Study. It is, however, very insensitive as it takes no account of the relative seriousness of different errors, or of near misses, although it does have respectable decision theoretic foundations.

(ii) The logarithmic score for a patient whose true category is, for example, Π_1 is

$$- \log_e p(\Pi_1 | y,D) = - \log p_1, \text{ say.}$$

This measure is sensitive to changes in the diagnostic probabilities. It has, however, one serious drawback from an applied point of view, namely, that if a probability of zero is attached to the actual category, then the penalty associated with this is infinite. In practice, there are methods of dealing with this if it poses a problem (Hilden et al., 1978b).

(iii) The quadratic or Brier score for the above patient is

$$(1 - p_1)^2 + \sum_{i=2}^{k} p_i^2.$$

This measure takes account of the distribution of probability to all outcome categories and not simply that assigned to the actual outcome.

Both the quadratic and logarithmic scores can be interpreted as the distance of the predicted outcome from the actual outcome. If the predicted outcome $\{p(\Pi_i | y,D) \quad i=1, \ldots, k\}$ is denoted by

$p = (p_1, \ldots, p_k)^T$ and the actual outcome by $q = (q_1, \ldots, q_k)^T$ where $q_i = 1$ if i is the actual outcome and 0 otherwise, then the Euclidean measure, $\Delta_Q(p,q)$, where

$$\Delta_Q(p,q) = \| p - q \|^2 = (p - q)^T(p - q)$$

and the Kullback-Leibler measure, $\Delta_{KL}(p,q)$, where

$$\Delta_{KL}(p,q) = \sum_{i=1}^{k} q_i \log \frac{q_i}{p_i}$$

give the quadratic and logarithmic scores respectively.

For good performance from the point of view of separation, all the above measures should be close to zero. A useful benchmark for comparison is to assign the prior probabilities to each individual and to evaluate the performance obtained using the measures described.

3.3.4 Missing Data

The problem of missing data often arises in practical applications of discriminant analysis. These missing values can arise for many different reasons. Studies extending over time are particularly vulnerable to missing observations. For example, a new test might be developed which is thought to provide discriminatory information, but if it is included in the study then its value would be missing in the early cases. Information can be lost through truncation. If a measuring device is only calibrated to give accurate results within a given range of values then any values outside that range would be missing. Other physical factors may also prevent information being recorded. For example, in head injured cases it is not possible to record the eye opening response in a patient with severely swollen eyes or the verbal response of

an intubated patient. These three examples differ in one important aspect. In the first example the data can be said to be missing at random in that, although they are missing in a systematic pattern, the fact that an item of information is missing gives no information about the value it might have taken. This is not the case in the second example, where if an item of information is not recorded, then it lies outside a possible range of values. In the third example it is much more difficult to ascertain whether the data are missing at random or not. Little (1979) gives the following definition of missing at random which is equivalent to that given by Rubin (1976). If n d-variate observations are denoted by the (n x d) data matrix $X = [x_{ij}]$, and the (n x d) random matrix $R = [r_{ij}]$ is defined so that $r_{ij} = 0$ or 1 according to whether x_{ij} is missing or observed, then any missing values are missing at random if the conditional distribution of R given X is independent of the missing values. In particular, the probability that a value x_{ij} is observed must not depend on the value x_{ij} (thus excluding truncation from the definition), although it may depend on the value of an observed variable x_{ik}. Rubin (1976) gives this as the weakest definition of missing at random which allows the mechanism generating the missing values to be ignored.

One method of dealing with missing data is to use only cases where the data are complete to obtain the parameter estimates for the specified model. This however is not always acceptable and other means have to be found to deal with the problem of parameter estimation with incomplete data. Assuming that the data are missing at random as defined previously, Murray (1979) compares different methods of dealing with missing data in the Head Injury Study. With many of the sophisticated statistical modelling techniques now being developed in discriminant analysis, a limiting

factor in their use is how well they can be adapted to cope with missing values.

In the Head Injury Study the mechanism by which a measurement became missing was ignored. In fact, the data were implicitly assumed to be missing at random within each prognostic category. It is fair to say that this is unrealistic. However, it is difficult to avoid this assumption by convenient realistic modelling and the incorporation of the incomplete data does add useful information (Murray, 1979).

3.4 Application of an Independence Model to the Head Injury Data Bank of 600 Cases

3.4.1 Introduction

By 1976 the data bank contained 428 cases from Glasgow and 172 cases from the Netherlands. Jennett et al. (1976) confirmed that the clinical features of the Dutch and Glasgow cases on entry to the study were very similar (Table 3.4). The main difference was that 30% of the Glasgow cases were admitted to the neurosurgical unit more than 24 hours after injury while almost all the Dutch cases were admitted to a neurosurgical or neurological unit within 24 hours. All patients were treated with the techniques and vigour which is normal in a fully equipped unit. In the two countries there were differences in the proportions of patients receiving various therapies, such as mannitol, steroids, and ventilation, and investigations such as angiography. Despite these differences the distributions of six month outcomes in the two centres were similar (Table 3.5). This suggested that, given the standard of care available in a specialised unit, the variations in details of management were not crucial in determining outcome.

Table 3.4 Initial features (24 hour best) of patients from Glasgow and Netherlands from the data bank of 600 patients

Feature	Glasgow (n=428)	Netherlands (n=172)
Mean age (years)	34	33
Lucid interval	31%	25%
Coma score 3-7	70%	73%
Eye movements impaired or absent	46%	42%
Non-reacting pupils	19%	29%
Hemiparesis	19%	21%

Table 3.5 Outcome of patients from Glasgow and Netherlands six months after injury in the data bank of 600 patients

	Glasgow (n=428)	Netherlands (n=172)
Death	52%	52%
Vegetative State	2%	1%
Severe Disability	8%	5%
Moderate Disability	17%	15%
Good Recovery	22%	27%

For the purpose of predicting prognosis the five categories in the Glasgow Outcome Scale were reduced to three:-

(i) death or vegetative state (D/V)

(ii) severe disability (SD)

(iii) moderate disability or good recovery (M/G)

The 600 cases from the two countries were divided at random into three groups of 200. Two of the groups were combined to produce a training data set of 400 cases. The remaining 200 cases acted as a test data set. Predictions were made at the end of the first three time periods i.e. at 24 hours, 3 days and 7 days after injury.

3.4.2 Independence Model

This model for unordered categorical data was chosen initially for its simplicity. For a given feature vector, y, Bayes' Theorem was used to give

$$p(\Pi_i \mid y,D) \propto p(y \mid \Pi_i,D)\ p(\Pi_i), \qquad i=1, \ldots, k.$$

The prior probabilities or relative incidences $\{p(\Pi_i),\ i=1, \ldots, k\}$ were estimated using the proportions among the training cases. For y complete, the density estimate for $p(y \mid \Pi_i,D)$ took the form

$$p(y \mid \Pi_i,D) = \prod_{r=1}^{d} p(y_r \mid \Pi_i,D)$$

$$= \prod_{r=1}^{d} \frac{n_i(y_r) + 1}{n_i(r) + c_r} \qquad (3.1)$$

where

d is the no. of variables,

y_r denotes the r^{th} component of y,

$n_i(y_r)$ is the no. of cases in the training set in outcome category Π_i with score y_r on variable r,

c_r is the number of categories in variable r,

and $n_i(r)$ is the number of patients in the training set in outcome category Π_i with variable r not missing.

Thus it was assumed that, within each outcome category, Π_i, variables were independent and $p(y|\Pi_i,D)$ was given by the product of the estimates of the marginal probabilities. The addition of 1 to the numerator and c_r to the denominator provided a small amount of smoothing to prevent a probability of zero resulting from an empty cell count. One of the most appealing features of this model was that it was trivial to deal with missing data. When y_r was missing, the appropriate factor on the right-hand side of Equation 3.1 was replaced by unity.

3.4.3 Variable Selection

Although many items of information about the patient had been recorded in the data bank, it was decided, after lengthy discussions with the clinicians, to restrict the number available for possible inclusion in the discriminant rule to around 25. Chosen for possible inclusion were those indicants which had already been shown to be related to outcome, such as coma score, pupil reaction, eye signs, motor response patterns, age, etc. If the indicant had a best and worst score available then both were included. The program used at this time to predict the outcome of the patients was provided by Dr Robin Knill-Jones who had used it in the diagnosis of jaundice (Knill-Jones, 1975). It had the following method of variable selection. From the list of indicants available, one was chosen at random, y_1, say. The prior probabilities were updated using Bayes' theorem to obtain the

posterior probabilities of each of the outcome categories

$$p(\Pi_i | y_1, D) \propto p(\Pi_i)\ p(y_1 | \Pi_i, D).$$

If any of the posterior probabilities, $p(\Pi_i | y_i)$, i=1, ..., k, exceeded a pre-determined level then no more information was added for that patient; if not, then another indicant was chosen at random and the probabilities updated again. This continued until either the pre-determined level was reached or all the indicants had been included. In much of the early work the pre-determined level was set arbitrarily at 0.97, and $p(\Pi_i) > 0.97$ for some i=1, ..., k was termed a confident prediction.

The increase in the probability of the actual outcome given by the inclusion of a particular indicant was termed the reduction in uncertainty. After all cases had been predicted, the average reduction in uncertainty for each indicant was given. During the development of the prognostic system I carried out many runs of the program using different training sets and data from different time periods. It became apparent that a relatively small number of indicants (about 4) were consistently useful in reducing uncertainty.

As well as comparing the performance of different variable sets using the error rate, I examined optimistic and pessimistic errors. An optimistic error is defined as a confident prediction of an outcome of M/G in a patient whose actual outcome is D/V, while a pessimistic error is a confident prediction of an outcome of D/V in a patient whose actual outcome is M/G. While optimistic errors as a result of the statistical methodology were acceptable to the clinicians, pessimistic errors were not. A closer examination of the pessimistic errors revealed that they were occurring in cases where the worst data were included rather than the best data. A

case with poor worst data and good best data would be predicted to have a poor prognosis if the worst data alone were chosen. As a result of the clinical unacceptability of pessimistic errors, it was decided at an early stage in the development to use only the best scores in the time periods.

By the end of the second time period almost twice the number of indicants were available for inclusion, with most clinical information being available as the 24 hour best score and the 2-3 day best score. When both the 24 hour and 2-3 day score were available for selection I invariably found that the reduction in uncertainty from the 2-3 day score was greater than that from the 24 hour score. Similarly, at 4-7 days the reduction in uncertainty was greater with the 4-7 day data than with that for the previous time periods. This reinforced the clinicians' conviction that the current state of the patient was more important than their past state. As fewer than eight variables were customarily all that was used to predict outcome, it was decided, for the time dependent variables, to use only the data for the current time period when predicting outcome. Thus, if a prediction was made at the end of 7 days, only the 4-7 day best data were used along with other variables such as age or time elapsed between injury and coma.

As a result of this developmental work, six variables (age, coma score, motor response pattern, pupil reaction, eye indicant and change in neurological function) were found to be consistently useful and were adopted in practice. If the indicant had more than three or four response levels, as for example with coma score, then these were grouped together in such a way as to retain as much of the prognostic information as possible. This was done using an entropy measure based on the conditional outcome given the (grouped) coma score (Teasdale et al., 1979a). As a result of

recovery and selection processes the marginal distribution changes through time so that the grouping used at 24 hours was 3-5, 6-7, 8-15 whereas at 28 days it was 3-10, 11-13, 14-15.

3.4.4 Results of the Predictions using the Independence Model

The six month outcomes of the test and training groups are given in Table 3.6. The classification matrices arising from the predictions using 24 hour, 2-3 day and 4-7 day data are given in Tables 3.7(a), 3.8(a) and 3.9(a) respectively. When predicting the outcome at 3 days after injury all cases who died in the first 24 hours were excluded from the test and training sets. Similarly, all cases who died within the first 3 days were excluded from the 7 day predictions. The error rates corresponding to these classification matrices and the error rates obtained by allocating the prior probabilities to each case are given in Table 3.10(a).

While it might be expected that more accurate predictions could be made at later time periods, the error rate increased from 24 hours to 3 days to 7 days. This was because the early deaths were excluded; these cases usually have correct predictions and so the problem of predicting prognosis becomes more difficult.

These error rates were unacceptable to the clinicians and so it was decided not to classify a patient unless he had a confident prediction of outcome as previously defined ($p(\Pi_i|y) > 0.97$ for some i). This reduced the number of cases being classified to 38%, 52% and 45% at 24 hours, 3 days and 7 days respectively. It also substantially reduced the error rate. The classification matrices for the first three time periods and their error rates for the confident cases are given in Tables 3.7(b) - 3.10(b).

By classifying only cases with confident predictions there was no pessimistic error and the few optimistic errors were acceptable

Table 3.6 Six month outcome of test and training data sets in the data bank of 600 patients

Outcome	Frequencies	
	Training Set	Test Set
Death or Vegetative State	213	106
Severe Disability	31	12
Moderate Disability or Good Recovery	156	82
Total	400	200

Table 3.7 Classification matrices for the independence model using the 24 hours best data with the data bank of 600 patients

a) for all cases predicted and

b) for cases with a confident prediction

Predicted Outcome	Actual Outcome			
	D/V	SD	M/G	Total
D/V	90	4	14	108
SD	2	1	0	3
M/G	14	7	68	89
Total	106	12	82	200

(a)

Predicted Outcome	Actual Outcome			
	D/V	SD	M/G	Total
D/V	45	1	0	46
SD	0	0	0	0
M/G	2	1	27	30
Total	47	2	27	76

(b)

Table 3.8 Classification matrices for the independence model using the 2-3 day best data with the data bank of 600 patients

a) for all cases predicted and

b) for cases with a confident prediction

Predicted Outcome	Actual Outcome			
	D/V	SD	M/G	Total
D/V	59	3	12	74
SD	2	0	0	2
M/G	10	9	70	89
Total	71	12	82	165

(a)

Predicted Outcome	Actual Outcome			
	D/V	SD	M/G	Total
D/V	37	0	0	37
SD	0	0	0	0
M/G	2	1	46	49
Total	39	1	46	86

(b)

Table 3.9 Classification matrices for the independence model using the 4-7 day best data with the data bank of 600 patients

a) for all cases predicted and

b) for cases with a confident prediction

Predicted Outcome	Actual Outcome			
	D/V	SD	M/G	Total
D/V	35	3	9	47
SD	3	0	3	6
M/G	7	9	70	86
Total	45	12	82	139

(a)

Predicted Outcome	Actual Outcome			
	D/V	SD	M/G	Total
D/V	21	1	0	22
SD	0	0	0	0
M/G	1	3	36	40
Total	22	4	36	62

(b)

Table 3.10 Error rates from classification matrices

a) for all cases predicted

b) for cases with a confident prediction

Time Period	Classification Error Rate	Prior Error Rate
24 hours	.205	.470
2-3 days	.218	.570
4-7 days	.245	.410

(a)

Time Period	Classification Error Rate	Proportion Confident
24 hours	.053	.380
2-3 days	.035	.521
4-7 days	.081	.446

(b)

on the grounds that complications can develop in patients who, soon after injury, appear to have the potential for recovery.

It should also be noted that few cases were predicted to have an outcome of severe disability. This can be largely explained by the low prior probability attached to this outcome and the fact that this group overlaps both of the other outcome categories.

3.4.5 Discussion

This early work produced encouraging results and was one of the first examples of such methodology to be published in the medical press. In spite of this success our use of the independence model was seen by some commentators (Becker, 1979; Stablein et al., 1980) to be simplistic. In particular they suggested that a logistic regression technique would obviate the problems of dependence and interaction amongst the variables. A comparative study of different discrimination techniques using a large data set is in itself an interesting statistical exercise. Partly for this reason and partly to answer the critics such a comparative study was carried out and this is described in Section 3.5.

3.5 Comparative Study of Discrimination Techniques

3.5.1 Introduction

As the comparative study was a major undertaking, all the statisticians who were involved in the Head Injury Study collaborated to make the report possible (Titterington et al., 1981). My main contribution was to the design of the study, and in particular to the selection of the variable subsets to be considered.

The purpose of this study was to compare statistical

methodology, and therefore certain standardisation in the data is needed so that the results of different methods would be based on equivalent information. Thus the five categories of the Glasgow Outcome Scale were reduced to three as in Section 3.4.

The various indicants considered are shown in Table 3.11. These factors have already been shown to be indicators of the degree of brain damage. For the purpose of this study, the indicants were based on the patient's best state during the first 24 hours after onset of coma, and the work was limited to estimating the probability of attaining one or other of the three outcome categories six months after injury. It can be seen that the variables are all categorical and are either binary or ordered. This means that methods based on continuous data might be considered as possible, albeit unsatisfactory, alternatives to categorical data techniques. Different subsets of these variables were chosen to compare how well the various methods were able to exploit the information in subsets of different sizes, and to see how the methods reacted to the degree of dependence among the variables as well as to the proportion of missing data. The four subsets used are given in Table 3.12. Set I consisted of four weakly dependent variables with appreciable missing data while set II consisted of four highly dependent variables with little missing data. Set III was an extension of I and set IV was obtained from set III by expanding the coma score and created eye indicant into their components. There was therefore high dependence and appreciable missing data within this set. The data bank had risen to 1000 cases by this time and these were split randomly into two groups of 500 to give separate test and training sets. The distribution of outcome in the two groups is given in Table 3.13.

Table 3.11 Feature variables used in the comparative study

Variable	Description
Age	Age, grouped into decades 0-9, 10-19, ..., 60-69, 70+
E score	Eye opening in response to simulation, graded 1 (nil) to 4 (normal), but grouped as 1 and 2 - 4 for these comparisons
M score	Motor response of best limb in response to stimulation, graded 1 (nil) to 6 (normal)
V score	Verbal response to stimulation, graded 1 (nil) to 5 (normal), but grouped as 1 and 2 - 5 for these comparisons
Coma score	The sum of the raw E, M and V scores, in the range 3 to 15, but grouped as 3, 4, 5, 6, 7, 8, 9 - 15 for these comparisons
MRP	Motor response pattern, an overall summary of the motor responses in all four limbs, graded 1 (nil) to 7 (normal)
Change	Change in neurological function over the first 24 hours, graded 1 (deteriorating), 2 (static) or 3 (improving)
Pupils	Pupil reaction to light, graded 1 (non-reacting) or 2 (reacting)
SEM	Spontaneous eye movements, graded 1 (nil) to 4 (normal)
OCS	Oculocephalics, graded 1 (nil) to 4 (normal)
OVS	Oculovestibulars, graded 1 (nil) to 4 (normal)
Eye indicant	A summary of SEM, OCS and OVS, graded 1 (bad), 2 (impaired) or 3 (good)

Table 3.12 Subsets of the feature variables used in the comparative study

Set	Variables
I	Age, Coma score, Change, Eye indicant
II	Age, E score, M score, V score
III	Age, Coma score, MRP, Change, Pupils, Eye indicant
IV	Age, E score, M score, V score, MRP, Change, Pupils, SEM, OCS, OVS

Table 3.13 Six month outcome of test and training data sets in the comparative study

	Frequencies	
	Training set	Test set
Death or Vegetative State	259	250
Severe Disability	52	48
Moderate Disability or Good Recovery	189	202
Total	500	500

3.5.2 Statistical Techniques

The statistical methods used can be brought together under the following general headings.

(i) Independence-based models for unordered categorical data, allowing for a single overall association factor.

(ii) Lancaster first-order interaction models for unordered categorical data.

(iii) Latent class models.

(iv) Kernel-based procedures for categorical data.

(v) Linear and quadratic discrimination based on normality assumptions.

(vi) Linear logistic discrimination.

All but (vi) involve density estimation in one form or another. Details of the different models and references are given in the paper.

(i) Independence based models

In these the density estimates took the form, for y complete,

$$p(y|\Pi_i,D) \propto \left[\prod_{r=1}^{d} \frac{n_i(y_r) + {}^1/c_r}{n_i(r) + 1} \right]^B$$

where d, y_r, $n_i(y_r)$, c_r and $n_i(r)$ are as defined in Section 3.4.2 and B is an overall assosciation factor. Three independence-based models, **INDEP1**, **INDEP2**, and **INDEP3** were used corresponding to the choices of 1.0, 0.8 and 0.5 respectively for the value of B. This factor B imposes some smoothing and represents the proportion of non-redundant information in the variables (Hilden and Bjerregaard, 1976). An association factor of 1.0 corresponds essentially to the model described in Section 3.4.2 and missing data were dealt with as described in that section.

(ii) Lancaster models

The structure of these models is such that a full range from basic independence to full multinomial models is permitted. Missing data treatment was the same as for the independence model, and when the independence model had to be used to avoid negative probabilities, the same three choices of association factor gave rise to the methods **LANC1**, **LANC2** and **LANC3** respectively.

(iii) Latent class models

In latent class analysis, mixture models are assumed for the density functions being estimated. Thus, for each Π_i, it is assumed that

$$p(y|\Pi_i) = \sum_{j=1}^{L} w_{ij}\, p_j(y),$$

where

L is the number of terms (latent classes) in the mixture,

$p_j(.)$, $j = 1, \ldots, L$ are the densities involved in the mixture

and w_{ij} are, for each i, a set of mixing weights (Fielding, 1977).

For each of the variable sets the two best consecutive numbers of latent classes gave rise to methods **LATCL1** and **LATCL2**.

(iv) Kernel-based procedures

With this procedure

$$p(y|\Pi_i,D) = \frac{1}{n_i} \sum_{j=1}^{n_i} K(y|x_{ij},\lambda),$$

where

n_i = number of patients in the training set in category Π_i,

x_{ij}, $j = 1, \ldots, n_i$ denote their feature vectors,

$K(.|x,\lambda)$ is a probability density over the sample space of y,

and λ describes the degree of smoothing of the relative frequencies.

The kernel methods used were:-

KERUN1: The kernel method of Murray & Titterington (1978) - unordered categories with the smoothing parameters chosen marginally.

KERUN2: As KERUN1 but with the smoothing parameters chosen by a multivariate pseudo-Bayesian technique.

KERORD1 and **KERORD2:** As KERUN1 and KERUN2 but assuming ordered categories.

KEREX1 and **KEREX2:** Marginal and multivariate choices of the smoothing parameters, treating 'missing' as an extra category.

KEREX3: 'Missing' treated as an extra category and a single smoothing parameter chosen for all dimensions.

(v) Normal Methods

These methods assume multivariate normality and estimate the mean vectors and covariance matrices by maximum likelihood. The methods used were:-

NORLIN1: Covariance matrices were assumed equal and sample means from available data were substituted for missing data.

NORLIN2: As for NORLIN1 but with proper maximum likelihood treatment for missing data via the EM algorithm (Dempster et al., 1977).

NORQUAD: As for NORLIN2 but without the assumption of equal covariance matrices.

With all three methods, incomplete test cases were classified on the basis of the relevant marginal distributions.

(vi) Linear Logistic Method

This is the only method in which $\{p(\Pi_i|y),\ i=1,\ \ldots,\ k\}$ is modelled directly. The models take the parametric form

$$p(\Pi_i|y) \ / \ p(\Pi_k|y) = \exp(\alpha_i + \beta_i{}^T y) \qquad i=1,\ \ldots,\ k-1$$

where $\{\alpha_i\}$ and $\{\beta_i\}$ are to be estimated. The technicalities are described by Anderson (1972). Missing data were replaced by group means in the training cases and grand means in the test cases, giving the method **LINLOG**.

3.5.2 Results of the Comparative Study

To provide a benchmark for the performance of the different methods the prior probabilities were assigned to each case and the error rate, average logarithmic score and average quadratic score were calculated for the test data set. This discriminant rule would score 0.500, 0.939 and 0.579 respectively on the three measures. The results for the four variable sets are given in Tables 3.14 - 3.17.

Many comparisons can be made. These were considered as follows:

(i) within groups of similar methods

(ii) among groups of similar methods

(iii) among the sets of variables.

(i) The discrete parametric models were considered first, namely the independence, Lancaster and latent class models. For variable set I INDEP1 and INDEP2 performed well, giving similar results; INDEP3 had poorer results, as had the latent class models. The Lancaster models all gave similar results which were also inferior to those of the independence model. The independence model still performed well with variable set II even though the variables were highly dependent. The Lancaster models again gave similar results which were superior to those of the independence model in terms of error rate but inferior in terms of the logarithmic score. The latent class results were poorer than the others with respect to the quadratic and logarithmic scores. For variable set III the

Table 3.14 Results of the comparative study for variable set I: Age, Coma score, Change, Eye indicant

Method	Measure of Separation		
	Error Rate	Average Logarithmic Score	Average Quadratic Score
INDEP1	.278	.685	.377
INDEP2	.268	.681	.379
INDEP3	.268	.708	.400
LANC1	.292	.737	.397
LANC2	.294	.735	.398
LANC3	.296	.742	.404
LATCL1	.264	.719	.390
LATCL2	.290	.752	.409
KERUN1	.316	.934	.467
KERUN2	.308	.925	.449
KERORD1	.292	.874	.443
KERORD2	.302	.900	.430
KEREX1	.320	.889	.453
KEREX2	.328	1.037	.477
KEREX3	.282	.800	.420
NORLIN1	.286	.707	.396
NORLIN2	.284	.702	.396
NORQUAD	.294	.779	.404
LINLOG	.290	.721	.400

Table 3.15 Results of the comparative study for variable set II: Age, E score, M score, V score

Method	Measure of Separation		
	Error Rate	Average Logarithmic Score	Average Quadratic Score
INDEP1	.338	.775	.438
INDEP2	.340	.762	.436
INDEP3	.338	.771	.445
LANC1	.298	.808	.435
LANC2	.298	.809	.437
LANC3	.296	.818	.445
LATCL1	.328	.819	.447
LATCL2	.310	.822	.446
KERUN1	.346	.924	.481
KERUN2	.328	.872	.463
KERORD1	.352	.905	.471
KERORD2	.332	.856	.454
KEREX1	.334	.953	.491
KEREX2	.326	.903	.475
KEREX3	.340	.852	.466
NORLIN1	.316	.760	.433
NORLIN2	.306	.757	.431
NORQUAD	.304	.884	.450
LINLOG	.314	.764	.436

Table 3.16 Results of the comparative study for variable set III: Age, Coma score, MRP, Change, Pupils, Eye indicant

Method	Measure of Separation		
	Error Rate	Average Logarithmic Score	Average Quadratic Score
INDEP1	.248	.686	.364
INDEP2	.246	.656	.358
INDEP3	.232	.652	.362
LANC1	.254	.738	.382
LANC2	.256	.728	.378
LANC3	.244	.727	.376
LATCL1	.298	.726	.412
LATCL2	.262	.718	.372
KERUN1	.332	1.103	.500
KERUN2	.338	1.267	.537
KERORD1	.328	1.030	.482
KERORD2	.316	1.270	.514
KEREX1	.310	1.013	.467
KEREX2	.344	1.412	.548
KEREX3	.278	.769	.395
NORLIN1	.256	.665	.368
NORLIN2	.258	.661	.367
NORQUAD	.276	.907	.411
LINLOG	.272	.676	.370

Table 3.17 Results of the comparative study for variable set IV: Age, E score, M score, V score, MRP, Change, Pupils, SEM, OCS, OVS

Method	Measure of Separation		
	Error Rate	Average Logarithmic Score	Average Quadratic Score
INDEP1	.272	.839	.399
INDEP2	.264	.757	.385
INDEP3	.264	.673	.368
LANC1	.286	.829	.410
LANC2	.286	.800	.403
LANC3	.280	.768	.395
LATCL1	.282	.726	.396
LATCL2	.244	.709	.381
KERUN1	.350	1.417	.566
KERUN2	.390	1.932	.645
KERORD1	.340	1.414	.543
KERORD2	.374	1.923	.628
KEREX1	.388	1.645	.634
KEREX2	.398	2.143	.652
KEREX3	.298	.806	.412
NORLIN1	.270	.804	.404
NORLIN2	.250	.663	.361
NORQUAD	.274	.947	.424
LINLOG	.286	.772	.412

independence models were better than the Lancaster models which in turn were better than the latent class results. With variable set IV the differences in results were most marked. INDEP3 clearly bettered INDEP2 which in turn bettered INDEP1. There was a similar pattern for the Lancaster models. INDEP3 was better than LANC3 and for this variable set alone the latent class models performed well.

The comparison among the discrete kernel methods was clear cut with KEREX3 being the best for all variable sets.

The continuous parametric models NORLIN1, NORLIN2, NORQUAD and LINLOG (strictly speaking this method is not restricted in application to continuous data) also gave a clear pattern of results over the four variable sets, with the quadratic method performing poorly. With the linear method it was always preferable to use the EM algorithm. The difference with the EM algorithm was small for sets I-III but marked for set IV. The results for the linear logistic method were slightly poorer than those of the linear methods, but better than NORQUAD. The results of LINLOG were particularly encouraging when the crude treatment of missing data is also considered.

(ii) Table 3.18 gives the results for the best method from each of the 3 groups of similar methods for each variable set. When it was not obvious which method was best they were ordered by the average quadratic score. Comparisons among groups of similar methods showed that the kernel methods had the most disappointing results, especially for the logarithmic and quadratic scores, and it was only for variable set II that they even approached the other methods. This was perhaps because a discrete kernel approach in this problem was too ambitious. The results for the linear methods were remarkably similar to those achieved with the discrete models. For sets I and III the discrete models had the edge while for sets

Table 3.18 Overall summary of the results of the comparative study

Method	Measure of Separation		
	Error Rate	Average Logarithmic Score	Average Quadratic Score
INDEP1	.278	.685	.377
KEREX3	.282	.800	.420
NORLIN2	.284	.702	.396
LANC1	.298	.808	.435
KERORD2	.332	.856	.454
NORLIN2	.306	.757	.431
INDEP2	.246	.656	.358
KEREX3	.278	.769	.395
NORLIN2	.258	.661	.367
INDEP3	.264	.673	.368
KEREX3	.298	.806	.412
NORLIN2	.250	.663	.361

II and IV this was true of the linear models. The differences were so small as to be of little importance in practice. However, with the linear method there was a single model which performed well for each variable set, whereas with the discrete methods the choice of model could be critical.

(iii) In the assessment of the overall performance of the different variable sets, it can be seen from Table 3.18 that the variation in performance among methods tended to be smaller than among the variable sets. The best overall set of results was obtained with method INDEP2 on variable set III and it was interesting that, although set IV contained strictly more information, the discrete model could not exploit this. In contrast, with the linear method, the performance improved going from set I to set III to set IV although the results for III and IV were very similar. This emphasised the robustness of the linear approach, which appeared to make sensible use of the available information, whereas the discrete parametric models had to be matched carefully to the variables being used.

This suggested that, while the linear approach with the EM algorithm was preferable for a quick, uninformed analysis, it was possible to achieve similar, if not better, performance with the much simpler independence model if prior, background information was used to combine groups of highly dependent variables into single created indicants.

3.5.4 Discussion and Conclusions of the Comparative Study

The results of this comparative study went some way towards defending the use of the independence model. Its robustness, together with the ease with which incomplete data are dealt, make it appealing even though the assumptions it makes are often

violated. This was partly explained by Hilden (1984), who describes weaker conditions than conditional independence under which the model is still valid. These findings highlight a fundamental feature of discriminant analysis, namely that any modelling involved is only an intermediate step and that methods should be assessed in terms of performance rather than in terms of goodness-of-fit.

The results of the more complex Kernel methods were disappointing and emphasise the need for more work on the choice of smoothing parameters.

No method was particularly successful in identifying patients who will be severely disabled. As these cases are in need of continuing medical and social care, this is an important practical aim. The lack of success is due, as stated previously, partly to the relatively low prior probability of this outcome and partly because, geometrically, the severe group is overlapped by those in both the other two outcome categories. A simple univariate example to illustrate how large these misclassification probabilities can be is given in Appendix 2. In view of the ordering of the outcome categories it is possible that, with further development, the methods of McCullagh (1980) and Anderson (1984a) might be useful. The class of McCullagh models is based on an underlying continuous latent variable, which may not be observable. The ordered outcomes correspond to adjacent grouped intervals on the latent scale. The 'stereotype regression' models of Anderson are more general in that they they do not assume an ordered structure, but they do allow one to test whether an ordered structure is appropriate.

3.6 Clinical Implications for Prognosis

One of the main features of the methods described in Sections 3.4 and 3.5 was the lack of success in predicting the outcome of severely disabled patients. Initially it was thought that making the prediction at a later time period might help solve the problem, but when the independence model was used with the data available at 28 days, as in Section 3.4, poor results were obtained (Tables 3.19 (a) and (b)). One of the reasons for this was that almost all survivors had by this time achieved the best score possible in the features used for the discriminant rule so the discriminatory power of these features had diminished. Table 3.20 illustrates this point.

The cases who are severely disabled are physically dependent to varying but significant extents on other individuals. This is particularly relevant in severe head injuries whose mean age is 34 years so that many still have the greater part of their life to live. It is thus of great clinical interest to be able to identify those cases who will be severely disabled soon after injury. On the one hand, intensified treatment at an early stage might reduce their dependence, on the other, the fruitlessness of long continued rehabilitation might be recognised and thus more emphasis put on the readjustments necessary to cope with handicap.

It was therefore thought to be important to pursue a new approach to try to identify these cases.

Table 3.19 Classification matrices for the independence model using the 15-28 day best data with the data bank of 600 patients

a) for all cases predicted and

b) for cases with a confident prediction

Predicted Outcome	Actual Outcome			
	D/V	SD	M/G	Total
D/V	18	1	4	23
SD	1	0	1	2
M/G	7	17	70	94
Total	26	18	75	119

(a)

Predicted Outcome	Actual Outcome			
	D/V	SD	M/G	Total
D/V	1	0	1	2
SD	0	0	0	0
M/G	0	1	25	26
Total	1	1	26	28

(b)

Table 3.20 Distribution of some features of the data set of 600 cases at 24 hours and 15 -28 days

Feature	Time Period	Proportion of those alive with feature
Coma score 13-15	24H	2%
	15-28D	71%
Normal eye movements	24H	55%
	15-28D	93%
Normal motor response pattern	24H	56%
	15-28D	77%

CHAPTER 4

MODELS USING TIME TRENDS

4.1 Introduction

A major criticism of all the methods used in the comparative study is that they ignore the dynamic nature of the recovery process. Over a long period of working with the data I gained the impression that the individuals who were ultimately severely disabled were characterised by a lack of detectable change in neurological function in the first few weeks after injury. They often did not have low coma scores on admission: indeed some cases with lower coma scores improved to make a good recovery while others with higher scores deteriorated and died. From this came the idea of trying to model the recovery trend through time in an attempt to use the different trends to identify those who would be severely disabled.

4.2 Exploratory Analysis of Recovery Trends

As the coma score had been shown previously to be consistently useful in predicting outcome, it was intuitively sensible and clinically acceptable to use this as the feature with which to try to develop the model. Up to ten scores can be available for each case from the best and worst coma scores at each of the five time periods within which the patient is monitored. Since much of the previous work was based on the best coma score within each time period this was chosen again for the analysis of trends. Thus each patient has up to five scores with which to model the trend in recovery.

The head injury data bank now had 1356 Glasgow cases and it was decided to use only those Glasgow cases to develop the model. Their outcomes are given in Table 4.1.

To obtain an overall impression of the recovery trends within each outcome category, each patient within that group had their coma score plotted at each of the 5 times with coma scores at successive times joined to give an impression of trend. To avoid the problem of missing data, only cases with complete data were used. The time scale used to plot the trends was chosen arbitrarily. As the time periods used in data collection were all unequal and the best coma score could occur at any time within the period there seemed no advantage or disadvantage in using an equally spaced time scale to try to model the trend. This scale was adopted throughout the study, but it must be emphasised that its choice was entirely arbitrary.

The results for each of the five outcome categories are shown in Figures 4.1(a) - 4.1(e) with each line representing a patient. A small random shift has been introduced to separate coincident lines. While the picture is somewhat confusing, the overall subjective impression from the figures seemed to reinforce the hypothesis that the coma scores of the severe disability group did indeed change more slowly than those of other survivors.

To summarise the data, the mean and standard deviation of the coma score at each time period for each outcome category are given in Table 4.2 and the mean trend is plotted in Figure 4.2.

From Figure 4.2 it appears that the severe disability, moderate disability and good recovery groups have similar recovery patterns while the vegetative survivors and deaths appear to follow a quite different pattern. Efforts were therefore concentrated on modelling the recovery trend in the severe disability, moderate

Table 4.1 Outcome of the 1356 Glasgow cases in the head injury data bank

Outcome	N	
Death	650	(48%)
Vegetative State	23	(2%)
Severe Disability	174	(13%)
Moderate Disability	211	(16%)
Good Recovery	291	(21%)
Lost to Follow-up	7	(1%)

Figure 4.1(a) Trend in coma score for patients in the good recovery group

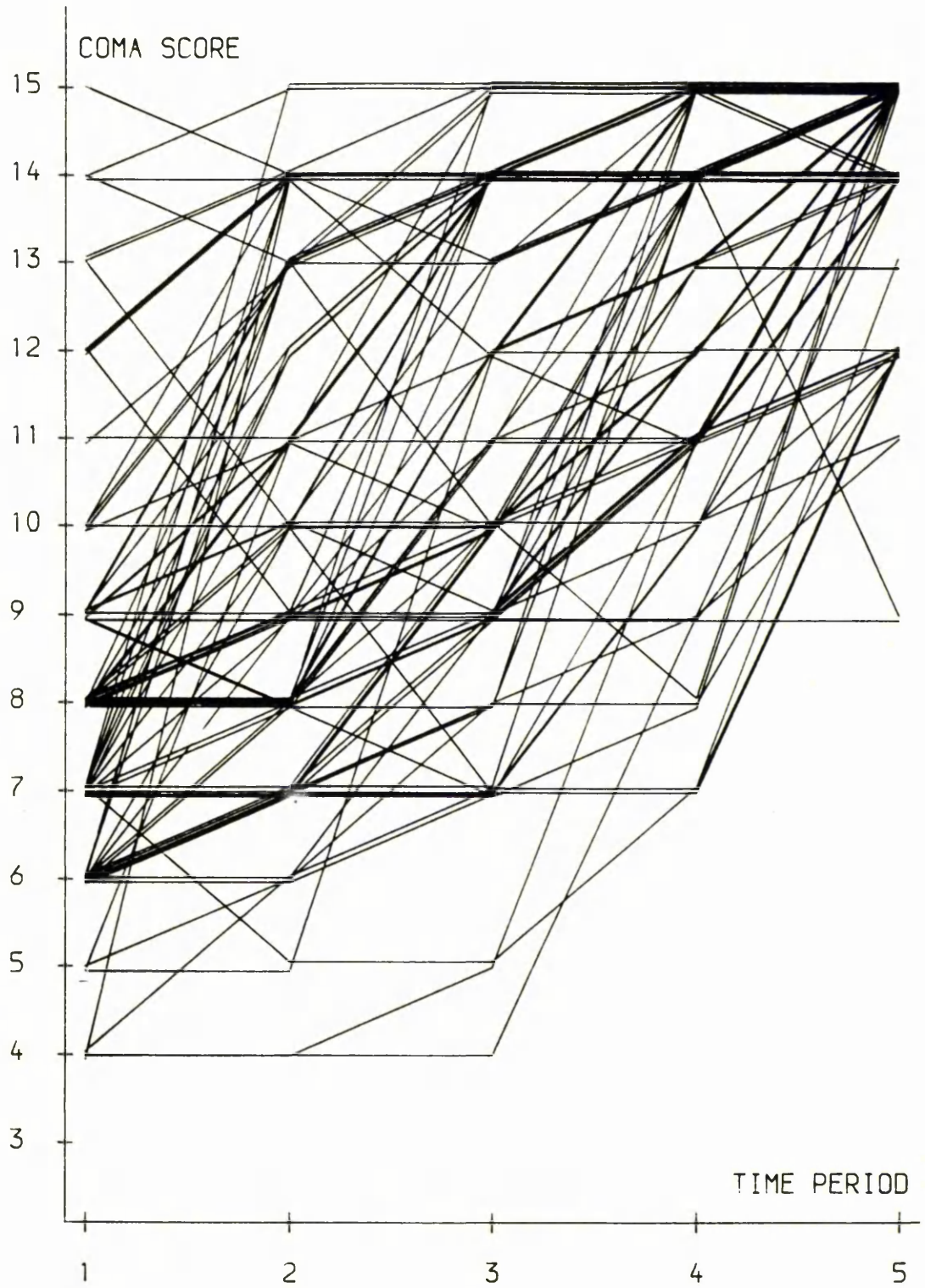


Figure 4.1(b) Trend in coma score for patients in the moderate disability group

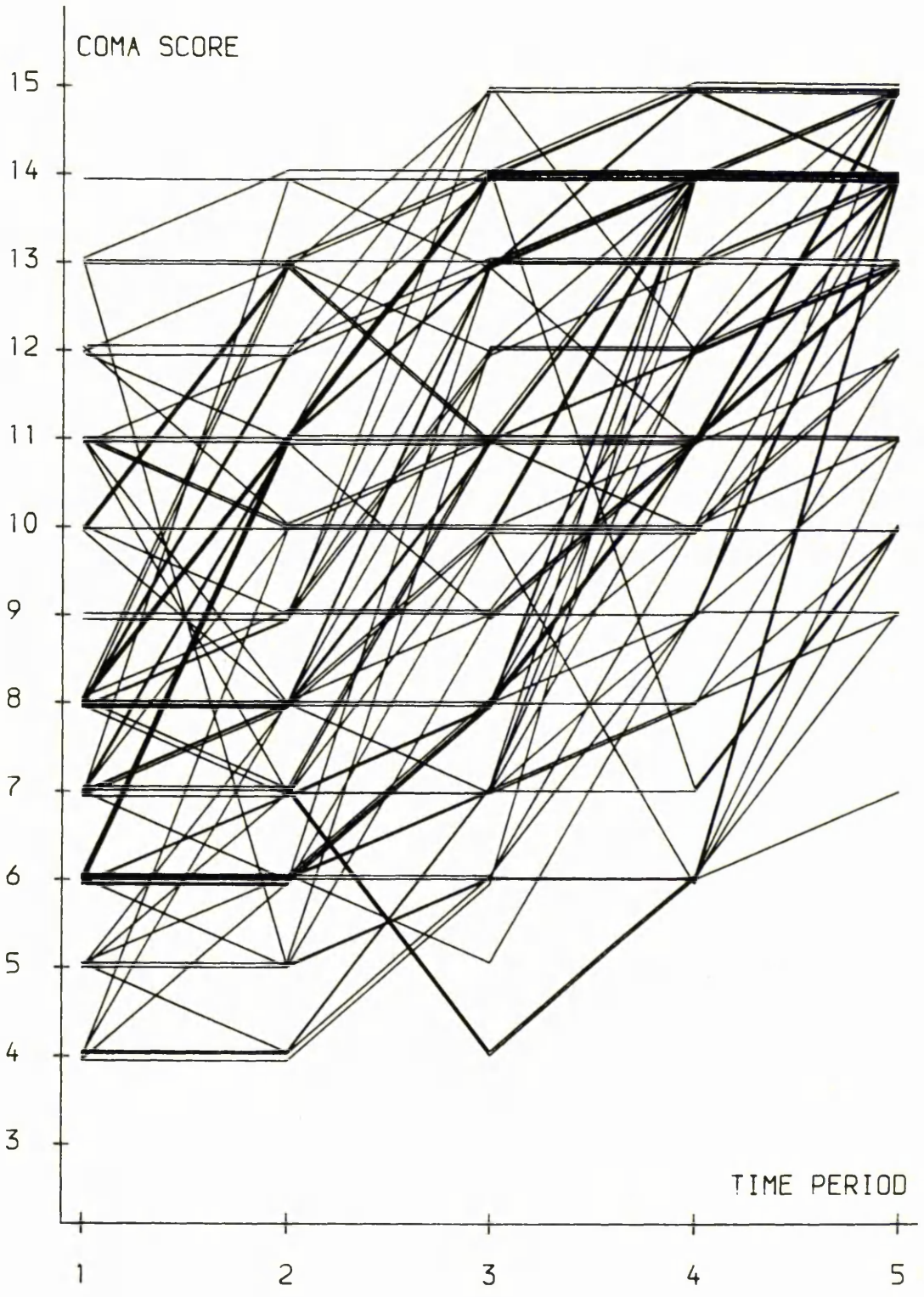


Figure 4.1(c) Trend in coma score for patients in the severe disability group

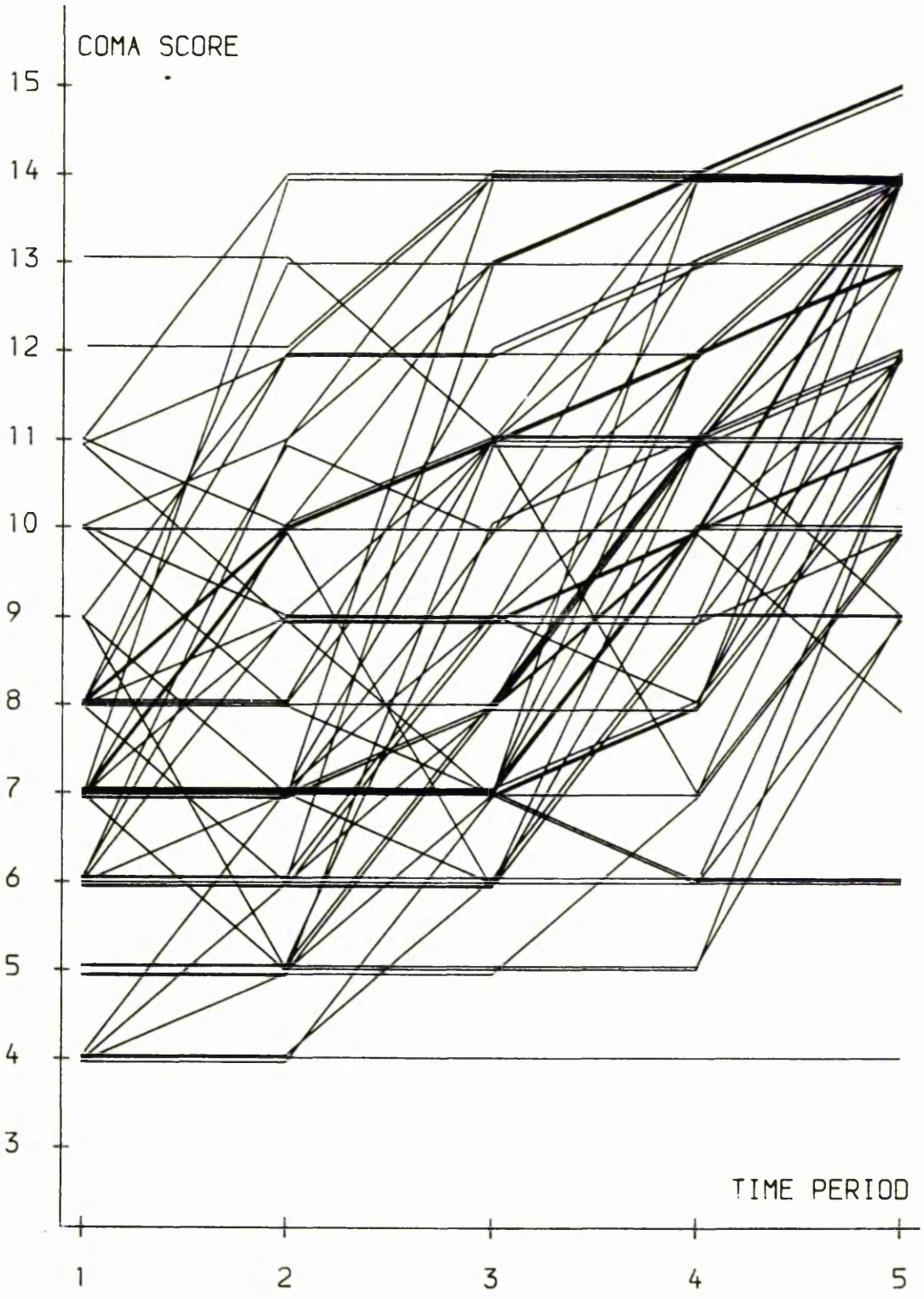


Figure 4.1(d) Trend in coma score for patients in the vegetative group

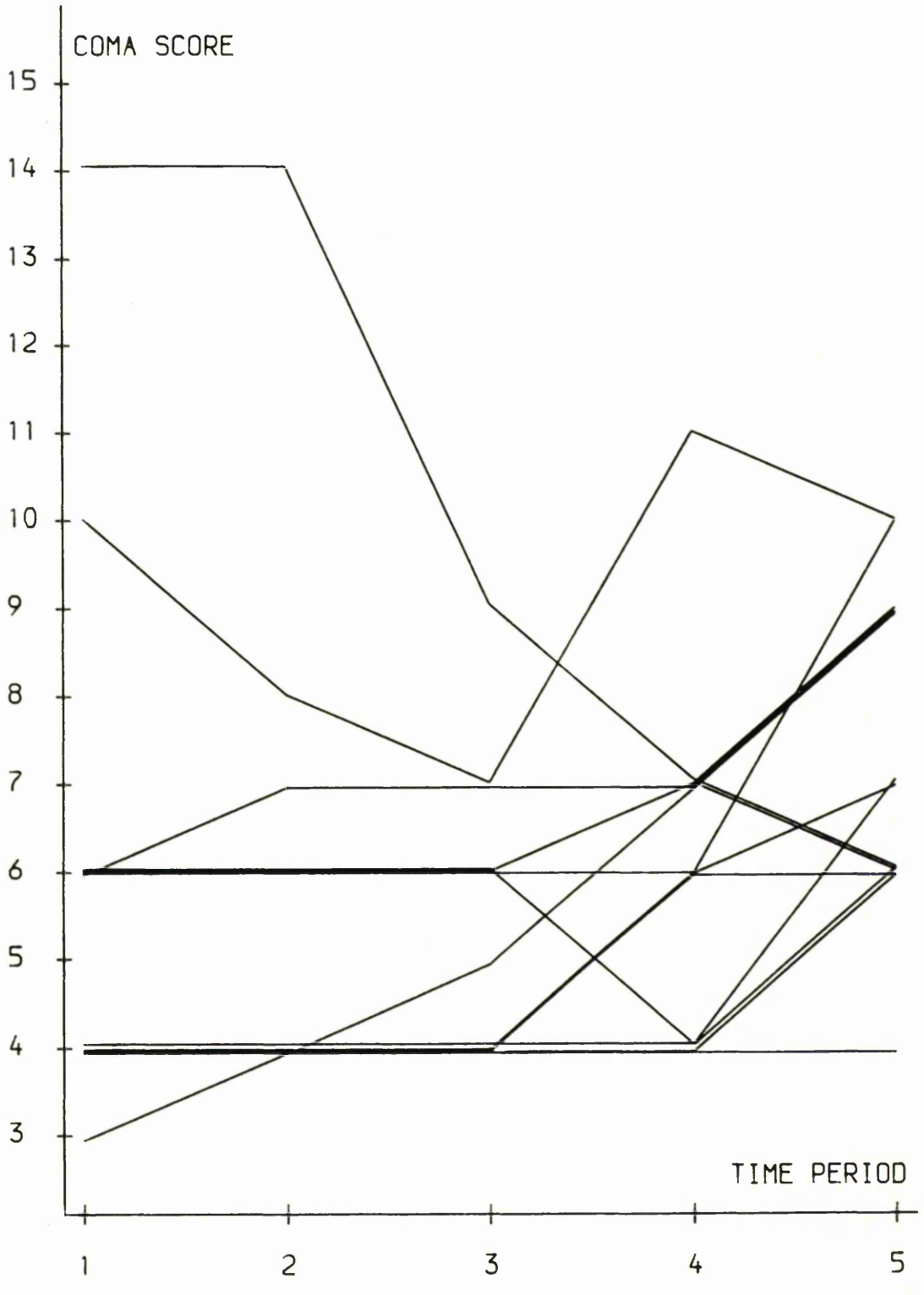


Figure 4.1(e) Trend in coma score for patients who died

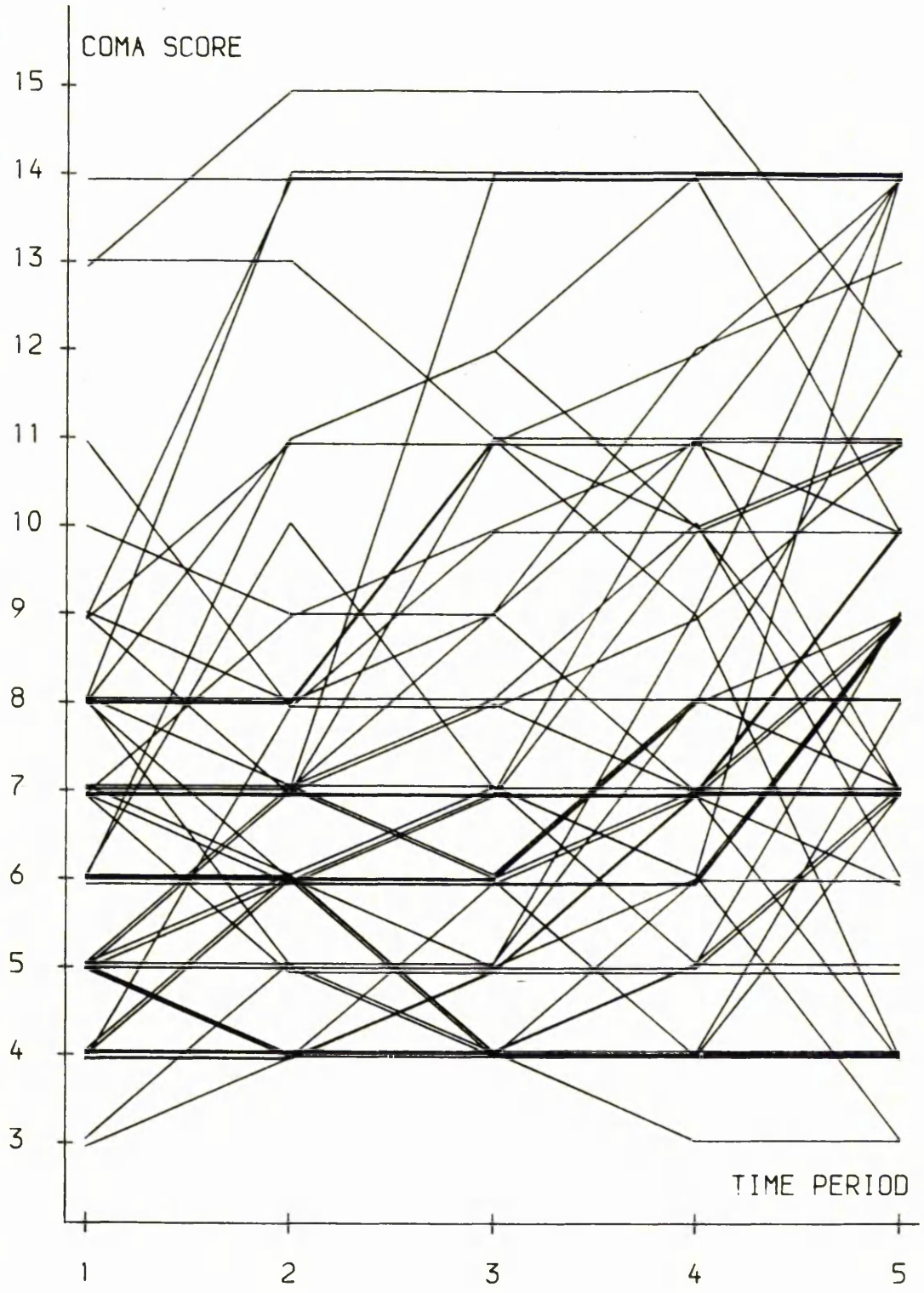


Table 4.2 Mean and standard deviation of the best coma score at each time period for each outcome category

Outcome	Mean Coma Score Standard Deviation				
	24 hours	2-3 days	4-7 days	8-14 days	15-28 days
Death	6.13 2.11	5.98 2.50	6.46 2.91	7.68 3.21	8.52 3.26
Vegetative State	6.00 2.57	6.00 2.43	5.87 2.03	6.28 2.40	7.12 2.23
Severe Disability	7.75 2.19	8.42 2.67	9.94 2.91	10.97 2.67	12.01 2.33
Moderate Disability	8.39 2.49	9.73 2.92	11.30 2.97	12.42 2.34	13.36 1.78
Good Recovery	8.85 2.51	10.94 2.84	12.46 2.61	13.41 1.95	14.01 1.50

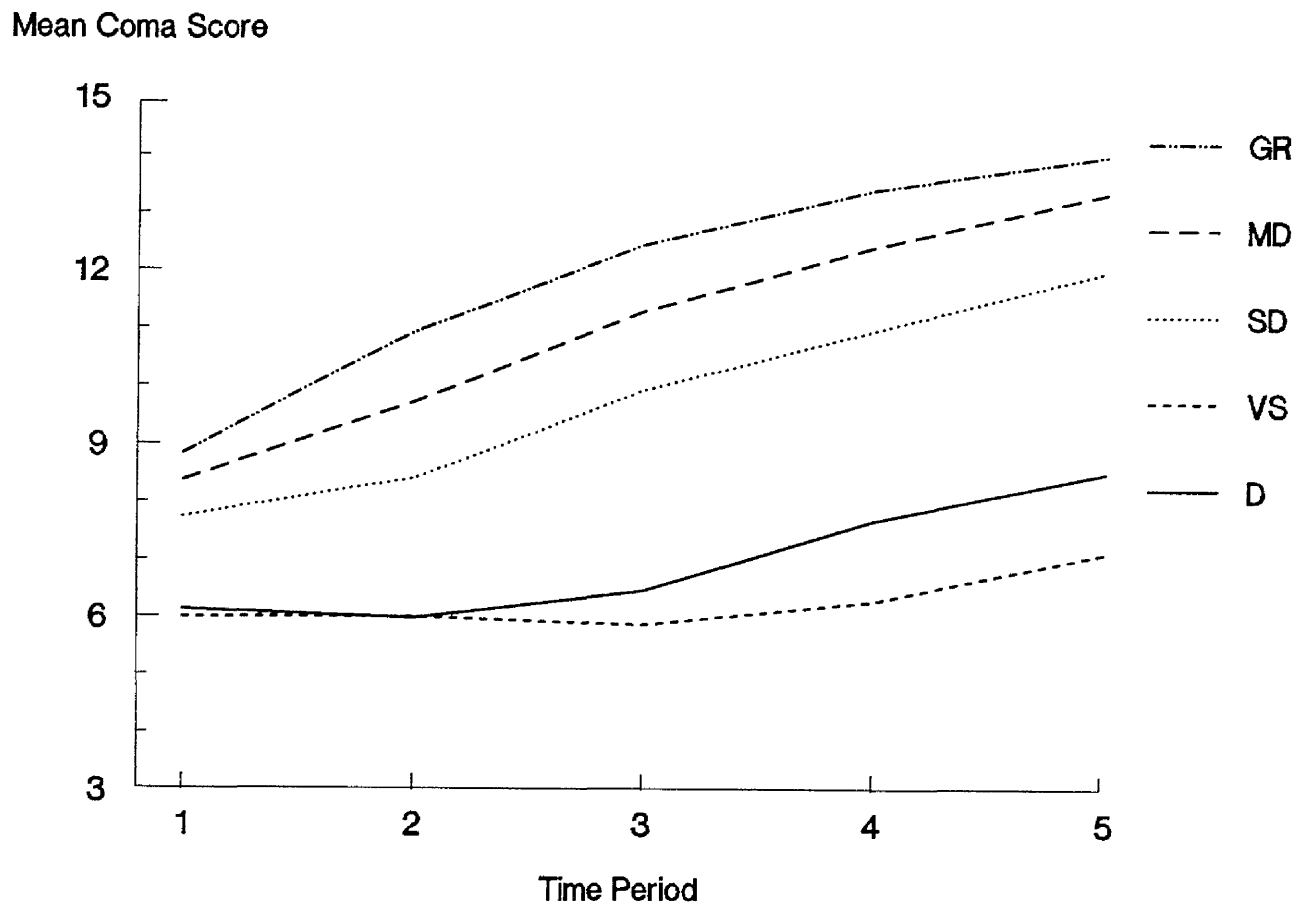


Figure 4.2 Trend in mean coma score for each outcome category

disability and good recovery groups.

4.3 Approaches to the Analysis of Repeated Measures Data

4.3.1 Introduction

The data displayed in Figures 4.1(a) - (e) have a typical repeated measures structure where the same variable is measured C times on each of the N individuals in the study. Thus the extensive literature on the analysis of repeated measures data can be exploited. Before proceeding with the analysis of the Head Injury Study data this literature is briefly reviewed.

There are two broad approaches to the analysis of repeated measures data, univariate and multivariate. These are distinguished by the basic unit for analysis. In the univariate approach each measurement of the variable forms the basic unit and is analysed individually whereas in the multivariate approach the vector of C measurements from each of the cases is the basic unit analysed. Overviews of the analysis of repeated measures data are given by Frane (1980), Davidson (1983) and Fleiss (1986, Chapter 8).

4.3.2 Univariate Approach

The univariate approach to repeated measures data essentially consists of performing a mixed model analysis of variance. The effect due to the individual is modelled as a random effect and usually any grouping variables and trial factors are modelled as fixed effects. This model imposes a strong circularity condition on the covariance structure of the vector of measurements from each individual, namely that all possible pairwise differences of measurements must have the same variance. When, as in the Head Injury Study, the repeated measurements are taken through time,

this assumption is almost certain to be untrue as the difference between two well separated measurements would be expected to be more variable than that between two successive ones.

A number of techniques have been developed which allow conservative hypothesis tests to be performed when the circularity conditions do not hold (Greenhouse and Geisser, 1959; Huynh and Feldt, 1976). However, these are not relevant to the Head Injury Study as these techniques adjust the degrees of freedom to compensate for the lack of fit of the underlying models, whereas in the Head Injury Study, the models themselves are of interest and not simply the derived test statistic.

4.3.3 Multivariate Approach

As mentioned in Section 4.3.1, the main feature of this approach is that the basic unit for analysis is the vector of measurements through time. This approach consists not of a single technique but of a spectrum of methods ranging from the restrictive univariate mixed model analysis of variance to the completely general multivariate model which imposes no structure on the mean vector or covariance matrix. Often the full multivariate model is too general, in that it involves a large number of parameters, and does not explicitly take account of the fact that the data are recorded through time. Therefore it is appealing to consider the models which lie between the extremes of the over restrictive univariate model and the full multivariate model.

One such approach is to summarise the data vector by a single number, for example, 'the area under the curve', or at least by a vector with fewer components than the original measurements. This includes the various growth curve models, where parametric models are fitted to each individual's data vector and the parameter

estimates from such an analysis are used in place of the original data. For example, a large number of measurements through time may be replaced by the gradient and intercept of the regression line through these points.

Another way to reduce the generality of the full multivariate model is to regard the data for each individual as a time series. There are many examples of such an approach in the medical literature, but in general they require a much longer series of observations (Smith and West, 1983). However a paper by Ulm (1984) described a way of parameterising the mean vector and covariance matrix in the multivariate approach by using a standard time series model on a data set with a short time series. This approach seemed to be particularly relevant and its application to the Head Injury Study data, and the modifications which were later found to be necessary, are described in Sections 4.4 - 4.6.

4.4 First Order Autoregressive Stochastic Model

4.4.1 Introduction

Ulm (1984) described a model for the classification of an individual into one of two disease categories on the basis of an enzyme level which was monitored at intervals over a period of time. The feature vector $x = \{x_1, \ldots, x_N\}$ thus consists of the enzyme level at times 1, ..., N. In our application the feature vector consists of the best coma score at each of the five time periods.

Ulm then models the feature vector using an autoregressive approach relating the component at time t to the component at time t-1 where $t = 2, \ldots, N$.

4.4.2 Derivation of the Model

Suppose that the distribution of feature vectors $X = \{X_1, \ldots, X_t\}$ of patients in category Π_i is $N(\mu_i, \Sigma_i)$ and that the coma score at time t, where $t = 2, \ldots, 5$, depends on the coma score at time t-1 in the following way

$$\alpha - X_t = \phi(\alpha - X_{t-1}) + \varepsilon_t \qquad (4.1)$$

where $E(\varepsilon_t) = 0$

and $0 < \phi < 1$.

Thus

$$E(\alpha - X_t) = E\phi(\alpha - X_{t-1}).$$

Applying this recursively gives

$$E(\alpha - X_t) = E\phi^{t-1}(\alpha - X_1).$$

If

$$E(X_1) = \alpha - \beta,$$

then

$$E(X_t) = \alpha - \phi^{t-1}\beta.$$

Thus

$$\mu_i = E\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{bmatrix} = \begin{bmatrix} \alpha - \beta \\ \alpha - \phi\beta \\ \alpha - \phi^2\beta \\ \alpha - \phi^3\beta \\ \alpha - \phi^4\beta \end{bmatrix}.$$

This is illustrated in Figure 4.3.

The usual assumptions made about the error term ε_t with this type of model are:-

(i) $E(\varepsilon_t) = 0$

(ii) $var(\varepsilon_t) = \sigma^2$ (constant)

(iii) $cov(\varepsilon_t, \varepsilon_{t'}) = 0, \quad t \neq t'.$

Wegman (1974) shows that the variance

Figure 4.3 Illustration of parameters in the autoregressive stochastic model

covariance structure of ε can still be derived if assumption (ii) is modified to:-

(ii) $\text{var}(\varepsilon_t) = \sigma_t$ (depending on t).

This means $\{\varepsilon_t\}$ is an uncorrelated sequence of random variables with expectation 0 and variance depending on t.

To derive the variance covariance structure, let

$$Y_t = \alpha - X_t.$$

Thus from Equation 4.1

$$Y_t = \phi Y_{t-1} + \varepsilon_t. \tag{4.2}$$

Multiplying both sides by Y_{t-s} gives

$$Y_t\, Y_{t-s} = \phi Y_{t-1}\, Y_{t-s} + \varepsilon_t\, Y_{t-s}.$$

Thus

$$E(Y_t\, Y_{t-s}) = \phi E(Y_{t-1}\, Y_{t-s}). \tag{4.3}$$

From Equation 4.2

$$E(Y_t) = \phi E(Y_{t-1}) \tag{4.4}$$

so

$$E(Y_t\, Y_{t-s}) - E(Y_t)E(Y_{t-s}) = \phi(E(Y_{t-1}\, Y_{t-s}) - E(Y_{t-1})E(Y_{t-s})).$$

Thus

$$\text{cov}(Y_t, Y_{t-s}) = \phi \text{cov}(Y_{t-1}, Y_{t-s}). \tag{4.5}$$

Applying Equation 4.5 recursively gives

$$\text{cov}(Y_t, Y_{t-s}) = \phi^s \text{var}(Y_{t-s}).$$

For the variances, from Equations 4.2 and 4.4,

$$Y_t - E(Y_t) = \phi Y_{t-1} + \varepsilon_t - \phi E(Y_{t-1}).$$

Thus

$$Y_t - E(Y_t) = \phi(Y_{t-1} - E(Y_{t-1})) + \varepsilon_t. \tag{4.6}$$

Applying Equation 4.6 recursively gives

$$Y_t - E(Y_t) = \sum_{j=1}^{t} \phi^{(t-j)}\, \varepsilon_j.$$

Thus

$$E(Y_t - E(Y_t))^2 = E\left[\sum_{j=1}^{t} \phi^{(t-j)}\, \varepsilon_j\right]^2.$$

Since, from the assumptions,

$$\text{var } \varepsilon_t = E(\varepsilon_t - E(\varepsilon_t))^2 = E(\varepsilon_t^2)$$

and

$$\text{cov}(\varepsilon_t, \varepsilon_{t'}) = E(\varepsilon_t - E(\varepsilon_t))(\varepsilon_{t'} - E(\varepsilon_{t'})) = 0,$$

$$E(Y_t - E(Y_t))^2 = \sum_{j=1}^{t} \phi^{2(t-j)} \text{var } \varepsilon_j.$$

Thus

$$\text{var } Y_t = \sum_{j=1}^{t} \phi^{2(t-j)} \sigma_j^2.$$

Since

$$X_t - E(X_t) = (X_t - \alpha) - (E(X_t) - \alpha)$$

$$= - (Y_t - E(Y_t)),$$

and similarly

$$X_{t-s} - E(X_{t-s}) = - (Y_{t-s} - E(Y_{t-s})),$$

$$\text{var } Y_t = \text{var } X_t$$

and

$$\text{cov } (Y_t, Y_{t-s}) = \text{cov}(X_t, X_{t-s}).$$

Thus

$$\text{var } X_t = \sum_{j=1}^{t} \phi^{2(t-j)} \sigma_j^2$$

and

$$\text{cov } (X_t, X_{t-s}) = \phi^s \text{ var } (X_{t-s}).$$

Thus

$$\text{var } X_1 = \sigma_1^2$$

$$\text{var } X_2 = \phi^2\sigma_1^2 + \sigma_2^2$$

$$\text{var } X_3 = \phi^4\sigma_1^2 + \phi^2\sigma_2^2 + \sigma_3^2$$

$$\text{var } X_4 = \phi^6\sigma_1^2 + \phi^4\sigma_2^2 + \phi^2\sigma_3^2 + \sigma_4^2$$

$$\text{var } X_5 = \phi^8\sigma_1^2 + \phi^6\sigma_2^2 + \phi^4\sigma_3^2 + \phi^2\sigma_4^2 + \sigma_5^2$$

and

$$\Sigma_i = \begin{bmatrix} \text{var}X_1 & \phi\text{var}X_1 & \phi^2\text{var}X_1 & \phi^3\text{var}X_1 & \phi^4\text{var}X_1 \\ \phi\text{var}X_1 & \text{var}X_2 & \phi\text{var}X_2 & \phi^2\text{var}X_2 & \phi^3\text{var}X_2 \\ \phi^2\text{var}X_1 & \phi\text{var}X_2 & \text{var}X_3 & \phi\text{var}X_3 & \phi^2\text{var}X_3 \\ \phi^3\text{var}X_1 & \phi^2\text{var}X_2 & \phi\text{var}X_3 & \text{var}X_4 & \phi\text{var}X_4 \\ \phi^4\text{var}X_1 & \phi^3\text{var}X_2 & \phi^2\text{var}X_3 & \phi\text{var}X_4 & \text{var}X_5 \end{bmatrix}.$$

Thus for the estimation of μ_i and Σ_i it is sufficient to estimate α, β and ϕ and the variances of the X_t where $t = 1, \ldots, 5$, a reduction from 20 parameters to 8 for the case where all 5 time periods are considered.

4.4.3 Parameter Estimation

A maximum likelihood approach was used to estimate the parameters for the mean and covariance matrix of each of the three outcome categories severe disability, moderate disability and good recovery.

Let x_{ijt} be the best coma score of the j^{th} patient in the i^{th} outcome group at the t^{th} time, so that x_{ij}, where $j = 1, \ldots, n_i$, are the feature vectors of the patients from category Π_i then, since the distribution of these feature vectors is $N(\mu_i, \Sigma_i)$ with μ_i and Σ_i constrained as specified in Section 4.4.2, the likelihood function is

$$L_i = \frac{1}{(2\Pi)^{\frac{1}{2}Cn_i}\left|\Sigma_i\right|^{\frac{1}{2}n_i}} \exp\left[-\frac{1}{2}\sum_{j=1}^{n_i}(x_{ij} - \mu_i)^T \Sigma_i^{-1}(x_{ij} - \mu_i)\right],$$

where C is the no. of time periods considered and $n_i > C$.

Since the exponent is written in terms of Σ_i^{-1}, the maximum likelihood estimates of μ_i and $\Sigma_i^{-1} = \psi_i$ say, were found. In the likelihood function the vectors x_{ij}, $j=1, \ldots, n_i$ are fixed at the sample values and L_i is a function of μ_i and ψ_i. Then the

logarithm of the likelihood function is

$$\log L_i = -\frac{1}{2} Cn_i \log(2\Pi) + \frac{1}{2} n_i \log|\psi_i| - \frac{1}{2} \sum_{j=1}^{n_i} (x_{ij} - \mu_i)^T \psi_i (x_{ij} - \mu_i).$$

Since $\log L_i$ is an increasing function of L_i, its maximum occurs at the same point in the space of μ_i, ψ_i as the maximum of L_i.

If $\bar{x}_i$ is the sample mean vector then

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} = \begin{bmatrix} \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij1} \\ \vdots \\ \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ijC} \end{bmatrix} = \begin{bmatrix} \bar{x}_{i.1} \\ \vdots \\ \bar{x}_{i.C} \end{bmatrix}$$

and the matrix of sums of squares and cross products about the mean is

$$A_i = \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)^T$$

so that

$$(A_i)_{mn} = \sum_{j=1}^{n_i} (x_{ijm} - \bar{x}_{i.m})(x_{ijn} - \bar{x}_{i.n})$$

where m,n = 1, ..., C.

Now

$$\sum_{j=1}^{n_i} (x_{ij} - \mu_i)(x_{ij} - \mu_i)^T$$

can be written as

$$\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)^T + n_i(\bar{x}_i - \mu_i)(\bar{x}_i - \mu_i)^T$$

$$= A_i + n_i(\bar{x}_i - \mu_i)(\bar{x}_i - \mu_i)^T$$

and so

$$\sum_{j=1}^{n_i} (x_{ij} - \mu_i)^T \psi_i (x_{ij} - \mu_i)$$

$$= \text{tr} \sum_{j=1}^{n_i} (x_{ij} - \mu_i)^T \psi_i (x_{ij} - \mu_i)$$

$$= \text{tr} \sum_{j=1}^{n_i} \psi_i (x_{ij} - \mu_i)(x_{ij} - \mu_i)^T$$

$$= \text{tr}\ \psi_i A_i + \text{tr}\ \psi_i n_i (\bar{x}_i - \mu_i)(\bar{x}_i - \mu_i)^T$$

$$= \text{tr}\ \psi_i A_i + n_i(\bar{x}_i - \mu_i)^T \psi_i(\bar{x}_i - \mu_i).$$

Thus $\log L_i$ can be written as

$$\log L_i = -\frac{1}{2} Cn_i \log(2\Pi) + \frac{1}{2} n_i \log|\psi_i| - \frac{1}{2} \text{tr}\ \psi_i A_i$$

$$- \frac{1}{2} n_i(\bar{x}_i - \mu_i)^T \psi_i (\bar{x}_i - \mu_i). \qquad (4.7)$$

This function was maximised numerically with respect to α, β, ϕ and the five variances, using the NAG routine E04JBF. Starting values were 15.0, 12.0, 0.8 and 5.0 for α, β, ϕ and the five variances respectively.

To avoid the problem of missing data, the model was initially fitted with cases who had a best coma score recorded at all 5 time periods. As the coma score has a maximum of 15, this was taken as the upper bound of α in each case. The parameter estimates for each of the three outcome categories are given in Table 4.3.

From these, estimates of $\hat{\mu}_i$ and $\hat{\Sigma}_i$ were calculated and these are given in Table 4.4. For comparison, the sample means and covariance matrices for the same data used to fit the model are given in Table 4.5 and the means from the fitted model and the data are plotted together in Figure 4.4.

As the estimate for α took the value of the upper bound set, it

Table 4.3 Parameter estimates for the first order autoregressive stochastic model

Parameter	Severe Disability	Moderate Disability	Good Recovery
α	15.00	15.00	15.00
β	7.76	7.06	6.80
ϕ	0.81	0.71	0.64
σ_1^2	4.18	6.35	5.85
σ_2^2	4.13	6.01	5.82
σ_3^2	3.74	3.94	3.62
σ_4^2	3.01	3.95	2.21
σ_5^2	3.14	2.46	2.13

Table 4.4 Estimates for the mean vector and covariance matrix for each outcome for the first order autoregressive stochastic model

Outcome Π_i	Mean Vector μ_i	Covariance Matrix Σ_i				
Severe Disability	7.24	4.18	3.39	2.75	2.23	1.81
	8.71	3.39	6.87	5.57	4.52	3.66
	9.90	2.75	5.57	8.26	6.69	5.43
	10.86	2.23	4.52	6.69	8.43	6.83
	11.65	1.81	3.66	5.43	6.83	8.61
Moderate Disability	7.94	6.63	4.51	3.20	2.27	1.61
	9.99	4.51	9.21	6.54	4.64	3.30
	11.44	3.20	6.54	8.59	6.10	4.33
	12.47	2.27	4.64	6.10	8.28	5.88
	13.20	1.61	3.30	4.33	5.88	6.63
Good Recovery	8.20	5.85	3.75	2.41	1.54	0.99
	10.64	3.75	8.23	5.28	3.38	2.17
	12.20	2.41	5.28	7.01	4.50	2.88
	13.21	1.54	3.38	4.50	5.09	3.26
	13.85	0.99	2.17	2.88	3.26	4.23

Table 4.5 Sample mean vector and covariance matrix for each outcome from the data used to fit the first order autoregressive stochastic model

Outcome Π_i	Sample Mean Vector μ_i	Sample Covariance Matrix Σ_i				
Severe Disability	7.24	4.18	3.57	2.92	2.19	1.97
	7.97	3.57	6.63	5.07	4.27	3.46
	9.11	2.92	5.07	7.56	5.92	4.20
	10.41	2.19	4.27	5.92	7.61	5.17
	11.94	1.97	3.46	4.20	5.17	6.09
Moderate Disability	7.94	6.35	4.28	3.12	1.36	0.41
	8.84	4.28	7.58	6.06	3.38	1.44
	10.39	3.12	6.06	8.67	5.18	2.10
	11.81	1.36	3.38	5.18	6.93	3.28
	13.39	0.41	1.44	2.10	3.28	3.20
Good Recovery	8.20	5.85	3.96	3.76	2.46	0.82
	9.93	3.96	8.00	6.25	3.92	0.98
	11.34	3.76	6.25	8.18	4.81	1.39
	12.78	2.46	3.92	4.81	5.00	1.52
	14.10	0.82	0.98	1.39	1.52	1.76

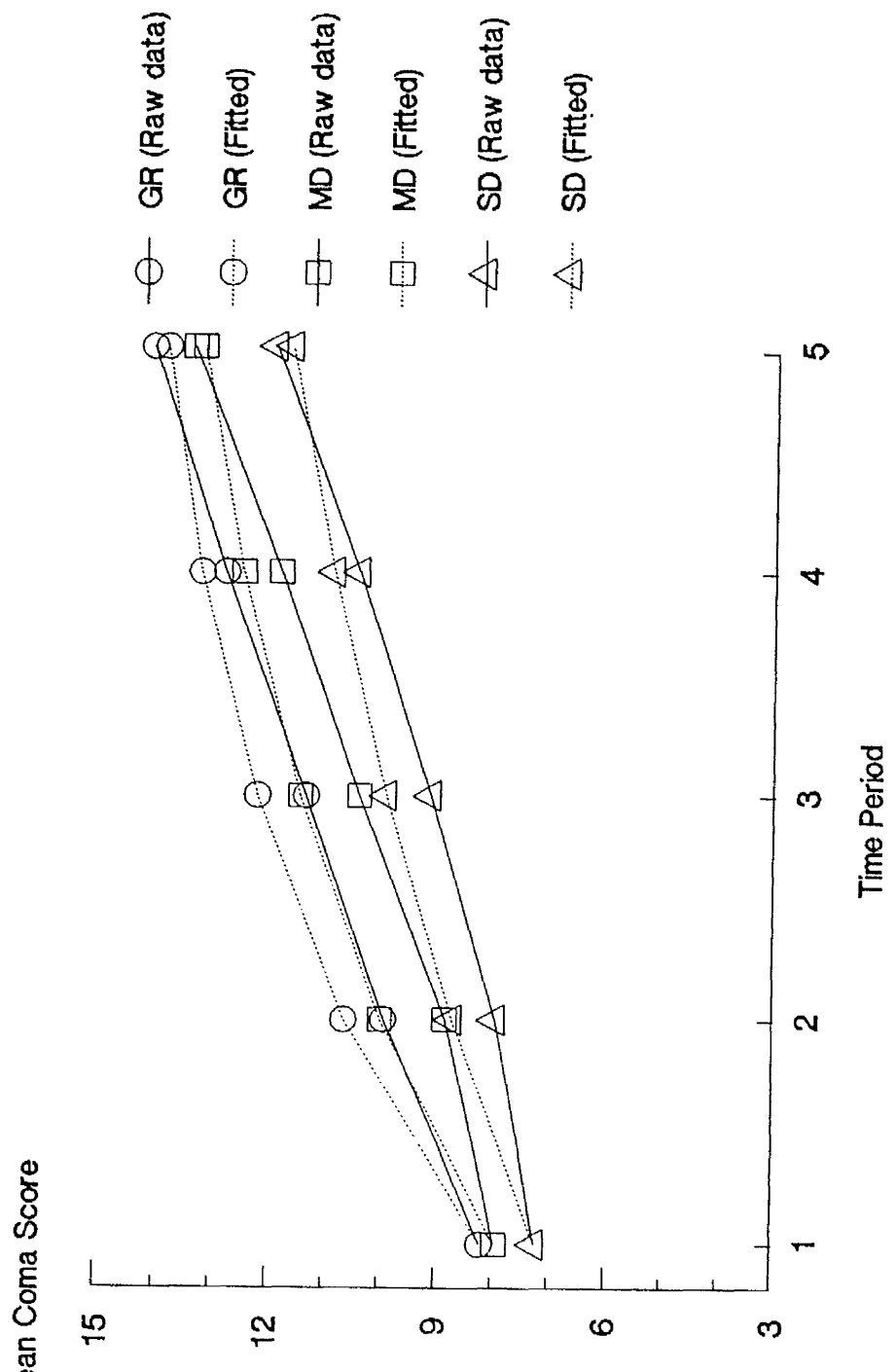


Figure 4.4 Mean coma score from the fitted autoregressive stochastic model (α bounded) and the corresponding raw data

was decided to increase the upper bound on the grounds that the cases were tending in the long run to some value over 15. The parameter estimates in this case are given in Table 4.6 and $\hat{\mu}_i$ and $\hat{\Sigma}_i$ from these estimates in Table 4.7, the means from the fitted model and the data are plotted in Figure 4.5.

4.4.4 Fit of Model to Data

The results in Figures 4.4 and 4.5 suggest that the model is a poor fit to the data. However, this was tested formally using an asymptotic likelihood ratio test.

Suppose under a more general model the distribution of the feature vectors of category Π_i is $N(\mu_i,\Sigma_i)$ with no constraints on μ_i and Σ_i. The maximised likelihood for outcome Π_i, L_i, is obtained from

$$\text{Log } L_i = \frac{1}{2} C\, n_i \log(2\Pi) + \frac{1}{2} n_i \log \left|\psi_i\right| - \frac{1}{2} \text{tr } \psi_i A_i$$

$$- \frac{1}{2} n_i (\bar{x}_i - \mu_i)^T \psi_i (\bar{x}_i - \mu_i).$$

Since ψ_i is positive semi-definite,

$$n_i(\bar{x}_i - \mu_i)^T \psi_i (\bar{x}_i - \mu_i) \geq 0$$

and is 0 if

$$\mu_i = \bar{x}_i.$$

Anderson (1984b, pp 62-63) shows that

$$\frac{1}{2} n_i \log \left|\psi_i\right| - \frac{1}{2} \text{tr } \psi_i A_i$$

is a maximum when

$$\psi_i = n_i A_i^{-1}$$

and takes the value

$$\frac{1}{2} n_i \log \left|n_i A_i\right| - \frac{1}{2} n_i C.$$

The maximised logarithm of the likelihood for outcome Π_i under

Table 4.6 Parameter estimates for the first order autoregressive stochastic model with the upper bound of 15 for α removed

Parameter	Severe Disability	Moderate Disability	Good Recovery
α	16.27	15.40	15.39
β	9.03	7.46	7.19
ϕ	0.84	0.73	0.66
σ_1^2	4.18	6.35	5.85
σ_2^2	4.11	6.01	5.80
σ_3^2	3.77	3.93	3.60
σ_4^2	3.01	3.97	2.22
σ_5^2	3.10	2.44	2.13

Table 4.7 Estimates of the mean vector and covariance matrix for each outcome for the first order autoregressive stochastic model with the upper bound of 15 for α removed

Outcome Π_i	Mean Vector μ_i	Covariance Matrix Σ_i				
Severe Disability	7.24	4.18	3.50	2.94	2.46	2.06
	8.71	3.50	7.05	5.90	4.95	4.14
	9.93	2.94	5.90	8.71	7.30	6.12
	10.96	2.46	4.95	7.30	9.13	7.65
	11.82	2.06	4.14	6.12	7.65	9.51
Moderate Disability	7.94	6.63	4.61	3.35	2.43	1.76
	9.98	4.61	9.36	6.79	4.93	3.58
	11.47	3.35	6.79	8.86	6.43	4.67
	12.55	2.43	4.93	6.43	8.65	6.28
	13.33	1.76	3.58	4.67	6.28	7.00
Good Recovery	8.20	5.85	3.88	2.57	1.70	1.13
	10.62	3.88	8.37	5.55	3.68	2.44
	12.23	2.57	5.55	7.27	4.82	3.20
	13.30	1.70	3.68	4.82	5.41	3.59
	14.00	1.13	2.44	3.20	3.59	4.51

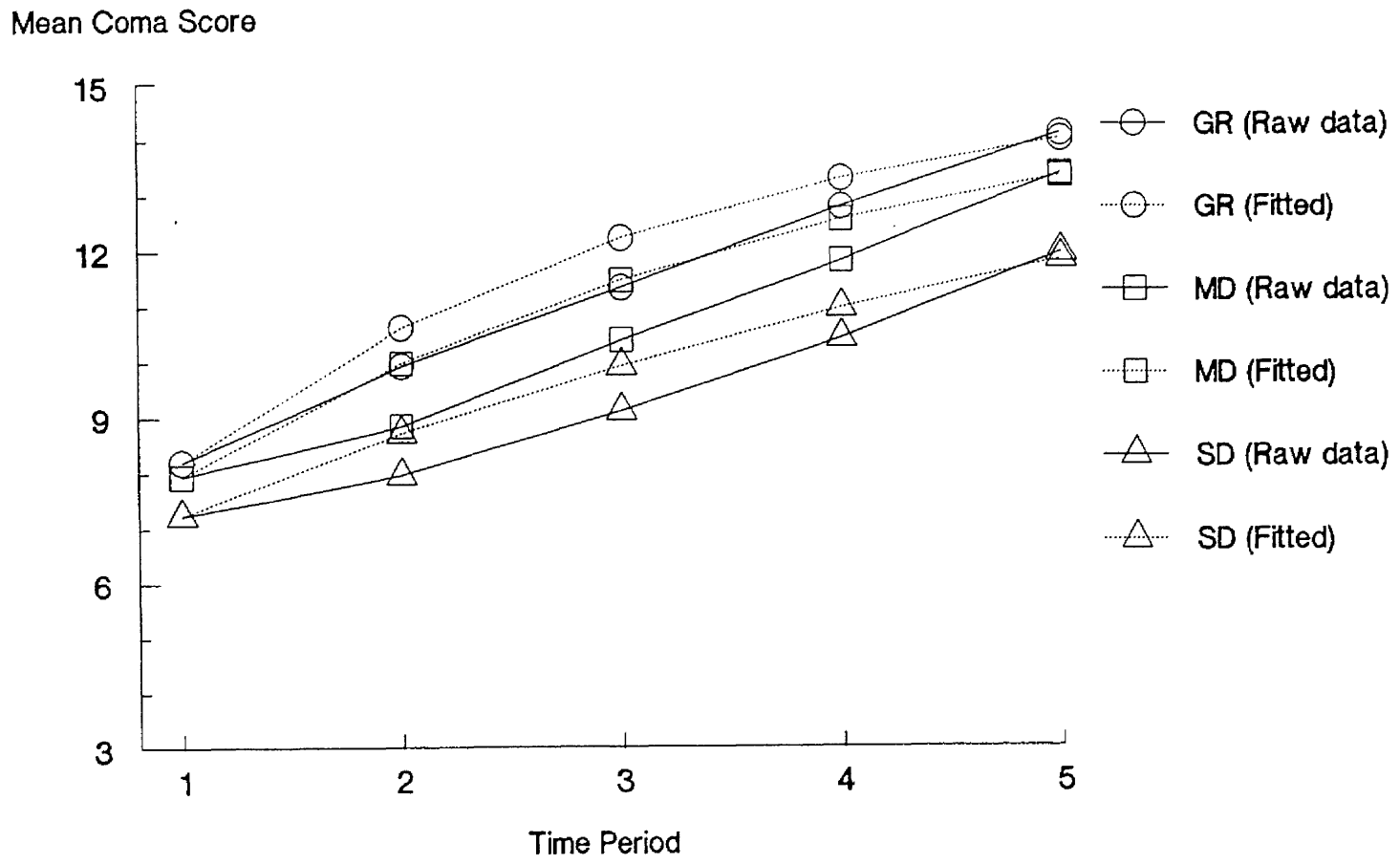


Figure 4.5 Mean coma score from the fitted autoregressive stochastic model (α unbounded) and the corresponding raw data

the general model with no constraints on μ_i and Σ_i is thus

$$\log L_i = -\frac{1}{2} Cn_i \log(2\Pi) + \frac{1}{2} n_i \log\left|n_i A_i\right| - \frac{1}{2} n_i C.$$

The results for the approximate likelihood ratio test for each outcome are given in Table 4.8 for both α bounded and unbounded. Constraining the mean and variance reduces the number of parameters to be estimated in this case from 20 to 8 so that under the null hypotheses the test statistic will be distributed approximately as $\chi^2(12)$. The rejection region at a significance level of 5% for the constrained model corresponds to values of the test statistic greater than 21.03 so the model was rejected in each case, thus endorsing the subjective impression of Figures 4.4 and 4.5.

4.4.5 Discussion

The fit of this model to the data was disappointing as the model was chosen initially from the pattern of recovery in Figure 4.2. On a re-examination of the approach it was found that, to avoid the problem of missing data, only cases with complete data at all five times had been used in fitting the model, while Figure 4.2 was based on all data available at each time. The main source of missing data is due to the fact that patients are discharged from the unit before the end of the monitoring period. One therefore might expect these cases to have, on average, higher coma scores than those who remain in hospital. This is indeed the case and Table 4.9 illustrates this fact.

It therefore seemed appropriate to use a different approach to constrain the general model.

Table 4.8 Test statistics for the approximate likelihood ratio test for each outcome for both α bounded and unbounded

Outcome	Test Statistic	
	α bounded	α unbounded
Severe Disability	26.07	25.53
Moderate Disability	62.13	61.81
Good Recovery	68.64	67.83

Table 4.9 Mean best coma score at 8 - 14 days for cases with complete data at all times and for cases with only the 15 - 28 day score missing

Outcome	Mean Coma Score	
	Data Complete	Data Missing at 15-28 days
Severe Disability	10.41	11.95
Moderate Disability	11.81	13.31
Good Recovery	12.78	13.99

4.5 Constrained Linear Model

4.5.1 Introduction

After the poor fit of the model described in Section 4.4 it was clearly necessary to employ a more appropriate data set when fitting a model. It was decided that if the model was being fitted at the end of the 15-28 day period then only cases with complete data at all times should be used, while if the model was being fitted at the end of the 4-7 day period then all cases with complete data at that time should be used, and so on. This was clinically and statistically acceptable in that it was not possible to have future information regarding the patient when making a prediction about prognosis. To simplify this description let D_C represent the data set of cases with a 6 month outcome of severe disability or better, who have best coma score available for all time periods up to and including time C (C = 1, ..., 5). Thus D_5 is the set of patients with complete data at all five times whose six month outcome is at least severe disability.

The mean best coma score at each time for each of the data sets D_3, ..., D_5 is given in Table 4.10 and plotted in Figure 4.6.

4.5.2 Derivation of the Model

Figure 4.6 suggested that a suitable model might take into account the following features:-

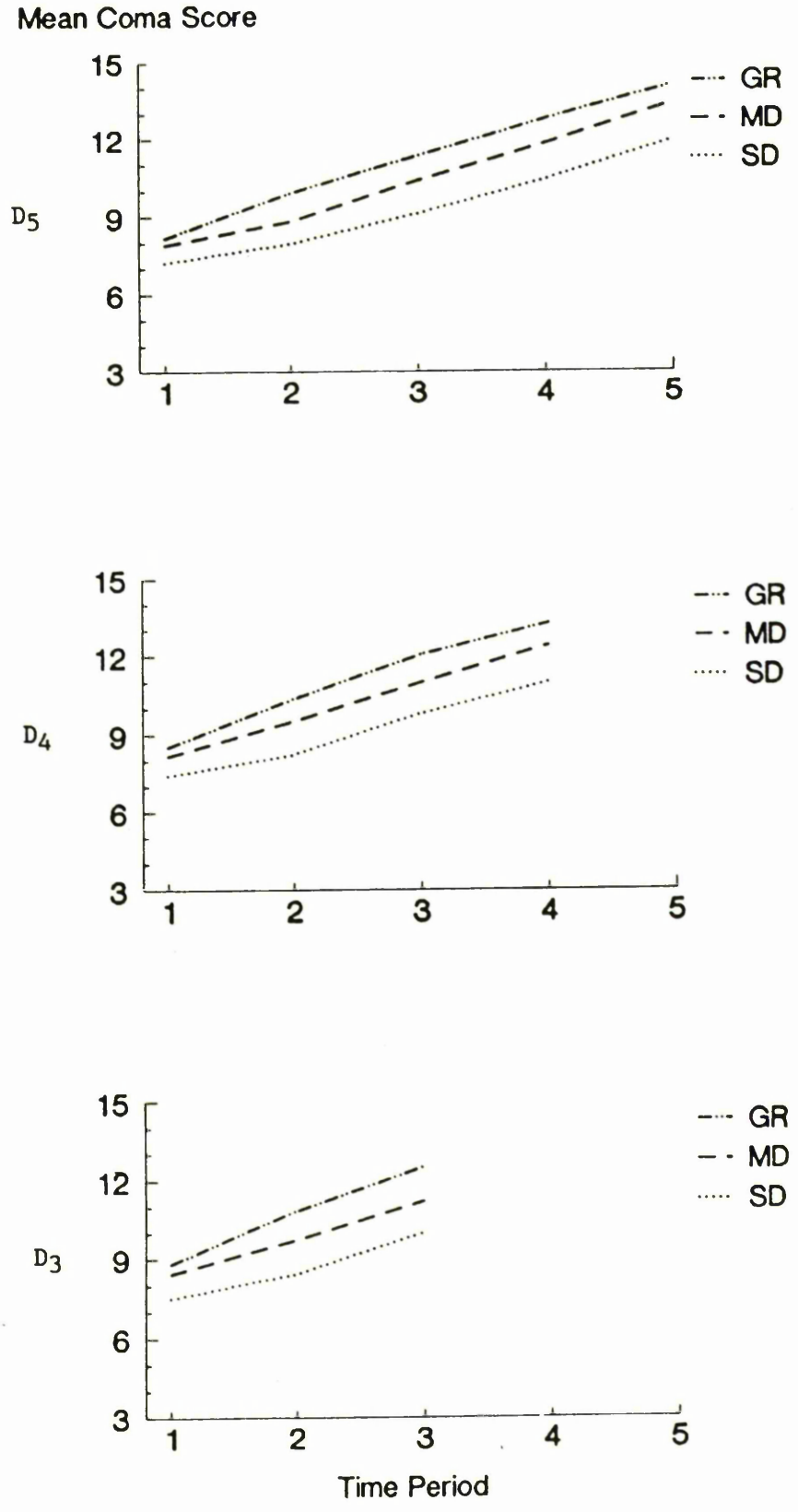
(i) The mean best coma scores for a given outcome category Π_i, i = 1,2,3, at a particular time period t, t = 1, ..., C, are not all equal for all data sets D_C, C = 1, ..., 5, for which they exist.

(ii) Within each data set D_C, C = 3,4,5, for any particular outcome category, the mean best scores at the different

Table 4.10 Mean best coma score at each time period for the outcome categories severe disability, moderate disability and good recovery and the data sets D_5, D_4 and D_3

Time Period	Data Set Number of Cases								
	D_5 236			D_4 406			D_3 528		
	SD 63	MD 83	GR 90	SD 103	MD 138	GR 165	SD 124	MD 168	GR 236
24 hours	7.24	7.94	8.20	7.44	8.20	8.55	7.54	8.46	8.84
2-3 days	7.97	8.84	9.93	8.26	9.54	10.39	8.46	9.76	10.87
4-7 days	9.11	10.39	11.34	9.82	11.02	12.06	10.04	11.24	12.52
8-14 days	10.41	11.81	12.78	11.01	12.41	13.28			
15-28 days	11.94	13.39	14.10						

Figure 4.6 Mean best coma score at each time period for each outcome with data sets D_5, D_4 and D_3



times, apart from that at 24 hours, are related in a simple linear manner.

(iii) Within each data set D_C, C = 3,4,5, the lines defined in (ii) for each outcome category are parallel.

From this the following model was proposed for data sets, D_C, C = 3,4,5.

Let x_{ijt} represent, as before, the best coma score of the jth patient in the ith category at time t, so that x_{ij}, $j = 1, \ldots, n_i$, are the feature vectors of the patients from category Π_i. Then if μ_{it} is mean score for category Π_i at time t, t = 3, ..., C

$$\mu_{it} = \mu_{it-1} + \delta \qquad \text{where } \delta = \text{constant.}$$

Thus, for each data set D_C, C = 3,4,5, the means can be specified using only 7 parameters

$$\mu_{11}, \mu_{21}, \mu_{31}, \mu_{12}, \gamma_{12}, \gamma_{13} \text{ and } \delta$$

where $\gamma_{12} = \mu_{22} - \mu_{12}$ and $\gamma_{13} = \mu_{32} - \mu_{12}$.
This is illustrated in Figure 4.7.

4.5.3 Parameter Estimation

If the x_{ij} ($j = 1, \ldots, n_i$; $i = 1,2,3$) are assumed to be normally distributed with mean $\mu_i = (\mu_{ii}, \ldots, \mu_{ic})^T$, where C = 3,4,5, and with μ_i structured as specified in section 4.5.2 and common covariance matrix, Σ, then a maximum likelihood approach can again be used to estimate the parameters. The likelihood function, L, is

$$L = \frac{1}{(2\Pi)^{\frac{1}{2}CN} \left|\Sigma\right|^{\frac{1}{2}N}} \exp\left[-\frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{n_i}(x_{ij}-\mu_i)^T \Sigma^{-1} (x_{ij}-\mu_i)\right],$$

where $N = n_1 + n_2 + n_3$

and so the logarithm of the likelihood function, log L, is

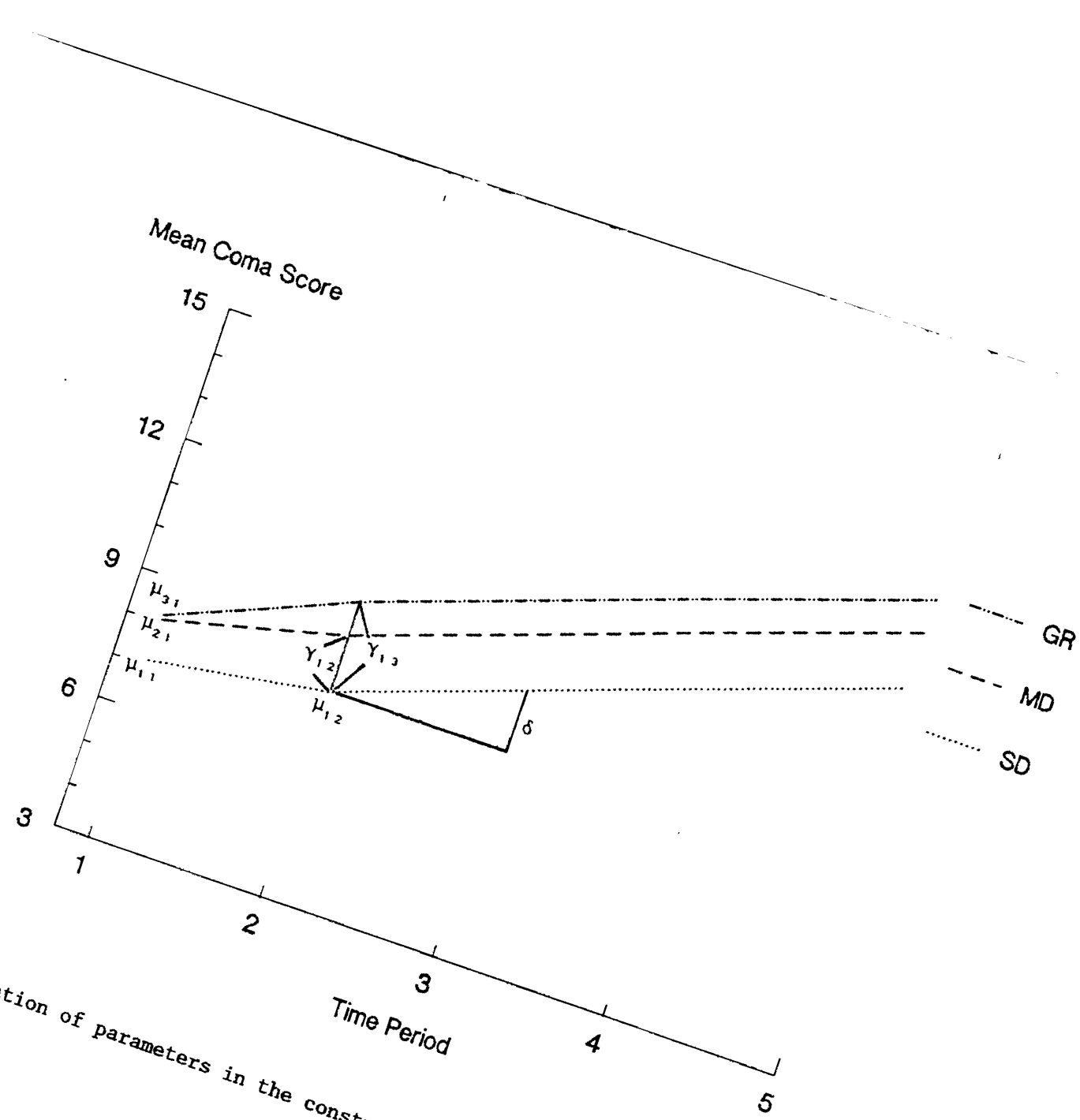


Figure 4.7 Illustration of parameters in the constrained normal model

$$\log L = -\frac{CN}{2}\log(2\Pi) - \frac{N}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{n_i}(x_{ij} - \mu_i)^T \Sigma^{-1} (x_{ij}-\mu_i).$$

If

$$\theta = (\mu_{11}, \mu_{21}, \mu_{31}, \mu_{12}, \gamma_{12}, \gamma_{13}, \delta)^T$$

then the constraint on μ_i is such that

$$\mu_i = M_i\theta, \qquad i=1, 2, 3$$

where

$$M_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 1 & 0 & 0 & C-2 \end{bmatrix}$$

$$M_2 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 1 & 1 & 0 & C-2 \end{bmatrix}$$

$$M_3 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 1 & 0 & 1 & C-2 \end{bmatrix}, \quad C = 3, 4, 5.$$

Thus

$$\sum_{i=1}^{3}\sum_{j=1}^{n_i}(x_{ij} - \mu_i)^T \Sigma^{-1} (x_{ij} - \mu_i)$$

$$= \sum_{i=1}^{3}\sum_{j=1}^{n_i}(x_{ij} - M_i\theta)^T \Sigma^{-1} (x_{ij} - M_i\theta)$$

$$= \sum_{i=1}^{3}\sum_{j=1}^{n_i}(x_{ij}^T\Sigma^{-1}x_{ij}^T - x_{ij}^T\Sigma^{-1}M_i\theta - \theta^TM_i^T\Sigma^{-1}x_{ij} + \theta^TM_i^T\Sigma^{-1}M_i\theta)$$

$$= \sum_{i=1}^{3}\left(\sum_{j=1}^{n_i}x_{ij}^T\Sigma^{-1}x_{ij} - n_i\bar{x}_i^T\Sigma^{-1}M_i\theta - n_i\theta^TM_i^T\Sigma^{-1}\bar{x}_i + n_i\theta^TM_i^T\Sigma^{-1}M_i\theta\right)$$

where $$\bar{x}_i = \frac{\sum_{j=1}^{n_i}x_{ij}}{n_i}$$

$$= \sum_{i=1}^{3} \sum_{j=1}^{n_i} x_{ij}^T \Sigma^{-1} x_{ij} - B^T\theta - \theta^T B + \theta^T A\theta$$

where $A = \sum_{i=1}^{3} n_i M_i^T \Sigma^{-1} M_i$

and $B = \sum_{i=1}^{3} n_i M_i^T \Sigma^{-1} \bar{x}_i$

$$= (\theta - A^{-1}B)^T A(\theta - A^{-1}B) + \sum_{i=1}^{3} \sum_{j=1}^{n_i} x_{ij}\Sigma^{-1}x_{ij} - B^T A^{-1}B.$$

Thus

$$\log L = -\frac{CN}{2}\log(2\Pi) - \frac{N}{2}\log\left|\Sigma\right| - \frac{1}{2}(\theta - A^{-1}B)^T A(\theta - A^{-1}B)$$

$$- \frac{1}{2}\left(\sum_{i=1}^{3} \sum_{j=1}^{n_i} x_{ij}^T \Sigma^{-1} x_{ij} - B^T A^{-1} B\right).$$

Since

$$-(\theta - A^{-1}B)^T A(\theta - A^{-1}B)$$

is a maximum (zero) when

$$\theta = A^{-1}B,$$

to find the maximum of the logarithm of the likelihood, it remains to find the value of Σ such that

$$-\frac{N}{2}\log\left|\Sigma\right| - \frac{1}{2}\left(\sum_{i=1}^{3} \sum_{j=1}^{n_i} x_{ij}^T \Sigma^{-1} x_{ij} - B^T A^{-1} B\right)$$

is maximised. That is

$$-\frac{N}{2}\log\left|\Sigma\right| - \frac{1}{2}\operatorname{tr}\left(\Sigma^{-1} \sum_{i=1}^{3} \sum_{j=1}^{n_i} x_{ij} x_{ij}^T - A^{-1}BB^T\right)$$

is maximised.

This was again done numerically using the NAG routine E04JBF. The starting values, Σ_o, taken in this routine were

$$\Sigma_o = [s_{ij}], \quad \text{where } s_{ij} = 8.0 \times 0.8^{|i-j|}.$$

The maximum likelihood estimates $\hat{\Sigma}$, $\hat{\theta}$ and $\hat{\mu}_i$ of Σ, θ and μ_i are given, for data set D_5, in Table 4.11, together with the

correlation matrix corresponding to $\hat{\Sigma}$ and the maximum of the logarithm of the likelihood, log L. The results for data sets D_4 and D_3 are given in Tables 4.12 and 4.13 respectively.

4.5.4 Fit of the Model

A comparison of $\hat{\mu}_i$ in Tables 4.11, 4.12 and 4.13 with the appropriate values in Table 4.10 suggests that the model is a good fit. This was tested formally again using an asymptotic likelihood ratio test. Suppose under a more general model, the distribution of the feature vectors in category Π_i, i = 1, 2, 3 is $N(\mu_i,\Sigma)$ with no constraints on μ_i. Under this model the maximum likelihood estimates $\hat{\mu}_i$ and $\hat{\Sigma}$ are given by

$$\hat{\mu}_i = \frac{\sum_{j=1}^{n_i} x_{ij}}{n_i} = \bar{x}_i$$

and

$$\hat{\Sigma} = \frac{1}{N} \sum_{i=1}^{3} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)^T$$

The maximum likelihood estimates for $\hat{\Sigma}$ (under this more general model) and the maximum of the logarithm of the likelihood are given in Table 4.14 for data sets D_5, D_4 and D_3.

From these the likelihood ratio test statistics were calculated and these are given in Table 4.15 for the three data sets considered.

Thus there was no evidence that the more general model with $x_{ij} \sim N(\mu_i,\Sigma)$ was a significantly better fit than that with structured means where $x_{ij} \sim N(M_i\theta,\Sigma)$ for any of the three data sets considered.

Table 4.11 Maximum likelihood estimates of Σ and θ, the corresponding correlation matrix and μ_i, and the maximimum log-likelihood for the model where $x_{ij} \sim N(M_i\theta,\Sigma)$ using data set D_5

Parameter	Estimate				
Σ	5.59	3.99	3.24	2.00	0.98
	3.99	7.52	5.87	3.82	1.79
	3.24	5.87	8.19	5.25	2.39
	2.00	3.82	5.25	6.39	3.12
	0.98	1.79	2.39	3.12	3.42
Correlation Matrix corresponding to Σ	1.00	0.67	0.48	0.33	0.22
	0.67	1.00	0.75	0.55	0.35
	0.48	0.75	1.00	0.73	0.45
	0.33	0.55	0.73	1.00	0.67
	0.22	0.35	0.45	0.67	1.00
θ	7.13				
	8.06				
	8.17				
	7.76				
	1.31				
	2.10				
	1.42				
	SD		MD		GR
μ_i	7.13		8.06		8.17
	7.76		9.08		9.87
	9.18		10.49		11.28
	10.60		11.91		12.70
	12.01		13.33		14.11
$\log L_{max}$	-247.87				

Table 4.12 Maximum likelihood estimates of Σ and θ, the corresponding correlation matrix and μ_i, and the maximimum log-likelihood for the model where $x_{ij} \sim N(M_i\theta,\Sigma)$ using data set D_4

Parameter	Estimate			
Σ	5.42	3.71	2.93	1.64
	3.71	7.57	5.60	3.48
	2.93	5.60	7.96	4.83
	1.64	3.48	4.83	5.30
Correlation Matrix corresponding to Σ	1.00	0.58	0.45	0.31
	0.58	1.00	0.72	0.55
	0.45	0.72	1.00	0.74
	0.31	0.55	0.74	1.00
θ	7.37			
	8.20			
	8.55			
	8.17			
	1.38			
	2.26			
	1.42			
	SD	MD	GR	
μ_i	7.37	8.20	8.55	
	8.17	9.55	10.43	
	9.59	10.97	11.85	
	11.02	12.39	13.27	
$\log L_{max}$	-435.97			

Table 4.13 Maximum likelihood estimates of Σ and θ, the corresponding correlation matrix and μ_i, and the maximimum log-likelihood for the model where $x_{ij} \sim N(M_i\theta,\Sigma)$ using data set D_3

Parameter	Estimate		
Σ	5.98	4.05	3.09
	4.05	8.04	5.84
	3.09	5.84	7.83
Correlation Matrix corresponding to Σ	1.00	0.58	0.45
	0.58	1.00	0.74
	0.45	0.74	1.00
θ	7.54		
	8.44		
	8.86		
	8.46		
	1.25		
	2.44		
	1.58		
	SD	MD	GR
μ_i	7.54	8.44	8.86
	8.46	9.71	10.91
	10.04	11.29	12.48
$\log L_{max}$	-585.74		

Table 4.14 Maximum likelihood estimates for Σ under the more general model where $x_{ij} \sim N(\mu_i, \Sigma)$ and the maximum log-likelihood using data sets D_5, D_4 and D_3

Data Set	$\hat{\Sigma}$					$\log L_{max}$
D_5	5.58	3.97	3.23	2.00	0.98	-246.25
	3.97	7.49	5.87	3.82	1.80	
	3.23	5.87	8.19	5.24	2.39	
	2.00	3.82	5.24	6.37	3.11	
	0.98	1.80	2.39	3.11	3.42	
D_4	5.42	3.71	2.92	1.64		-433.65
	3.71	7.57	5.60	3.48		
	2.92	5.60	7.93	4.82		
	1.64	3.48	4.82	5.30		
D_3	5.98	4.05	3.09			-585.40
	4.05	8.09	5.87			
	3.09	5.87	7.87			

Table 4.15 Test statistics for the likelihood ratio test of the constrained model where $x_{ij} \sim N(M_i\theta,\Sigma)$ within the more general model where $x_{ij} \sim N(\mu_i,\Sigma)$

	Data Set D_5	D_4	D_3
$2 \log \frac{L \text{ General}}{L \text{ Constrained}}$	3.250	4.636	0.649
number of parameters in the general model (n_g)	30	22	15
number of parameters in the constrained model (n_c)	22	17	13
critical value of $X^2(n_g - n_c)$ at 5% significance level	15.51	11.07	5.99

4.6 Further Developments of the Constrained Linear Model

4.6.1 Model Fitting by Pseudo Maximum Likelihood

The computation involved in obtaining the maximum likelihood estimate of Σ in the constrained linear model was substantial. It was therefore decided to investigate, in this constrained model, the effect of the substitution of the maximum likelihood estimate of Σ by an alternative estimate which could be calculated more easily. In this case the maximum likehood estimate under the more general model $x_{ij} \sim N(\mu_i,\Sigma)$ was used, as it was easily calculated.

In Appendix 3 it is shown that, in general with this approach, testing hypotheses that imposed linear constraints on the parameters not assosciated with Σ leads to conservative goodness - of - fit tests. Thus a constrained model which was not rejected by this approach would not be rejected by the true likelihood ratio test. Indeed for the special case of the models under consideration in this chapter it can be shown that the approximate likelihood ratio test is asymptotically equivalent to the exact likelihood ratio test.

The estimates for μ_i and the logarithm of the likelihood are given in Table 4.16. Comparison of the results in Table 4.16 with those in Tables 4.11, 4.12 and 4.13 did not show that it was beneficial to use the full maximum likelihood estimate of Σ. The estimates for μ_i and Σ, using full and pseudo maximum likelihood estimation, differed only after the fifth and first decimal places respectively. It would therefore have been worthwhile to use this pseudo maximum likelihood approach in the first instance to test the fit of the model.

Table 4.16 Pseudo maximum likelihood estimates for the mean vector under the model where $x_{ij} \sim N(M_i\theta,\Sigma)$ and the maximum log-likelihood using data sets D_5, D_4 and D_3

Data Set	Mean Vector			log L_{max}
	SD	MD	GR	
D_5	7.13	8.06	8.17	-247.88
	7.76	9.08	9.87	
	9.18	10.49	11.28	
	10.60	11.91	12.70	
	12.01	13.33	14.11	
D_4	7.37	8.20	8.55	-435.98
	8.17	9.55	10.43	
	9.59	10.97	11.85	
	11.02	12.39	13.27	
D_3	7.54	8.44	8.86	-585.74
	8.46	9.71	10.91	
	10.04	11.29	12.48	

4.6.2 Increasing the Structure on Σ

In an attempt to reduce further the number of parameters to be estimated, the effect of imposing some structure on the covariance matrix under the constrained linear model was investigated. Since the correlation between best coma scores at different time periods decreases as the time difference increases (see Tables 4.11 - 4.13), the following structures on the covariance matrix were proposed:-

(i) $\Sigma = SRS$

where $S = \text{diag}(s_1, \ldots, s_C)$,

$C = 3, 4, 5,$

and $R = \begin{bmatrix} 1 & \rho & . & . & . & \rho^C \\ \rho & 1 & . & . & . & \rho^{C-1} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \rho^C & \rho^{C-1} & . & . & . & 1 \end{bmatrix}$

(ii) As (i) but with $s_1 = s_2 = \ldots = s_C = s$

so that

$\Sigma = s^2R.$

The maximum likelihood estimates from fitting the model $x_{ij} \sim N(M_i\theta, SRS)$ described in (i) and the likelihood ratio test results for the fit of this model in comparison to the more general model with $x_{ij} \sim N(\mu_i, \Sigma)$ are given in Table 4.17. Thus although there was no evidence that the structure imposed upon the covariance matrix gave a significantly worse fit than the general model for data set D_5, this was not the case with data sets D_4 and D_3.

When the number of parameters involved in the estimation of Σ was further reduced as described in (ii) the maximum value of the logarithm of the likelihood for data set D_5 was -283.91. This gave

Parameter	Estimate		
	D_5	D_4	D_3
S	diag {5.94, 7.48, 7.56, 6.13, 3.45}	diag {6.02, 7.92, 7.37, 4.90}	diag {6.39, 8.04, 7.35}
ρ	0.685	0.683	.662
$M_i\theta$	SD MD GR 7.11 8.08 8.16 7.56 9.07 9.96 9.17 10.49 11.28 10.59 11.90 12.70 12.01 13.32 14.11	SD MD GR 7.39 8.20 8.59 8.19 9.55 10.45 9.61 10.97 11.87 11.03 12.40 13.29	SD MD GR 7.54 8.43 8.86 8.46 9.70 10.91 10.04 11.28 12.49
log L_c	-254.47	-449.26	-598.70
2 log $\frac{L_g}{L_c}$	16.45	31.23	26.61
Reduction in number of parameters estimated (n_r)	17	10	4
Critical value for $\chi^2(n_r)$ at 5% significance level	27.59	18.31	11.07

Table 4.17 Maximum likelihood estimates for model $x_{ij} \sim N(M_i\theta, SRS)$ and the likelihood ratio test results from the comparison with the more general model $x_{ij} \sim N(\mu_i, \Sigma)$

a test statistic of 75.33 for the likelihood ratio test. Since this was a significantly worse fit than the general model, no further constraints on the covariance matrix were considered.

4.6.3 Different Covariance Matrices for Each Group

Up to this point a common covariance structure for the three groups had been assumed. However, the model $x_{ij} \sim N(M_i\theta,\Sigma_i)$ with different covariance matrices for each outcome category was then considered. A maximum likelihood approach was adopted as before and this gave the logarithm of the likelihood to be

$$\log L = -\frac{CN}{2}\log(2\Pi) - \sum_{i=1}^{3}\frac{n_i}{2}\log\left|\Sigma_i\right| - \frac{1}{2}(\theta - A^{-1}B)^T A (\theta - A^{-1}B)$$

$$- \frac{1}{2}\left(\sum_{i=1}^{3}\sum_{j=1}^{n_i} x_{ij}^T\Sigma_i^{-1}x_{ij} - B^TA^T\right)$$

$$\text{where } A = \sum_{i=1}^{3} n_i\, M_i^T\, \Sigma_i^{-1}\, M_i$$

$$\text{and} \quad B = \sum_{i=1}^{3} n_i\, M_i^T\, \Sigma_i^{-1}\, \bar{x}_i.$$

Thus the maximum likelihood is attained when $\hat{\theta} = A^{-1}B$ and is given by

$$\log L = -\frac{CN}{2}\log(2\Pi) - \sum_{i=1}^{3}\frac{n_i}{2}\log\left|\Sigma_i\right| - \frac{1}{2}\left(\sum_{i=1}^{3}\sum_{j=1}^{n_i} x_{ij}^T\Sigma_i^{-1}x_{ij} - B^TAB\right)$$

To evaluate this numerically $\hat{\Sigma}_1$, $\hat{\Sigma}_2$ and $\hat{\Sigma}_3$ have to be estimated so that log L is a maximum. Computationally this involves estimating 45 parameters.

Since the value of the likelihood function decreased only slightly when the maximum likelihood estimates of Σ from the model $x_{ij} \sim N(\mu_i,\Sigma)$ were used as estimates for Σ in the model $x_{ij} \sim N(M_i\theta,\Sigma)$, this approach was adopted again.

Instead of evaluating the maximum likelihood estimates of Σ_i from the model $x_{ij} \sim N(M_i\theta,\Sigma_i)$ in the calculation of the logarithm

of the likelihood, the maximum likelihood estimates of Σ_i from the more general model $x_{ij} \sim N(\mu_i, \Sigma_i)$ were used.

Within the general model $x_{ij} \sim (\mu_i, \Sigma_i)$ the maximum likelihood estimates for μ_i and Σ_i are (Anderson, 1984b, pp 63-64):-

$$\hat{\mu}_i = \bar{x}_i$$

and

$$\hat{\Sigma}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)^T.$$

The maximum likelihood estimates of Σ_i, i=1, 2, 3 within this model are given in Table 4.18.

With substitution of these values as estimates of Σ_i in both the general model $x_{ij} \sim N(\mu_i, \Sigma_i)$ and the constrained model $x_{ij} \sim N(M_i\theta, \Sigma_i)$ the logarithm of the likelihood was calculated. The mean values $\hat{\mu}_i = M_i\hat{\theta}$ were also calculated for the constrained model and the fit of the constrained model compared with that of the more general model using a likelihood ratio test. These results are given in Table 4.19. Thus there was no evidence that the constrained model $X_{ij} \sim N(M_i\theta, \Sigma_i)$ was a significantly worse fit than the more general model $x_{ij} \sim N(\mu_i, \Sigma_i)$.

A summary of all the models fitted, the number of parameters estimated and the logarithm of the likelihood of the fitted model are given in Table 4.20. From this table it can be seen that, in each case, the model with different covariance matrices for each group does give a significantly better fit than that with a common covariance matrix, but that, for a specified covariance structure the structure on the mean provides a good description of the data.

Outcome Π_i	$\hat{\Sigma}$													
	D_5					D_4					D_3			
Severe Disability	4.18	3.57	2.92	2.19	1.97	4.40	3.33	2.40	1.99		4.47	3.73	2.75	
	3.57	6.63	5.07	4.27	3.46	3.33	6.81	4.91	3.98		3.73	7.52	5.57	
	2.92	5.07	7.56	5.92	4.20	2.40	4.91	8.09	6.05		2.75	5.57	8.62	
	2.19	4.27	5.92	7.61	5.17	1.99	3.98	6.05	6.96					
	1.97	4.46	4.20	5.17	6.09									
Moderate Disability	6.35	4.28	3.12	1.36	0.41	5.62	4.03	3.38	1.54		6.37	4.30	3.48	
	4.28	7.58	6.06	3.38	1.44	4.03	7.99	6.66	3.78		4.30	8.36	6.88	
	3.12	6.06	8.67	5.18	2.10	3.38	6.66	8.76	5.03		3.48	6.88	8.76	
	1.36	3.38	5.18	6.93	3.28	1.54	3.78	5.03	5.73					
	0.41	1.44	2.10	3.28	3.20									
Good Recovery	5.85	3.96	3.76	2.46	0.82	5.88	3.68	2.87	1.52		6.50	4.04	2.99	
	3.96	8.00	6.25	3.92	0.98	3.68	7.68	5.15	2.92		4.04	8.09	5.25	
	3.76	6.25	8.18	4.81	1.39	2.87	5.15	7.12	3.88		2.99	5.25	6.74	
	2.46	3.92	4.81	4.00	1.52	1.52	2.92	3.88	3.90					
	0.82	0.98	1.39	1.52	1.76									

Table 4.18 Maximum likelihood estimates for Σ_i for model $x_{ij} \sim N(\mu_i, \Sigma_i)$ for data sets D_5, D_4 and D_3

Model	Parameter	Estimate								
		D_5			D_4			D_3		
		SD	MD	GR	SD	MD	GR	SD	MD	GR
$x_{ij} \sim N(M_i\theta, \Sigma_i)$	μ_i	7.14	8.09	8.16	7.42	8.20	8.55	7.54	8.44	8.86
		7.80	9.09	9.86	8.22	9.55	10.45	8.46	9.72	10.92
		9.22	10.50	11.27	9.63	10.97	11.87	10.03	11.29	12.49
		10.63	11.92	12.69	11.05	12.38	13.28			
		12.05	13.34	14.11						
	$\log L_c$	-219.71			-418.92			-572.98		
	n_c (no. of parameters)	52			37			25		
$x_{ij} \sim N(\mu_i, \Sigma_i)$	$\log L_g$	-218.02			-416.49			-572.60		
	n_g (no. of parameters)	60			42			27		
	$2 \log \frac{L_g}{L_c}$	3.38			4.85			0.78		
	$\chi^2 (n_g - n_c, 0.95)$	15.51			11.07			5.99		

Table 4.19 Maximum likelihood estimates for mean vector for the model where $x_{ij} \sim N(M_i\theta, \Sigma_i)$ and likelihood ratio test results from the comparison with the more general model $x_{ij} \sim N(\mu_i, \Sigma_i)$

Table 4.20 Summary of the log-likelihoods and degrees of freedom of the models discussed

Model	Maximum Likelihood Method	Log L Degrees of Freedom		
		D_5	D_4	D_3
$x_{ij} \sim N(\mu_i, \Sigma_i)$	Exact	-218.02 60	-416.49 42	-572.60 27
$x_{ij} \sim N(M_i\theta, \Sigma_i)$	Pseudo	-219.71 52	-418.92 37	-572.98 25
$x_{ij} \sim N(\mu_i, \Sigma)$	Exact	-246.25 30	-433.65 22	-585.40 15
$x_{ij} \sim N(M_i\theta, \Sigma)$	Exact	-247.87 22	-435.97 17	-585.74 13
$x_{ij} \sim N(M_i\theta, \Sigma)$	Pseudo	-247.88 22	-435.98 17	-585.74 13
$x_{ij} \sim N(M_i\theta, SRS)$	Exact	-254.47 13	-449.26 12	-598.70 11

4.7 Assumptions of Multivariate Normality

4.7.1 Introduction

All the models fitted so far assumed that the data had a multivariate normal distribution. This assumption was examined in two ways.

(i) by looking at the univariate marginal distribution of the scores at the different times.

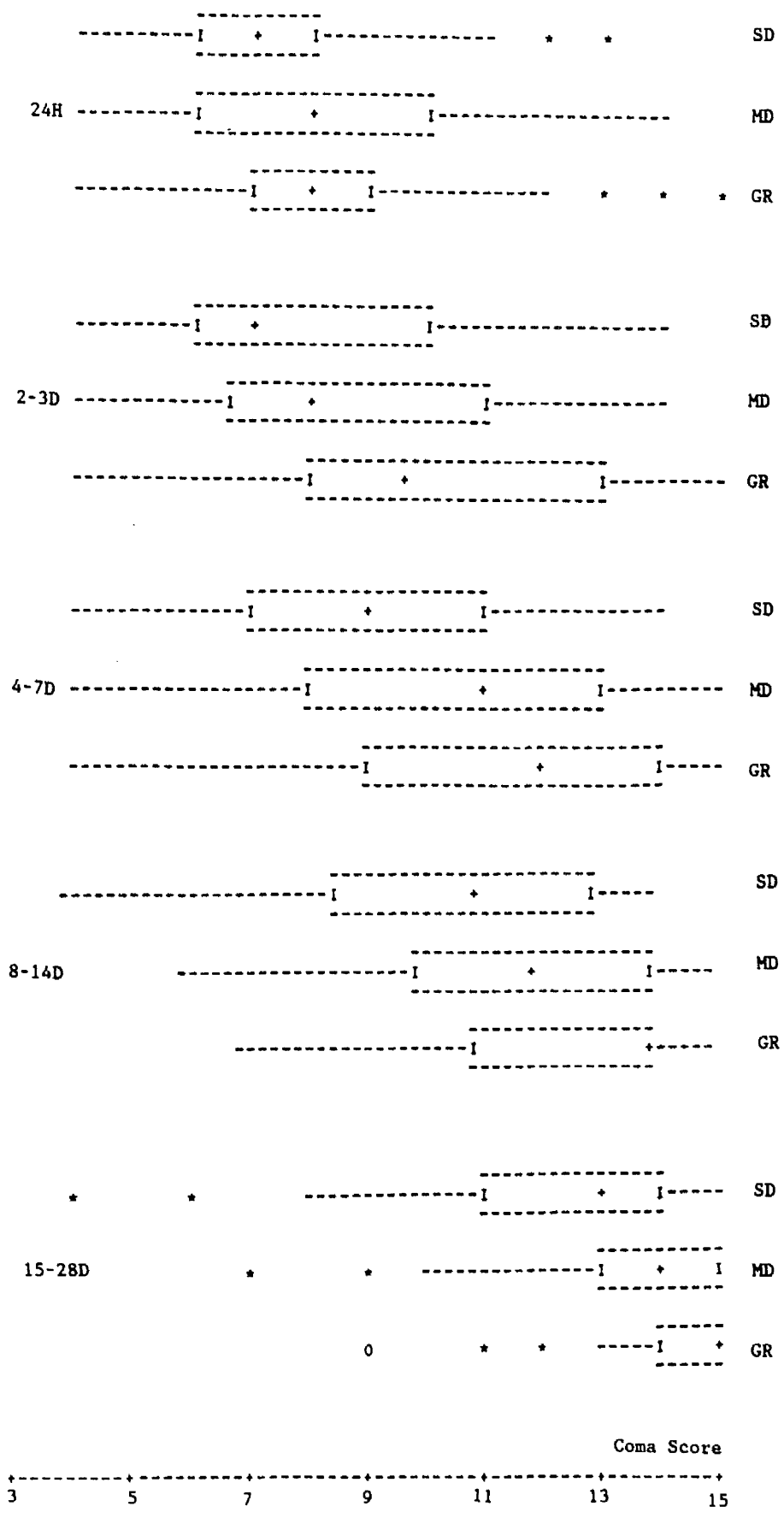
(ii) by considering the Andrews curves of the data.

4.7.2 Marginal Distribution of Scores

If the feature vector has a multivariate normal distribution then each of the components of the feature vector should also be normally distributed. To look at this assumption, box and whisker plots of the scores at each time period were drawn for each outcome category for the three data sets D_3, D_4 and D_5 considered. All data sets showed a similar pattern. In the early stages after injury the plots looked symmetrical about the median value; this suggested that, at least marginally, the components were normally distributed. However, the data became increasingly skewed as time progressed and the patients improved towards the top of the coma scale. The box and whisker plots for the data set D_5 are given in Figure 4.8 and illustrate these points. The same features were seen in normal probability plots of the data, in which deviations from linearity occurred most markedly with the 15-28 day scores of those who made a good recovery.

In an attempt to improve the marginal normality, various transformations of the data were considered but none was particularly successful, so the multivariate normality assumption was adopted for the original variables.

Figure 4.8 Box and whisker plots of the coma scores at the five time periods with data set D_5



4.7.3 Assessment of Multivariate Normality using Andrews Curves

A better means of assessing multivariate normality is the Andrews curve (Andrews, 1972). Each feature vector is denoted by a curve, $f_x(t)$, which, for the feature vector $x = \{x_1, \ldots, x_C\}$ is

$$f_x(t) = \frac{1}{\sqrt{2}} x_1 + x_2\sin t + x_3\cos t + x_4\sin 2t + x_5\cos 2t + \ldots \text{ to C terms,}$$

$$\text{where} \quad -\pi < t < \pi.$$

One interpretation of a set of Andrews curves from a data set is as an infinity (as t varies) of sets of univariate projections of the data, and an informal test of multivariate normality is to assess univariate normality simultaneously for all t. With a large set of data the plots of all the Andrews curves are difficult to interpret but Gnanadesikan (1977) used a quantile contour plot approach to simplify the procedure. The values of a few sample percentiles are evaluated for a large number of t values giving contour curves for the percentiles chosen. If for each t chosen, the univariate data are standardised, then, if the original data were multivariate normal, the resulting standardised quantile contours should be roughly horizontal straight line plots at the levels indicated by the standard normal quantiles. The contour plots corresponding to the 5, 25, 50, 75 and 95 percentiles for each outcome category were produced for data sets D_5, D_4 and D_3. These contours should then be roughly horizontal lines at -1.645, -0.675, 0.000, 0.675 and 1.645. The plots for data sets D_5, D_4 and D_3 are shown in Figures 4.9, 4.10 and 4.11 respectively. These plots are still difficult to interpret and therefore a comparison was made with corresponding plots of simulated normal data. These had means and covariance matrices equal to the maximum liklihood estimates of the mean and covariance matrix from the data set with which they were to be compared. Also the number of simulations was

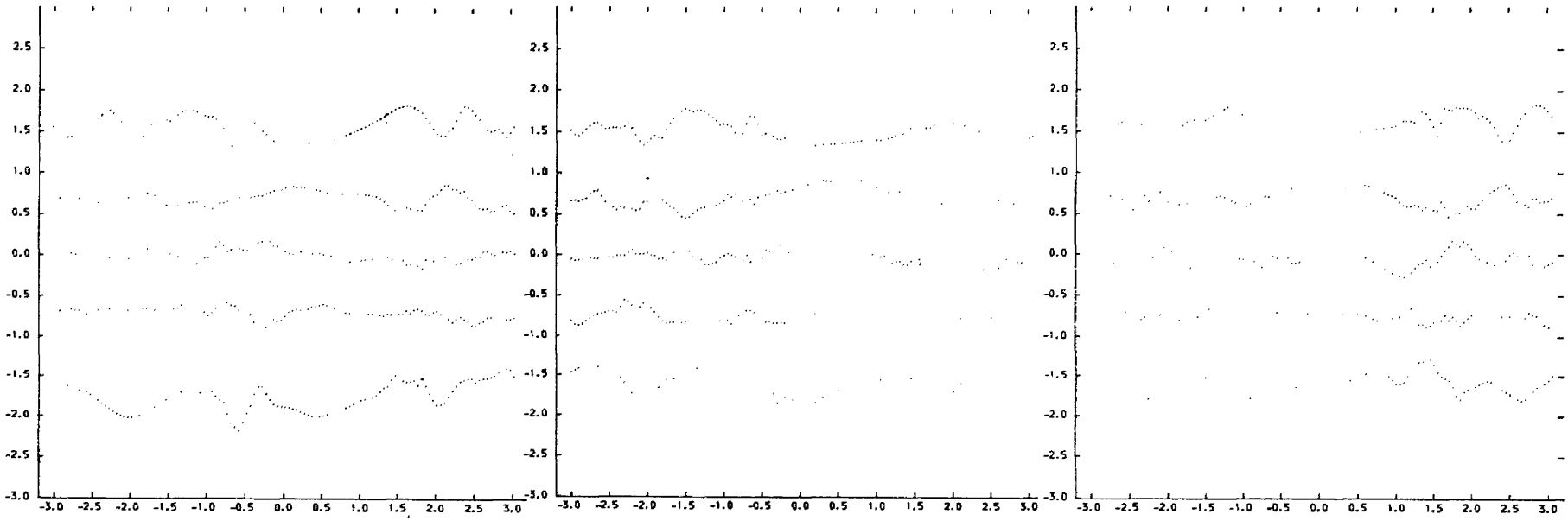


Figure 4.9 Quantile contour plots from Andrews curves for data set D_5

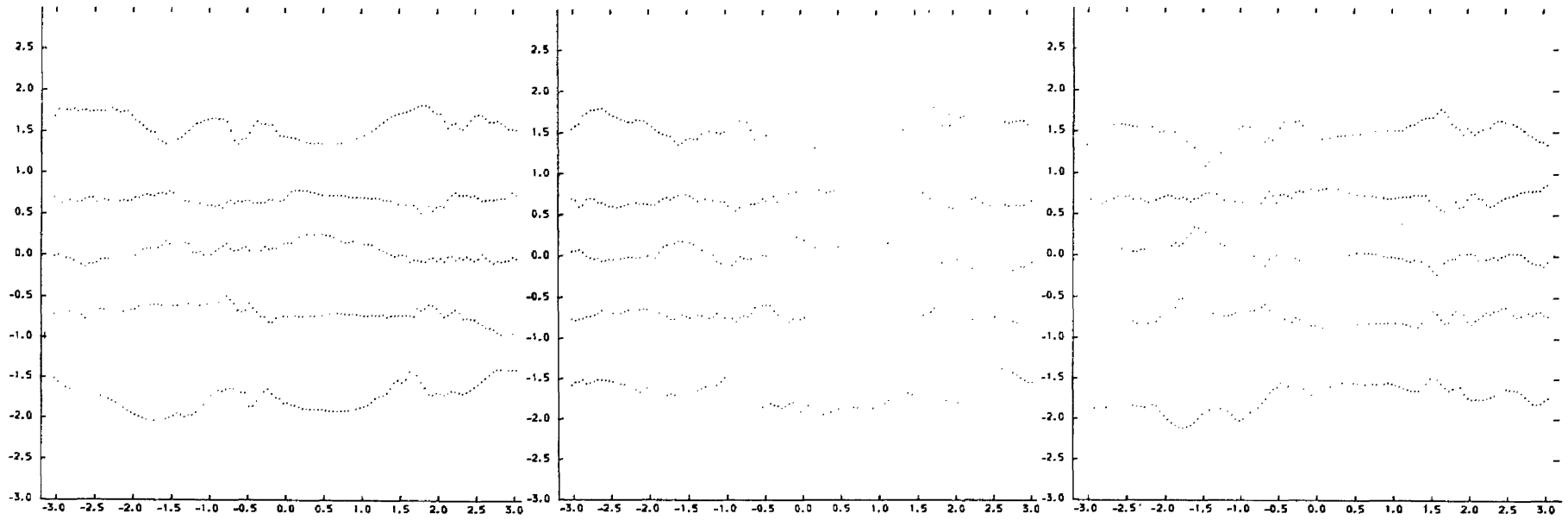


Figure 4.10 Quantile contour plots from Andrews curves for data set D_4

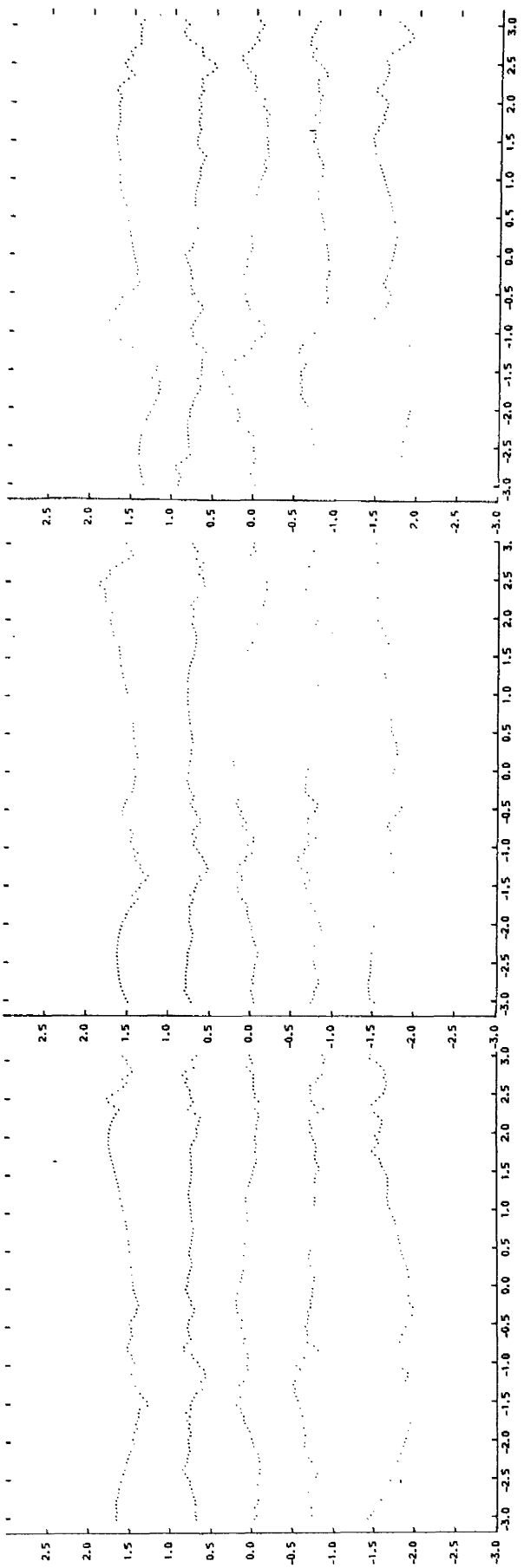


Figure 4.11 Quantile contour plots from Andrews curves for data set D_3

equal to the number of cases in that data set. The quantile contour plots of the Andrews curves of the simulated normal data are given in Figures 4.12, 4.13 and 4.14 and can be compared with Figures 4.9, 4.10 and 4.11 respectively. Subjectively the plots in Figures 4.12, 4.13 and 4.14 appeared no more 'horizontal' than those in Figures 4.9, 4.10 and 4.11.

Finally, a hundred simulations were produced of the quantile contour plots of 63 data simulations from a normal distribution with mean and covariance matrix from the severe disability group in data set D_5 were produced. These are shown in Figure 4.15(a). Figure 4.15(b) shows the bands produced by the plots in Figure 4.15(a) together with the quantile contour plots of the Andrews curves of the severe disability group in data set D_5. While these cannot be interpreted as confidence bands it is encouraging to see that the quantile contour plots for the actual data lie within the band produced from the 100 simulations of the corresponding normal data.

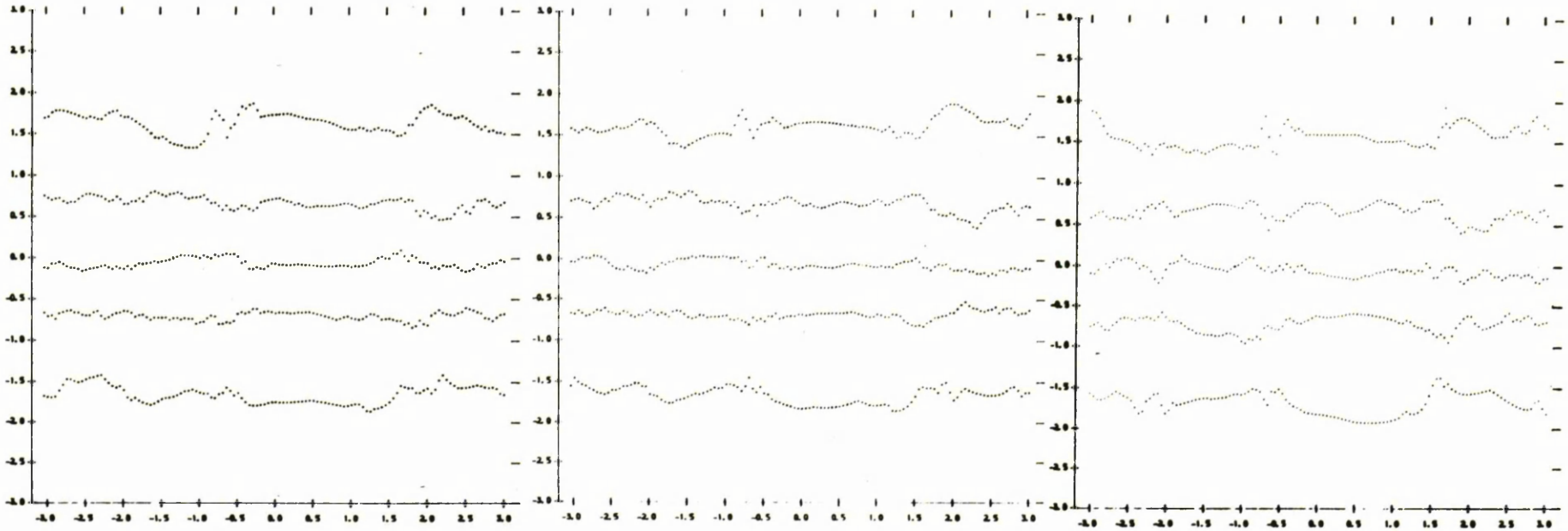
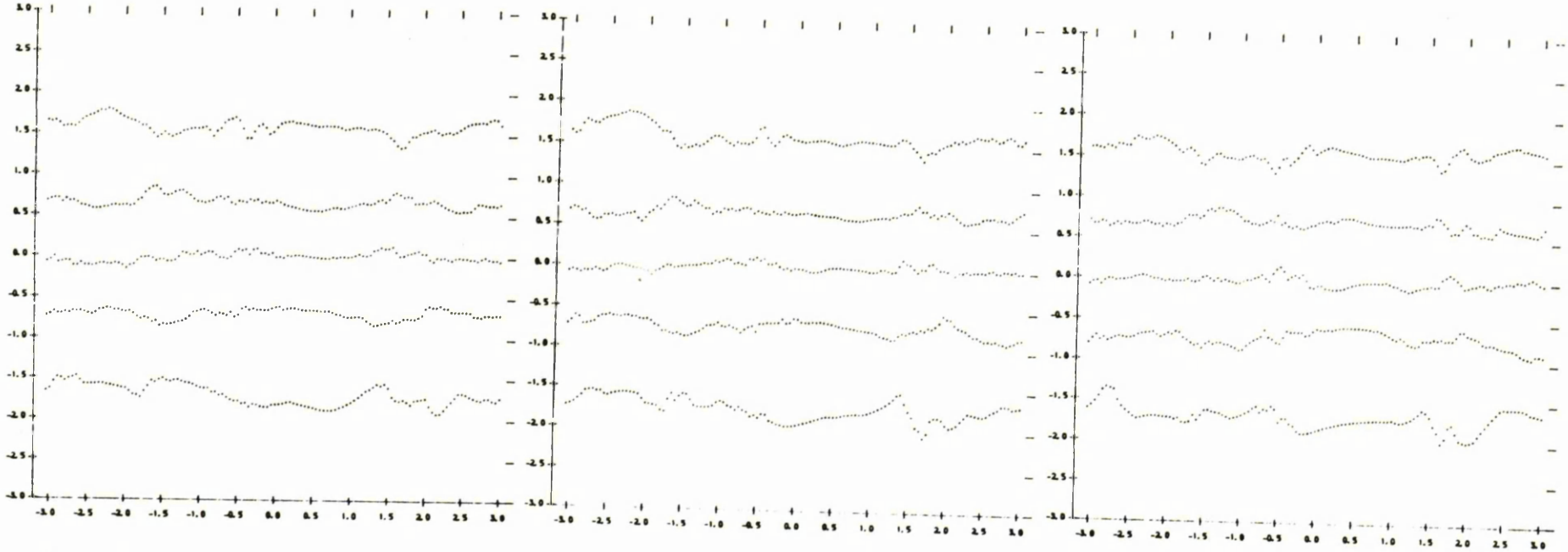


Figure 4.12 Quantile contour plots from Andrews curves of simulated normal data for comparison with those from data set D_5



Good recovery

Moderate disability

Severe disability

Figure 4.13 Quantile contour plots from Andrews curves of simulated normal data for comparison with those from data set D_4

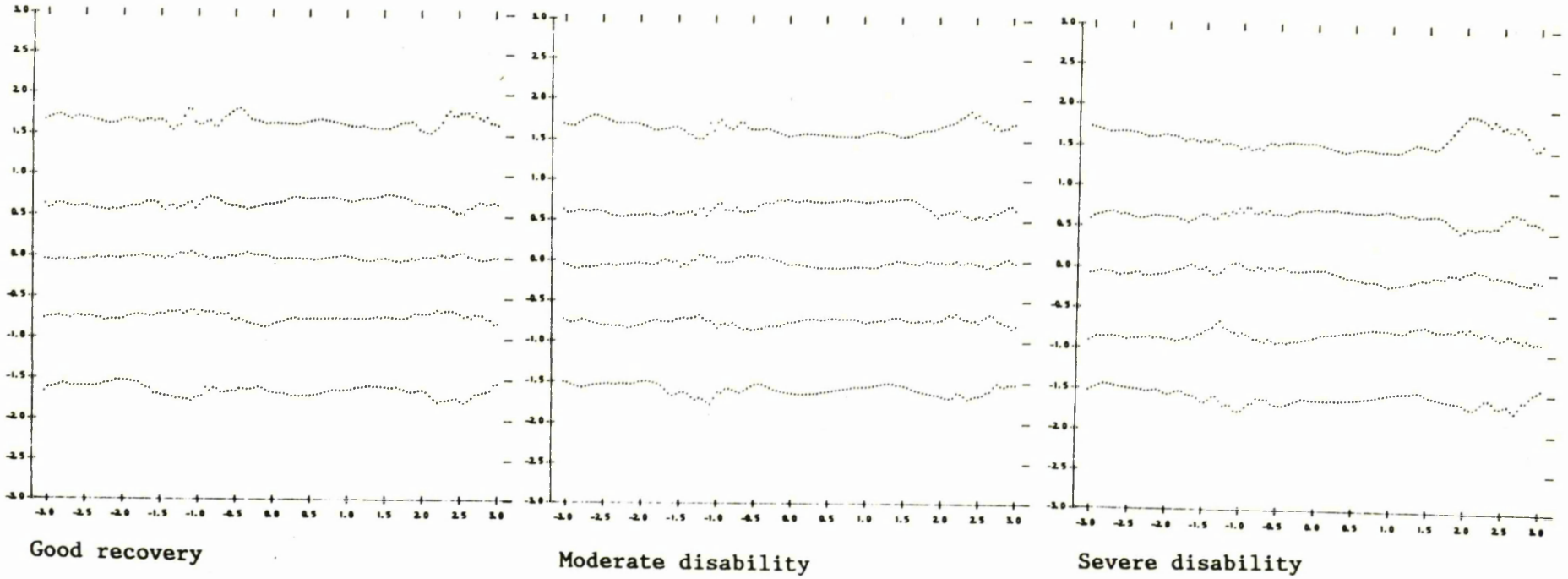
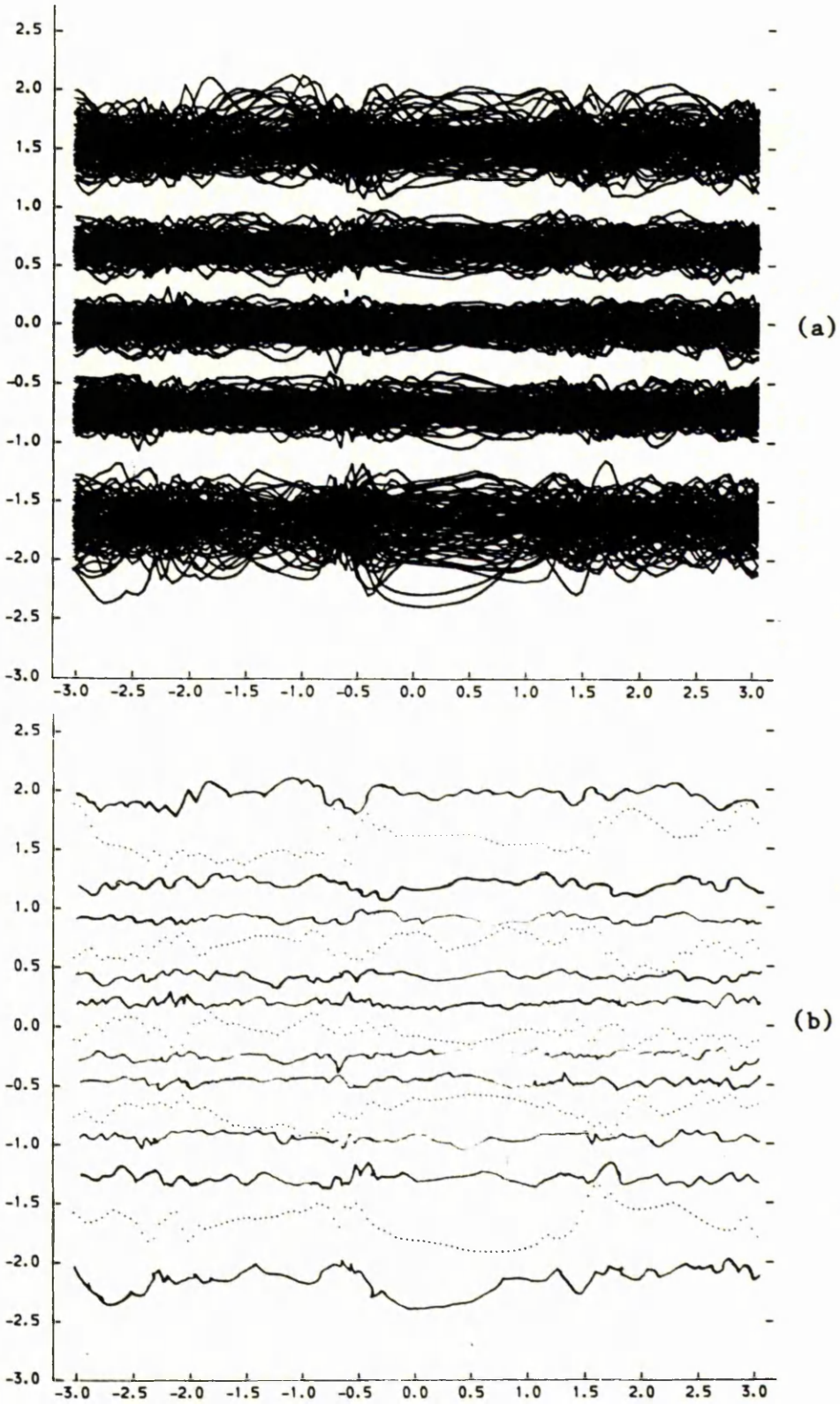


Figure 4.14 Quantile contour plots from Andrews curves of simulated normal data for comparison with those from data set D_3

Figure 4.15 Quantile contour plots of Andrews curves

a) 100 simulations of normal data (from SD group, D_3)

b) bands produced by a) and plot of SD group from D_3



CHAPTER 5

ASSESSMENT OF PERFORMANCE

5.1 Introduction

The original purpose of this work was to discover if data collected over time could be used to improve prediction of the six month outcome. This chapter describes the use of some of the models developed in Chapter 4 to make such predictions, and compares the results of these predictions with those from other available models.

The data sets used were the same as those used to develop the models in Sections 4.5 and 4.6, namely D_C, where C = 5, 4 and 3. The main emphasis was on the assessment of the performance of the predictions made at 28 days; but the results of predictions at 14 days and 7 days were also considered.

The same test and training data sets were used. To overcome the bias thereby introduced, the cross-validatory technique suggested by Lachenbruch and Mickey (1968) was used. With their 'jack-knife' technique each case is omitted from the training data set in turn and the coefficients of the classification rule are recalculated. The omitted patient is then classified on the basis of the remaining patients. Unless excessive computation is required for the jack-knife procedure, the results are given with and without the jack-knife procedure.

Two different sets of prior probabilities were used in each case. The relative incidence or arrival rate associated with each of the three outcome categories gave one set of prior probabilities which for the three data sets were:-

D_5: p(SD) = .267 p(MD) = .352 p(GR) = .381

D_4: p(SD) = .254 p(MD) = .340 p(GR) = .406

D_3: p(SD) = .235 p(MD) = .318 p(GR) = .447.

If one outcome group has an incidence substantially less than the others this can lead to that outcome seldom having the highest probability attached to it. The separation achieved solely on the basis of the patients' feature vectors was therefore assessed by the results obtained using equal prior probabilities.

To evaluate the various models, the three scores described in Section 3.3.3 were used, namely, the error rate, the average logarithmic score and the average quadratic score. However, to give a further indication of the problems associated with ordered outcome categories, the classification matrices have also been given.

5.2 Constrained Linear Model

5.2.1 Practical Aspects

Predictions of outcome were made using three of the models described in Sections 4.5 and 4.6. The data, x, were assumed to have the following distributions

(i) $x \sim N(M_i\theta,\Sigma)$ where M_i is as stated in Section 4.5 and estimates of θ and Σ were obtained using maximum likelihood estimation as described in Section 4.5.

(ii) $x \sim N(M_i\theta,\Sigma)$ where M_i is as stated in Section 4.5, and the estimates of θ and Σ were obtained using the pseudo maximum likelihood methods described in Section 4.6.1.

(iii) $x \sim N(M_i\theta,\Sigma_i)$ where M_i is as stated in Section 4.5 and the pseudo maximum likelihood approach described in Section 4.6.3 was again used to obtain estimates of θ and Σ_i.

For convenience these methods were called **CONOR1, CONOR2** and **CONOR3** respectively. With the models of CONOR1 and CONOR2 the probability that a new patient with feature vector y belongs to category Π_i is estimated by

$$p(\Pi_i|y,D) = \frac{p(\Pi_i)\exp\left[-\frac{1}{2}(y-M_i\hat{\theta})^T\hat{\Sigma}^{-1}(y-M_i\hat{\theta})\right]}{\sum_{j=1}^{3}p(\Pi_j)\exp\left[-\frac{1}{2}(y-M_j\hat{\theta})^T\hat{\Sigma}^{-1}(y-M_j\hat{\theta})\right]},$$

where $p(\Pi_i)$ is the prior probability of outcome Π_i

and $\hat{\theta}$ and $\hat{\Sigma}$ are the appropriate estimates of θ and Σ.

With the model CONOR3 the above probability for the new patient, y, is estimated by

$$p(\Pi_i|y,D) = \frac{p(\Pi_i)\left|\hat{\Sigma}_i\right|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(y-M_i\hat{\theta})^T\hat{\Sigma}_i^{-1}(y-M_i\hat{\theta})\right]}{\sum_{j=1}^{3}p(\Pi_j)\left|\hat{\Sigma}_j\right|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(y-M_j\hat{\theta})^T\hat{\Sigma}_j^{-1}(y-M_j\hat{\theta})\right]},$$

where $\hat{\Sigma}_i$ is the appropriate estimate of Σ_i.

To use the jack-knife technique for the predictions with CONOR1 would involve a complete recalculation of the maximum likelihood estimates $\hat{\theta}$ and $\hat{\Sigma}$ for every case. However with CONOR2 and CONOR3 a recursive approach can be adopted.

For CONOR2 let $\hat{\Sigma}_N$ and $\hat{\theta}_N$ be the estimates for Σ and θ involving all N patients and $\hat{\Sigma}_{N-1}$ and $\hat{\theta}_{N-1}$ be the estimates with 1 patient removed. If the last patient in the first category say, is removed (that is, the patient with feature vector x_{1j}, where $j=n_1$), then $\hat{\Sigma}_{N-1}$ can be obtained from $\hat{\Sigma}_N$ as follows:-

$$(N-1)\hat{\Sigma}_{N-1} = \sum_{j=1}^{n_1-1}(x_{1j}-\tilde{x}_1)(x_{1j}-\tilde{x}_1)^T + \sum_{i=2}^{3}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)(x_{ij}-\bar{x}_i)^T,$$

where $$\tilde{x}_1 = \frac{1}{n_1-1}\sum_{j=1}^{n_1-1}x_{ij}$$

$$= \sum_{j=1}^{n_1-1}(x_{1j} - \bar{x}_1 + \bar{x}_1 - \tilde{x}_1)(x_{1j} - \bar{x}_1 + \bar{x}_1 - \tilde{x}_1)^T + \sum_{i=2}^{3}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)^T$$

$$= \sum_{j=1}^{n_1-1}(x_{1j} - \bar{x}_1)(x_{1j} - \bar{x}_1)^T + \sum_{j=1}^{n_1-1}(\bar{x}_1 - \tilde{x}_1)(x_{1j} - \bar{x}_1)^T + \sum_{j=1}^{n_1-1}(x_{1j} - \bar{x}_1)(\bar{x}_1 - \tilde{x}_1)^T$$

$$+ \sum_{j=1}^{n_1-1}(\bar{x}_1 - \tilde{x}_1)(\bar{x}_1 - \tilde{x}_1)^T + \sum_{i=2}^{3}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)^T$$

$$= N\hat{\Sigma}_N - (x_{1n_1} - \bar{x}_1)(x_{1n_1} - \bar{x}_1)^T - (\bar{x}_1 - \tilde{x}_1)(x_{1n_1} - \bar{x}_1)^T$$

$$- (x_{1n_1} - \bar{x}_1)(\bar{x}_1 - \tilde{x}_1)^T + (n_1 - 1)(\bar{x}_1 - \tilde{x}_1)(\bar{x}_1 - \tilde{x}_1)^T$$

since $\sum_{j=1}^{n_1-1}(x_{1j} - \bar{x}_1) = -(x_{1n_1} - \bar{x}_1)$,

$$= N\hat{\Sigma}_N - \left(1 + \frac{1}{n_1 - 1}\right)(x_{1n_1} - \bar{x}_1)(x_{1n_1} - \bar{x}_1)^T$$

since $(x_{1n_1} - \bar{x}_1) = (n_1 - 1)(\bar{x}_1 - \tilde{x}_1)$.

Thus omitting the q^{th} patient from the p^{th} group with feature vector x_{pq} where $p = 1, 2, 3$ and $q = 1, \ldots, n_p$ gives

$$\hat{\Sigma}_{N-1} = \frac{N}{N-1}\hat{\Sigma}_N - \frac{n_p}{(N-1)(n_p-1)}(x_{pq} - \bar{x}_p)(x_{pq} - \bar{x}_p)^T.$$

So

$$\hat{\Sigma}_{N-1} = c\hat{\Sigma}_N - v\,v^T,$$

where $c = \frac{N}{N-1}$ and $v = \sqrt{\frac{n_p}{(N-1)(n_p-1)}}\,(x_{pq} - \bar{x}_p)$,

and $\hat{\Sigma}_{N-1}^{-1} = c^{-1}\hat{\Sigma}_N^{-1} + \dfrac{c^{-1}\,\hat{\Sigma}_N^{-1}\,v\,v^T\,c^{-1}\,\hat{\Sigma}_N^{-1}}{1 - c^{-1}\,v^T\,\hat{\Sigma}_N^{-1}\,v}$

(Draper and Smith, 1981, p127).

From this estimate for Σ a new estimate $\hat{\theta}_{N-1}$ for θ can be obtained as follows.

Since

$$\hat{\theta}_N = A_N^{-1}\,B_N$$

$$\text{where} \qquad A_N = \sum_{i=1}^{3} n_i \, M_i^T \, \hat{\Sigma}_N^{-1} M_i$$

$$\text{and} \qquad B_N = \sum_{i=1}^{3} n_i \, M_i^T \, \hat{\Sigma}_N^{-1} \bar{x}_i,$$

$$\hat{\theta}_{N-1} = A_{N-1}^{-1} \, B_{N-1}$$

$$\text{where} \quad A_{N-1} = \sum_{i=1}^{3} n_i \, M_i^T \, \hat{\Sigma}_{N-1}^{-1} \, M_i - M_p^T \, \hat{\Sigma}_{N-1}^{-1} \, M_p$$

$$\text{and} \quad B_{N-1} = \sum_{i=1}^{3} n_i \, M_i^T \, \hat{\Sigma}_{N-1}^{-1} \, \bar{x}_i - M_p^T \, \hat{\Sigma}_{N-1}^{-1} \, x_{pq}.$$

Similarly for CONOR3, if the q^{th} patient from the p^{th} group is omitted and $\hat{\Sigma}_{i,N}$ and $\hat{\Sigma}_{i,N-1}$ represent the estimates of Σ_i, and $\hat{\theta}_N$ and $\hat{\theta}_{N-1}$ the estimates of θ, with N patients and N-1 patients respectively then

$$\hat{\Sigma}_{i,N} = \hat{\Sigma}_{i,N-1} \qquad \text{if } i \neq p,$$

$$\hat{\Sigma}_{p,N-1} = c\hat{\Sigma}_{p,N} - v \, v^T,$$

$$\text{where} \quad c = \frac{n_p}{n_p - 1} \quad \text{and} \quad v = \sqrt{\frac{n_p}{(n_p-1)^2}}(x_{pq} - \bar{x}_p),$$

and

$$\hat{\Sigma}_{p,N-1}^{-1} = c^{-1} \, \hat{\Sigma}_{p,N}^{-1} + \frac{c^{-1} \, \hat{\Sigma}_{p,N}^{-1} \, v \, v^T \, c^{-1} \, \hat{\Sigma}_{p,N}^{-1}}{1 - c^{-1} \, v^T \, \hat{\Sigma}_{p,N}^{-1} \, v}.$$

Since

$$\hat{\theta}_N = A_N^{-1} B_N$$

$$\text{where} \qquad A_N = \sum_{i=1}^{3} n_i \, M_i^T \, \hat{\Sigma}_{i,N}^{-1} \, M_i$$

$$\text{and} \qquad B_N = \sum_{i=1}^{3} n_i \, M_i^T \, \hat{\Sigma}_{i,N}^{-1} \, \bar{x}_i,$$

$$\hat{\theta}_{N-1} = A_{N-1}^{-1} \, B_{N-1}$$

$$\text{where} \qquad A_{N-1} = \sum_{i=1}^{3} n_i \, M_i^T \, \hat{\Sigma}_{i,N-1}^{-1} \, M_i - M_p^T \, \hat{\Sigma}_{p,N-1}^{-1} \, M_p$$

$$\text{and} \qquad B_{N-1} = \sum_{i=1}^{3} n_i \, M_i^T \, \hat{\Sigma}_{i,N-1}^{-1} \, \bar{x}_i - M_p^T \, \hat{\Sigma}_{p,N-1}^{-1} \, x_{pq}.$$

As the results for the estimates of θ and Σ in CONOR1 and CONOR2 were almost identical, it was not felt that it was computationally

worthwhile to use the jack-knife technique with CONOR1. However, the jack-knife technique was used with CONOR2 and CONOR3.

5.2.2 Results for the Constrained Linear Model

In the comparison of the results for the different constrained linear models there were three main considerations: the effect of the jack-knife technique; the results obtained from the two different priors; and, most important, the comparison of the three models. In particular, it was important to see if the better fit associated with the model CONOR3, where $x \sim N(M_i\theta, \Sigma_i)$ was translated into improved performance. These considerations will be discussed with regard to first the classification matrices and then the measures of separation.

The classification matrices for the data sets D_5, D_4 and D_3 are given in Tables 5.1, 5.2 and 5.3 respectively. The separation measures for these data sets are given in Tables 5.4, 5.5 and 5.6 respectively.

The classification matrices in Tables 5.1, 5.2 and 5.3 show that the jack-knife procedure had little effect in changing the allocation of a patient to a particular category. However, any effect was more noticeable with CONOR3, where there are more parameters to estimate, than with CONOR2. The magnitude of the effect of the jack-knife technique gives an indication of the variability of the parameter estimates, so these results are not surprising, reflecting the greater variability in the parameter estimates of CONOR3.

The result of the comparison of the effect of the different priors on the classification matrices reflected the higher probability of good recovery and the lower probability of severe disability, relative to the probability of moderate disability, in

Table 5.1 Classification matrices for the constrained normal methods and data set D_5

Priors	Method		No Jack-knife			Jack-knife		
		Predicted Outcome	Actual Outcome			Actual Outcome		
			SD	MD	GR	SD	MD	GR
	CONOR1	SD	30	17	10			
		MD	15	23	17			
		GR	18	43	63			
Arrival		SD	30	17	10	30	17	10
	CONOR2	MD	15	23	17	15	23	17
Rate		GR	18	43	63	18	43	63
		SD	27	14	7	27	14	9
	CONOR3	MD	12	24	17	11	25	16
		GR	24	45	66	25	44	65
		SD	31	21	16			
	CONOR1	MD	17	25	15			
		GR	15	37	59			
		SD	31	21	16	31	21	16
Equal	CONOR2	MD	17	25	15	16	23	15
		GR	15	37	59	16	39	59
		SD	32	17	12	32	19	14
	CONOR3	MD	10	23	15	11	23	15
		GR	21	43	63	20	41	61

Table 5.2 Classification matrices for the constrained normal methods and data set D_4

Priors	Method		No Jack-knife			Jack-knife		
		Predicted Outcome	Actual Outcome			Actual Outcome		
			SD	MD	GR	SD	MD	GR
	CONOR1	SD	42	31	17			
		MD	19	16	21			
		GR	42	91	127			
Arrival		SD	42	31	17	41	31	17
	CONOR2	MD	19	16	21	20	16	22
Rate		GR	42	19	127	42	91	126
		SD	42	23	18	38	24	18
	CONOR3	MD	18	29	24	21	29	23
		GR	43	86	123	44	85	124
		SD	59	45	34			
	CONOR1	MD	20	20	25			
		GR	24	73	106			
		SD	59	45	34	59	46	34
Equal	CONOR2	MD	20	20	25	20	19	27
		GR	24	73	106	24	73	104
		SD	60	35	43	51	32	41
	CONOR3	MD	16	31	18	22	34	17
		GR	27	72	104	30	72	107

Table 5.3 Classification matrices for the constrained normal methods and data set D_3

Priors	Method	Predicted Outcome	No Jack-knife Actual Outcome			Jack-knife Actual Outcome		
			SD	MD	GR	SD	MD	GR
	CONOR1	SD	47	41	28			
		MD	16	12	16			
		GR	61	115	192			
Arrival		SD	47	41	28	47	42	29
	CONOR2	MD	16	12	16	16	11	15
Rate		GR	61	115	192	61	115	192
		SD	53	43	36	52	40	37
	CONOR3	MD	12	16	13	15	22	14
		GR	59	109	187	57	106	185
		SD	69	65	60			
	CONOR1	MD	23	24	33			
		GR	32	79	143			
		SD	69	65	60	69	65	61
Equal	CONOR2	MD	23	24	33	23	24	32
		GR	32	79	143	32	79	143
		SD	77	68	80	70	68	81
	CONOR3	MD	10	14	13	19	14	11
		GR	37	86	143	35	86	144

Table 5.4 Measures of separation for the constrained normal methods and data set D_5

Measure	Method	Arrival Rate Priors		Equal Priors	
		Jack-knife		Jack-knife	
		No	Yes	No	Yes
Error Rate	CONOR1	.508	-	.513	-
	CONOR2	.508	.508	.513	.521
	CONOR3	.504	.504	.500	.509
Average Logarithmic Score	CONOR1	.989	-	.995	-
	CONOR2	.989	1.010	.995	1.016
	CONOR3	1.026	1.118	1.030	1.122
Average Quadratic Score	CONOR1	.599	-	.603	-
	CONOR2	.599	.611	.603	.615
	CONOR3	.612	.643	.612	.644

Table 5.5 Measures of separation for the constrained normal methods and data set D_4

Measure	Method	Arrival Rate Priors		Equal Priors	
		Jack-knife		Jack-knife	
		No	Yes	No	Yes
Error Rate	CONOR1	.544	-	.544	-
	CONOR2	.544	.549	.544	.552
	CONOR3	.522	.530	.520	.527
Average Logarithmic Score	CONOR1	1.007	-	1.020	-
	CONOR2	1.007	1.018	1.020	1.032
	CONOR3	1.030	1.072	1.042	1.085
Average Quadratic Score	CONOR1	.611	-	.619	-
	CONOR2	.611	.618	.619	.626
	CONOR3	.619	.635	.623	.637

Table 5.6 Measures of separation for the constrained normal methods and data set D_3

Measure	Method	Arrival Rate Priors		Equal Priors	
		Jack-knife		Jack-knife	
		No	Yes	No	Yes
Error Rate	CONOR1	.524	-	.553	-
	CONOR2	.524	.526	.553	.553
	CONOR3	.515	.510	.557	.568
Average Logarithmic Score	CONOR1	1.001	-	1.030	-
	CONOR2	1.001	1.009	1.030	1.038
	CONOR3	1.012	1.032	1.044	1.065
Average Quadratic Score	CONOR1	.605	-	.623	-
	CONOR2	.605	.610	.623	.628
	CONOR3	.610	.619	.629	.639

the arrival rate priors. Thus more patients were predicted to make a good recovery and fewer to have a severe disability when the arrival rate priors were used as opposed to equal priors.

The comparison between the results of the three methods was interesting and showed that CONOR1 and CONOR2 gave identical results. The reason for this was that the estimates of θ and Σ differed only after the fifth and first decimal places respectively. CONOR3 was not obviously superior to CONOR1 or CONOR2. In no case did it allocate more or fewer cases correctly in all three outcome categories. Indeed, there did not even seem to be a pattern over the three data sets D_5, D_4 and D_3. For example, when comparing the results based on arrival rate priors, more cases were predicted correctly to have a good recovery with CONOR3 than with CONOR1 or CONOR2 for data set D_5, while the reverse was true for data sets D_4 and D_3. The classification matrices did, however, show the large overlaps between the groups, and that the ordered nature of the outcomes resulted in very few of the moderate disability group being allocated to the correct category.

To look more quantitatively at the results, the separation measures in Tables 5.4, 5.5 and 5.6 were examined. A benchmark for comparison of the measures was obtained by assigning the prior probabilities to each case and these are given in Table 5.7.

The jack-knife technique has already been shown to have little effect on the classification matrix and thus had little effect on the error rate. The same was also true for the average logarithmic and quadratic scores. The jack-knife again had most effect with CONOR3, where the largest number of parameters had to be estimated: this was most marked with data set D_5, which had the smallest number of cases of the three data sets and also the largest number of parameters to be fitted. However, in all cases the jack-knife

Table 5.7 Measures of separation obtained by assigning the prior probabilities

Data Set	Measure	Arrival Rate Priors	Equal Priors
	Error Rate	.619	.667
D_5	Average Logarithmic Score	1.088	1.110
	Average Quadratic Score	.660	.674
	Error Rate	.594	.667
D_4	Average Logarithmic Score	1.081	1.117
	Average Quadratic Score	.655	.678
	Error Rate	.553	.667
D_3	Average Logarithmic Score	1.064	1.133
	Average Quadratic Score	.644	.690

method gave slightly worse average logarithmic and quadratic scores than the same methods without jack-knife. The jack-knife procedure with CONOR3 and data set D_3 gave an error rate which was lower than that without this procedure. This was because six more moderate disability cases were correctly classified while only three fewer severe disability and good recovery cases were correctly classified. All other error rates were either the same or slightly worse with the jack-knife technique.

The arrival rate priors gave slightly better average logarithmic and quadratic scores than the corresponding equal prior results. The extra information contained in the arrival rate priors contributed to this improvement. Also the error rates using the arrival rate priors were generally better than those using equal priors. However, with CONOR3 there were again examples where the reverse was true.

In a comparison of the three methods, CONOR1 and CONOR2 again gave identical scores. With the average logarithmic and quadratic scores, CONOR3 always had a higher score than CONOR1 and CONOR2. Even though the fit of CONOR3 to the data was shown to be better than CONOR1 or CONOR2, this was not reflected in the results of the predictions. This result is in accord with the findings of the comparative study described in Section 3.5. In terms of the error rate, CONOR3 did better than CONOR1 and CONOR2 except with equal priors and data set D_3. The error rates with CONOR3 did not follow the same pattern as that of the other separation measures. This was because, with this particular model, small changes in the probabilities were giving a different classification of outcome. When this occurs, the practical implications of the use of the error rate as a measure of assessment have to be considered. The usual criticism of the error rate is that of insensitivity to small

changes in probabilities, but here the opposite seemed to be the case.

Comparison of Tables 5.4, 5.5 and 5.6 with Table 5.7 shows that, while these methods did not solve all the problems of predicting the outcome in survivors, they did give a worthwhile improvement over simply allocating the prior probabilities to each case.

5.3 Normal Linear Model

5.3.1 Fisher's Linear Discriminant Function

The classical linear discriminant function is a natural comparator for the models in Section 5.2. The linear discriminant function was originally obtained by Fisher (1936) as that linear function which maximises the ratio of 'between' to 'within' sums of squares. However exactly the same discriminant function results from the normal linear model assuming no structure other than the equality of the covariance matrices. Relaxation of the equality of the covariance matrices leads to the standard quadratic discriminant model. However, in view of the poor performance of the quadratic methods in the comparative study and in the results described in Section 5.2, this method was not included for further comparison. With the normal linear model the probability of outcome Π_i for a new individual with feature vector y is estimated by

$$p(\Pi_i|y,D) = \frac{p(\Pi_i)\ \exp\left[-\frac{1}{2}(y-\hat{\mu}_i)^T\ \hat{\Sigma}^{-1}(y-\hat{\mu}_i)\right]}{\sum_{j=1}^{3} p(\Pi_j)\ \exp\left[-\frac{1}{2}(y-\hat{\mu}_j)^T\ \hat{\Sigma}^{-1}(y-\hat{\mu}_j)\right]}$$

where $p(\Pi_i)$ is the prior probability of outcome Π_i and $\hat{\mu}_i$ and $\hat{\Sigma}$ are the maximum likelihood estimates of μ_i and Σ. This method was called **NORLIN** and corresponds to the methods NORLIN1 and NORLIN2 described in Section 3.5. The biomedical computer program BMDP P7M by Dixon et al. (1985) was used to calculate the probalilities.

5.3.2 Results for the Normal Linear Model

The effects of the jack-knife technique and of using the different priors were studied, first in terms of the classification matrices and then by the measures of separation. The classification matrices and the separation measures for the various predictions made using this model are given in Tables 5.8 and 5.9 respectively.

The classification matrices in Table 5.8 show that the jack-knife technique led to fewer cases being correctly classified. This occurred with all the data sets and with either arrival rate or equal priors.

The use of arrival rate priors again led to more cases being predicted to make a good recovery and fewer to be severely disabled than with equal priors. The moderate disability group still proved difficult to predict.

The separation measures in Table 5.9 show that not only was the error rate worse with the jack-knife technique, but that the other separation measures also gave poorer results for all data sets and both prior probabilities. The use of arrival rate priors improved all scores for the measures of separation with one exception. The error rate for data set D_4 with no jack-knife was slightly less with equal priors than arrival rate priors.

Table 5.8 Classification matrices for method NORLIN

Data Set	Priors		No Jack-knife			Jack-knife		
		Predicted Outcome	Actual Outcome			Actual Outcome		
			SD	MD	GR	SD	MD	GR
D_5	Arrival Rate	SD	29	16	13	29	17	14
		MD	18	29	16	18	26	21
		GR	16	38	61	16	40	55
	Equal	SD	31	18	14	31	18	14
		MD	17	32	23	17	29	27
		GR	15	33	53	15	36	49
D_4	Arrival Rate	SD	39	30	17	38	30	18
		MD	23	21	22	23	19	21
		GR	41	87	126	42	89	126
	Equal	SD	59	45	35	57	47	35
		MD	16	21	20	18	17	24
		GR	28	72	110	28	74	106
D_3	Arrival Rate	SD	48	39	24	46	39	24
		MD	13	16	22	14	14	23
		GR	63	113	190	64	115	189
	Equal	SD	70	65	57	68	66	58
		MD	18	20	26	20	18	28
		GR	36	83	153	36	84	150

Table 5.9 Measures of separation for method NORLIN

Data Set	Measure	Arrival Rate Priors		Equal Priors	
		Jack-knife		Jack-knife	
		No	Yes	No	Yes
D_5	Error Rate	.496	.534	.508	.538
	Average Logarithmic Score	.987	1.031	.992	1.036
	Average Quadratic Score	.600	.628	.602	.630
D_4	Error Rate	.542	.549	.532	.557
	Average Logarithmic Score	1.006	1.027	1.019	1.040
	Average Quadratic Score	.611	.624	.618	.631
D_3	Error Rate	.519	.528	.540	.553
	Average Logarithmic Score	1.000	1.011	1.029	1.040
	Average Quadratic Score	.604	.611	.622	.629

5.4 Independence Model

5.4.1 Practical Aspects

In the Head Injury Study it was the independence model, described in Section 3.4, which was first used to make the predictions of outcome. The comparative study described in Section 3.5 showed it performed well when compared with other methods. It was also well understood by the clinicians and is the method currently being used to provide the medical staff with on-line predictions of prognosis in the early stages after head injury (Barlow et al., 1984; Barlow et al., 1987). It was therefore important to compare its performance with that of the normal models described in Sections 5.2 and 5.3. A feature of the independence model is that unlike, for example, the linear discriminant model the ordered nature of the coma score is not explicitly incorporated.

Four sets of variables were considered and data sets D_5, D_4 and D_3 were used as the test and training data for the predictions at the end of 28 days, 14 days and 7 days respectively. The model used was that described in Section 3.5.2 with B=1, and the predicted probability of each outcome Π_i was obtained using the program INDEP-SELECT by Habbema and Gelpke (1981).

The four variable sets used were as follows:

INDEP1 - The best coma score at 24 hours

INDEP2 - The best coma score within the latest available time period. For example, the best coma score within the 8-14 day period was used with data set D_4. It was anticipated that the results from INDEP2 would be better than those of INDEP1 thereby showing that the current state of the patient is more important than the initial state.

INDEP3 - All best coma scores available for the particular data set. Here it was anticipated that utilising all information would provide better results than those purely using the current state of the patient.

INDEP4 - For the latest time period available the best coma score, motor response pattern, pupil reaction, created eye indicant along with the change in neurological function and the patient's age. This is the variable set III described in Section 3.5 and is the one currently in use for the on-line predictions of prognosis in the early stages after injury. It therefore gave a standard against which to judge all other methods.

With INDEP1, INDEP2 and INDEP3 the raw coma scores were used, whereas with INDEP4 the data were suitably collapsed to reduce the number of categories. For example, age was collapsed into decades (0-9, 10-19, ...) and the coma score into 3 groups (3-10, 11-13 and 14-15) for data set D_4 and D_5, and 4 groups (3-4, 5-7, 8-11 and 12-15) for data sets D_3. The splits were those currently used in practice and were chosen using an entropy measure as described in Section 3.4.3.

There was no missing data with INDEP1, INDEP2 and INDEP3, but there was considerable missing data in the additional variables included in INDEP4. This was dealt with as in Section 3.4.2.

5.4.2 Results for the Independence Models

In the discussion of the results of the four independence models the approach taken in Section 5.2.2 with the constrained normal models will be used again. Thus, the effect of employing the jack-knife technique and of the use of two different sets of prior probabilities is considered before the comparing the models.

The classification matrices are given in Tables 5.10, 5.11 and

5.12, for the data sets D_5, D_4 and D_3 respectively and the corresponding separation measures are given in Tables 5.13, 5.14 and 5.15.

The jack-knife technique again gave fewer correct classifications in all but two comparisons in which the same number were correctly classified. In many cases this technique had a marked effect on classification. The number of parameters estimated and the number of cases in the data set should both influence the magnitude of the effect of the jack-knife technique. With data set D_5 it was difficult to see a pattern in the results, but with data sets D_4 and D_3, INDEP1 and INDEP2 were indeed less affected by the technique than INDEP3 and INDEP4, each of which had more parameters to estimate.

The different priors also had a marked effect on the classification and this was most noticeable with data set D_3. The probabilities were obviously such that the changes induced by the arrival rate priors led to a different classification in many cases.

The instability of the classification matrices meant that a comparison of the different independence models was not easy. Subjectively INDEP2 did appear better than INDEP1. INDEP3 was not obviously better than INDEP2 and the results of INDEP4 seemed comparable to those of INDEP3 and INDEP2.

When the separation measures were considered, the jack-knife technique gave poorer average quadratic and logarithmic scores in every case. With these scores, the number of parameters to be estimated and the number of cases in the data set had a noticeable effect. The jack-knife technique had more effect with INDEP3 and INDEP4 which needed more parameters to be estimated. Data set D_5, with fewest cases, showed more effect than data set D_4, which in

Table 5.10 Classification matrices for the independence methods and data set D_5

Priors	Method		No Jack-knife			Jack-knife		
		Predicted Outcome	Actual Outcome			Actual Outcome		
			SD	MD	GR	SD	MD	GR
Arrival Rate	INDEP1	SD	19	11	16	19	17	16
		MD	27	39	27	27	33	37
		GR	17	33	47	17	33	37
	INDEP2	SD	19	8	4	19	13	4
		MD	34	48	28	34	43	28
		GR	10	27	58	10	27	58
	INDEP3	SD	42	19	12	34	26	14
		MD	14	37	9	19	28	16
		GR	7	27	69	10	29	60
	INDEP4	SD	32	18	8	27	19	9
		MD	22	29	17	25	26	17
		GR	9	36	65	11	38	64
Equal	INDEP1	SD	34	28	28	25	28	29
		MD	26	51	45	35	27	45
		GR	3	4	17	3	28	16
	INDEP2	SD	31	22	7	24	23	17
		MD	22	34	25	29	33	25
		GR	10	27	58	10	27	48
	INDEP3	SD	46	24	16	37	30	19
		MD	10	33	8	18	25	15
		GR	7	26	66	8	28	56
	INDEP4	SD	36	20	9	32	21	11
		MD	19	28	17	22	27	18
		GR	8	35	64	9	35	61

Table 5.11 Classification matrices for the independence methods and data set D_4

Priors	Method		No Jack-knife			Jack-knife		
		Predicted Outcome	Actual Outcome			Actual Outcome		
			SD	MD	GR	SD	MD	GR
Arrival Rate	INDEP1	SD	45	35	36	45	35	36
		MD	5	14	11	5	14	11
		GR	53	89	118	53	89	118
	INDEP2	SD	50	33	31	50	33	34
		MD	17	23	9	17	23	9
		GR	36	82	125	36	82	122
	INDEP3	SD	61	39	26	54	44	28
		MD	25	49	30	30	31	37
		GR	17	50	109	19	63	100
	INDEP4	SD	54	30	16	43	34	17
		MD	21	35	15	32	25	24
		GR	28	73	134	28	79	124
Equal	INDEP1	SD	60	55	57	45	55	59
		MD	31	64	65	46	51	65
		GR	12	19	43	12	32	41
	INDEP2	SD	66	49	43	61	49	43
		MD	34	73	68	39	73	68
		GR	3	16	54	3	16	54
	INDEP3	SD	71	53	38	61	57	41
		MD	22	46	31	28	36	43
		GR	10	39	96	14	45	81
	INDEP4	SD	66	41	25	60	43	26
		MD	20	44	29	24	38	42
		GR	17	53	111	19	57	97

Table 5.12 Classification matrices for the independence methods and data set D_3

Priors	Method		No Jack-knife			Jack-knife		
		Predicted Outcome	Actual Outcome			Actual Outcome		
			SD	MD	GR	SD	MD	GR
Arrival Rate	INDEP1	SD	20	15	13	20	15	13
		MD	7	19	18	7	19	18
		GR	97	134	205	97	134	205
	INDEP2	SD	30	22	13	30	22	13
		MD	22	37	26	22	37	42
		GR	72	109	197	72	109	181
	INDEP3	SD	64	46	38	58	50	40
		MD	24	42	33	29	26	44
		GR	36	80	165	37	92	152
	INDEP4	SD	49	29	10	41	32	10
		MD	32	51	33	38	44	47
		GR	43	88	193	45	92	179
Equal	INDEP1	SD	52	40	50	52	40	51
		MD	54	95	107	54	95	107
		GR	18	33	79	18	33	78
	INDEP2	SD	74	65	53	74	68	71
		MD	19	39	36	19	36	36
		GR	31	64	147	31	64	129
	INDEP3	SD	78	59	48	65	61	52
		MD	27	64	57	39	53	61
		GR	19	45	131	20	54	123
	INDEP4	SD	66	42	22	59	52	24
		MD	32	68	58	39	52	58
		GR	26	58	156	26	64	154

Table 5.13 Measures of separation for the independence methods and data set D_5

Measure	Method	Arrival Rate Priors		Equal Priors	
		Jack-knife		Jack-knife	
		No	Yes	No	Yes
Error Rate	INDEP1	.555	.623	.568	.712
	INDEP2	.470	.492	.479	.555
	INDEP3	.373	.483	.386	.500
	INDEP4	.466	.504	.458	.492
Average Logarithmic Score	INDEP1	1.026	1.142	1.036	1.153
	INDEP2	.924	.992	.931	.998
	INDEP3	.916	1.339	.931	1.356
	INDEP4	.922	1.021	.928	1.027
Average Quadratic Score	INDEP1	.620	.679	.627	.687
	INDEP2	.565	.605	.570	.609
	INDEP3	.527	.704	.540	.713
	INDEP4	.561	.625	.565	.626

Table 5.14 Measures of separation for the independence methods and data set D_4

Measure	Method	Arrival Rate Priors		Equal Priors	
		Jack-knife		Jack-knife	
		No	Yes	No	Yes
Error Rate	INDEP1	.564	.564	.589	.663
	INDEP2	.512	.520	.525	.537
	INDEP3	.461	.544	.475	.561
	INDEP4	.451	.527	.456	.520
Average Logarithmic Score	INDEP1	1.040	1.106	1.057	1.124
	INDEP2	.969	1.019	.983	1.033
	INDEP3	.988	1.212	1.010	1.234
	INDEP4	.899	.978	.911	.989
Average Quadratic Score	INDEP1	.629	.664	.640	.676
	INDEP2	.587	.618	.590	.626
	INDEP3	.591	.692	.606	.704
	INDEP4	.547	.599	.555	.605

Table 5.15 Measures of separation for the independence methods and data set D_3

Measure	Method	Arrival Rate Priors		Equal Priors	
		Jack-knife		Jack-knife	
		No	Yes	No	Yes
Error Rate	INDEP1	.538	.538	.572	.574
	INDEP2	.500	.530	.508	.547
	INDEP3	.487	.553	.483	.544
	INDEP4	.445	.500	.451	.498
Average Logarithmic Score	INDEP1	1.026	1.077	1.058	1.110
	INDEP2	.962	1.017	.990	1.045
	INDEP3	.970	1.114	1.003	1.147
	INDEP4	.912	.981	.934	1.003
Average Quadratic Score	INDEP1	.620	.647	.641	.670
	INDEP2	.581	.608	.599	.625
	INDEP3	.584	.650	.603	.669
	INDEP4	.549	.591	.560	.602

turn showed more effect than data set D_3.

The arrival rate priors gave better average logarithmic and quadratic scores in every case. With the error rate there were cases where the reverse was true, particularly with INDEP4.

Comparison of the models on the basis of the separation measures showed that INDEP2 always performed better than INDEP1. When the average quadratic and logarithmic scores were used as a basis for comparison, INDEP3 performed less well than INDEP2 with data sets D_4 and D_3. With data set D_5 INDEP3 performed less well than INDEP2 with the jack-knife technique but better without it. However, for data sets D_4 and D_3, INDEP4 had the best overall performance. It had the lowest score in every case, with the exception of data set D_4, in which, using arrival rate priors and jack-knife, it performed slightly worse than INDEP2. With data set D_5, the performance of INDEP4 was between that of INDEP2 and INDEP3, again with one exception, when assessed on the basis of the average logarithmic and quadratic scores.

Thus the general conclusions of the results of the independence model were that

(i) INDEP2 was better than INDEP1

(ii) INDEP3 performed less well than expected and was in general not better than INDEP2

(iii) INDEP4 performed surprisingly well and for data set D_4 and D_3 gave the best results.

5.5 Comparison of Models

5.5.1 Comparison of Models Assuming Normality

The relative performance of those models in Sections 5.2, 5.3 and 5.4 based on normality assumptions will be discussed first.

The classification matrices of the constrained methods CONOR1, CONOR2 and CONOR3 in Tables 5.1, 5.2 and 5.3 were compared with those of the standard method NORLIN in Table 5.8 and the overall impression was that NORLIN seemed to give very similar results to the other methods; certainly no striking difference was apparent.

When the separation measures for CONOR1, CONOR2 and CONOR3 in Tables 5.4, 5.5 and 5.6 were compared with those for NORLIN in Table 5.9, the differences with the different models were very small indeed. In almost all comparisons in which no jack-knife technique was used, NORLIN gave a lower average logarithmic and quadratic score than the constrained methods. The exception was the average quadratic score with data set D_5 and arrival rate priors, for which NORLIN gave a score between CONOR2 and CONOR3. With the jack-knife technique, the average logarithmic and quadratic scores for NORLIN were between those for CONOR2 and CONOR3. A particular pattern did not emerge for the error rates and the results of the four methods were very similar.

5.5.2 Comparison of Independence and Normal Based Models

To simplify the comparison of the independence based methods with those based on normality assumptions, only the separation measures will be discussed. The results in Tables 5.13 and 5.4 and for data set D_5 in Table 5.9 are compared. Similar comparisons are made for data sets D_4 (Tables 5.14, 5.5 and 5.9) and D_3 (Tables 5.15, 5.6 and 5.9). Since INDEP1 was consistently worse than INDEP2 for all data sets it was excluded.

When no jack-knife technique was used, all the independence methods were better than any of of the normal based methods except with data set D_4 and equal priors. Here the error rate for CONOR3 was better than that of INDEP2. When the jack-knife technique was

used, one of the independence methods always gave the best performance but no one method was consistently better than the normal methods. With data sets D_4 and D_3, INDEP4 performed better than any of the normal methods, but with data set D_5, CONOR2 performed better than INDEP4 in terms of the average logarithmic and quadratic scores.

5.6 Discussion

The results obtained were disappointing on several accounts. First, attempts to use the constrained normal models to model the trend in the coma score through time did not appear to improve performance when compared with the model NORLIN. In the latter the coma scores at the different time periods were simply assumed to have a multivariate normal distribution. The constrained models CONOR1 and CONOR2 were slightly better when the jack-knife technique was used and the converse was true with no jack-knife. This may be because NORLIN requires the estimation of more parameters so that the effect of the jack-knife was more marked. The results using the jack-knife approach are likely to give a more reliable measure of performance and should be given more weight than those without this approach.

Secondly, when the normal based methods were compared with the independence model INDEP4, it was found that INDEP4 for the most part performed better. As INDEP4 provided the standard for comparison, this too was disappointing and led to further scrutiny of the independence models.

The finding that INDEP3, which used the coma scores for all time periods, did not perform better than INDEP2, which used only the latest available coma score seemed unrealistic. However, an explanation for the poor performance of INDEP3 can be found in the

number of parameters that needed to be estimated. Each of variables included had 13 categories; these were not reduced so that a direct comparison could be made with the methods based on normality assumptions. The small numbers of cases in the data sets (especially D_5) meant that the estimates of the parameters had large variability and this resulted in inaccurate estimates. This view is supported by the marked change in performance seen when the jack-knife technique was used with INDEP3.

To examine the performance of the independence models more closely, the number of cases with each coma score and each outcome was tabulated, and the conditional probability of outcome, given the coma score, was calculated. Many of the cells in these tables contained small numbers of cases. This obscured the pattern in the conditional probability of the outcome, given the coma score, that had been expected in view of the ordering in the coma score. This is clearly shown in Table 5.16 which gives the numbers involved and the conditional probabilities of the outcomes, given the 4-7 day best coma score, for data set D_3. The problem was even more marked with data sets D_4 and D_5 which contained fewer cases. This problem was compounded in INDEP3 because more variables were included.

When the indicants used in INDEP4 were tabulated as above it was seen that pupil reaction, motor response pattern and created eye indicant did not contribute much to discrimination: this was because the data were mostly either missing or in the highest category. Discrimination was therefore made on the basis of the coma score, change in neurological condition and age. The conditional probabilities of the outcomes, given age, and the numbers involved are given in Table 5.17. The conditional probabilities here reflect the relationship between age and

		Coma Score (CS)												
		3	4	5	6	7	8	9	10	11	12	13	14	15
	SD	0	2	5	11	14	12	9	9	16	16	8	21	1
		-	.333	.500	.500	.424	.364	.265	.214	.333	.348	.174	.152	.014
Outcome (O)	MD	0	3	2	10	10	14	7	16	16	13	20	46	11
		-	.500	.200	.455	.303	.424	.206	.381	.333	.283	.435	.333	.157
	GR	0	1	3	1	9	7	18	17	16	17	18	71	58
		-	.167	.300	.045	.273	.212	.529	.405	.333	.370	.391	.514	.829

Table 5.16 Number of cases and conditional probability of outcome (O) given the 4-7 day best coma score (CS) for data set D_3

		Age (A)							
		0-9	10-19	20-29	30-39	40-49	50-59	60-69	≧70
Outcome (O)	SD	8 .099	17 .139	14 .192	22 .306	28 .364	19 .288	11 .393	5 .556
	MD	18 .222	35 .287	23 .315	30 .417	33 .429	19 .288	8 .286	2 .222
	GR	55 .679	70 .573	36 .493	20 .278	16 .208	28 .424	9 .321	2 .222

Table 5.17 Number of cases and conditional probability of outcome (O) given age (A) for data set D_3

outcome, older cases having a higher probability of poor outcome. Thus, the inclusion of age had made a significant contribution to the results of INDEP4.

The results of this chapter suggested ways in which performance might be further improved. Age had a major effect on the results of INDEP4 and incorporation of age into the models based on trends would hopefully improve their performance. In conjunction with this, it seemed appropriate to attempt to improve the performance of INDEP3 in relation to INDEP2 by reducing the number of parameters to be estimated. This is carried out in Chapter 6.

CHAPTER 6

THE INTRODUCTION OF AGE INTO THE MODELS

6.1 Introduction

The effect of age on outcome after severe head injury is well recognised (Carlson et al., 1968; Teasdale et al., 1979b). It was included as an indicant to predict outcome both in the early research described in Section 3.4 and the comparative study described in Section 3.5. While my main interest was to use the trend in the patient's state through time to try to identify the individuals who ultimately would be severely disabled, the contribution which age made in the model INDEP4 could not be ignored. It was therefore decided to incorporate age into the formal structure of the other models in order to examine if performance could be improved further.

6.2 Constrained Normal Model with Age

Age was incorporated into the constrained normal models in two different ways. In the first it was assumed that, within outcome categories, age and the coma scores were independent, while in the second age was incorporated using linear regression on the coma score.

6.2.1 Model Assuming Independence of Age and Coma Score

One of the simplest ways of introducing age into the constrained normal model was to assume that age and the coma scores were independent within outcome categories. Thus if x is the feature vector of coma scores and x_a is age, then the joint

distribution of the new feature vector (x,x_a) is given by the product of the two marginal distributions

$$p(x,x_a) = p(x)\ p(x_a).$$

The average quadratic and logarithmic scores of the model CONOR3, where the distribution of the feature vector was assumed to be $N(M_i\theta,\Sigma_i)$, gave consistently poorer results than the models CONOR1 and CONOR2, where a single covariance matrix was assumed. CONOR3 was therefore not considered here. The density $p(x)$ was thus assumed to be that of a $N(M_i\theta,\Sigma)$ random vector, as initially described in Section 4.5. Age was assumed to be normally distributed with mean λ_i and variance τ_i for outcome category Π_i, where i=1, 2, 3. Boxplots of the distribution of age within each outcome for the three data sets (Figure 6.1) suggest that this assumption is not unreasonable. The maximum likelihood estimates $\hat{\lambda}_i$ and $\hat{\tau}_i$ of the parameters λ_i and τ_i were calculated. Thus, as in Section 5.2, the probability that a new patient with feature vector (y,y_a) belongs to category Π_i is estimated by

$$p(\Pi_i \mid y,y_a,D) = \frac{p(\Pi_i)\hat{\tau}_i^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left[(y-M_i\hat{\theta})^T\hat{\Sigma}^{-1}(y-M_i\hat{\theta})+(y_a-\hat{\lambda}_i)^2\hat{\tau}_i^{-1}\right]\right\}}{\sum_{j=1}^{3} p(\Pi_j)\hat{\tau}_j^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left[(y-M_j\hat{\theta})^T\hat{\Sigma}^{-1}(y-M_j\hat{\theta})+(y_a-\hat{\lambda}_j)^2\hat{\tau}_j^{-1}\right]\right\}}$$

where $p(\Pi_i)$ is the prior probability of outcome Π_i,

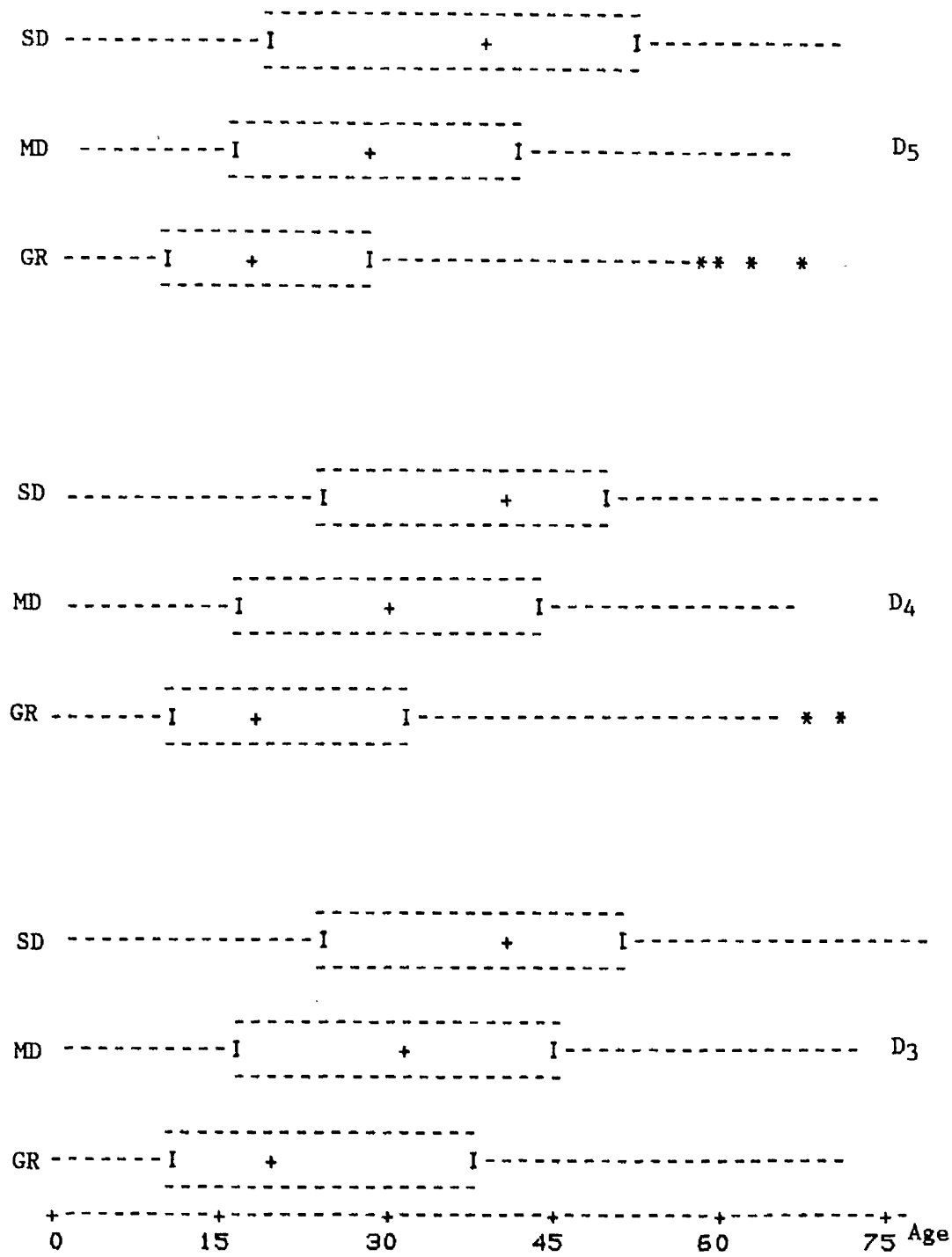
$\hat{\theta}$ and $\hat{\Sigma}$ are the appropriate estimates of θ and Σ,

and $\hat{\lambda}_i$ and $\hat{\tau}_i$ are the maximum likelihood estimates of λ_i and τ_i.

The models corresponding to CONOR1 and CONOR2 are referred to as **CONOR1A1** and **CONOR2A1** and have six more parameters to estimate as a result of the introduction of age. When the jack-knife technique was used with CONOR2A1 a recursive technique was again used to re-estimate λ_i and τ_i as in Section 5.2.

This method of incorporating age into the model assumed

Figure 6.1 Boxplots of the distribution of age within each outcome category for data set D_5, D_4 and D_3



independence between age and coma score within outcome categories. However, when these relationships were examined this assumption was clearly violated. This does not necessarily imply that such a model will not lead to good discrimination, as is discussed in Section 3.5.4. The relationship between age and coma score was such that for a given degree of recovery the lower coma scores were associated with younger patients. Table 6.1 shows the relationship between age and coma score at 28 days in the severe disability group. A similar pattern was obtained with the other outcome categories and other time periods.

To exploit the dependence between age and coma score, age was then incorporated into the model using linear regression on the coma score.

6.2.2 Model Using Linear Regression on the Coma Score

In this model age was related to coma score in the following way. The joint distribution of the feature vector (x,x_a) could be expressed as follows

$$p(x,x_a) = p(x_a|x)\ p(x)$$

The distribution of x was as described in the constrained normal models of Section 5.2 and $p(x_a|x)$ obtained by regressing age as a linear function of coma score. Initially, for each data set D_C, C = 5, 4, 3 and outcome category Π_i, i = 1, 2, 3, a stepwise procedure, using the BMDP program P2R (Dixon et al., 1985), was carried out to regress age on all coma scores available up to the time period considered. With the stepwise procedure at most two, and more often only one, coma scores were included in the regression equation before the stopping criterion was met. The particular score included in the equation varied according to the data set and outcome category considered. However, the best coma

Table 6.1 Relationship between age and the best coma score in the 15-28 day period for the severe disability group

Frequency		Age in years 0 - 29	30 - 49	≧ 50
Coma Score	3-10	10	2	3
	11-13	11	7	5
	14-15	4	10	11

score at 4-7 days was seen to be consistently and significantly related to age in all cases. As a result, the conditional distribution, $p(x_a|x)$, was modelled as $N(\alpha_i + \beta_i x_3, \sigma_i^2)$ where x_3 is the 4-7 day best coma score. A likelihood ratio test for each data set showed that modelling $p(x_a|x_3)$ as $N(\alpha_i + \beta_i x_3, \sigma^2)$ did not give a significantly poorer fit so this assumption of common variance was adopted for incorporation into the model. As age was now being introduced into the model in a far more structured fashion, the more general model CONOR3 was again included in order to confirm that, as before, it performed poorly relative to the models with a common covariance matrix. The models corresponding to CONOR1, CONOR2 and CONOR3 are referred to as **CONOR1A2, CONOR2A2** and **CONOR3A2** and had seven more parameters as a result of the introduction of age. When the jack-knife technique was used with CONOR2A2 and CONOR3A2 a recursive technique was again used to re-estimate α_i and β_i.

6.2.3 Results for the Constrained Linear Model with Age

Three aspects of the results for the constrained linear models with age will be presented. First, the results obtained with CONOR1A2, CONOR2A2 and CONOR3A2 are discussed; these would be expected to be similar to the pattern of results with CONOR1, CONOR2 and CONOR3 in Section 5.2. Next, the relative performance of the two different methods of introducing age, A1 and A2, will be considered, and finally, the effect of the introduction of age on performance examined by comparing the results including age to the corresponding results obtained without the inclusion of age.

The results of the five constrained normal models including age i.e. CONOR1A1, CONOR2A1, CONOR1A2, CONOR2A2, CONOR3A2 are given in Tables 6.2 - 6.7. The classification matrices for data sets D_5, D_4

and D_3 are given in Tables 6.2, 6.3 and 6.4 respectively and the separation measures in Tables 6.5, 6.6 and 6.7 respectively.

These results confirmed that the general pattern of results within the constrained models including age was similar to that for the results without age. The results with the jack-knife technique were worse than those obtained without it. The use of the arrival rate priors once more gave slightly better results than did equal priors. The performance of CONOR1A2 and CONOR2A2 were identical. With the jack-knife technique, both were better than CONOR3A2 in terms of the average logarithmic and quadratic scores. Without the jack-knife, the three were similar.

When the method A1 of incorporating age into the model was compared with A2, where age was introduced as a function of the 4-7 day best coma score, the classification matrices were similar. However when the separation measures were considered the average logarithmic and quadratic scores for A2 were consistently slightly better than those for the corresponding A1 model. In most cases the error rate for A2 was also lower than that for A1.

The effect on performance of the introduction of age into the model was first studied using the classification matrices. The results in Tables 6.2 - 6.4 were compared with those in Tables 5.1 - 5.3 respectively and showed that the addition of age considerably improved the classification matrices, with more cases being classified correctly, particularly in the middle, moderate disability group. Also, fewer good recoveries were classified as severe disability and vice versa.

The separation measures also confirmed the improvement in performance. A comparison of Tables 6.5 - 6.7 with Tables 5.4 - 5.6 respectively, showed that either method of incorporating age gave markedly better results, no matter which measure was

Table 6.2 Classification matrices for the constrained normal methods including age and data set D_5

Priors	Method	Predicted Outcome	No Jack-knife Actual Outcome			Jack-knife Actual Outcome		
			SD	MD	GR	SD	MD	GR
	CONOR1A1	SD	35	13	7			
		MD	18	32	17			
		GR	10	38	66			
	CONOR2A1	SD	35	13	7	31	15	7
		MD	18	32	17	22	29	17
		GR	10	38	66	10	39	66
Arrival	CONOR1A2	SD	36	14	5			
		MD	19	33	18			
Rate		GR	8	36	67			
	CONOR2A2	SD	36	14	5	36	14	5
		MD	19	33	18	19	33	21
		GR	8	36	67	8	36	64
	CONOR3A2	SD	41	15	5	39	19	5
		MD	13	33	17	16	25	21
		GR	9	35	68	9	39	64
	CONOR1A1	SD	43	22	8			
		MD	11	27	17			
		GR	9	34	65			
	CONOR2A1	SD	43	22	8	42	23	9
		MD	11	27	17	12	25	19
		GR	9	34	65	9	35	62
Equal	CONOR1A2	SD	47	24	9			
		MD	9	28	19			
		GR	7	31	62			
	CONOR2A2	SD	47	24	9	44	25	10
		MD	9	28	19	12	24	18
		GR	7	31	62	7	34	62
	CONOR3A2	SD	43	18	8	43	21	8
		MD	11	32	16	11	26	21
		GR	9	33	66	9	36	61

Table 6.3 Classification matrices for the constrained normal methods including age and data set D_4

Priors	Method		No Jack-knife			Jack-knife		
		Predicted Outcome	Actual Outcome			Actual Outcome		
			SD	MD	GR	SD	MD	GR
Arrival Rate	CONOR1A1	SD	54	19	12			
		MD	32	51	24			
		GR	17	68	129			
	CONOR2A1	SD	54	19	12	53	21	13
		MD	32	51	24	33	46	26
		GR	17	68	129	17	71	126
	CONOR1A2	SD	58	25	12			
		MD	33	55	36			
		GR	12	58	117			
	CONOR2A2	SD	58	25	12	56	25	14
		MD	33	55	36	35	55	37
		GR	12	58	117	12	58	114
	CONOR3A2	SD	60	21	12	57	21	12
		MD	26	56	38	28	55	43
		GR	17	61	115	18	62	110
Equal	CONOR1A1	SD	68	38	24			
		MD	29	50	29			
		GR	6	50	112			
	CONOR2A1	SD	68	38	24	67	40	24
		MD	29	50	29	30	48	31
		GR	6	50	112	6	50	110
	CONOR1A2	SD	70	39	21			
		MD	25	49	38			
		GR	8	50	106			
	CONOR2A2	SD	70	39	21	67	44	21
		MD	25	49	38	28	43	38
		GR	8	50	106	8	51	106
	CONOR3A2	SD	72	36	21	70	36	17
		MD	18	48	40	20	57	44
		GR	13	54	105	13	45	104

Table 6.4 Classification matrices for the constrained normal methods including age and data set D_3

Priors	Method		No Jack-knife			Jack-knife		
		Predicted Outcome	Actual Outcome			Actual Outcome		
			SD	MD	GR	SD	MD	GR
Arrival Rate	CONOR1A1	SD	53	31	12			
		MD	30	29	20			
		GR	41	108	204			
	CONOR2A1	SD	53	31	12	51	32	12
		MD	30	29	20	32	25	21
		GR	41	108	204	41	111	203
	CONOR1A2	SD	61	36	14			
		MD	31	44	40			
		GR	32	88	182			
	CONOR2A2	SD	61	36	14	60	36	14
		MD	31	44	40	32	42	42
		GR	32	88	182	32	90	180
	CONOR3A2	SD	64	31	17	62	33	18
		MD	27	52	43	28	51	47
		GR	33	85	176	34	84	171
Equal	CONOR1A1	SD	78	54	37			
		MD	29	51	48			
		GR	17	36	151			
	CONOR2A1	SD	78	54	37	77	57	38
		MD	29	51	48	30	45	54
		GR	17	63	151	17	66	144
	CONOR1A2	SD	77	51	31			
		MD	30	57	57			
		GR	17	60	148			
	CONOR2A2	SD	77	51	31	75	54	31
		MD	30	57	57	32	53	57
		GR	17	60	148	17	61	148
	CONOR3A2	SD	79	52	31	78	57	29
		MD	25	55	57	26	52	62
		GR	20	61	148	20	59	145

Table 6.5 Measures of separation for the constrained normal methods including age and data set D_5

Measure	Method	Arrival Rate Priors		Equal Priors	
		Jack-knife		Jack-knife	
		No	Yes	No	Yes
Error Rate	CONOR1A1	.436	-	.428	-
	CONOR2A1	.436	.466	.428	.453
	CONOR1A2	.424	-	.420	-
	CONOR2A2	.424	.436	.420	.449
	CONOR3A2	.398	.462	.402	.449
Average Logarithmic Score	CONOR1A1	.917	-	.923	-
	CONOR2A1	.917	.952	.923	.958
	CONOR1A2	.897	-	.901	-
	CONOR2A2	.897	.936	.901	.941
	CONOR3A2	.911	1.023	.914	1.025
Average Quadratic Score	CONOR1A1	.554	-	.556	-
	CONOR2A1	.554	.574	.556	.575
	CONOR1A2	.537	-	.540	-
	CONOR2A2	.537	.558	.540	.559
	CONOR3A2	.538	.582	.537	.580

Table 6.6 Measures of separation for the constrained normal methods including age and data set D_4

Measure	Method	Arrival Rate Priors		Equal Priors	
		Jack-knife		Jack-knife	
		No	Yes	No	Yes
Error Rate	CONOR1A1	.424	-	.434	-
	CONOR2A1	.424	.446	.434	.446
	CONOR1A2	.434	-	.446	-
	CONOR2A2	.434	.446	.446	.468
	CONOR3A2	.431	.453	.446	.431
Average Logarithmic Score	CONOR1A1	.918	-	.930	-
	CONOR2A1	.918	.937	.930	.949
	CONOR1A2	.902	-	.912	-
	CONOR2A2	.902	.922	.912	.932
	CONOR3A2	.908	.955	.916	.963
Average Quadratic Score	CONOR1A1	.554	-	.560	-
	CONOR2A1	.554	.566	.560	.572
	CONOR1A2	.543	-	.547	-
	CONOR2A2	.543	.555	.547	.558
	CONOR3A2	.542	.565	.542	.564

Table 6.7 Measures of separation for the constrained normal methods including age and data set D_3

Measure	Method	Arrival Rate Priors		Equal Priors	
		Jack-knife		Jack-knife	
		No	Yes	No	Yes
Error Rate	CONOR1A1	.458	-	.470	-
	CONOR2A1	.458	.472	.470	.496
	CONOR1A2	.456	-	.466	-
	CONOR2A2	.456	.466	.466	.477
	CONOR3A2	.447	.462	.466	.479
Average Logarithmic Score	CONOR1A1	.922	-	.947	-
	CONOR2A1	.922	.936	.947	.962
	CONOR1A2	.911	-	.935	-
	CONOR2A2	.911	.926	.935	.950
	CONOR3A2	.912	.940	.936	.964
Average Quadratic Score	CONOR1A1	.552	-	.568	-
	CONOR2A1	.552	.562	.568	.577
	CONOR1A2	.547	-	.558	-
	CONOR2A2	.547	.555	.558	.567
	CONOR3A2	.546	.560	.558	.572

considered. The improvement was so marked that, with one exception, all the constrained models which included age were better than the best model without age. With data set D_5 CONOR2 is slightly better than CONOR3A2 when the jack-knife technique is used. It is thus clearly important to incorporate age in some form into the model.

6.3 Normal Linear Model with Age

6.3.1 Practical Aspects

The incorporation of age into the classical linear discriminant model gave the model **NORLINA**. This involved the estimation of three more parameters for the mean age and four, five or six more parameters in the common covariance matrix, depending on whether data set D_3, D_4 or D_5 was being used, than with the corresponding model without age. In the constrained normal models, six extra parameters were required with method A1 and seven with A2. Thus with one exception, more additional parameters had to be estimated than with the inclusion of age into the constrained model. Again the biomedical computer program BMDP P7M was used to calculate the probabilities.

6.3.2 Results for the Normal Linear Model with Age

The classification matrices of the three data sets for the model NORLINA are given in Table 6.8 and the separation measures in Table 6.9. The effect of the jack-knife technique and the use of different priors was considered first, then the models NORLIN and NORLINA were compared.

The results of Tables 6.8 and 6.9 confirm that, as with NORLIN, the jack-knife technique gave poorer results for all the separation

Table 6.8 Classification matrices for method NORLINA

Data Set	Priors		No Jack-knife			Jack-knife		
		Predicted Outcome	Actual Outcome			Actual Outcome		
			SD	MD	GR	SD	MD	GR
D_5	Arrival Rate	SD	36	14	5	33	17	5
		MD	18	38	20	21	32	22
		GR	9	31	65	9	34	63
	Equal	SD	45	19	8	42	21	8
		MD	12	37	19	13	33	21
		GR	6	27	63	8	29	61
D_4	Arrival Rate	SD	60	25	13	57	28	13
		MD	33	53	28	35	47	30
		GR	10	60	124	11	63	122
	Equal	SD	71	43	22	70	43	22
		MD	24	43	35	25	43	35
		GR	8	52	108	8	52	108
D_3	Arrival Rate	SD	59	37	15	59	38	17
		MD	32	37	31	32	35	33
		GR	33	94	190	33	95	186
	Equal	SD	83	56	40	81	57	40
		MD	27	50	42	28	46	44
		GR	14	62	154	15	65	152

Table 6.9 Measures of separation for method NORLINA

Data Set	Measure	Arrival Rate Priors		Equal Priors	
		Jack-knife		Jack-knife	
		No	Yes	No	Yes
D_5	Error Rate	.411	.458	.386	.424
	Average Logarithmic Score	.884	.942	.888	.947
	Average Quadratic Score	.532	.566	.533	.567
D_4	Error Rate	.416	.443	.453	.458
	Average Logarithmic Score	.896	.923	.907	.934
	Average Quadratic Score	.538	.554	.542	.558
D_3	Error Rate	.458	.470	.456	.472
	Average Logarithmic Score	.907	.924	.932	.948
	Average Quadratic Score	.544	.554	.556	.565

measures and each data set. The arrival rate priors also gave better average quadratic and logarithmic scores than were obtained with equal priors.

Comparison of the results with those of the normal linear model without age in Tables 5.8 and 5.9 again showed that the inclusion of age produced a substantial improvement in all measures. As with the constrained models, more of the moderate disability group were correctly classified and fewer good recoveries were classified as severe disability and vice versa. Thus, the extra information that age provided in addition to that provided by the coma scores once more made a significant contribution to performance.

6.4 The Independence Model with Age

6.4.1 Practical Aspects

Because the incorporation of age substantially improved the results of the other methods, it should have had a similar beneficial effect on the independence model. However, an additional problem with the independence model that was highlighted in Chapter 5 was the large effect of the jack-knife technique.

The effect of the jack-knife was because the probability estimates are based on individual cell counts rather than on all the cases within the outcome category. Thus, the removal of one case was having a marked effect on the results. In an attempt to minimise this problem the cell counts were increased by grouping categories together. The aim was to bring the results obtained with the jack-knife technique closer to those in which it was not, without substantially worsening the latter. By grouping categories, a method that made use of the coma scores at all time

periods and age might also prove an improvement on the use of only the latest coma score available and age. Five sets of results with age incorporated into the independence model were thus obtained

(i) **INDEP2A:** As INDEP2 in Section 5.4 but with age included. The coma score was incorporated in exactly the same way as in that section and the age cells were taken from successive decades, 0-9 years, 10-19 years, etc.

(ii) **INDEP3A:** As INDEP3 in Section 5.4 but with age included as in INDEP2A.

(iii) **INDEP5A:** As INDEP3A but with only four age cells used, namely 0-19 years, 20-39 years, 40-59 years and ≥ 60 years.

(iv) **INDEP6A:** The number of coma score cells was reduced from thirteen to three at each time point and three cells were used for age. The category splits were chosen as follows so that there were approximately equal numbers in each of the splits:-

24 hour best coma score	3-6, 7-8, 9-15
2-3 day best coma score	3-7, 8-10, 11-15
4-7 day best coma score	3-9, 10-13, 14-15
8-14 day best coma score	3-10, 11-13, 14-15
15-28 day best coma score	3-13, 14, 15
Age	0-19, 20-39, ≥ 40

(v) **INDEP7A:** An intermediate between INDEP3A and INDEP6A where six cells for the coma score and seven for age were used as follows:-

24 hour best coma score	3-5, 6, 7, 8, 9-10, 11-15

2-3 day best coma score	3-6, 7, 8, 9-10, 11-12, 13-15
4-7 day best coma score	3-6, 7-8, 9-10, 11-13, 14, 15
8-14 day best coma score	3-8, 9-10, 11, 12-13, 14, 15
15-28 day best coma score	3-8, 9-10, 11, 12-13, 14, 15
Age	0-9, 10-19, ... , 50-59, ≧ 60

6.4.2. Results of the Independence Model with Age

The classification matrices for the above models corresponding to data sets D_5, D_4 and D_3 are given in Tables 6.10, 6.11 and 6.12 respectively and the separation measures in Tables 6.13, 6.14 and 6.15 respectively.

Two aspects were considered in results for the independence models with age: first, the pattern of results within the five models specified; and second, the comparison of these results with those of the independence models in Section 5.4.

When the classification matrices were studied the same general pattern seen with previous independence models was noted. Thus, the jack-knife technique gave fewer cases correctly classified and the arrival rate priors gave more cases predicted to make a good recovery and fewer to be severely disabled. A comparison of the different models on the basis of the classification matrices was difficult to make as these were very similar. However INDEP6A and INDEP7A, where there were fewer parameters to estimate, were less affected by the jack-knife technique. With INDEP2A more of the moderate disability group were classified correctly than with other methods while with INDEP6A fewer of this group were classified correctly than with other methods.

The measures of separation confirmed that the pattern of results in the models with age was similar to those without. Thus the results using the jack-knife technique were consistently worse

Table 6.10 Classification matrices for the independence methods including age and data set D_5

Priors	Method	Predicted Outcome	No Jack-knife: Actual Outcome SD	MD	GR	Jack-knife: Actual Outcome SD	MD	GR
Arrival Rate	INDEP2A	SD	32	14	7	21	18	7
		MD	26	44	24	35	33	27
		GR	5	25	59	7	32	56
	INDEP3A	SD	43	19	12	37	27	12
		MD	16	39	7	22	25	15
		GR	4	25	71	4	31	63
	INDEP5A	SD	42	18	13	38	26	13
		MD	16	39	7	19	27	14
		GR	5	26	79	7	30	63
	INDEP6A	SD	45	29	17	44	29	17
		MD	12	23	17	13	21	17
		GR	6	31	45	6	33	56
	INDEP7A	SD	43	20	12	41	27	13
		MD	15	38	7	17	24	14
		GR	5	25	71	5	32	63
Equal	INDEP2A	SD	45	28	12	39	31	12
		MD	13	30	19	18	26	22
		GR	5	25	59	6	26	56
	INDEP3A	SD	45	23	13	41	30	14
		MD	15	37	8	19	24	14
		GR	3	23	69	3	29	62
	INDEP5A	SD	46	22	13	40	28	15
		MD	13	37	8	18	27	17
		GR	4	24	69	5	28	58
	INDEP6A	SD	46	30	20	46	30	20
		MD	12	26	14	12	21	14
		GR	5	27	56	5	32	56
	INDEP7A	SD	47	26	15	43	30	16
		MD	11	33	7	15	23	16
		GR	5	24	68	5	30	58

Table 6.11 Classification matrices for the independence methods including age and data set D_4

Priors	Method	Predicted Outcome	No Jack-knife: Actual Outcome			Jack-knife: Actual Outcome		
			SD	MD	GR	SD	MD	GR
Arrival Rate	INDEP2A	SD	45	16	9	33	23	9
		MD	38	60	34	45	50	43
		GR	20	62	122	25	65	113
	INDEP3A	SD	67	40	20	52	44	23
		MD	27	51	28	38	33	37
		GR	9	47	117	13	61	105
	INDEP5A	SD	66	39	23	56	45	26
		MD	27	48	29	35	39	42
		GR	10	51	113	12	54	97
	INDEP6A	SD	67	41	26	66	44	26
		MD	21	27	16	22	19	19
		GR	15	70	123	15	75	120
	INDEP7A	SD	69	43	21	65	44	24
		MD	24	42	22	27	38	29
		GR	10	53	122	11	56	112
Equal	INDEP2A	SD	64	32	18	56	34	20
		MD	28	65	52	34	59	56
		GR	11	41	95	13	45	89
	INDEP3A	SD	73	48	27	64	49	31
		MD	23	51	32	32	42	41
		GR	7	39	106	7	47	93
	INDEP5A	SD	75	46	30	62	50	33
		MD	22	53	37	35	39	40
		GR	6	39	98	6	49	92
	INDEP6A	SD	72	49	32	72	49	32
		MD	20	30	22	20	29	22
		GR	11	59	111	11	60	111
	INDEP7A	SD	74	48	30	71	48	30
		MD	22	43	29	25	39	39
		GR	7	47	106	7	51	96

Table 6.12 Classification matrices for the independence methods including age and data set D_3

Priors	Method	Predicted Outcome	No Jack-knife: Actual Outcome SD	MD	GR	Jack-knife: Actual Outcome SD	MD	GR
			SD	MD	GR	SD	MD	GR
	INDEP2A	SD	47	25	12	40	30	13
		MD	46	65	36	50	57	39
		GR	31	78	188	34	81	184
	INDEP3A	SD	65	40	27	55	46	29
		MD	37	57	28	44	47	38
		GR	22	71	181	25	75	169
Arrival	INDEP5A	SD	65	43	29	58	48	32
		MD	36	62	31	42	50	41
Rate		GR	23	63	176	24	70	163
	INDEP6A	SD	73	48	36	73	49	36
		MD	22	31	16	22	20	16
		GR	29	89	184	29	99	184
	INDEP7A	SD	69	42	29	57	50	32
		MD	30	53	22	41	40	27
		GR	25	73	185	26	78	177
	INDEP2A	SD	68	39	29	64	47	30
		MD	38	70	44	42	63	48
		GR	18	59	163	18	59	158
	INDEP3A	SD	90	59	42	72	63	44
		MD	21	67	48	37	54	57
		GR	13	42	146	15	51	135
	INDEP5A	SD	90	56	46	74	62	49
Equal		MD	23	67	46	37	57	60
		GR	11	45	144	13	49	127
	INDEP6A	SD	80	60	43	80	60	43
		MD	27	36	48	27	35	29
		GR	17	72	165	17	73	164
	INDEP7A	SD	83	62	38	77	64	39
		MD	26	51	42	31	45	46
		GR	15	55	156	16	59	151

Table 6.13 Measures of separation for the independence methods including age and data set D_5

Measure	Method	Arrival Rate Priors		Equal Priors	
		Jack-knife		Jack-knife	
		No	Yes	No	Yes
Error Rate	INDEP2A	.428	.534	.432	.487
	INDEP3A	.352	.470	.360	.462
	INDEP5A	.360	.458	.356	.470
	INDEP6A	.470	.487	.458	.479
	INDEP7A	.356	.458	.373	.475
Average Logarithmic Score	INDEP2A	.878	.992	.886	.999
	INDEP3A	.829	1.279	.843	1.294
	INDEP5A	.838	1.267	.852	1.283
	INDEP6A	1.011	1.105	1.031	1.124
	INDEP7A	.930	1.161	.946	1.177
Average Quadratic Score	INDEP2A	.536	.605	.539	.607
	INDEP3A	.480	.679	.492	.689
	INDEP5A	.485	.671	.498	.681
	INDEP6A	.582	.629	.595	.641
	INDEP7A	.527	.636	.537	.644

Table 6.14 Measures of separation for the independence methods including age and data set D_4

Measure	Method	Arrival Rate Priors		Equal Priors	
		Jack-knife		Jack-knife	
		No	Yes	No	Yes
Error Rate	INDEP2A	.441	.517	.448	.498
	INDEP3A	.421	.532	.434	.510
	INDEP5A	.441	.527	.443	.525
	INDEP6A	.466	.495	.475	.478
	INDEP7A	.426	.479	.451	.493
Average Logarithmic Score	INDEP2A	.881	.962	.893	.974
	INDEP3A	.864	1.108	.884	1.127
	INDEP5A	.878	1.105	.898	1.125
	INDEP6A	.970	1.016	.990	1.036
	INDEP7A	.916	1.030	.936	1.050
Average Quadratic Score	INDEP2A	.534	.585	.539	.589
	INDEP3A	.521	.641	.534	.652
	INDEP5A	.528	.639	.542	.651
	INDEP6A	.579	.605	.591	.617
	INDEP7A	.543	.606	.556	.617

Table 6.15 Measures of separation for the independence methods including age and data set D_3

Measure	Method	Arrival Rate Priors		Equal Priors	
		Jack-knife		Jack-knife	
		No	Yes	No	Yes
Error Rate	INDEP2A	.432	.468	.430	.462
	INDEP3A	.426	.487	.426	.506
	INDEP5A	.426	.487	.430	.511
	INDEP6A	.455	.475	.468	.472
	INDEP7A	.419	.481	.451	.483
Average Logarithmic Score	INDEP2A	.883	.961	.908	.986
	INDEP3A	.864	1.026	.895	1.056
	INDEP5A	.884	1.030	.915	1.061
	INDEP6A	.940	.969	.967	.996
	INDEP7A	.894	.970	.922	.997
Average Quadratic Score	INDEP2A	.531	.574	.545	.588
	INDEP3A	.520	.599	.538	.617
	INDEP5A	.531	.603	.550	.621
	INDEP6A	.563	.580	.579	.596
	INDEP7A	.534	.577	.551	.593

than those without, with the difference in results being larger when there were more parameters to estimate or fewer cases in the data set. Arrival rate priors gave better results than equal priors.

In the comparison of the different models with age INDEP2A and INDEP3A were first compared. INDEP3A performed better than INDEP2A without the jack-knife technique but was worse than INDEP2A with it. However when these results were compared to the corresponding models without age, INDEP2 and INDEP3, the results with age were, in general, better. Surprisingly, with INDEP2A and data set D_5 the results with the jack-knife technique gave almost identical average logarithmic and quadratic scores with and without age, even though the actual probabilities and classification matrices were quite different. This was the one instance in which the results with age were not markedly better than those without.

The original motivation for the inclusion of age into the model was the performance of INDEP4. With no jack-knife technique the results of INDEP2A and INDEP3A were always better than INDEP4. With the jack-knife technique INDEP2A always performed better than INDEP4 but INDEP4 was usually better than INDEP3A.

As INDEP3 did not perform better than INDEP2, the models INDEP5A, INDEP6A and INDEP7A were considered. In these models the variables were grouped to reduce the number of parameters to be estimated in the hope that the performance of INDEP3A would be improved when the jack-knife technique was used. A pattern emerged when this was carried out. With no jack-knife technique the pattern in the average logarithmic and quadratic scores was such that, as the number of parameters was reduced the performance dateriorated. Thus the order of performance from best to worst was – INDEP3A, INDEP5A, INDEP7A then INDEP6A. However, as mentioned

previously, more weight should be placed on the results based on the jack-knife technique. When these results were considered the pattern was, in general, reversed with the order of performance from best to worst being – INDEP6A, INDEP7A, INDEP5A then INDEP3A. Thus grouping the variables had improved the performance when the jack-knife technique was used. The results of INDEP6A and INDEP7A were similar. The same was true for INDEP3A and INDEP5A.

With the jack-knife technique, INDEP2A, as well as being better than INDEP3A, was also better than INDEP5A, INDEP6A and INDEP7A. Without the jack-knife technique, INDEP2A gave a performance which was comparable to that of INDEP5A.

In summary, these results showed that the inclusion of age into the independence model was successful in improving performance, but attempts at bettering the performance of INDEP2A by a model including the coma scores at all time periods was unsuccessful. If only one independence model had to be chosen then INDEP2A, which used the latest available coma score and age, would be the model of choice.

6.5 Comparison of Models Including Age

6.5.1 Comparison of Models Assuming Normality

As in the previous chapter, when assessing the relative performance of the models in Sections 6.2, 6.3 and 6.4, the models based on normality assumptions were compared first.

The results of the classification matrices for the model NORLINA in Table 6.8 can be seen to be very similar to the corresponding matrices for the constrained models in Tables 6.2, 6.3 and 6.4. As CONOR2A2 had the best performance of the constrained models in terms of the average logarithmic and

quadratic scores, this constrained model was compared with the model NORLINA. The results of both methods proved to be very close indeed. NORLINA performed better than CONOR2A2 when no jack-knife technique was used. With the jack-knife technique and data set D_5 CONOR2A2 performed slightly better than NORLINA, with data set D_4 the results were very similar, while with data set D_3 NORLINA performed slightly better than CONOR2A2.

If only one normal based model were to be used NORLINA would be the clear choice. It was consistently better than CONOR2A2 when no jack-knife technique was used and although the results with the jack-knife technique should be given more weight, here it did perform slightly better than CONOR2A2 with data set D_3.

6.5.2 Comparison of Independence and Normal Based Models

Here, as before, only the separation measures are used for this comparison. As in the previous section the normal models considered are CONOR2A2 and NORLINA.

The jack-knife technique produced clear results in terms of the average quadratic and logarithmic scores: both normal based methods performed better than any of the independence based methods. This was also true for the error rates with the exception of the result for data set D_3 and equal priors where INDEP2A had a lower error rate than CONOR2A2 but not NORLINA. With no jack-knife technique the performance of the normal based models was similar to that of INDEP7A.

If an overall choice of model had to be made then again NORLINA would be the choice. The fact that it performed consistently better than the independence methods when the jack-knife technique was used outweighs its poorer performance without this technique. Some idea of the size of the bias introduced by using the same test

and training set is obtained by comparing the results with and without the jack-knife technique. It is obvious from the results that with independence models, where there are more parameters to estimate, this bias is sufficiently large to make it essential to use a jack-knife or split-sample approach to assess performance.

6.6 Discussion

The main aim in this chapter was to expand the models in Chapter 5 to add age to the trend in coma score in order to see if performance was improved. This was very successful in that the performance of all models was substantially improved when age was included. It is difficult to find a reason for the one anomalous set of results with data set D_5 and the jack-knife technique, where the performance of INDEP2A was almost identical to that of INDEP2. Neither the small size of the data set nor the number of parameters provides the explanation; INDEP3A, for example, has many more parameters and does improve on the performance of INDEP3 with the same data set.

Thus these results showed that age was associated with quality of survival even after conrolling for trend in coma score. This association was in the expected direction with younger patients having a better prognosis and agrees with the conclusions of Teasdale et al. (1982b) who looked at the effect of age on survival after severe head injury and also found that, even controlling for the best 24 hour coma score, age affected survival with younger patients more likely to have a good outcome than older ones.

The grouping of categories to reduce the number of parameters in the independence model was less successful in improving performance. INDEP2A still performed better than the other

independence models which included all coma scores available. With the independence models the bias introduced by using the same test and training sets was larger than with the normal based models. This was to be expected as the independence models, even with categories merged, had more parameters to estimate.

Overall the results were such that the normal based methods were preferable. When it came to the choice of normal based model, the normal linear model NORLINA had similar but slightly better performance than the constrained model CONOR2A2. Moreover the fitting of such a specific model as CONOR2A2 involves considerable programming whereas a range of statistical packages are widely available to perform the analysis of NORLINA. This is clearly a factor in favour NORLINA.

Although the best models in this chapter which include all coma scores and age do give considerable improvemant in predicting the quality of survival over the method currently used, they have not given a clinically practical improvement. The probabilities produced are frequently in the range 0.3 - 0.6 with no outcome being predicted 'confidently'. This results in poor separation of the outcome categories. The different approaches which might be adopted to solve this problem along with some wider practical aspects are considered in the final chapter.

CHAPTER 7

PRACTICAL IMPLICATIONS AND FURTHER WORK

7.1 Introduction

The aim at the outset of this work was to use data collected through time to predict the quality of survival in individuals after severe head injury and in particular to identify cases who would be severely disabled. The results obtained have fallen short of this target and before proceeding further it is important to consider carefully any modifications that might lead to improvements.

There are basically two factors that determine performance, namely the statistical model and the patient data included. The ways in which these factors could be altered to improve performance will now be considered.

7.2 Review of Statistical Models

The range of statistical models considered here was less comprehensive than that of the comparative study (Titterington et al., 1981) described in Section 3.5. However the performance of the highly specific model developed to take account of the trend in coma score through time was compared with that of the models which gave good performance in the comparative study.

In Chapter 5 the differences in the results obtained with these various models were considered. In practice, these differences were small relative to the effect of the introduction of age. This agrees with the finding of the comparative study that the choice of model is less important than the choice of data included, provided

that account is taken of the assumptions on which the model is based. Thus in any new work it would be worthwhile to look at the results available from standard discriminant methods before embarking on fitting complex models which require extensive programming. The fact that pseudo maximum likelihood methods gave almost identical results to the full maximum likelihood method further emphasises this point.

7.3 Review of Data

7.3.1 The Data Bank at Present

The coma score had proved over the years to be the item of clinical data which provided the most information about prognosis. It therefore seemed appropriate to choose the coma score to model trend in recovery. Its lack of success in predicting quality of survival prompts a scrutiny of the components of the coma score and how these relate to neurological function.

Figure 7.1 shows the major parts of the brain and an overview of these provides one possible reason for the relative lack of success of the coma score in predicting quality of survival. The brain consists mainly of two large convoluted masses or cerebral hemispheres, which together form the cerebrum, the structures in the 'posterior fossa' or brain stem (which is formed by the midbrain, pons and medulla oblongata), and the cerebellum. Each hemisphere is composed of an outer layer of grey matter called the cortex and an inner core of white matter with embedded grey matter constituting the corpus striatum. The cortex is essential for "higher functions" such as volitional movement, sensory perception, speech and especially aspects of mental performance and personality, for example, memory, intelligence, emotion and

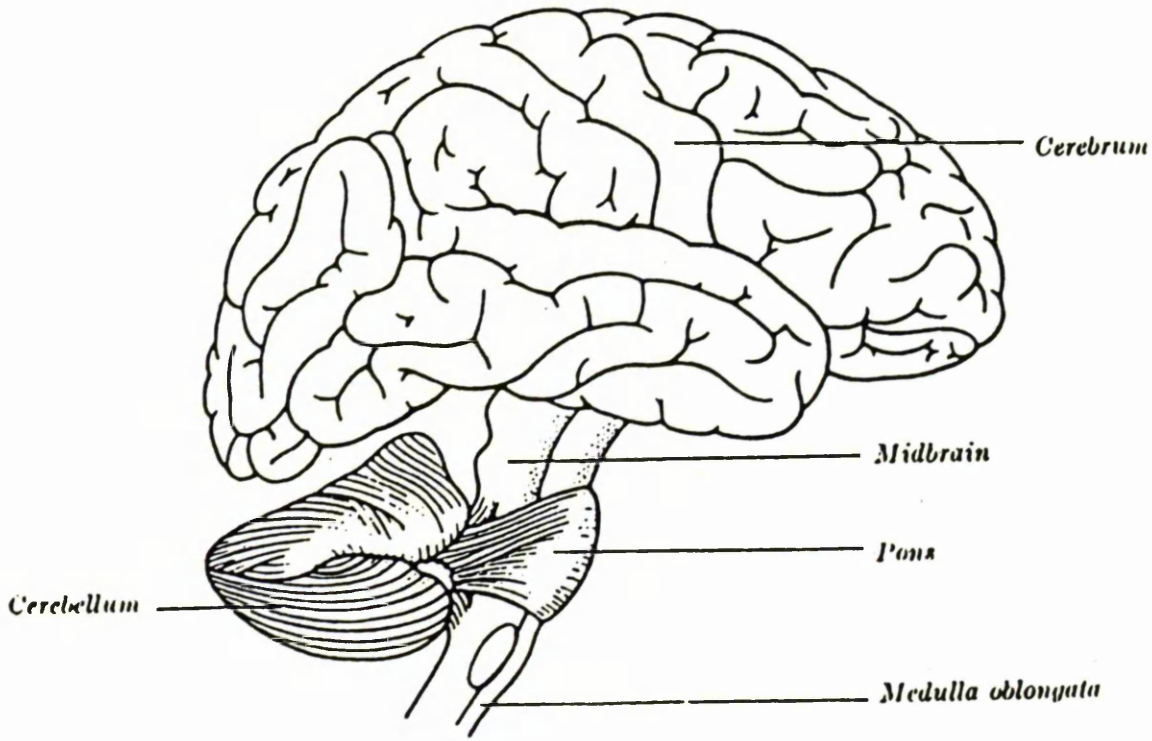


Figure 7.1 Illustration of the major parts of the brain (adapted from Gray's Descriptive and Applied Anatomy, 32nd edition, 1958. Eds. Johnston TB, Davies DV and Davies F. Longmans Green and Co. London.)

behaviour. Nervous transmission from the central hemispheres to the rest of the body occurs via the brain stem which also plays a more basic role in maintaining consciousness. Thus, the indices of brain stem responsiveness reflect the degree of impairment of consciousness, and whether or not a patient is in coma, while the cortical responses indicate the 'content of consciousness'. The latter is reflected only by the upper levels of the coma scale: a motor score of 6 and verbal scores of 3, 4 or 5. Lower scores and impaired eye opening, that is, overall coma scores between 3 and 11, mainly indicate brain stem dysfunction. The coma score may thus be successful in the prediction of death or survival, but less able to predict the quality of survival if consciousness is regained.

However, there are clinical and physiological reasons to expect that the depth and duration of coma, and the rate of improvement over the range of responsiveness reflected by the GCS, provide valid indices of the degree of diffuse brain damage. Studies in Glasgow using the new sensitive techniques of magnetic resonance imaging (MRI) confirm that such diffuse damage underlies both the depth of coma in the earlier stages (Jenkins et al., 1986) and the severity of disability in survivors (Wilson et al., 1988). However the balance between such diffuse damage and lesions in specific areas of the cortex in causing disability is still a topic of controversy. Furthermore, studies using electrophysiological methods showed that the brain stem response was not as good as the cortical response in predicting outcome (Cant et al., 1986; Lindsay et al., 1988).

Can other features recorded in the Head Injury Study data bank give a better performance than the coma score in the prediction of the quality of survival? The pupil reaction and eye signs are even

more closely related to brain stem dysfunction and are unlikely to help. As mentioned above, the feature that most obviously reflects cortical activity is the verbal component of the coma scale. To discover if the other components were masking information in the verbal score, predictions of outcome were made using a normal linear model and a feature vector restricted to verbal scores and age. This model was called **NORLINVA.** The classification matrices and separation measures for the three data sets are given in Tables 7.1 and 7.2 respectively. However, when these are compared with the results for the corresponding model NORLINA in Tables 6.8 and 6.9 respectively it is clear that the results are similar, with NORLINVA being in general slightly worse than NORLINA. This result may mean that even more sensitive and specific measures of cortical function than provided by the verbal response are needed. Such approaches include the more detailed prospective testing of orientation and amnesia as described by Levin et al. (1979) and more complex analysis of mild impairment of consciousness (Sano et al., 1983).

An alternative view is that the biological process of recovery is inevitably so variable that any measure of brain damage and dysfunction over the first month will be unable to provide an accurate guide to outcome six months or one year later in individual cases. The influence of rehabilitation may also need to be taken into account.

7.3.2 The Data Bank of the Future

At this point it is relevant to remember that two decades have elapsed since the Head Injury Study data bank was initiated. It was designed in such a way that it was based on readily available clinical data. Other investigations are now available (Figure

Table 7.1 Classification matrices for method NORLINVA

Data Set	Priors		No Jack-knife			Jack-knife		
		Predicted Outcome	Actual Outcome			Actual Outcome		
			SD	MD	GR	SD	MD	GR
D_5	Arrival Rate	SD	35	17	6	34	21	6
		MD	22	32	22	21	28	25
		GR	6	34	62	8	34	59
	Equal	SD	44	28	9	42	30	9
		MD	14	25	21	15	21	22
		GR	5	30	60	6	32	59
D_4	Arrival Rate	SD	56	25	14	52	26	14
		MD	36	57	32	40	56	34
		GR	11	56	119	11	56	117
	Equal	SD	71	49	24	71	50	24
		MD	22	39	31	22	38	33
		GR	10	50	110	10	50	108
D_3	Arrival Rate	SD	53	28	13	52	31	15
		MD	40	46	40	41	40	40
		GR	31	94	183	31	97	181
	Equal	SD	85	58	34	82	62	34
		MD	26	52	47	29	47	47
		GR	13	58	155	13	59	155

Table 7.2 Measures of separation for method NORLINVA

Data Set	Measure	Arrival Rate Priors		Equal Priors	
		Jack-knife		Jack-knife	
		No	Yes	No	Yes
D_5	Error Rate	.453	.487	.453	.483
	Average Logarithmic Score	.881	.938	.887	.944
	Average Quadratic Score	.535	.570	.537	.572
D_4	Error Rate	.429	.446	.458	.466
	Average Logarithmic Score	.905	.931	.918	.944
	Average Quadratic Score	.543	.558	.548	.564
D_3	Error Rate	.466	.483	.447	.462
	Average Logarithmic Score	.913	.928	.939	.955
	Average Quadratic Score	.546	.555	.558	.568

7.2). Biochemical tests can reflect the extent of tissue damage and many different enzymes have been claimed to be associated with outcome. Rabow and Hedman (1985) and Hans et al. (1983) related the enzyme creatine kinase $_{BB}$ to outcome in head injury while Rao et al. (1978) used serum lactate dehydrogenase and Thomas et al. (1978) serum myelin basic protein. In general these studies were based on small numbers of patients and showed a correlation with outcome. Thomas et al. (1979) concluded abnormal test results are associated with a decrease in the coma score, but that more extensive studies were needed to discover the value of biochemical results in prognosis if clinical details such as the GCS of the patient were known.

Lindsay et al. (1988) found that, while electrophysiological results were useful as a prognostic guide in paralysed or sedated patients, they were of little value over the clinical information and the small benefit did not justify the effort involved in data collection.

At the time the Head Injury Study was initiated, computerised tomography (CT) was in its infancy. Since then CT scanning has become widely available; this has improved the detection of secondary intracranial haematomas but it has been relatively insensitive to primary brain damage. New scanning techniques which further enhance the images of the brain have since been developed and Jenkins et al. (1986) and Hadley et al. (1988) concluded that MRI can provide a striking picture of the effects of a head injury on the brain. While MRI looks at the structure of the brain, positron emission tomography (PET) looks at function by measuring glucose metabolism. Langfitt et al. (1986) have begun to evaluate the extra advantages of PET and MRI over CT but point out the problems associated with the large amount of data produced by one

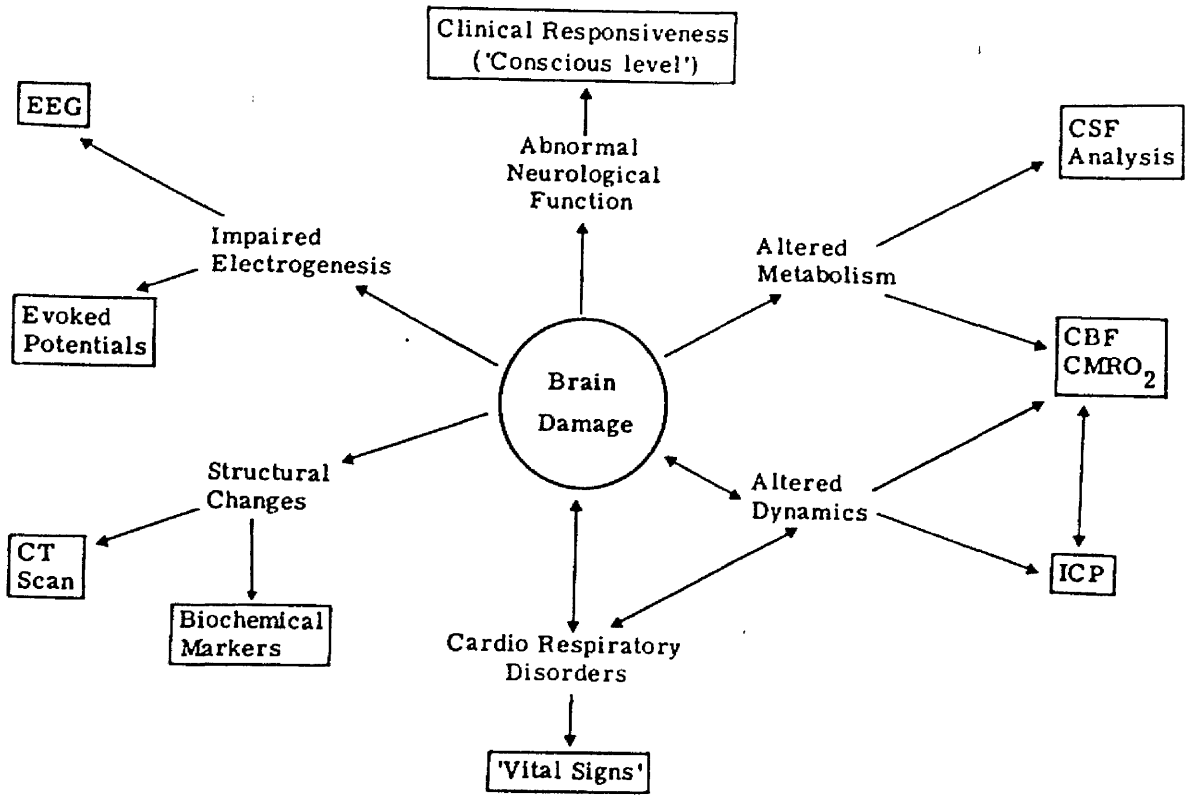


Figure 7.2 Consequences of brain damage and the methods by which these can be detected (adapted from Management of Head Injuries, 1981. Jennett B and Teasdale G. FA Davis. Philadelphia.)

image. Development of these and other scanning techniques is still progressing and if the results obtained can be summarised to give accurate measures of cerebral function then these measures may provide more accurate predictions of the quality of survival.

7.3.3 Choice of Outcome Categories

None of the models was outstandingly successful in identifying those individuals who would remain disabled after their injury. However the task set was itself a difficult one because of the ordered nature of the outcome categories. The problems of classification with three ordered outcomes, discussed in Section 3.5.4, were apparent in the results of Chapter 5 and Chapter 6, with the moderate disability group always having a large number of cases misclassified as severe disability or good recovery. While at the outset of this research it seemed appropriate to try to separate all three categories, I now feel that, when the ordered nature of the outcome is considered, this was too ambitious a goal, and with hindsight it might have been better to have used only two outcome categories. This could have been achieved by merging the moderate disability and good recovery groups to form an independent survivors group, with the severe disability patients making up the dependent survivors. The burden imposed on society by the dependent survivors was discussed in Sections 1.4 and 2.3.1 and makes the dependent-independent split of the survivors the clinically and socially relevant one to make. Indeed in all the initial work on discriminant analysis in the Head Injury Study the moderate disability and good recovery patients were considered as one category so I would suggest this approach in preference to trying to incorporate the ordered structure of the outcome into the model.

The vegetative patients and those who died after 28 days have not been considered in this work, although, if trying to make a prospective prediction of outcome these are indeed possibilities. One reason for excluding them was that the number of these cases was too small to estimate reliably any parameters involved in model fitting. Figures 4.1(a) - 4.1(e) however showed that these cases had a quite different coma score profile from the other outcomes. If these cases were of concern then a two stage approach could be adopted. The first step would be to discriminate between those likely to be dead or vegetative versus some better outcome, and only then to predict the quality of survival of those who had higher than a predetermined probability of this better outcome.

7.3.4 Missing Data

This work was based only on cases with complete data up to a particular time point but this is clinically unrealistic. With this data set, for example, patients increasingly are being ventilated for short periods during their stay and such patients often have their best verbal score missing. Indeed in some units it is hospital policy to ventilate certain head injured patients routinely. No matter what data are used, always, some will be found missing. Thus blood may not be sampled, charts and biochemical results go astray and machines malfunction. Some method to cope with missing data must be developed if a model is to be practicable.

With data collected through time it is clearly dangerous to extrapolate beyond the last observed value. Nevertheless it should be possible to accommodate data missing at earlier time periods. This might be done simply by using group means to substitute missing values; a more sophisticated approach is to use the EM

algorithm to fit the model, using maximum likelihood based on tha available (incomplete) data. An appealing method of dealing with missing values, if the same variable is measured repeatedly, is to use a growth curve approach to fit an appropriate curve to the data present for each individual; the missing values for that individual could then be interpolated from the curve.

7.3.5 Timing of Predictions

In Chapters 5 and 6 no comment was made on the time of prediction beyond saying that the data from the latest time period (INDEP2) gave a better performance than that from the first 24 hours (INDEP1). This was because the predictions made at the end of the different time periods were based on different cases. To examine whether there is any gain in waiting till 14 days or 28 days to obtain extra data, predictions were made at 7 days, 14 days amd 28 days with data set D_5. The normal linear model with the feature vector consisting af all verbal scores up to the time of prediction and age was used. The average logarithmic and quadratic scores are given in Table 7.3. These results showed that there was a consistent benefit in making the predictions at 28 days rather than 14 or 7 days. The only exception was that the average quadratic scores using the jack-knife technique and arrival rate priors differed only slightly at the three times. It is doubtful whether this particular benefit is clinically useful but it offers hope that when new data are collected through time, the additional data will provide additional information. This will be worthwhile so long as the cost of obtaining the extra information is less than its benefit.

Table 7.3 Average quadratic and logarithmic scores for data set D_5 when predictions are made at 4-7 days, 8-14 days and 15-28 days using the normal linear model NORLINVA

Measure	Time of Prediction	Arrival Rate Priors		Equal Priors	
		Jack-knife		Jack-knife	
		No	Yes	No	Yes
Average Logarithmic Score	15-28D	.881	.938	.887	.944
	8-14D	.919	.945	.936	.962
	4-7D	.937	.951	.958	.972
Average Quadratic Score	15-28D	.535	.570	.537	.572
	8-14D	.555	.570	.565	.581
	4-7D	.561	.570	.576	.585

7.4 Predictions in Practice

All the work on statistical modelling would be of little more than academic interest if the predictions were not acceptable to the clinicians involved. Currently a study examining the reliablity and acceptability of such predictions, and their impact on clinical practice, is underway in four British centres (Glasgow, Edinburgh, Liverpool and Southampton). This study has three distinct phases.

In the first phase the current practice and resource utilisation are monitored so that a baseline is established. In the second phase, the computer prediction of outcome at various stages after injury is brought to the attention of those caring for the patient. The program was modified to make it 'user friendly' and run on a microcomputer in the intensive care ward by any of the medical staff involved in the care of the patients. In the last phase the predictions are withdrawn.

During all three phases, the treatment given and use of resources are monitored to see if the provision of the predictions influences management practices. The withdrawal period is an attempt to determine if any change in practice is sustained. At present the study is nearing the end of phase three. A formal analysis of the results has still to be carried out but many of the clinicians have asked for the computer predictions of certain patients during the withdrawal phase. Thus, whether or not there has been any measurable effect on management, clinicians are interested in predictive information about a particular patient. Their interest might be solely to attach a figure to the probability of the various outcomes on a discharge summary or to inform relatives about the likely outcome, and these are considered

useful applications. Whether there are further reaching consequences, leading to an effect on outcome, or at least more efficient or appropriate, consistent use of resources, remains to be seen.

It could be argued that in the short term prognosis is not important as all patients should have the same high standard of care offered. Long term care and rehabilitation are perhaps in even shorter supply than acute intensive care and hence have a high premium. With resources limited, the predictions after the first week become increasingly important as these are the ones which may identify possible groups of patients who might benefit from treatment. Alternatively it would be useful to know that prolonged rehabilitation will be fruitless, so that efforts can be directed to coping with and adjusting to limitations.

The work in this thesis was the first attempt to use the recovery trend over the first month to predict a patient's ultimate degree of recovery. The results demonstrate how difficult the problem is. The coma score has been successful in predicting death or survival after a severe head injury, but modelling trend through time did not successfully predict the quality of survival. Such an ability would have considerable value in planning the management of individual patients and in the rigorous, efficient comparison of different methods of rehabilitation and late care.

Perhaps the most important implication of the present study, for those who will undoubtedly work towards this goal in future, is that the quality of information contained in the data and an adequate number of cases are of primary importance. Severe head injuries fortunately are relatively rare events and a data bank of this type will be essential if the original aim is to be achieved. Shortcomings in information can not be compensated for by

sophisticated statistical techniques; moreover there is a clear implication from this work that simple methods are to be preferred. Whatever new features are analysed, the practical principles described in this thesis and embodied in the methods used to collect data about early severity and outcome from coma will still be relevant.

APPENDIX 1

HEAD INJURY STUDY DATA COLLECTION FORM

Coma Study No. | 1 | | | | | | |

SEVERE HEAD INJURY STUDY

University Department of Neurosurgery,
Institute of Neurological Sciences,
Southern General Hospital,
Glasgow, G51 4TF.

University Department of Neurosurgery,
Institute of Neurological Sciences,
Southern General Hospital,
Glasgow, G51 4TF.

SEVERE HEAD INJURY STUDY

Identifying Characteristics

Coma Study Number

a) First box = card number (printed).
b) Second and third boxes = centre code (issued by Glasgow).
c) Remaining boxes = consecutive study number allocated by centre to each patient (first will be 0001).

The coma study number occurs again at the beginning of each card, and must be filled in for identification.

Name - fill last name from left till boxes or name complete; leave blank if confidentiality rules require this.

Unit Number - additional identification - usually is hospital record number (Each centre should keep careful cross-tabulation of study number, unit number and name).

Date of injury - note order .....

1	7	0	1	7	8
Day		Month		Year	

Study type - original data bank was limited to patients in coma for 6 hours; to allow inclusion of other cases, fill in Box 31 as follows:—

Coma > 6 hours	= 1
Coma < 6 hours	= 2
Died (or brain death) < 6 hours	= 3
Never in coma	= 4

Note definition of coma = EMV 1/5/2 or worse, i.e. no eye opening, not obeying commands and not uttering words.

Time Related Data

Time Epochs (24H, 2-3D etc.) refer to time from onset of coma; if coma is delayed, indicant 17 records this; if no coma, take epochs as time from injury.

Best/Worst within epochs - if no change then B = W.

Coma Score

The coma score in No. 55 is obtained by adding the three components of the coma scale (No. 138-173).

HEAD INJURY CODING FOR COMA PROGNOSIS

1. Coma Study No. (including Centre Code) [1][ ][ ][ ][ ][ ][ ] 1-7

2. Name [ ][ ][ ][ ][ ][ ][ ][ ][ ][ ] 8-18

   Unit No. [ ][ ][ ][ ][ ][ ] 19-24

3. Date of Injury [ ][ ][ ][ ][ ][ ] 25-30

4. Study Type (see opposite) [ ] 31

PERSONAL DATA

5. Sex

   Male = 1
   Female = 2 [ ] 32

6. Age (years) [ ][ ] 33-34

7. Handedness

   Right = 1
   Left = 2 [ ] 35

8. Pre-existing Medical Conditions causing Continuing Disability

   Cardiovascular = 1
   Respiratory = 2
   Renal = 3
   Gastrointestinal = 4
   Nervous System = 5
   Skeletal = 6
   Multiple = 7
   Other (including psychiatric) = 8 Specify _____
   None = 9 [ ] 36

9. Previous Head Injury

   No = 1
   Mild (PTA < 24 hrs) = 2
   Severe (PTA>24 hrs) = 3
   Indefinite PTA = 4 [ ] 37

10. Previous Epilepsy

    No = 1
    <1 year = 2
    >1 year = 3
    Frequency not known = 4 [ ] 38

11. Type of Injury

Motor vehicle occupant	= 1	☐	39
Pedestrian	= 2		
RTA other (or unknown)	= 3		
Sport	= 4		
Work	= 5		
Assault	= 6		
Domestic (+ fall from window)	= 7		
Fall under influence of alcohol	= 8		
Other (includes gunshot)	= 9		

12. Recent alcohol

No	= 1	☐	40
Suspected	= 2		
Definite	= 3		

13. Alcohol level

Actual value (not done = 999)		☐☐☐	41-43

14. Influence of other drugs on initial assessment

No	= 1	☐	44
Suspected	= 2		
Definite	= 3 Specify _____		

15. Lucid interval (= talked)

None	= 1	☐	45
Partial - words/confused	= 2		
Total - sensible/orientated	= 3		

Coding for 16, 17, 19, 20.

<6 hours	= 1
6-12 hours	= 2
13-24 hours	= 3
0-24 hours (unspecified)	= 4
2-3 days	= 5
4-7 days	= 6
>1 week	= 7
Gradual (undefined time)	= 8
Not known	= 9

16. Time to onset of deterioration (since injury) ☐ 46

17. Time from injury to onset of coma (= EMV 1/5/2 or worse) ☐ 47

18. If coded 1, 2 or 3 give exact hours, if possible ☐☐ 48-49

19. Time from injury to first admission to any hospital ☐ 50

20. Time from injury until admitted to neuroservice ☐ 51

CRANIAL INJURY

21. **Linear Fracture of Skull (X-ray or operation)** ☐ 52

None = 1
Vault = 2
Base = 3
Both = 4

22. **Depressed Fracture** ☐ 53

None = 1
Closed (no related scalp wound) = 2
Compound (dura intact) = 3
Compound (dura torn) = 4
Not known = 5

23. **Vault Fracture Site (linear or depressed)** ☐ 54

None = 1
Frontal = 2
Temporal = 3
Parieto-occipital = 4
>1 site = 5
Not known = 6

24. **Side of Linear or Depressed Fracture** ☐ 55

Right = 1
Left = 2
Bilateral = 3
Not known = 4

25. **Signs of Basal Fracture** ☐ 56

Mastoid haematoma = 1
CSF or blood otorrhoea = 2
1 + 2 = 3
Orbital haematoma = 4
CSF rhinorrhoea = 5
4 + 5 = 6
(1 or 2) + (4 or 5) = 7

EXTRACRANIAL COMPLICATIONS

26. **Chest injury** ☐ 57
27. **Other trunk injury** ☐ 58
28. **Limb injury** ☐ 59
29. **Facial injury** ☐ 60

No = 1
Minor = 2
Major = 3 (requiring hospital admission itself)

30-32. **G.I. Bleeds** 61-63

No = 1
Minor = 2
Major = 3 (blood transfusion needed)

0-3D	4-7D	8-28D
☐	☐	☐

33-36. **Shock (B.P.<90/60)** 64-67

No = 1
Yes = 2

24H	2-3D	4-7D	8-28D
☐	☐	☐	☐

37-41. **Chest Complication** 68-72

No = 1
Minor = 2 (limited respiratory infection)
Major = 3

24H	2-3D	4-7D	8-14D	15-28D
☐	☐	☐	☐	☐

Coma Study No. [2 | | | | | |] 1-7

INTRACRANIAL COMPLICATIONS

Supratentorial haematoma

			Operated	Known Unoperated	P.M. Only	
42-44.	Subdural		☐	☐	☐	8-10
45-47.	Intracerebral	No = 1 Right = 2 Left = 3 Both = 4	☐	☐	☐	11-13
48-50.	Extradural		☐	☐	☐	14-16

51-53. Infratentorial haematoma

No = 1
Subdural = 2
Intracerebellar/stem = 3
Extradural = 4

☐ 17-19

54. Pre-Operative Course (before haematoma operation)

Static = 1
Deteriorating = 2
Unknown = 3

☐ 20

55. Effect of Haematoma Operation (first 24 hours after surgery)

Improvement in coma score = 1
Improvement in pupils only = 2
No improvement in coma score or pupils = 3
Deterioration in coma score = 4

☐ 21

56. Post-traumatic Epilepsy (one fit counts)

None = 1
Early (<7 days) = 2
Late = 3
Both = 4

☐ 22

57. Intracranial Infection (during current hospital admission)

No = 1
Meningitis = 2
Abscess = 3
2 + 3 = 4
Other = 5

☐ 23

INVESTIGATIONS

58-61. **Angiogram**

Normal = 1
Displacement = 2
Spasm = 3
2 + 3 = 4
Other = 5

24H ☐ 2-3D ☐ 4-7D ☐ 8-28D ☐ 24-27

62-65. **Mean Intraventricular Pressure**

<20 mm.Hg = 1
20-40 mm.Hg = 2
>40 mm.Hg = 3

24H ☐ 2-3D ☐ 4-7D ☐ 8-28D ☐ 28-31

66-69. **Electroencephalogram**

Normal = 1
Focal abnormality = 2
Diffuse = 3
Both = 4

24H ☐ 2-3D ☐ 4-7D ☐ 8-28D ☐ 32-35

70-73. **Ventriculogram or Air Encephalogram**

Normal = 1
Hydrocephalus = 2
Shift = 3
Other = 4

24H ☐ 2-3D ☐ 4-7D ☐ 8-28D ☐ 36-39

74-77. **EMI scan**

Normal = 1
Contusion (high + low density) = 2
Haematoma = 3
Ventricular displacement = 4
2 + 3 = 5
2 + 4 = 6
3 + 4 = 7
Other = 8 Specify _____

24H ☐ 2-3D ☐ 4-7D ☐ 8-28D ☐ 40-43

Coma Study No. [3][ ][ ][ ][ ][ ][ ] 1-7

TREATMENT

78-81. **Burr hole/craniotomy/craniectomy** 8-11

	24H	2-3D	4-7D	8-28D
	☐	☐	☐	☐

No = 1
Burr hole only = 2
Craniectomy (< 5 cm) = 3
Craniotomy = 4
(2 or 3) + 4 = 5

82-85. **Ventricular tap** 12-15

	24H	2-3D	4-7D	8-28D
	☐	☐	☐	☐

No = 1
Ventricular tap = 2
Drain = 3
Both = 4

86-89. **Surgical Decompression** 16-19

	24H	2-3D	4-7D	8-28D
	☐	☐	☐	☐

No = 1
External (bony) = 2
Internal (lobectomy) = 3
Both = 4

90-93. **Tracheostomy/tube/ventilation (excluding temporary or terminal)** 20-23

	24H	2-3D	4-7D	8-28D
	☐	☐	☐	☐

No = 1
Intubated = 2
Tracheostomy = 3
2 + controlled ventilation = 4
2 + patient triggered = 5
3 + controlled ventilation = 6
3 + patient triggered = 7

94-97. **Steroids** 24-27

	24H	2-3D	4-7D	8-28D
	☐	☐	☐	☐

None = 1
< 20 mg. Dexamethazone or equivalent daily = 2
> 20 mg. Dexamethazone or equivalent daily = 3
Unknown or shock dose = 4

98-101. **Osmotics** 28-31

	24H	2-3D	4-7D	8-28D
	☐	☐	☐	☐

None = 1
One dose = 2
Repeated dosage = 3
Unspecified = 4

102-105. **Drugs possibly affecting observations** 32-35

	24H	2-3D	4-7D	8-28D
	☐	☐	☐	☐

No = 1
Yes = 2 Specify _____

LOCALISATION

Coding for 106-117.

No = 1
Suspect = 2
Definite = 3

				24H	2-3D	4-7D	8-28D	
106-109.	Right Hemisphere	suspect	= hemiparesis, vault fracture					36-39
		definite	= dysphasia or epilepsy or radiological/ operation evidence					
110-113.	Left Hemisphere							40-43
114-117.	Localisation Post Fossa	suspect	= basal fracture signs, ataxia, dysarthria					44-47
		definite	= bilateral motor abnormalities, dysconjugate eye movements, autonomic abnormalities					

CRANIAL NERVES — (RECOGNISED PALSIES)

Coding for 118-137.

No = 1
Right = 2
Left = 3
Both = 4

		24H	2-3D	4-7D	8-28D	
118-121.	II					48-51
122-125.	III					52-55
126-129.	VI					56-59
130-133.	VII					60-63
134-137.	VIII					64-67

Coma Study No. [4][ ][ ][ ][ ][ ][ ] 1-7

COMA SCALE

For coma observations — B = best response during epoch
W = worst response during epoch

Coding for 138-173.

Eye Opening (138-149)		Best Motor Response (150-161)		Verbal Response (162-173)	
Spontaneous	= 4	Obey	= 6	Orientated	= 5
To Sound	= 3	Localise	= 5	Confused	= 4
To Pain	= 2	Normal flexion	= 4	Words	= 3
Nil	= 1	Abnormal flexion *	= 3	Sounds	= 2
		Extension	= 2	Nil	= 1
		Nil	= 1		

* Score abnormal flexion (3) if either:

1) preceding extension movement in arms

or 2) extension in a leg

or 3) two of these:
- i) stereotyped flexion posture
- ii) extreme wrist flexion
- iii) adduction of arm
- iv) fingers flexed over thumb

If in doubt, score 4

		adm. to 1st hosp.	best after injury before coma	24H B	24H W	2-3D B	2-3D W	4-7D B	4-7D W	8-14D B	8-14D W	15-28D B	15-28D W	
138-149.	Eye Opening	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	8-19
150-161.	Motor Response	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	20-31
162-173.	Verbal Response	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	32-43

174-178. Temporal Order of B/W Observations after Onset of Coma

Improving = 1
Deteriorating = 2
No Change = 3
Fluctuating = 4

24H	2-3D	4-7D	8-14D	15-28D	
[ ]	[ ]	[ ]	[ ]	[ ]	44-48

Coma Study No. [5][ ][ ][ ][ ][ ][ ] 1-7

MOTOR RESPONSE PATTERNS

		Adm. to 1st hosp.	best after injury before coma	24H B	24H W	2-3D B	2-3D W	4-7D B	4-7D W	8-14D B	8-14D W	15-28D B	15-28D W	
179-190.	Right side													8-19
191-202.	Left side													20-31

Coding for 179-202.

No response	= 1
Extension	= 2
Abnormal flexion (spastic, decorticate)	= 3
Better type of response, but weaker than other side	= 4
Better type of response and normal strength	= 5

Code response in arms; if arm flexes and leg extends code 3. If doubt whether 3 or 4/5, then code latter. If two types of response are found in a limb at one examination, code both using best and worst boxes.

203-213. Tonic Spasms (spontaneous and generalised)

	Adm.	24H B	24H W	2-3D B	2-3D W	4-7D B	4-7D W	8-14D B	8-14D W	15-28D B	15-28D W	
Absent = 1 Present = 2												32-42

Coma Study No. [6][ ][ ][ ][ ][ ][ ] 1-7

EYE SIGNS

214-224. Pupils 8-18

	Adm.	24H		2-3D		4-7D		8-14D		15-28D	
		B	W	B	W	B	W	B	W	B	W
	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]

Both reacting - equal = 1
Both reacting - unequal = 2
One reacting = 3
Non-reacting equal <2mm = 4
Non-reacting equal 2-4mm = 5
Non-reacting equal >4mm = 6
Non-reacting unequal = 7

225-230. Pupil side/size (for cases coded 2, 3 or 7 above) 19-24

	Adm.	24H	2-3D	4-7D	8-14D	15-28D
	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]

If 2 or 7: R > L = 1
L > R = 2
If 3: R non-reacting < 4 = 3
R non-reacting > 4 = 4
L non-reacting < 4 = 5
L non-reacting > 4 = 6

231-236. Local Factors Affecting Pupils 25-30

	Adm.	24H	2-3D	4-7D	8-14D	15-28D
	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]

None = 1
Right = 2
Left = 3
Both = 4

237-247. Spontaneous Eye Movements 31-41

	Adm.	24H		2-3D		4-7D		8-14D		15-28D	
		B	W	B	W	B	W	B	W	B	W
	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]

Orientating = 1
Roving conjugate = 2
Roving dysconjugate = 3
Lateral deviation = 4
None = 5
Other = 6

248-258. Oculocephalics 42-52

	Adm.	24H		2-3D		4-7D		8-14D		15-28D	
		B	W	B	W	B	W	B	W	B	W
	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]

Nil (normal) = 1
Full = 2
Minimal = 3
Absent = 4

259-269. Oculovestibulars 53-63

	Adm.	24H		2-3D		4-7D		8-14D		15-28D	
		B	W	B	W	B	W	B	W	B	W
	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]

Nystagmus (normal) = 1
Conjugate tonic = 2
Dysconjugate = 3
No OV* = 4

* No response scored only after 100 ml. iced water delivered into clear canal.

Coma Study No. [7][ ][ ][ ][ ][ ][ ] 1-7

AUTONOMIC ABNORMALITIES (Predominant or persistant abnormalities)

270-274. Respiratory patterns

Regular = 1
Periodic = 2
Ataxic = 3
1 + 2 = 4
1 + 3 = 5
2 + 3 = 6
1 + 2 + 3 = 7
Ventilation = 8

24H [ ] 2-3D [ ] 4-7D [ ] 8-14D [ ] 15-28D [ ] 8-12

275-279. Episode of Apnoea (long enough to require at least temporary ventilation and not induced by relaxant drugs).

No = 1
Yes = 2

24H [ ] 2-3D [ ] 4-7D [ ] 8-14D [ ] 15-28D [ ] 13-17

280-284. Respiratory Frequency

<20 = 1
20-30 = 2
>30 = 3
1 + 2 = 4
1 + 3 = 5
2 + 3 = 6
1 + 2 + 3 = 7

24H [ ] 2-3D [ ] 4-7D [ ] 8-14D [ ] 15-28D [ ] 18-22

285-289. Pulse

Normal = 1
High (>120) = 2
High + Low = 3
Low (<60) = 4

24H [ ] 2-3D [ ] 4-7D [ ] 8-14D [ ] 15-28D [ ] 23-27

290-294. Systolic Blood Pressure

Normal = 1
High (>160) = 2
High + Low = 3
Low (<90) = 4

24H [ ] 2-3D [ ] 4-7D [ ] 8-14D [ ] 15-28D [ ] 28-32

295-299. Temperature

Normal = 1
High (>39° C) = 2
High + Sweating = 3
High + Low = 4
Low (<35° C) = 5

24H [ ] 2-3D [ ] 4-7D [ ] 8-14D [ ] 15-28D [ ] 33-37

RECOVERY PROCESS

300-302. Speech

Normal	= 1
Mild dysphasia	= 2
Severe dysphasia	= 3
Untestable	= 4

1-7D ☐ 8-14D ☐ 15-28D ☐ 38-40

303. Time to speak (V = 3 or more) — Actual week ☐☐ 41-42

304. Time to obey (M = 6) ☐☐ 43-44

305. Time to spontaneous eye opening (E = 4) ☐☐ 45-46

Time to Disappearance of Extension Responses

Right arm
- 306. Supraorbital stimulus ☐☐ 47-48
- 307. Finger stimulus ☐☐ 49-50

Left arm
- 308. Supraorbital stimulus ☐☐ 51-52
- 309. Finger Stimulus ☐☐ 53-54

310-311. Post-traumatic amnesia — Period ☐ Actual week ☐☐ 55-57

312-313. Leave intensive care unit ☐ ☐☐ 58-60

Periods

24 hrs.	= 1
2-3 days	= 2
4-7 days	= 3
8-14 days	= 4
15-28 days	= 5
>28 days	= 6

314-315. Return home ☐ ☐☐ 61-63

OUTCOME

		Outcome Categories	Actual	Best	
316-317.	1 month outcome	Death = 1 Vegetative state = 2 Severe disability = 3 Moderate disability = 4 Good recovery = 5 Out of hospital, lost to follow-up = 6 If 2/3 indistinguishable at 1 month = 7	☐	☐	64-65
318-319.	3 month outcome		☐	☐	66-67
320-321.	6 month outcome		☐	☐	68-69
322-323.	12 month outcome		☐	☐	70-71

N.B. Severe - conscious but dependent i.e. requiring help of another person during every 24 hrs.
Moderate - independent but disabled.

DEATH

324-325. Time to death

<24 hours = 1
2-3 days = 2
4-7 days = 3
8-14 days = 4
15-28 days = 5
>28 days = 6

(i) after injury ☐ 72

(ii) after coma ☐ 73

326. Post-Mortem

Yes = 1
No = 2

☐ 74

CAUSE OF DEATH

(Allocate a total of 5 points between the four causes)

327. Primary brain damage ☐ 75

328. Expanding intracranial lesion ☐ 76

329. Other intracranial complication ☐ 77

330. Extracranial complications ☐ 78

APPENDIX 2

NUMERICAL EXAMPLE OF THE MISCLASSIFICATION PROBABILITIES ASSOCIATED WITH ORDERED OUTCOME CATEGORIES

Suppose that Π_1, Π_2 and Π_3 are ordered categories with equal prior probabilities and the model is such that, for an individual with feature vector y,

$$p(y|\Pi_1) \sim N(\varepsilon,1)$$

$$p(y|\Pi_2) \sim N(0,1)$$

and $$p(y|\Pi_3) \sim N(\delta,1) \qquad \text{where } \varepsilon < 0 \text{ and } \delta > 0.$$

Since the distributions all have a common variance, it can be seen from Figure A2.1 that the individual will be allocated to Π_1 if $y < \varepsilon/2$ and to Π_3 if $y > \delta/2$; if $\varepsilon/2 < y < \delta/2$ then they will be allocated to Π_2, the middle category. Thus the probabilities of correctly classifying an individual are:-

$$\text{(i)} \quad p(\text{classify as } \Pi_1|\Pi_1) = \int_{-\infty}^{\varepsilon/2} \frac{1}{\sqrt{(2\Pi)}} \exp\{-\tfrac{1}{2}(y-\varepsilon)^2\}dy \quad \text{where } \varepsilon<0,$$

$$\text{(ii)} \quad p(\text{classify as } \Pi_2|\Pi_2) = \int_{-\varepsilon/2}^{\delta/2} \frac{1}{\sqrt{(2\Pi)}} \exp\{-\tfrac{1}{2}\, y^2\}dy$$

$$\text{and (iii)} \quad p(\text{classify as } \Pi_3|\Pi_3) = \int_{\delta/2}^{\infty} \frac{1}{\sqrt{(2\Pi)}} \exp\{-\tfrac{1}{2}(y-\delta)^2\}dy \quad \text{where } \delta>0.$$

The probability of correctly classifying an individual from group Π_2 is given in Table A2.1 for a range of values of ε and δ. The probabilities of correctly classifying an individual from groups Π_1 and Π_3 are given in Table A2.2.

For example, with $\varepsilon = -3$ and $\delta = 3$ then the probabilities of correctly classifying an individual from groups 1 and 3 are both 0.933 while for group 2 it is 0.866. However, if $\varepsilon = -1$ and $\delta = 1$,

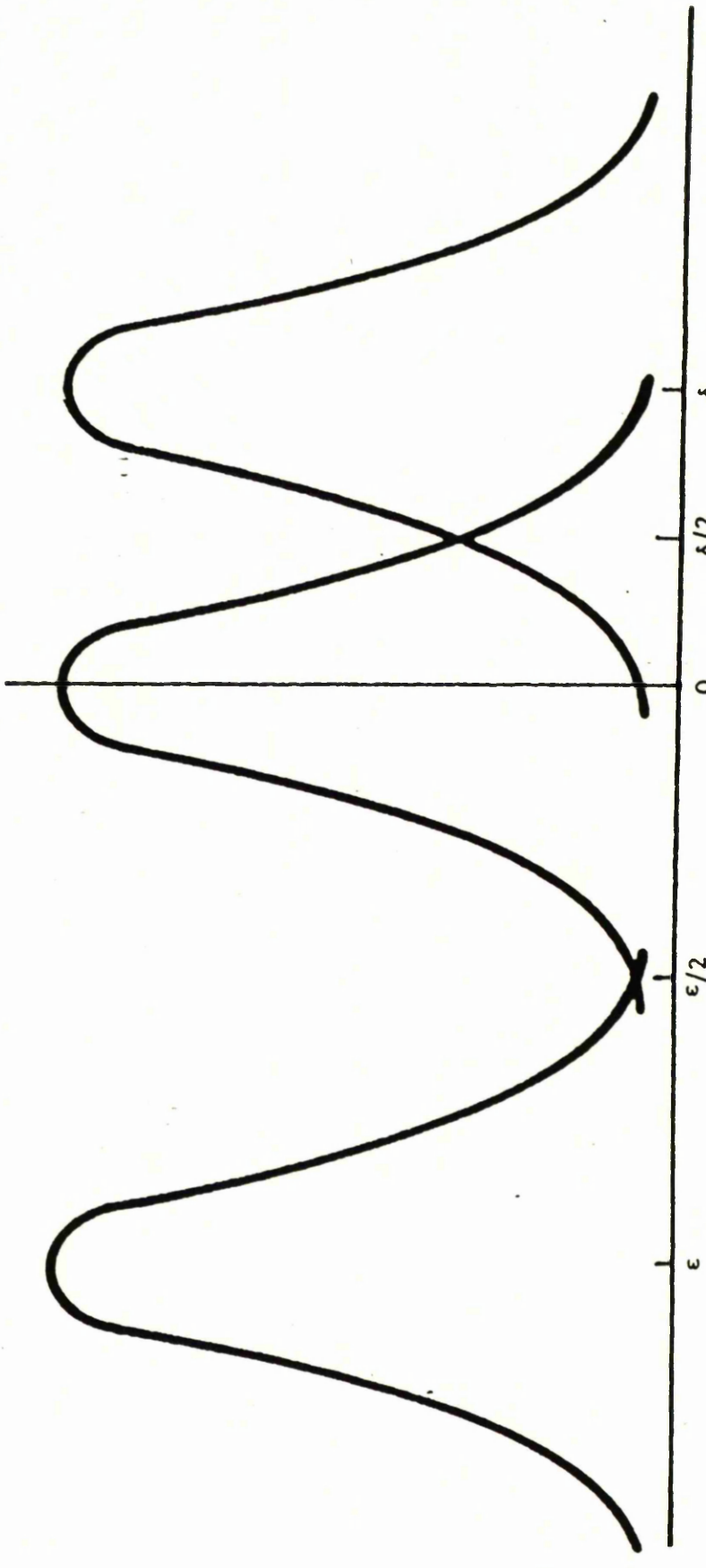


Figure A2.1 Probability density functions of N(ε,1), N(0,1) and N(δ,1) distributions, where ε < 0 and δ > 0

$\delta \setminus \varepsilon$	-3.00	-2.50	-2.00	-1.50	-1.00	-0.90	-0.80	-0.70	-0.60	-0.50	-0.30	-0.10
3.00	.866	.828	.775	.707	.625	.607	.589	.570	.551	.532	.493	.453
2.50	.828	.789	.736	.668	.586	.568	.550	.531	.512	.493	.454	.414
2.00	.775	.736	.683	.615	.533	.515	.497	.478	.459	.440	.410	.361
1.50	.707	.668	.615	.547	.465	.447	.429	.410	.391	.372	.333	.293
1.00	.625	.586	.533	.465	.383	.365	.347	.328	.309	.290	.251	.211
0.90	.607	.568	.515	.447	.365	.347	.329	.310	.292	.272	.233	.194
0.80	.589	.550	.497	.429	.347	.329	.311	.292	.273	.254	.215	.175
0.70	.570	.531	.478	.410	.328	.310	.292	.274	.255	.236	.196	.157
0.60	.551	.512	.459	.391	.309	.292	.273	.255	.236	.217	.178	.138
0.50	.532	.493	.440	.372	.290	.233	.254	.236	.217	.197	.158	.119
0.30	.493	.454	.401	.333	.251	.213	.215	.196	.178	.158	.119	.080
0.10	.453	.414	.361	.293	.211	.194	.175	.157	.138	.119	.080	.040

Table A2.1 Probability of correctly classifying an individual from category Π_2 for a range of values of δ and ε

$-\delta,\varepsilon$	-3.00	-2.50	-2.00	-1.50	-1.00	-0.90	-0.80	-0.70	-0.60	-0.50	-0.30	-0.10
	.933	.894	.841	.773	.691	.673	.656	.637	.618	.599	.560	.520

Table A2.2 Probabilities of correctly classifying an individual from categories Π_1 and Π_3 for a range of values of δ and ε

so that the overlap between the groups is much greater, the probabilities of correctly classifying an individual from groups 1 and 3 are both 0.691 while for group 2 it drops to 0.383.

So it can be seen, that unless there is a large difference between the means of the categories or a small variance in each, it will be difficult to allocate a member of the middle group to the correct category.

APPENDIX 3

MODEL FITTING BY PSEUDO MAXIMUM LIKELIHOOD

In Section 4.6.1 models imposing linear constraints upon the means of multivariate normal populations were fitted by pseudo maximum likelihood (PML) estimation to avoid the computational difficulties associated with full maximum likelihood (ML) estimation of the covariance matrices. In practice this meant that the covariance matrices were estimated simply by the sample covariance matrices rather than by full ML under the constrained model.

The following results, based upon Parke (1986), justify this approach by showing that "pseudo likelihood ratio tests" are conservative in that any model rejected by the full ML approach would also be rejected by PML.

Let $\ell(\theta,\pi)$ denote the log-likelihood function and (θ_o,π_o) the true values of the parameters. The following argument holds under very general conditions, but in this application the vector θ denotes the parameters associated with the mean vectors and the vector π denotes the parameters associated with the covariance matrices.

Let $\tilde{\pi}_n$ denote a consistent sequence of estimators of π.

Let $\hat{\theta}_n(\pi)$ denote the MLE of θ, for fixed π.

let $\theta_n^*(\pi)$ denote the MLE of θ, for fixed π, and subject to the constraints $h(\theta) = 0$.

Section 3 of Parke establishes that

$$\sqrt{n}\,(\hat{\theta}_n(\pi_o) - \theta_o) = -\sqrt{n}\,(I^o_{11})^{-1}\, D_\theta\, \ell(\theta_o,\pi_o) + o_p(1). \qquad \text{(A3.1)}$$

Here $D_\theta \ell(\theta_o,\pi_o)$ denotes the vector of partial derivatives of ℓ with respect to θ, evaluated at (θ_o,π_o) and I^o_{11} denotes the θ-block of the Fisher information matrix for (θ,π), evaluated at (θ_o,π_o).

Differentiating with respect to π and evaluating at π_o then gives

$$\sqrt{n}\,\frac{\delta\hat{\theta}_n(\pi_o)}{\delta\pi} = \sqrt{n}\,(I^o_{11})^{-1}\,I^o_{12} + o_p(1), \qquad \text{(A3.2)}$$

where I^o_{12} denotes the off-diagonal block of the matrix, evaluated at (θ_o,π_o).

Finally expanding $\sqrt{n}\,(\hat{\theta}_n(\tilde{\pi}_n) - \theta_o)$ gives

$$\sqrt{n}\,(\hat{\theta}_n(\tilde{\pi}_n) - \theta_o) = \sqrt{n}\,(\hat{\theta}_n(\pi_o) - \theta_o) + \sqrt{n}\,\frac{\delta\hat{\theta}_n(\pi_o)}{\delta\pi}\,(\tilde{\pi}_n - \pi_o) + o_p(1)$$

$$= \sqrt{n}\,(I^o_{11})^{-1}D_\theta\,\ell(\theta_o,\pi_o) + \sqrt{n}\,(I^o_{11})^{-1}I^o_{12}(\tilde{\pi}_n - \pi_o) + o_p(1). \qquad \text{(A3.3)}$$

In the constrained case, provided that the null hypothesis is true (i.e. provided that $h(\theta_o) = 0$), there are three corresponding results:

$$\text{(i)}\quad \sqrt{n}\,(\theta^*_n(\pi_o) - \theta_o) = -\sqrt{n}\,P_o\,D_\theta\,\ell(\theta_o,\pi_o) + o_p(1), \qquad \text{(A3.4)}$$

where $P_o\,I^o_{11}\,P_o = P_o$ (Silvey, 1975, §4.7).

Note that if θ is of length k and if $h(\theta) = 0$ represents s constraints, then P_o is of rank k - s.

$$\text{(ii)}\quad \sqrt{n}\,\frac{\delta\theta^*_n(\pi_o)}{\delta\pi} = \sqrt{n}\,P_o\,I^o_{12} + o_p(1) \qquad \text{(A3.5)}$$

$$\text{(iii)}\quad \sqrt{n}\,(\theta^*_n(\tilde{\pi}_n) - \theta_o) = \sqrt{n}\,P_o\,D_\theta\,\ell(\theta_o,\pi_o) + \sqrt{n}\,P_o\,I^o_{12}\,(\tilde{\pi}_n - \pi_o) + o_p(1). \qquad \text{(A3.6)}$$

Thus from equations A3.3 and A3.6

$$\sqrt{n}\ (\hat{\theta}_n(\tilde{\pi}_n) - \theta_o) = \sqrt{n}\ (I^o_{11})^{-1}\ Y + o_p(1)$$

and

$$\sqrt{n}\ (\theta^*_n(\tilde{\pi}_n) - \theta_o) = \sqrt{n}\ P_o\ Y + o_p(1)$$

$$\text{where}\quad Y = D_\theta\ \ell(\theta_o,\pi_o) + I^o_{12}\ (\tilde{\pi}_n - \pi_o),$$

so that

$$\sqrt{n}\ (\hat{\theta}_n(\tilde{\pi}_n) - \theta^*_n(\tilde{\pi}_n)) = \sqrt{n}\ \{\ (I^o_{11})^{-1} - P_o\}\ Y + o_p(1) \qquad \text{(A3.7)}$$

and, asymptotically,

$$Y \sim N(\ 0\ ,\ I^o_{11} + I^o_{12}\ I^o_{22}\ I^{o\,T}_{12}\ ).$$

Now considering the log-likelihood ratio statistic

$$2\{\ \ell(\hat{\theta}_n(\tilde{\pi}_n),\tilde{\pi}_n) - \ell(\theta^*_n(\tilde{\pi}_n),\tilde{\pi}_n)\ \}$$

$$= 2\{\ \ell(\hat{\theta}_n(\tilde{\pi}_n),\tilde{\pi}_n) - \ell(\hat{\theta}_n(\tilde{\pi}_n),\tilde{\pi}_n) - (\hat{\theta}_n - \theta^*_n)^T\ D_\theta\ \ell(\hat{\theta}_n,\tilde{\pi}_n)$$
$$- \tfrac{1}{2}\ (\hat{\theta}_n - \theta^*_n)^T\ D^2_\theta\ (\hat{\theta}_n,\tilde{\pi}_n)\ (\hat{\theta}_n - \theta^*_n) + \ldots\ \}$$

$$= -\ (\hat{\theta}_n - \theta^*_n)^T\ D^2_\theta\ (\hat{\theta}_n,\tilde{\pi}_n)\ (\hat{\theta}_n - \theta^*_n) + \ldots$$

$$= (\hat{\theta}_n - \theta^*_n)^T\ I^o_{11}\ (\hat{\theta}_n - \theta^*_n) + \ldots \quad \text{,under the appropriate regularity conditions.}$$

Thus it can be seen for equation A3.7 that, provided $h(\theta_o) = 0$, the test statistic takes the form of a $\chi^2(s)$ random variable plus an asymptotically independent positive random variable. Thus with the critical region $2\log\lambda > \chi^2(s,1-\alpha)$, any model which would be rejected by the correct ML approach would also be rejected by PML, i.e. the PML approach is conservative.

REFERENCES

AITCHISON J and DUNSMORE IR (1975) Statistical Prediction Analysis. Cambridge University Press, Cambridge.

ANDERSON JA (1972) Separate sample logistic discrimination. Biometrika, 59, 19-36.

ANDERSON JA (1984a) Regression and ordered categorical variables. Journal of the Royal Statistical Society, Series B, 46, 1-30.

ANDERSON TW (1984b) An Introduction to Multivariate Statistical Analysis. Second Edition. Wiley, New York.

ANDREWS DF (1972) Plots of high-dimensional data. Biometrics, 28, 125-136.

AVEZAAT CJJ, van de BERGE JH and BRAAKMAN R (1979) Eye movements as a prognostic factor. Acta Neurochirurgica, Suppl 28, 26-28.

BARLOW P, MURRAY L and TEASDALE G (1987) Outcome after severe head injury - the Glasgow Model. In: Medical Applications of Microcomputers, pp. 105-126. Ed. Corbett WA. Wiley, Chichester.

BARLOW P and TEASDALE G (1986) Prediction of outcome and the management of severe head injuries: the attitudes of neurosurgeons. Neurosurgery, 19, 989-991.

BARLOW P, TEASDALE G, JENNETT B, MURRAY L, DUFF C and MURRAY G (1984) Computer-assisted prediction of outcome of severely head-injured patients. Journal of Microcomputer Applications, 7, 271-277.

BECKER DP (1979) Comments on Jennett et al. (1979). Neurosurgery, 4, 289.

BECKER DP, MILLER JD, WARD JD, GREENBERG RP, YOUNG HF and SAKALAS R (1977) The outcome from severe head injury with early diagnosis and intensive management. Journal of Neurosurgery, 47, 491-502.

BRAAKMAN R, GELPKE GJ, HABBEMA JDF, MAAS AIR and MINDERHOUD J (1980) Systematic selection of prognostic features in patients with severe head injury. Neurosurgery, 6, 362-370.

BRIGGS M, CLARKE P, CROCKARD A, CUMMINS B, GALBRAITH S, GARFIELD J, JEFFREYS R, JENNETT B, KALBAG R, MENDELOW AD, MILLER JD, PRICE D, TAYLOR J, TEASDALE G and LITTLEY D (1984) Guidelines for initial management after head injury in adults. Suggestions from a group of neurosurgeons. British Medical Journal, 288, 983-985.

CANT BR, HUME AL, JUDSON JA and SHAW NA (1986) The assessment of severe head injury by short latency somatosensory and brain stem auditory evoked potentials. Electroencephalography and Clinical Neurophysiology, 65, 188-195.

CARLSON C-A, von ESSEN C and LOFGREN J (1968) Factors affecting the clinical course of patients with severe head injuries. Part 1: Influences of biological factors. Part 2: Significance of post traumatic coma. Journal of Neurosurgery, 29, 242-251.

DAVIDSON ML (1983) The multivariate approach to repeated measures. BMDP Technical Report No. 75. BMDP Statistical Software, Los Angeles.

DAWID AP (1976) Properties of diagnostic data distributions. Biometrics, 32, 647-658.

DEMPSTER AP, LAIRD N and RUBIN DB (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). Journal of the Royal Statistical Society, Series B, 39, 1-38.

DIXON WJ, BROWN MB, ENGLEMAN L, FRANE JW, HILL MA, JENNRICH RI and TOPOREK JD (1985) BMDP Statistical Software Manual. University of California Press, Berkeley.

de DOMBAL FT (1978) Medical diagnosis from a clinician's point of view. Methods of Information in Medicine, 17, 28-35.

DRAPER N and SMITH H (1981) Applied Regression Analysis. Second Edition. Wiley Interscience, New York.

DUDA RO and HART PE (1973) Pattern Classification and Scene Analysis. Wiley, New York.

FIELDING A (1977) Latent structure analysis. In: Exploring Data Structures, pp 125-157. Eds. O'Muircheartaigh CA and Payne C. Wiley, New York.

FISHER RA (1936) The use of multiple measurement in taxonomic problems. Annals of Eugenics, 7, 179-188.

FLEISS JL (1986) The Design and Analysis of Clinical Experiments. Wiley, New York.

FRANE JW (1980) The univariate approach to repeated measures - foundation, advantages and caveats. BMDP Technical Report No. 69. BMDP Statistical Software, Los Angeles.

FRIES JF (1976) A data bank for the clinician? New England Journal of Medicine, 294, 1400-1402.

FROWEIN RA (1976) Classification of Coma. Acta Neurochirurgica, 34, 5-10.

GENTLEMAN D and TEASDALE G (1981) Adoption of Glasgow Coma Scale in the British Isles. British Medical Journal, 283, 408.

GORE SM and ALTMAN DG (1982) Statistics in Practice. British Medical Association Publication, London.

GNANDESIKAN R (1977) Methods for Statistical Data Analysis of Multivariate Observations. Wiley, New York.

GREENHOUSE SW and GEISSER S (1959) On methods in the analysis of profile data. Psychometrika, 24, 95-112.

HABBEMA JDF and GELPKE GJ (1981) A computer program for selection of variables in diagnostic and prognostic problems. Computer Programs in Biomedicine, 13, 251-270.

HABBEMA JDF and HILDEN J (1981) The measurement of performance in probabilistic diagnosis IV. Utility considerations in therapeutics and prognosis. Methods of Information in Medicine, 20, 80-96.

HABBEMA JDF, HILDEN J and BJERREGAARD B (1978) The measurement of performance in probabilistic diagnosis I. The problem, descriptive tools and measures based on classification matrices. Methods of Information in Medicine, 17, 217-226.

HABBEMA JDF, HILDEN J and BJERREGAARD B (1981) The measurement of performance in probabilistic diagnosis V. General recommendations. Methods of Information in Medicine, 20, 97-100.

HADLEY DM, TEASDALE GM, JENKINS A, CONDON B, MACPHERSON P, PATTERSON J and ROWAN JO (1988) Magnetic resonance imaging in acute head injury. Clinical Radiology, 39, 131-139.

HAND DJ (1981) Discrimination and Classification. Wiley, New York.

HANS P, BORN JD, CHAPELLE J-P and MILBOUW G (1983) Creatine kinase isoenzymes in severe head injury. Journal of Neurosurgery, 58, 689-692.

HILDEN J (1984) Statistical diagnosis based on conditional independence does not require it. Computers in Biology and Medicine, 14, 429-435.

HILDEN J and BJERREGAARD B (1976) Computer-aided diagnosis and the atypical case. In: Decision Making and Medical Care: Can Information Science Help? pp. 365-378. Eds. de Dombal FT and Gremy FD. North Holland, Amsterdam.

HILDEN J, HABBEMA JDF and BJERREGAARD B (1978a) The measurement of performance in probabilistic diagnosis II. Trustworthiness of the exact values of the diagnostic probabilities. Methods of Information in Medicine, 17, 227-237.

HILDEN J, HABBEMA JDF and BJERREGAARD B (1978b) The measurement of performance in probabilistic diagnosis III. Measures based on continuous functions of the diagnostic probabilities. Methods of Information in Medicine, 17, 238-246.

HUYNH H and FELDT LS (1976) Estimation of the Box correction for degrees of freedom from sample data in the randomised block and split-plot designs. Journal of Educational Statistics, 1, 69-82.

JENKINS A, TEASDALE G, HADLEY MDM, MACPHERSON P and ROWAN JO (1986) Brain lesions detected by magnetic resonance imaging in mild and severe head injuries. Lancet, 2, 445-446.

JENNETT B (1975) Who cares for head injuries? British Medical Journal, 3, 267-270.

JENNETT B and BOND M (1975) Assessment of outcome after severe brain damage. A practical scale. Lancet, 1, 480-484.

JENNETT B, MURRAY A, MACMILLAN R, MACFARLANE J, BENTLEY C, STRANG I and HAWTHORNE V (1977a) Head injuries in Scottish hospitals. Lancet, 2, 696-698.

JENNETT B and PLUM F (1972) Persistent vegetative state after brain damage. Lancet, 1, 734-737.

JENNETT B and TEASDALE G (1981) Management of Head Injuries. FA Davis, Philadelphia.

JENNETT B, TEASDALE G, BRAAKMAN R, MINDERHOUD J, HEIDEN J and KURZE T (1979) Prognosis of patients with severe head injury. Neurosurgery, 4, 283-289.

JENNETT B, TEASDALE G, BRAAKMAN R, MINDERHOUD J and KNILL-JONES R (1976) Predicting outcome in individual patients after severe head injury. Lancet, 1, 1031-1034.

JENNETT B, TEASDALE G, GALBRAITH S, PICKARD J, GRANT H, BRAAKMAN R, AVEZAAT C, MAAS A, MINDERHOUD J, VECHT CJ, HEIDEN J, SMALL R, CATON W and KURZE T (1977b) Severe head injuries in three countries. Journal of Neurology, Neurosurgery and Psychiatry, 40, 291-298.

KNILL-JONES R (1975) The diagnosis of jaundice by the computation of probabilities. Journal of the Royal College of Physicians, London, 9, 205-210.

LACHENBRUCH PA (1975) Discriminant Analysis. MacMillan, New York.

LACHENBRUCH PA and MICKEY MR (1968) Estimation of error rates in discriminant analysis. Technometrics, 10, 1-10.

LANGFITT TW (1978) Measuring the outcome from head injuries. Journal of Neurosurgery, 48, 673-678.

LANGFITT TW, OBRIST WD, ALAVI A, GROSSMAN RI, ZIMMERMAN R, JAGGI J, UZZELL B, REIVICH M and PATTON D (1986) Computerised tomography, magnetic resonance imaging and positron emission tomography in the study of brain trauma. Journal of Neurosurgery, 64, 760-767.

LEVIN HS, O'DONNELL VM and GROSSMAN RG (1979) The Galveston Orientation and Amnesia Test. A practical scale to assess cognition after head injury. Journal of Nervous and Mental Disease, 67, 675-684.

LINDSAY K, PASAOGLU A, HIRST D, ALLARDYCE G, KENNEDY I and TEASDALE G (1988) Somatosensory and auditory brainstem conduction after head injury: a comparison with clinical features in prediction of outcome. Submitted for publication.

LITTLE RJA (1979) Maximum likelihood inference for multiple regression with missing values: a simulation study. Journal of the Royal Statistical Society, Series B, 41, 76-87.

McCULLAGH P (1980) Regression models for ordinal data. Journal of the Royal Statistical Society, Series B, 42, 109-142.

MACMILLAN R, STRANG I and JENNETT B (1979) Head injuries in primary surgical wards in Scottish hospitals. Health Bulletin, 37, 75-81.

MAAS AIR, BRAAKMAN R, SCHOUTEN HJA, MINDERHOUD JM and van ZOMEREN AH (1983) Agreement between physicians on assessment of outcome following severe head injury. Journal of Neurosurgery, 58, 321-325.

MARSHALL LF, BECKER DP, BOWERS SA, CAYARD C, EISENBERG H, GROSS CR, GROSSMAN RG, JANE JA, KUNITZ SC, RIMEL R, TABADDOR K and WARREN J (1983) The National Traumatic Data Bank. Part 1: Design, purpose, goals and results. Journal of Neurosurgery, 59, 276-284.

MARSHALL LF, SMITH RW and SHAPIRO HM (1979) The outcome with aggressive treatment in severe head injuries: Part 1. The significance of intracranial pressure monitoring. Journal of Neurosurgery, 50, 20-25.

MILLER JD, BECKER DP, WARD JD, SULLIVAN HG, ADAMS WE and ROSNER MJ (1977) Significance of intracranial hypertension in severe head injury. Journal of Neurosurgery, 47, 503-516.

MILLER JD and TEASDALE GM (1985) Clinical trials for assessing treatment. In: Central Nervous System Trauma - Status Report, pp. 17-32. Eds. BECKER DP and POVLISHOCK J. NIH, Bethesda.

MITCHELL DE and ADAMS JH (1973) Primary focal impact to the brain stem in blunt head injuries. Does it exist? Lancet, 2, 215-218.

MURRAY GD (1979) Missing Data Problems in Discriminant Analysis. Ph.D. Thesis, University of Glasgow.

MURRAY GD and TITTERINGTON DM (1978) Estimation problems with data from a mixture. Applied Statistics, 27, 325-334.

OVERGAARD J, CHRISTENSEN S, HVID-HANSEN O, HAASE J, LAND AM, HEIN O, PEDERSEN KK and TWEED WA (1973) Prognosis after head injury based on early clinical examination. Lancet, 2, 631-635.

PAGNI CA (1973) The prognosis of head injured patients in a state of coma with decerebrated posture: Analysis of 471 cases. Journal of Neurosurgical Sciences, 17, 289-295.

PARKE WR (1986) Pseudo maximum likelihood estimation: the asymptotic distribution. Annals of Statistics, 14, 355-357.

PAZZAGLIA P, FRANK G, FRANK F and GAIST G (1975). Clinical course and prognosis of acute post traumatic coma. Journal of Neurology, Neurosurgery and Psychiatry, 38, 149-154.

PLUM F and POSNER JB (1972) Diagnosis of Stupor and Coma. Second Edition. FA Davis, Philadelphia.

RABOW L and HEDMAN G (1985) Creatine kinase$_{BB}$-activity after head trauma related to outcome. Acta Neurochirurgica, 76, 137-139.

RAO CJ, SHUKLA PK, MOHANTY S and REDDY YJV (1978) Predictive value of serum lactate dehydrogenase in head injury. Journal of Neurology, Neurosurgery and Psychiatry, 41, 948-953.

RUBIN DB (1976) Inference and missing data. Biometrika, 63, 581-592.

SANO K, MANAKA S, KITAMURA K, KAGAWA M, TAKEYCHI K, OGASHIWA M, KAMEYAMA M, TOHGI H and YAMADA H (1983) Statistical studies on evaluation of mild disturbance of consciousness. Study of a simpler scale for clinical use. Journal of Neurosurgery, 58, 223-230.

SCHEIN M (1988) Acute surgical disease and scoring systems in daily surgical practice. British Journal of Surgery, 75, 731-732.

SILVEY SD (1975) Statistical Inference. Chapman and Hall, London.

SMITH AFM and WEST M (1983) Monitoring renal transplants. An application of the multiprocess Kalman filter. Biometrics 39, 867-878.

STABLEIN DM , MILLER JD, CHOI SC and BECKER DP (1980) Statistical methods for determining prognosis in severe head injury. Neurosurgery, 6, 243-248.

STRANG I, MACMILLAN R and JENNETT B (1978) Head injuries in accident and emergency departments at Scottish hospitals. Injury, 10, 154-159.

TEASDALE G (1975) Acute impairment of brain function I: assessing "conscious level". Nursing Times, 71, 914-917.

TEASDALE G, GALBRAITH S and CLARKE K (1975) Acute impairment of brain function II: observation record chart. Nursing Times, 71, 972-973.

TEASDALE G, GALBRAITH S, MURRAY L, WARD P, GENTLEMAN D and McKEAN M (1982a) Management of traumatic intracranial haematoma. British Medical Journal, 285, 1695-1697.

TEASDALE G and JENNETT B (1974) Assessment of coma and impaired consciousness: a practical scale. Lancet, 2, 81-84.

TEASDALE G and JENNETT B (1976) Assessment and prognosis of coma after severe head injury. Acta Neurochirurgica, 34, 45-55.

TEASDALE G, JENNETT B, MURRAY L and MURRAY G (1983) Glasgow Coma Scale: To sum or not to sum? Lancet, 2, 678.

TEASDALE G, KNILL-JONES R and van der SANDE J (1978) Observer variability in assessing impaired consciousness and coma. Journal of Neurology, Neurosurgery and Psychiatry, 41, 603-610.

TEASDALE G, MURRAY G, PARKER L and JENNETT B (1979a) Adding up the Glasgow Coma Score. Acta Neurochirurgica, Suppl. 28, 13-16.

TEASDALE G, SKENE A, PARKER L and JENNETT B (1979b) Age and outcome of severe head injury. Acta Neurochirurgica, Suppl. 28, 2, 140-143.

TEASDALE G, SKENE A, SPEIGELHALTER D and MURRAY L (1982b) Age, severity and outcome of head injury. In: Head Injury: Basic and Clinical Aspects, pp. 213-220. Eds. Grossman RG and Gildenberg PL. Raven Press, New York.

TITTERINGTON DM, MURRAY GD, MURRAY LS, SPEIGELHALTER DJ, SKENE AM, HABBEMA JDF and GELPKE GJ (1981) Comparison of discriminant techniques applied to a complex data set of head injured patients (with discussion). Journal of the Royal Statistical Society, Series A, 144, 145-175.

THOMAS DE, RABOW L and TEASDALE G (1979) Serum myelin basic protein, clinical responsiveness and outcome of severe head injury. Acta Neurochirurgica, Suppl. 28, 93-95.

THOMAS DGT, PALFREYMAN JW and RATCLIFFE JG (1978) Serum myelin basic protein assay in daignosis and prognosis of patients with head injury. Lancet, 1, 113-115.

ULM K (1984) Classification on the basis of successive observations. Biometrics, 40, 1131-1136.

WAGNER G, TAUKU P and WOLBER U (1978) Problems of medical diagnosis - a bibliography. Methods of Information in Medicine, 17, 55-74.

WEGMAN EJ (1974) Some results on new stationary first order autoregression. Technometrics, 16, 321-322.

WILSON JTL, WIEDMANN KD, HADLEY DM, CONDON B, TEASDALE G and BROOKS DN (1988) Early and late magnetic resonance imaging and neuropsychological outcome after head injury. Journal of Neurology, Neurosurgery and Psychiatry, 51, 391-396.