Assessment of Control and Performance of Biomedical Systems

Katie Harris

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

Division of Biostatistics

June 2010

Abstract

Introduction: In biomedical systems repeated measurements are often collected, thus presenting a statistical challenge due to high temporal correlation. This research investigates the potential utility of two distinct statistical methodologies in their application.

Application: Two clinically diverse biomedical systems, linked by the common methodological interest of assessing control and performance, are considered. (i) An application to renal anaemia aims to investigate the stability of haemoglobin levels, measured monthly for 13 months within 151 patients, with the ultimate goal of improving patient control; (ii) the second an application concerns cerebral autoregulation (a stable cerebral blood flow over a range of arterial blood pressure), to maintain patient safety during a surgical procedure to prevent stroke. Repeated measurements of cerebral blood flow and arterial blood pressure were collected on 36 patients, yielding a total of 4519 cerebral blood flow and 4574 arterial blood pressure measurements (note that the number of observations vary between patients).

Statistical methodology: Functional data analysis and multilevel modelling are utilised in the investigation of these two biomedical systems. Functional data analysis considers observations as a function rather than a highly correlated sequence of measurements. Multilevel modelling assumes that measurements are clustered and that within clusters, measurements are scattered about a trend in an uncorrelated manner.

Results: Assessment of control within the renal anaemia system and knowledge of the relationship within the cerebral autoregulation system, has been achieved through the successful application of functional data analysis. Loess curves were used as means of exploring the cerebral blood flow – arterial blood pressure relationship in the cerebral autoregulation application. B-splines and phase plots were used to explore haemoglobin control in the renal anaemia system. Further, multilevel modelling

incorporating autoregressive correlation structures appropriately models the dependency amongst model residuals due to temporal correlation. Both functional data analysis and multilevel modelling have demonstrated their utility in the application to model control in biomedical systems.

Conclusions: The novel application of these statistical methodologies has successfully provided contemporary insight into these biomedical systems and shows strong prospects for further applications.

Acknowledgements

Foremost, I would like to thank my supervisors Professors Robert West and Mark S Gilthorpe for their expert guidance, continued encouragement and confidence building throughout this PhD. Robert always had time and an open door, even when extremely busy. His support has been considerable and for this I will always be grateful. Mark for his refreshing insight and critical questioning, a trait which he has assisted me to develop. To both for their feedback throughout, especially during the final 'writing up' stage.

This work has been funded by The Emma and Leslie Reid Scholarship and the Yorkshire Kidney Research Fund. The Emma and Leslie Reid Scholarship was awarded by the University of Leeds for research in the area brain disease. The Yorkshire Kidney Research Fund grant supported work in the field of renal anaemia.

Many thanks go to my clinical collaborators, Dr Eric J Will and Dr Simon Howell. Their sharing of data and primarily clinical expertise, has been invaluable in my understanding of the clinical applications. Their interest and questioning attitude towards the statistical aspects being hugely influential in developing my ability to present a variety of ways in which to explain somewhat complex statistical issues.

Undertaking this PhD brought many wonderful opportunities to visit different parts of the UK, Europe and Canada. From attending and presenting at national and international conferences and training courses.

I would like to thank my colleagues in the Division of Biostatistics and fellow PhD students past and present. I will particularly hold fond memories of the numerous social events we have shared over the last four years.

Last but not least I would like to thank my family and friends for their support and encouragement. Thank you for being interested in my work and asking inquiring questions. Special thanks go to my Grandparents Pat and Raymond for asking "when

are you going to be a Professor? ”. And to my Nan, who after all this time continues to ask “remind me what it is that you are doing? ”.

Contents

List of figures

List of tables

Abbreviations

CA – Cerebral autoregulation

CBF – Cerebral blood flow

ABP – Arterial blood pressure

CAS – Carotid artery stenosis

CEA – Carotid endarterectomy

TCD – TransCranial Doppler ultrasound

TIA – Transient ischaemic attack

Hb – Haemoglobin

CDSS – Computerised Decision Support System

EB – Epoetin Beta

DA – Darbepoetin Alpha

FDA – Functional data analysis

MLM – Multilevel modelling / Multilevel model

ACF – Autocorrelation function

$AR(p)$ **–** Autoregressive model of order p

AIC – Akaike's information criterion

BIC – Bayesian information criterion

LRT – Likelihood ratio test

R – Statistical software package

MLwiN – Statistical software package (specifically for MLM)

ML – Maximum likelihood

REML – Restricted maximum likelihood

SD – Standard deviation

SE – Standard error

DF – Degrees of freedom

loglik – log likelihood

Chapter 1

Introduction

A biomedical system shall be defined as: *a collection of interrelated elements connected as a unified and dynamic whole, related to maintaining health and its preservation in the treatment of disease.*

1.1 Background and motivation

This research is motivated by the novel application of unexploited statistical techniques, which could provide more informative analyses to a wide range of biomedical systems. It is often the case in medical research that repeated measurements are collected on patients over a period of time; this might be over the course of a treatment or drug, or during a surgical procedure. A difficulty posed by such datasets is the high degree of correlation that may occur amongst the repeated measures. This means that the traditional statistical assumption, that observations (or model residuals) are independent of each other, is erroneous [1]. A solution would be to bypass this issue, although any conclusions drawn from the analysis could be largely incorrect and thus the conclusions are potentially erroneous. A more appropriate approach would be to embrace the complex and interesting

features of the data and encompass these in the statistical analysis, and exploit these additional modelling features.

Two clinically diverse biomedical systems have been investigated, where repeated measurements were taken. The choice of datasets was influenced by the funding sources of this PhD, these are; the Emma and Leslie Reid Scholarship which was awarded by the University of Leeds to study brain disease. A further grant was obtained from the Yorkshire Kidney Research Fund by Katie Harris, Robert West and Eric Will to carry out research into renal disease [2]. In summary the systems were:

1. **Cerebral Autoregulation system –** where measurements of arterial blood pressure and cerebral blood flow were collected during a surgical procedure which is undertaken to prevent stroke. The key issue was to monitor a phenomenon known as cerebral autoregulation in patients before and after surgery. The aim was to determine whether a change could be detected in cerebral autoregulation from before to after surgery, i.e. whether there is an improvement or deterioration in autoregulation due to the procedure.

2. **Renal Anaemia system –** where measurements on blood samples were collected monthly for the monitoring of anaemia in patients with end stage renal disease. The aim was to determine whether a patient's haemoglobin can be adequately managed when administered with a synthetic, recombinant agent (epoetin) as the natural hormone (erythropoietin) is no longer produced by the kidneys.

The two systems posed similarities and differences. In both instances the primary interest was in assessing whether the biomedical systems were adequately controlled. An intriguing difference occurs in the mechanism by which control was implemented. In the renal anaemia system the patients were externally managed by the clinician. This system is driven by the administration of drugs to replace erythropoietin whose production declined due to failure of the kidneys. The cerebral autoregulation application was a

biomedical system that is self-regulatory, where there are biological mechanisms within the human body. Within this system the ideal behaviour for patients is that a constant blood flow to the brain is maintained over a range of arterial blood pressure. The key issue is to determine whether this 'natural mechanism' occurs in patients throughout a surgical procedure – a situation which could possibly alter the intrinsic control of blood flow and blood pressure.

In both systems multiple time series were collected, where a number of repeated measurements for each patient. In the renal anaemia system, measurements were taken over the period of 12 months. During this period the patients undergo a long-term course of haemodialysis treatment, as immediate alleviation of symptoms due to renal anaemia is not possible. In this situation each patient had the same number of measurements recorded. The cerebral autoregulation systems differs in that measurements were taken during surgery for each patient, which was a relatively short time frame in comparison to the annual course of treatment in the renal anaemia system. Throughout surgery measurements were collected from patients every 15 seconds for the duration of the procedure, yielding a large number of measurements. The timing of the surgical procedure varied between patients, thus resulting in different amounts of data points per patient.

It was not appropriate to analyse either dataset using traditional time series analysis as there was the added dimension of multiple time series. It was more apt to analyse all series simultaneously to maximise the power of the statistical model, rather than analyse each time series individually. Furthermore, time series analysis is often used for analysis of data collected over a long period, e.g. to analyse weather patterns over many years in order to identify seasonal variation or a long-term trend in the data [3]. This suggests that for these biomedical systems, time series analysis would not have been entirely appropriate, as identifying features such as trend, seasonal variation or cyclical behaviour are not primarily relevant in the these data. Time series methodology was not totally

disregarded, however, as there were some aspects that were successfully exploited.

Analysing data from all patients collectively not only provided more powerful statistical analysis but also yielded more robust findings about the patient groups, as opposed to the individual. The statistical analysis in this thesis was population focused and sought to find statistical models which represented the patient groups as a whole. It would have been implausible to find a model that represented all patients accurately due to random error, possibly due to unknown features about the patients that were not incorporated in the modelling. It was, however, clinically important to recognise the patients who deviated from the norm, i.e. responded differently to surgery or treatment. The aim was to determine whether these patients could be identified from analysis of the population group.

This thesis aimed to demonstrate that statistical methodologies, namely functional data analysis and multilevel modelling, can be used to analyse the data from both biomedical systems, albeit with the variation in time-scales from a few balanced measurements taken over the long term to a large number of unbalanced measurements taken in the short term. These statistical techniques are relatively unexploited, thus development of the methodologies was necessary to address fully and understand the clinical questions which were posed. Moreover, these methodologies were applied in a novel way for the specific biomedical area. The two biomedical systems may be diverse in respect of their clinical area, though it is the need to address control in both applications, which links the two. In particular, the common research question was: can ‘control’ be assessed in statistical terms and if so, is this possible using functional data analysis and multilevel modelling? This will highlight the usefulness of the two methodologies, as they will be applied and yield practical and meaningful clinical results.

Assessment of control is an important issue within the clinical domain. For example, it is of interest to identify how patient symptoms are managed with a certain drug. Initially, the clinician may consult the medical field as a means of addressing their research question. If

this is not plausible, or the clinician prefers to take an alternative approach, assistance may be sought from a biostatistician. The biostatistician will help to determine whether there is a 'statistical answer' to the clinical problem. Furthermore, by involving a biostatistician in the clinical research, additional findings may be gained than originally sought.

1.2 Aims and Objectives

The primary aim of this thesis was to develop unexploited statistical techniques through their novel application to diverse biomedical systems. In particular with interest in assessing control within the systems and identifying how to manage patients, in order for them to achieve and preserve a good quality of life. A number of statistical challenges that were posed by the two biomedical applications needed to be overcome in order to address successfully the specific clinical aims about the systems; these clinical aims have been discussed in full in their respective chapters (Chapters 2 and 3).

In summary, the statistical issues to be considered were:

- Analysis of control - of patients either undertaking a surgical procedure or course of treatment. This is an aspect which is also of clinical importance, but, as yet, is no gold standard statistical technique to assess control.
- Clustering - of repeated measurements within patients. The assumption will be made that measurements within an individual are not independent. Measurements collected within an individual will be more alike than measurements from other individuals, albeit the individuals share characteristics which make them more alike than from the general population.
- Autocorrelation - due to measurements collected relatively close together in time, which violates the assumption that the error terms are uncorrelated. Adjacent and nearby measurements will be more similar than measurements taken far apart.

- Varying lengths of time series - in particular in the cerebral autoregulation system due to differing durations of surgical procedure amongst patients.

1.3 Outline of thesis

This thesis comprises the following chapters:

1. Introduction – The aim of Chapter 1 was to ‘set the scene’ and explain the motivation for this research. A challenge for a statistician in the biomedical domain is to consider the presentation of somewhat complex statistics to a non-specialist audience. This thesis highlights how this can be achieved and illustrates how the methods are received by a clinical audience, whilst retaining statistical validity and plausibility.

2. Cerebral Autoregulation – Chapter 2 introduces the clinical background of the cerebral autoregulation application.

3. Renal Anaemia – Chapter 3 introduces the clinical background of the renal anaemia application.

   In addition to introducing the clinical background of the biomedical applications Chapters 2 and 3 give explanation of where the datasets for this thesis arose. The clinical motivation behind studying the particular systems is described. Following introductions of the respective datasets, there is discussion into the statistical issues that arose within these systems, together with how they have been developed from the clinical questions that were posed. These chapters contain a literature review, which comprise critiques of the relevant clinical issues together with discussion of how the data in the relevant fields have been analysed previously.

4. Functional data analysis – See below

5. Multilevel modelling – Chapters 4 and 5 consist of a general discussion of the statistical methods, with particular focus on the specific areas that were considered in this thesis. Following the general discussions, there are details of the reasons why these particular methods are relevant and useful for the data, together with how and why these statistical methods are able to bring novelty to the application and statistical area.

6. Cerebral Autoregulation Results – See below

7. Renal Anaemia Results – Chapters 6 and 7 presents the main results. The choice has been made to present the results as separate chapters for each application area rather than in terms of the statistical methods; this is a more chronological approach and will also show the development of the methods.

8. General Discussion

   Particular features of interest for this chapter were:

   - The first point to address was: has it been possible to successfully analyse the two biomedical systems using the chosen statistical methodology (functional data analysis and multilevel modelling)?

   Further discussion points:

   - Has it been possible to assess and quantify control?
   - How were the methods initially chosen for this thesis?
   - How has the applicability of statistical methods been developed?
   - Contrast and comparison of the two statistical methodologies.
   - Contrast and comparison of the biomedical systems.
   - What did the statistical results mean in clinical terms?
   - How has this thesis contributed to knowledge and understanding in the medical domain?

- What were the statistical issues that have been raised (and how were they overcome)?
- How might the work from this thesis be further developed?

9. Conclusions and Future Work – The conclusions of this research will be presented in this chapter. Ideas for future work arising from the analysis undertaken and knowledge gained from the research will also be suggested.

Chapter 2

Cerebral Autoregulation

2.1 Background

Every year in the UK it is estimated that 150,000 people have a stroke [4]. A stroke occurs when the blood supply to the brain is interrupted [5]. This may be due to either a lack of blood supply, resulting in deprivation of oxygen to the brain (ischaemic stroke) or accumulation of blood within the skull which occurs when a blood vessel bursts within the brain (haemorrhagic stroke). Ischaemic stroke accounts for 80% of cases and haemorrhagic stroke for up to 20% of cases. A stroke is defined by the World Health Organization as a “neurological deficit of cerebrovascular cause that persists beyond 24 hours or is interrupted by death within 24 hours” [6]. This definition reflects that stroke is the second most common cause of death [7] and the leading cause of adult disability worldwide [8]. It is therefore evident that preventing and identifying signs of stroke is of great importance.

Carotid endarterectomy (CEA) is a well-established surgical procedure for preventing stroke and is recognised as the gold standard for both symptomatic and asymptomatic patients with high-grade, extracranial carotid artery stenosis (CAS) [9]. In CAS the

carotid arteries (see figure 2.1) are narrowed due to atheroma; a build up of fatty plaque in the inner lining of the artery, which can cause complete cerebral blood flow (CBF) blockage or break off into small fragments (embolism). CEA is performed in order to prevent stroke; potentially caused by the small fragments of plaque blocking arteries and arterioles in the brain. CEA is not performed to remove the blockage or where complete stenosis occurs. A transient ischaemic attack (TIA) is often referred to as a 'mini stroke' as the individual will experience stroke like symptoms but not experience any lasting damage. TIA's are often a warning that a stroke may happen, though the time in which this occurs is uncertain. In approximately 50 % of cases, stroke occurs within 24 hours of the TIA [10]. There are over 25,000 TIA's in the UK annually and clinical guidelines indicate that it is necessary for CEA to be undertaken within two weeks in patients who have suffered TIA [11], especially if CAS is detected [12].

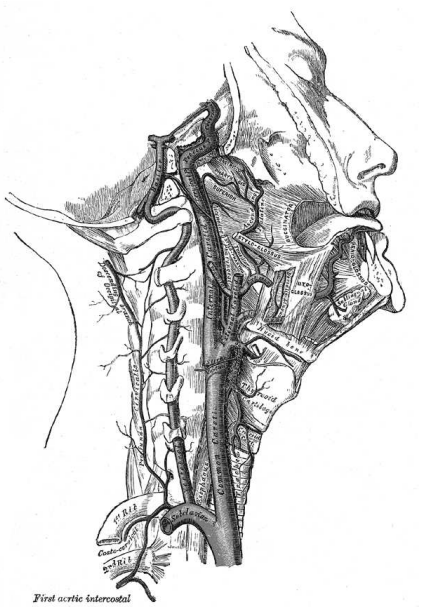

Figure 2.1: Diagram of the Carotid Artery [13].

Cerebral autoregulation (CA) is known to be disordered in patients with CAS, which is associated with an increased risk of cerebrovascular ischaemic events [14]. Cerebral

autoregulation is an intrinsic mechanism of the body whereby a constant cerebral blood flow (CBF) is maintained to the brain over a specific range of arterial blood pressure (ABP) [15]. A significant change in CBF, due to an increase or decrease in ABP indicates that CA is impaired. Conversely, if there is no significant change in CBF with ABP then CA is intact.

Carotid endarterectomy can be considered in three phases: preoperative (phase 1), intraoperative (phase 2) and postoperative (phase 3). The preoperative phase refers to the period where the neck is prepared for the procedure and the incision is made. During the intraoperative phase a carotid shunt is inserted into the diseased artery, to re-route blood flow to the brain. This is a difficult procedure as the surgeon will need to assure that blood flow continues around the circle of Willis [16](Figure 2.2). The circle of Willis constitutes part of the cardiovascular system, which is a circle of several arteries (the largest being internal carotid and vertebral, left and right) that supply the brain. If one of the arteries becomes obstructed, or the circle is damaged, this may be mediated since CBF can be redirected [17]. Throughout phase 2, ABP is maintained at a controlled high level in order to keep perfusion around the circle of Willis. The postoperative phase is where the carotid shunt is removed and the artery is stitched. CEA has been shown to reduce the risk of stroke, approximately 3-6 months following the procedure [18], however the effect of CEA on CA in the immediate post-operative period remains unclear [19].

After removal of the atheroma, the brain can be exposed to high systolic blood pressure and an increase in CBF, which is typical post-surgery behaviour and steps are taken to manage this. Occasionally patients develop cerebral hyperperfusion syndrome [20], with very marked increases in flow that may be associated with neurological signs, fits and cerebral oedema [21]. The opposite effect may also occur during surgery (particularly phase 2), where there is risk of hypoperfusion (decreased blood flow) [20], which can lead to ischaemic stroke.

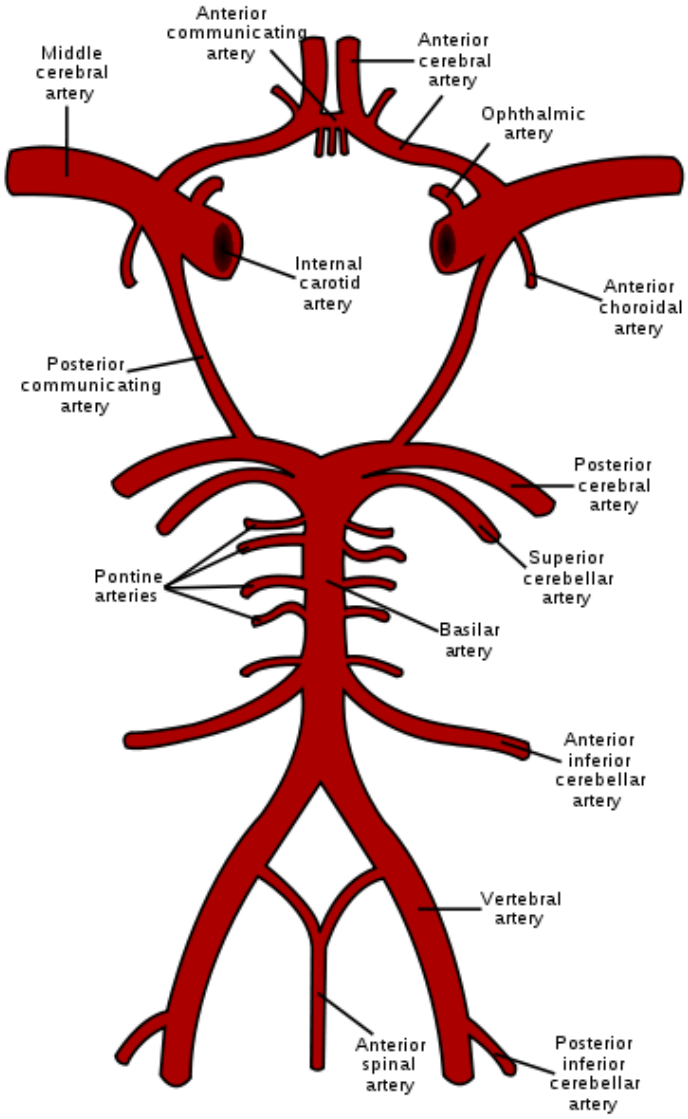

Figure 2.2: Circle of Willis [13].

2.2 Assessment of Autoregulation

There are two general approaches for the assessment of CA; dynamic and static autoregulation. Each considers the response of CBF in relation to ABP, albeit with a distinction made between the circumstances in which the measurements are taken. Static autoregulation examines the relationship between CBF and ABP without reference to the time course of changes in flow in response to changes in pressure. The stabilised response of CBF is considered with a consistent change in ABP, in this situation time is irrelevant as stabilisation of both measures occurs before the subsequent measurement is taken. The technique of static autoregulation has dominated studies in past decades, however with the advance in technology which allows for continuous measurement of CBF and ABP, interest in the field of dynamic behaviour has become popular. Dynamic CA is the immediate response of CBF to ABP where frequent or rapid changes of ABP have occurred. Rapid in this context is not defined in the literature; Reinhard [22] and Tiecks [23] simply state the word without stating the time period which is defined as rapid.

Assessment of dynamic CA has become achievable with the development of the TransCranial Doppler ultrasound (TCD) and non-invasive beat-to-beat blood pressure monitors (i.e. Finapres), which enable CBF and ABP measurements to be taken every 5 to 10 seconds [24]. The TCD technique allows for continuous measurement of CBF, which consists of insonating (exposing to ultrasound waves) the basel section of the major cerebral arteries, around the circle of Willis, through different ”windows” found at various locations in the skull. The windows commonly used for TCD monitoring include areas of the skull that are relatively thin . The transtemporal window is used to insonate the vessels of the circle of Willis, through the thinnest portion of the temporal bone. A probe is used to detect the Doppler signal which is generated by blood flow in the middle cerebral artery. Other windows include the transorbital and sub-occipital, which are used when other areas of the brain are examined [25].

There is uncertainty reflected in the literature as to whether the response of CBF to spontaneous changes in ABP may be considered static CA [26] or dynamic CA [14]. Thus, it is not clear whether methods for static or dynamic assessment should be adopted. One approach has been to examine the circumstances under which the measurements are taken and, for instance, ask whether repeated measurements over time have enjoyed a sufficient time-interval to permit physiological stability, as in static CA, or might ABP fluctuate, as with dynamic CA.

In the static regime it is assumed that there is no autocorrelation between adjacent measurements since there is a long time period between them. In the dynamic regime however, measurement of CBF might not have zero autocorrelation. This is due to the transient behaviour being observed before CBF settles into an equilibrium of a new ABF environment. Given that the timescale of this transient behaviour has not been defined, it is not clear if autocorrelation in CBF should be modelled. Should autocorrelation be present, then this must be accounted for in the statistical modelling, in order to estimate the underlying relationship without bias. Moreover, CA has not previously been explored during CEA, which further demonstrates the ambiguity as to whether static or dynamic assessment should be preferred. There have been a number of studies examining the changes in cerebral autoregulation in the days, weeks and months after CEA [27] [14] [28] , when it is possible to collect data on patients after they have recovered from surgery. Collecting data for CA assessment during surgery is more challenging.

2.2.1 Assessment and measurement of static autoregulation

The approach employed to assess CA is to investigate the relationship of CBF against ABP. The ideal model for static cerebral autoregulation is attributed to Lassen [29], who initially described the concept as a piecewise linear curve comprising two segments (see figure 2.3). The model was subsequently revised to consist of three linear segments (see figure 2.4), where the plateau represents where CBF is constant over the range of ABP,

between the lower(L) and upper(U) limits of CA - this shall be referred to as the CA curve. The plateau region with a zero slope represents perfect autoregulation [26].

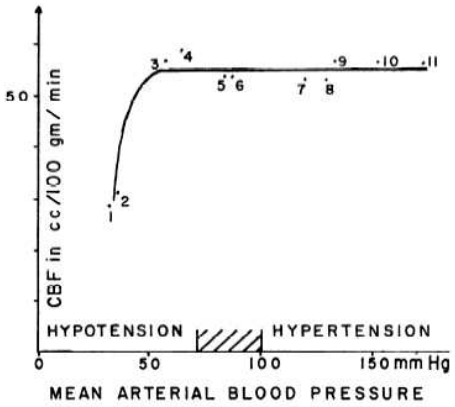

FIG. 1. Cerebral blood flow and blood pressure. Mean values of 11 groups of subjects reported in 7 studies have been plotted. Various acute and chronic conditions have been selected, characterized by a change in blood pressure. In all, this figure is based on 376 individual determinations.

1 and *2*, Drug-induced severe hypotension (81). *3* and *4*, Drug-induced moderate hypotension (206). *5* and *6*, Normal pregnant women and normal young men (206, 173). *7*, Drug-induced hypertension (230). *8*, Hypertensive toxemic pregnancy (206). *9*, *10*, *11*, Essential hypertension (229, 131, 228).

Figure 2.3: The original Lassen cerebral autoregulation curve displaying idealised piecewise linear behaviour [29].

Theoretically, it is anticipated that the CA curve is closely followed by most individuals for a range L-U within a healthy population. The L-U plateau region contains ABP values necessary for the brain tissue to be perfused with an adequate supply of blood, though the numerical values of L and U are subject to much debate [26]. An approximate range is thought to be 50 - 150 mmHg [15]. Values of ABP below L, i.e. less than 50 mmHg, are regarded as hypotensive and CBF is decreased due to low ABP, whereas values of ABP above U, i.e. greater than 150 mmHg, are regarded as hypertensive and CBF is increased due to high ABP. If there is no significant change in CBF with ABP then CA is regarded as being intact. The range of ABP covered by the CA curve is very large, and yet spontaneous variation in ABP may be more restricted; some variation is expected however, due to the stress induced by the surgical procedure. Thus, patients are unlikely

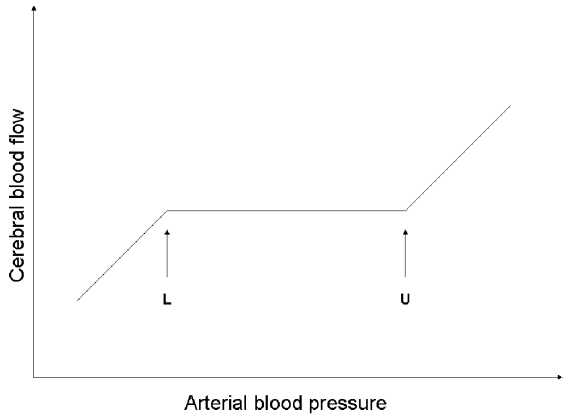

Figure 2.4: The theoretical CA curve of CBF plotted against ABP [26]: lower (L) and upper (U) limits of CA flank a plateau region where perfect CA occurs, adapted from the curve proposed by Lassen [29].

to experience the whole range of measurements to fit the whole curve, or even the range of perfect autoregulation. If the idealised piecewise linear curve is a plausible model for CA, then it is necessary to determine upon which segment (or segments) of the curve each patient lies.

The CA curve (figure 2.4) differs from the Lassen curve (figure 2.3) in that there is a positive slope for which high ABP occurs. The reason for this might be that the data used by Lassen to construct the curve did not extend beyond the upper limit U. The curve attributed to Lassen (figure 2.3) is constructed from 11 mean values, which raises doubt over the robustness of the curve due to the small sample of data. The 11 mean values are calculated from different patient groups with a variety of medical conditions that affect CBF and ABP. Thus, CA is calculated in groups of patients. Although the CA curve has been used to assess patients on an individual basis [30], few studies have attempted to do this [26].

It has been suggested that a slight slope exists between the limits L and U [31], as opposed

to a plateau, though this has not been substantiated by any patient or population study. Arguably, this observation might be true, and over simplicity has sought to hypothesise a flat slope where a modest slope may exist. Modest deviation from a zero gradient in the CBF-ABP relationship, which would normally suggest that CA is not intact (and may be impaired), may not be as problematic as considered previously.

The CA curve seems a rather simplistic view of such a complex entity as CA, which is viewed as not being completely understood [32]. Studies have calculated the gradient using standard linear regression [33] [34] and calculated Pearson's correlation coefficient [35] [36] [37] for the response of CBF to changes in ABP. Panerai [26] discusses the large number of studies which have adopted these approaches. A slope equal to zero is taken to represent the plateau region and hence perfect CA. A positive slope is thought to represent either of the linear sections of the CA curve and hence impaired CA.

In the clinical literature [23], Pearson's correlation coefficient ρ is usually calculated in addition to linear regression. For $\rho > 0$, this reflects impaired CA, and $\rho = 0$ represents the plateau region, hence perfect CA. It is noted that the calculation of Pearson's correlation coefficient has the implied assumption of linearity and that to observe a value close to zero within a individual with intact CA measurements must be confined to the plateau region. It is also implied that the patients with impaired CA will demonstrate a linear relationship with a positive slope for the relationship between CBF and ABP. If the autoregulating region is merely compressed with U being closer to L and the range of ABP extend beyond this interval, then a positive value of Pearson's coefficient will be observed. A similar value might be seen for a patient with impaired CA since the value is strongly dependent upon the scatter about the true relationship, which is influenced by measurement accuracy. Hence, this approach has its limitations

2.2.2 Assessment and measurement of dynamic autoregulation

Currently there is no consistent way to assess or model dynamic CA in the way that the Lassen curve forms the basis for analysis of static CA [38]. There are a number of methods presented in the literature such as: autoregulation indices [23], transfer function analysis [39] and correlation coefficients [40], although these have yet to be accepted as an established method to assess dynamic CA.

Experimentally, dynamic CA can be assessed by inducing transient changes in ABP in order to determine the response of CBF over a wide range of ABP. The thigh cuff method [41] and the lower body negative pressure [42] have previously been used to induce oscillations and rapid changes in ABP. These methods are not ideal as they have been found to cause patients pain and discomfort, thus alternative methods are preferred. Transfer function analysis is often used to analyse data produced from the thigh cuff method [41]. This methodology will not be discussed in detail as transfer function analysis makes the assumption that CA is measured in the frequency domain, where oscillations of ABP and CBF are analysed, whereas such oscillatory behaviour in ABP and CBF is not expected during CEA .

Tilt-tables in particular allow a wide range of ABP to be experienced by the individual [43], which permits the investigation of the CBF-ABP relationship over a wide range of ABP values. The sit-to-stand procedure is another method used which induces oscillatory and step ABP changes [38]. This technique might be preferred over the tilt table approach as it is easier to perform and tilt tables can provoke syncope (fainting).

Inducing changes in ABP for patients undergoing carotid endarterectomy would not be possible since patients undergoing a surgical procedure are required to lie supine throughout. Therefore, spontaneous fluctuations in ABP are relied upon. Methods such as the thigh cuff would clearly be inappropriate since it would be unsafe to delay surgery, particularly at the time of clamping the carotid artery.

It is likely that throughout the procedure patients will experience ABP values mostly within the L-U range, where intact CA should occur; as ABP is closely monitored and to a certain extent controlled during CEA, it is unlikely that the anaesthetists would allow the patients to be subject to extreme ABP values. This may not be realistic in all patients however, as the clinician may find themselves caught between the head and the heart [44]. Patients who become hypotensive during the procedure, and thus experience abnormally low ABP, are a particular concern. Increasing the blood pressure to a safer level in these patients, may lead to dramatic and possibly dangerous increases in CBF, especially after the stenosis has been relieved. It is vital to treat a patient's blood pressure that is persistently lower than their mean ABP would be in a non surgical environment, since this may compromise blood flow to the heart. There may be dramatic changes in CA therefore, as a consequence of CEA, if the range of ABP moves outside the range of intact autoregulation.

Similar to static CA, dynamic CA is viewed as a linear relationship between CBF and ABP, thus subsequent statistical analyses have been based on this assumption. Panerai [26] justifies this, suggesting that the CBF-ABP relationship does not show significant departure from the linear hypothesis. Transfer function analysis [22] [39] and various dynamic autoregulation indices [23] [45] assume linearity between CBF and ABP. There are a small number of papers, however, that have applied nonlinear modelling techniques [46] [47].

One measure occasionally encountered in CA studies is cerebrovascular resistance (CVR), which is defined as: $\mathbf{CVR} = \frac{\mathbf{Mean(ABP)}}{\mathbf{Mean(CBF)}}$. CVR is a mechanism that regulates the constriction and dilation of the smaller vessels (arterioles) in the brain, such that it quantifies the extent of which CBF is affected by changes in ABP. This measure is useful because one can then define the dynamic autoregulation index (DARi) as: $DARi = \frac{\Delta CVR}{(\Delta T)(\Delta ABP)}$ [26], which is sometimes referred to as the rate of recovery [38]. The variation in the language and labels adopted for the same quantities suggest that there

may be confusion amongst authors in regards to analysis of CA. The DARi essentially grades CA between 0 and 9; with 0 indicating no autoregulation and 9 indicating very fast autoregulation. This might be useful to summarise CA, however point estimates (mean values) only are used to define CA, which have limited utility without a confidence interval. A more informative measure would incorporate the variation in ABP and CBF values.

2.2.3 General vs Local anaesthesia

Carotid endarterectomy may be performed under local and general anaesthesia. The anaesthetic technique maybe an underlying explanation for those patients who become hypotensive during CEA, as blood pressure generally falls after induction of general anaesthesia [48]. However, patients receiving both local and general anaesthesia have displayed clinically significant hypotension and hypertension in the post operative period [49]. Therefore, since anaesthesic technique has been shown to have an effect of ABP, it is reasonable to suggest that the choice of anaesthetic technique for CEA may have an impact on perioperative CA.

McCleary and colleagues demonstrated a decrease in oxygen supply to the brain in patients receiving local and general anaesthesia for CEA, although recovery was more likely in patients undergoing surgery under local anaesthesia [50]. This observation was one of the justifications of the GALA trial [51]; an international multicentre randomised controlled trial to compare local and general anaesthesia for CEA. The GALA trial, however, failed to demonstrate a clinically or statistically significant difference between local and general anaesthesia in terms of the risk of stroke, myocardial infarction, and death as a result of carotid endarterectomy. A suggestion for this finding is that ABP was manipulated (increased) by anesthesiologists in 43 % of general anaesthesia patients compared with only 17 % of local anaesthesia patients, which compensated the failure of CA in the general anaesthesia group [52]. This result may not have occurred if ABP had

not been manipulated.

There may be patient and surgeon preference for anaesthesia type. For example, patients can find CEA under local anaesthetic stressful and uncomfortable, since they must lie still with their heads turned to one side for ninety minutes or longer if the operation is difficult. Further, the patient may feel claustrophobic due to the positioning of the surgical drapes and being surrounded by numerous people throughout surgery. A high dose of local anaesthesia may be required for CEA due to the invasiveness of the surgery. In large doses local anaesthesia can have a toxic effect, leading to systemic toxicity, whereby toxins are absorbed into the body through the bloodstream. Infection, swelling and bruising may also occur at the injection site. If a patient is particularly agitated under local anaesthesia it is possible to convert from local anaesthesia to general anaesthesia; which would mean further complication for patient, anaesthetist and surgeon, which would have been avoided if general anaesthesia had been administered in the first instance. Conversion may be required if the patient experiences pain at the operative side, general discomfort and anxiety, physiological instability, or neurological deterioration. It is suggested that the preference of all parties is for CEA to be carried out under general anaesthesia [9].

An alternative view is that performing CEA under local anaesthesia, rather than general anaesthesia, may be safer [52]. A benefit of CEA under local anaesthesia is the increase in ABP which occurs in phase 3 of surgery, which in turn better maintains CBF. This natural phenomena that occurs with local anaesthesia is mimicked by the anaesthetist for patients undergoing CEA with general anaesthesia, and is hence seen as a benefit to occur naturally rather than by manipulation. A second benefit is that the medical team may converse and interact with the patient, which would enable a quicker response if they notice a change in patient behaviour.

In summary, there appears to be advantages and disadvantages to both anaesthesia methods. Upon reflection it seems that CEA should be made available under local and general anaesthetic, in order to cater for patient preference, or medical reasons why one

may be preferred over the other. It is important, however, that the effect of anaesthesia on CEA and CA is investigated further, so patients can be advised accordingly and patient safety is preserved.

2.3 Clinical aims and objectives

Assessment of CA is considered in several clinical conditions, such as: head injury [53], respiratory distress of newborn babies [33], and carotid artery disease [54]. These conditions are known to severely impair autoregulation. Patients in whom autoregulation is impaired are at risk of brain tissue ischaemia; where the brain does not receive an adequate supply of blood and oxygen and hence suffers ischaemic stroke. It is extremely important and of clinical interest that CA is assessed in many (clinical) settings.

CAS is a disease known to be associated with impaired CA [28]. CAS accounts for 15-20 % of ischaemic strokes [55], although patients with CAS may undergo CEA to remove the atheroma and thereby reduce the risk of detachment of the emboli and hence reduce the risk of stroke. This suggests strong clinical interest in directly assessing the patient benefit of CEA.

Repeated measurements of patient ABP and CBF were collected for a group of patients, with CAS, whilst undergoing CEA, with the aim of assessing CA throughout the procedure. There is particular clinical interest in assessing the immediate impact on CA of undergoing CEA, which might therefore directly demonstrate benefit to the patient of the surgical procedure.

A number of techniques are available to examine CA, although previously CA has not been assessed during CEA. Methods will therefore need to be explored to find an appropriate way of modelling these data.

A further cause for concern in these patients is that CEA may be associated with large

changes in ABP, which in turn may cause changes in CBF. The choice of anaesthetic technique may also have an impact on CA following CEA, specifically CA may be better preserved with local anaesthesia than with general anaesthesia.

In summary there are two types of aims and objectives to consider, namely the purely clinical issues together with issues relating to measurement and assessment. The clinical and measurement issues are also linked because of the need to develop an approach to model dynamic CA, since there is no clear way of assessing this entity. In this research dynamic CA will be explored as part of a complex modelling process.

Clinical

- to determine whether CA can be assessed in an operating theatre
- to devise an approach for dynamic CA (under CEA), if appropriate
- to investigate whether there is any improvement in CA immediately following CEA
- to investigate whether CBF and ABP are affected by CEA
- to determine whether the choice of anaesthesia has any impact on CA

Measurement

- to use appropriate statistical methodology to model CA
- to distinguish between static and dynamic autoregulation using statistical methodology (see next section), thereby addressing a clinical requirement.

2.4 Data

A prospective observational study was conducted in patients undergoing CEA at the Leeds General Infirmary between February 2004 and May 2006. The proposal for this study was to investigate immediate changes in CA in patients undergoing CEA. Approval for the study was granted by the Leeds West Research Ethics Committee and informed consent obtained from all patients.

Sixty-two patients who presented for CEA, to be performed under local and general anaesthesia, were approached for participation. All patients had experienced recent TIA and evidence of CAS. Exclusion criteria included patient refusal or withdrawal from the study, absence of a temporal window for TCD monitoring and the presence of atrial fibrillation or other arrhythmias (where the heart beats at irregular intervals), this yielded 26 patients who were not eligible for participation.

Data were collected in the immediate perioperative period of CEA, starting after induction of general anaesthesia or completion of local anaesthesia and concluding with the end of surgery. Repeated measurements of CBF and ABP were recorded for 36 patients, together with the respective phase of surgery: preoperative (phase 1), intraoperative (phase 2) and postoperative (phase 3) . Measurements were collected concurrently at 15 second intervals. It was not possible to control the number of measurements within each phase, as the timing of the whole procedure and each particular phase varied across patients.

Patients in the study were elderly with a median age of 73 years (range 65 - 82), which may have an impact on a number of factors considered in the data collection and analysis of CA. Measurements of CBF velocity were obtained using a TCD ultrasound sensor probe, fixed in place at the temple with a metal frame or plastic headband. CBF velocity was used as an equivalent of CBF, since accurate measurement of CBF is difficult [45]. In children and young adults it is particularly easy to obtain good signals from the desired vessels. However, these signals weaken as age increases. In elderly patients it is difficult

to obtain an adequate signal, with no signal detected in 20 % of patients [9]. In those patients where an adequate signal could not be obtained it was not possible to record CBF measurements (8 patients), this therefore reduced the number of patients who were eligible to participate in the study (54 out of 62). A further limitation to TCD monitoring is that the probe is placed relatively near to the surgical site and may impede the surgeon, especially if constant adjustment is necessary. In some patients, CBF measurements were collected from the left and right sides of the head, but these were not complete. There were complete ipsilateral (same side) measurements for 36 patients. ABP measurements were obtained invasively using a radial arterial line for routine intraoperative monitoring. Invasive methods are the most accurate way to measure ABP, though this involves a higher risk than non invasive methods, such as when using a sphygmomanometer (device to measure blood pressure).

Data were also provided on anaesthetic technique. The choice of general or local anaesthesia was made according to clinician and patient preference. Twenty-three patients received local anaesthesia and 13 patients received general anaesthesia. Initially the aim was to randomly assign all patients to anaesthesia type, which would have allowed an unbiased assessment of how anaesthesia method affects CA. It was not possible to implement this, however, as in certain cases there was clear medical indication for anaesthesia type, i.e. if a patient is considered unfit for a particular method due to certain medical conditions that increase the risk of complications from anaesthesia. Other factors also need to be considered, such as age, weight, medication, etc; as well as patient preference.

It is also thought that CA may be adversely affected by age [38]. Authors have presented opposing views: Heckmann and colleagues [56] suggest that CA is delayed in elderly individuals, although numerous authors have concluded that autoregulation is not affected by age, and that older subjects autoregulated as good as younger subjects [57] [58] [59]. Studies which demonstrated that CA is preserved in the elderly examined healthy patients,

thus not reflecting the particular scenario that occurs in this study. It is possible, therefore, that since these patients are elderly they may not exhibit the perfect CA as displayed in the CA curve; though this is unlikely to be due to age alone and the factors of CAS, undergoing CEA, and anaesthetic are all more influential.

2.5 Statistical challenges

The primary statistical challenge is to use appropriate statistical methodology to model CA during CEA. This was motivated by the clinical aim: to determine whether there is an improvement in CA immediately following CEA. Before this can be addressed fully, it is necessary to ascertain the most suitable model for CA, which in itself poses challenges. These will be discussed in the following sections.

2.5.1 Relationship between blood pressure and blood flow

The exact relationship between ABP and CBF for the patient group of interest is unknown, which was revealed by the literature review. Patients are suffering with CAS, which means it is possible that typical autoregulatory behaviour is not exhibited, together with assessment being made in the conditions of an operating theatre. Anaesthetic technique and age may also influence whether or not CA occurs. It would therefore be informative to investigate the relationship between ABP and CBF without preconceptions or assumptions of the underlying relationship. A better understanding of the CBF-ABP relationship during CEA may inform patient care.

Investigating the ABP-CBF relationship is similar to previous approaches, such as linear regression or the CA curve. There is more statistical power, however, in investigating the ABP-CBF relationship without making assumptions about the data or relationship. By applying linear regression the assumption is made that the measurements of ABP and

CBF are independent, although it is not possible to make this assumption without further investigation. Secondly the aim is not to replicate the whole CA curve, but to determine where the patient measurements would lie on this curve. It would not be possible to do this for each patient as it is unlikely they will cover the whole range of ABP that would be necessary. Similarly it would not be possible to do this across all patients as it is unlikely that the range of ABP across the whole group would be large enough to determine the L and U limits of the curve.

2.5.2 Data quality/ limitations

Measurements of ABP and CBF are collected on patients during a surgical procedure which makes data collection relatively difficult. The CBF measurements may not be so reliable, as the TCD probe would need to be in constant contact with the temple and remain fixed for accurate measurement to take place. The probe was attached to the patient by a headband; so it may have been difficult to ensure that the probe was correctly in place throughout CEA, especially as it was situated so close to the site of surgery.

A large number of measurements are collected on all the patients, which is different and substantially better than past CA studies, where in some cases the regression line has been calculated from only two measurements [26].

There is potential bias in the data since patients were not randomly assigned to anaesthesia technique; therefore any conclusion drawn from investigating anaesthesia needs to be interpreted cautiously.

2.5.3 Autocorrelation

Measurements of ABP and CBF are recorded with a short time between them; it is possible that measurements closer together in time will be more correlated than

measurements farther apart. Due to the biological nature of CBF and ABP, it is likely that they vary smoothly, opposed to erratically changing every 15 seconds. Should correlation between measurements be present, this must be accounted for in the modelling in order to estimate the underlying relationship without bias. Methods that assume independence of observations will be invalid if autocorrelation is present. Dynamic autoregulation would need to assume autocorrelation between measurements.

2.5.4 Clustering

These data form a natural hierarchical structure, with repeated measurements clustered within patients. A further level of the hierarchy may be phase of CEA; such that repeated measurements are clustered within phases which in turn are clustered within patients. A statistical technique is required that will exploit the full power of the data, such that all available data are used appropriately to fit the model, i.e. all patients, phases and measurements. In particular, it is the multiple patients that will give the most value to the modelling. Measurements of ABP cannot be assumed to be independent in each of the phases, and the phases cannot be assumed to be independent within a particular patient.

2.5.5 Spontaneous blood pressure measurements

It is unclear whether static or dynamic autoregulation should be investigated, due to spontaneous blood pressure readings. This means that available methods for assessing CA cannot be used with confidence. It will therefore be more appropriate to use statistical techniques which consider the particular issues raised with these data, rather than incorrectly using existing methods. Where feasible, methods to assess dynamic CA may need to be developed.

The range of ABP values within an individual will possibly be limited, which may be problematic because the slope estimate (of a linear model) will be less accurate with a

larger standard error than for ABP covering a larger range. This can be resolved, however, by using all measurements from all patients.

2.5.6 Phases of Carotid endarterectomy

The phases of CEA must be acknowledged in the modelling, as there will most likely be a clear distinction in the behaviour of ABP and CBF between phases. A better understanding of the separate phases would certainly be of value to the clinician to improve patient care and maintenance during surgery.

The statistical technique should also be able to handle unbalanced data, due to the number of measurements varying between patients and phases. It is not possible to assume a continuous trajectory for CBF and ABP throughout CEA; it is more likely that there will be step changes between the three phases.

Other, prior approaches to CA assessment, such as transfer function analysis, would not be applicable due to the distinct phases because the technique makes the assumption that the underlying form of ABP and CBF (over time) is sinusoidal [22], suggesting that ABP and CBF vary periodically over time. This assumption might be suitable for CBF and ABP measured on a day to day basis, though it is not plausible to make such assumptions considering CBF and ABP during CEA. A sine decomposition would not sufficiently represent the anticipated step changes in ABP and CBF between the phases of CEA, suggesting that transfer function analysis is an inappropriate technique when assessing CA during CEA.

2.6 Summary

This is a novel scenario and analysis for cerebral autoregulation. There are a number of clinical, measurement, and statistical issues to consider when investigating CA, which

contribute to the novelty of this work. For clarity, these are summarized below:

Clinical

- to determine CA for this setting
- to investigate whether there is any improvement in CA following CEA
- to investigate whether CBF and ABP are affected by CEA
- to determine whether the choice of anaesthesia is associated with CA

Measurement

- To determine whether dynamic or static autoregulation is being assessed
- Assessment in a non-laboratory environment
- Frequent measurements collected repeatedly
- To account for the three distinct phases of CEA
- To account for the clustering of measurements within patients

Statistical

- To determine whether CA can be assessed during a surgical procedure, and develop methods appropriate to achieve this
- To determine whether CA can be assessed using spontaneous CBF and ABP measurement, thereby devising a method of dynamic CA
- To determine whether autocorrelation needs to be accounted for in the modelling, and if so make appropriate modelling changes

- To find a suitable generic model for the CBF-ABP relationship

Chapter 3

Renal Anaemia

3.1 Background

End–stage renal disease is a severe illness where kidney function has completely failed, which is associated with a high risk of morbidity and mortality [60] [61]. It is classified as the fifth (and final) stage of chronic kidney disease. At this stage, patients require permanent renal replacement therapy, which usually includes either: haemodialysis, peritoneal dialysis, haemofiltration, or renal transplant. The patients in this study have end– stage renal disease and undergo haemodialysis to filter harmful wastes, salts and fluid from the blood, since the kidneys are no longer able to carry out this functionality. As renal disease develops, an associated problem is renal anaemia. Patients at any stage of chronic kidney disease may have renal anaemia, with prevalence increasing with the severity of disease [62]. Almost all patients with end–stage renal disease experience anaemia [63]. The significant feature of renal anaemia is the reduction in the oxygen carrying capacity of blood, which means haemoglobin (Hb) levels are reduced. Hb is the iron–containing protein within an erythrocyte (red blood cell) that transports oxygen around the body.

Erythropoietin is the protein hormone from the kidneys that promotes the production of red blood cells: erythroblasts (nucleated cells normally found only in bone marrow that develop into erythrocytes) evolve into reticuliocytes (immature red blood cells), which are then released into the bone marrow and develop into red blood cells containing Hb. A deficiency in Hb occurs when the kidneys do not produce sufficient erythropoietin, resulting in renal anaemia[64]. Erythropoietin deficiency is considered the most important cause of renal anaemia in patients with chronic kidney disease [65].

Low Hb can be detrimental to a patient's health [66] particularly those with comorbidities, since oxygen supply to the tissues is diminished. Insufficient oxygen induces lethargy which leads to: reduction of a persons general well being; impaired cognitive function; and worsening of quality of life. Furthermore, renal anaemia also increases the risk of cardiac failure and can cause hypertrophy (abnormal enlargement of a body part or organ) and tachycardia (rapid heartbeat).

Other contributors to renal anaemia are reduced red blood cell lifespan and blood loss. Blood loss contributes to renal anaemia as the uremic toxins (which are usually released in the urine and are retained in the blood) mean that the blood clotting mechanism is defective, since the platelets, which stimulate the production of a blood clot, do not work correctly. This means that patients are more susceptible to bleeding. If a patient suffers a fall for example, and experiences rapid or heavy bleeding, this would result in a rapid decrease in Hb. Further, the patient may suffer from certain medical conditions which makes them susceptible to bleeding, such as haemophilia (reduction in the blood's ability to clot) and thrombocytopenia (low platelet count). In addition to the increase in a patient's sensitivity to bleeding, a major cause of blood loss is through haemodialysis [64]; due to frequent blood sampling, gastrointestinal ulcers, bleeding from the dialysis access site, and the haemodialysis procedure[67].

The average lifetime of a red blood cell is approximately 100-120 days [68]; although this is reduced by approximately one third in patients with renal disease [69], due to uremic

toxin and decreased flexibility of the red blood cells. Regardless of the technological advances within renal medicine, red blood cell loss remains as a major contributor to renal anaemia [70].

Although there are many complications associated with renal anaemia, it can be effectively treated with doses of recombinant erythropoietin, which are more commonly known as epoetin agents. The FDA (Food and Drug Agency, USA) state that epoetin is one of the most important medical advances in treating patients with kidney disease [71]. Epoetin agents may also slow down the progression of renal failure [72]. A number of epoetin agents are available to treat patients (types and brands), which are manufactured by various pharmaceutical companies. The choice of agent will usually be made by the hospital administering the dose, rather than the individual patient. The patient–specific dose of the agent is managed with the assistance of a computerised decision support system [73], which adjusts the dose each month by examining current Hb concentration, which is based upon the previous months epoetin dose.

A Computerised Decision Support System (CDSS) is a computer application designed to aid clinicians when making diagnostic or therapeutic decisions in patient care [74]. CDSS are particularly useful in diagnosing a patient with a disease based on particular symptoms, or in the case of this application, suggesting a drug dose to be prescribed for the patient. A CDSS is not designed to remove the clinician from the decision making procedure but to provide a ‘recommendation’ for the clinician. When an epoetin dose is calculated by the CDSS, after Hb results have been entered into the system, the clinician will either accept or override the the suggested prescription [73]. The final decision will generally be made by the clinician. There has been some resistance to the CDSS, however, whereby the general belief by those opposing it is that a better decision would be made by a human rather than a computer, and that patients should be treated as individuals. This CDSS opposition group also stipulates that the algorithm should be perfect. The opinion on behalf of the clinicians is that the CDSS is a clinical aid, which is required to manage

the large number of patients receiving renal therapy and without it they would not be effectively managed.

Doses of the epoetin agents are available in pre-filled syringes which are available in varying strengths; 25, 40, 60, 100, 150, 200, 300 IU/kg. Dose units are measured in IU/week/kg body weight. IU is the abbreviation of international units, which is a standard measurement for a pharmaceutical drug [75]. A maximum dose is 300 IU/kg per week. A patient undergoing epoetin therapy will always be required to receive some of the agent, even if their Hb is on target, as complete cessation of the agent may cause a sudden breakdown of the new cells. The patient's dose each month is adjusted using an epoetin dose ladder, which increases or decreases the dose if required based on the predetermined steps of the dose ladder.

3.1.1 Subcutaneous vs. Intravenous

A patient's response to the epoetin agent depends on their red blood cell lifespan and production rate [76]. It is important to identify those patients whose Hb does not respond to the epoetin, as death is more likely in these patients. The two main methods are by subcutaneous (just under the skin) and intravenous (within a vein) injection. A number of studies have been undertaken to compare subcutaneous and intravenous administration of epoetin [77] [78] [79] [80].

The optimal route of administration remains undetermined [78]. In general, the intravenous route is the fastest way to deliver fluids and medications throughout the body as there is a direct route to the blood stream, compared with the subcutaneous route which is usually used for medication that needs to work slowly. The subcutaneous route is thought to be more prolonged than the intravenous method [79], though in the case of the epoetin agent, where the maximum effect is seen in Hb in 2 to 6 weeks, this time difference would have negligible consequences as the drug does not have an immediate

effect.

It has been shown that administering epoetin using subcutaneous injection rather than intravenous can allow the patient to receive lower doses (Henderson, 1998 and [77]. Further, subcutaneous and intravenous administration of epoetin was found to be equivalent in terms of mean Hb [81] and with doses significantly lower in the subcutaneous group together with a shorter time to stabilisation, thus being advantageous in clinical and cost terms. Aggarwal and colleagues found a greater increase in Hb in the subcutaneous group than the intravenous group [78]; they also state that subcutaneous injection is more effective [78] and [80].

3.1.2 History of Epoetin Agents

Anaemia was discovered as a serious complication of renal failure by English physician Dr. Richard Bright in 1836. He also first described the common symptoms of kidney disease in 1836 [82]. The term erythropoietin was established in 1948 for the erythropoiesis (the process in which red blood cells are produced) stimulating hormone [83]. In 1977, the human gene for erythropoietin was isolated by Miyake and colleagues [84], which was later cloned in 1985 by Lin and colleagues [85]. This meant recombinant human erythropoietin (epoetin) could be manufactured for treatment of anaemia in chronic renal failure. Eschbach and colleagues undertook the first clinical trial in 1987 investigating the effect of epoetin on anaemia in patients with end stage renal disease [69]. Subsequently, results were reported in a series of publications, indicating that epoetin was successful in its aim of treating anaemia in end stage renal disease patients [86].

Prior to epoetin agents being available for treatment, approximately 25% of patients with chronic kidney disease required transfusion of red blood cells on a regular basis [69]. This exposed the patient to unnecessary risk in terms of undergoing the transfusion and also the possibility of experiencing severe bacterial infection and sepsis. More common adverse

Age / Sex group	Haemoglobin threshold (g/dL)
Children (0.50 – 4.99 yrs)	11.0
Children (5.00 – 11.99 yrs)	11.5
Children (12.00 – 14.99 yrs)	12.0
Non-pregnant women (> 15.00 yrs)	12.0
Pregnant women	11.0
Men (> 15.00 yrs)	13.0

Table 3.1: Haemoglobin thresholds used to define anaemia [88]

reactions to blood transfusions include fever, pain, tachycardia, and hypotension. Patients suffering with renal anaemia and not receiving epoetin agents experienced disabling fatigue and greatly reduced quality of life.

3.1.3 Haemoglobin Guidelines

It is accepted that patients with renal anaemia will experience a lower Hb concentration than a 'normal' healthy population [87], where approximately <12.0 g/dL is considered anaemic, albeit with caveats dependent on sex, age and gravidness (pregnancy status). Table 3.1 shows the Hb thresholds for anaemia [**?**].

Hb should be controlled within certain limits, as both high and low levels increase susceptibility to adverse health effects. There is conflicting guidance as to whether Hb should be maintained at an optimum value, above or below a specific value, or within a particular range [89] [90] [91] [92] [93]. The anticipated mean Hb concentration for patients with chronic kidney disease patients is 11.8 g/dL [94], though some authors suggest a target range of 10.5 – 12.5 g/dL and others 11.0 – 12.0 g/dL. Previous European best practice advice was that Hb should be maintained above 11.0 g/dL. Table 3.2 shows published guideline levels on Hb for patients with chronic kidney disease. However,

Guideline	Country	Haemoglobin target
British Renal Association	UK	> 10.0
National Kidney Foundation-Dialysis Outcome Quality Initiative	USA	11.0 – 12.0
Canadian Society of Nephrology	Canada	11.0 – 12.0
European Best Practice Guidelines	Europe	> 11.0
Health Care and Financing Administration	USA	10.3 – 12.0
Care for Australians with Renal Impairment	Australia	> 12.0

Table 3.2: Published Guidelines on Haemoglobin targets in patients with chronic kidney disease

regardless of the various different guidelines, it is not feasible to specify an optimum range of values [71].

It is highly desirable that a patient's Hb does not fall too low, as this may lead to cardiac failure, vascular complications and possible death [66]. Keeping Hb $\leq$ 10.0 g/dL leads to an increased risk of seizures compared with maintaining $>$10.0 g/dL.

A generous Hb is also problematic in patients with end-stage renal disease, since for these patients Hb above 12.5 g/dL are considered unsafe and may be associated with an increased risk of hypertension, cardiovascular events and death [71]. There are also major cost implications in reaching and maintaining high Hb levels [90]. A suspected cause of elevated Hb is that the patient has received too much of the epoetin agent [95].

An alternative view has been suggested that a target range is inappropriate and providing that patients are well and responding to treatment, then Hb may be maintained at any 'reasonable' level (Personal communication - Dr. E.J. Will). For instance the clinician may attempt to raise a patients Hb from 8g/dL to 12g/dL, when in fact this patient could attain a reasonable quality of life with a Hb of 8g/dL. It has been shown that there are adverse consequences of targeting high Hb levels [92] . There would be undue stress for

the patient and unnecessary effort on behalf of the clinician to raise the patients Hb, in addition to administering a higher dose than necessary thus incurring greater costs.

3.1.4 Haemoglobin Cycling

A related concern for patients, in addition to maintaining Hb within the specified limits, is to sustain Hb stability. This is beneficial for two reasons: patient health and cost [96] [97]. A stable dose would be beneficial since this would reduce the overall epoetin dose administered, which would therefore reduce the cost of treatment. A fluctuating dose is more expensive than a stable one due to asymmetry of the dose ladder. If in addition, a patient's Hb levels remain stable (and possibly within limits) then patient care is improved.

Hb cycling is a frequent occurrence in patients who receive epoetin as treatment for renal anaemia, though this is considered dangerous for the patient [98]. A number of reasons are proposed as to why this phenomenon occurs, such as: a narrow target range of Hb, the patient's iron status, and inflexible dose adjustments. Those of particular relevance to this work are: incorrect dose adjustment and protocols not accounting for an individual patient's response.

A possible reason for Hb cycling is that the epoetin dose is not adjusted immediately each month; it could be between 5 to 12 days or at worst 6 weeks before the adjustment is made. This means that the patient would receive the wrong dose of epoetin until the dose is changed. Guidelines state that from administration of the agent the first evidence of a response is in an increase in reticulocyte count within 10 days, subsequently an increase in erythrocyte count, followed by an increase in Hb within 2 to 6 weeks [86] [69]. This response will vary due to patients experiencing intercurrent complications, which will prolong the production of Hb [99]. If the dose is adjusted monthly and the effect of epoetin is not seen in the Hb for 6 weeks, then it is possible that cycling occurs since the

drug has not been given sufficient time to take effect.

3.2 Data

The data used for this analysis were provided from a randomised controlled trial [99] with the aim of comparing two different types of epoetin agent: Epoetin beta (EB) or Darbepoetin alpha (DA). The trial initially ran for 9 months, though data collection was extended for a subsequent 3 months. Patients were selected from a haemodialysis population with end–stage renal disease at St James University Hospital, Leeds. Patients were randomly assigned to two groups; Epoetin beta or Darbepoetin alpha. In the original 9 month trial there were 217 patients across the two groups, at the end of this time 162 remained on protocol. There were 151 patients who provided complete data from the extended trial: with 77 having received EB and 74 having received DA, both injected subcutaneously on a weekly basis. The groups were balanced according to age, sex and time on dialysis. These data do not include those patients who dropped out, due to either mortality or transplant.

Blood samples were collected monthly and the dose of the epoetin agent was adjusted monthly with the standard dosage ladders using the computerised decision support system. Including the 3 month extension this gave 13 data points per patient, comprising an initial baseline value and 12 more from the end of each month of treatment.

The usual approach of managing anaemia was to administer EB on a thrice weekly basis. However, the interest of the weekly regime governed by this trial was promoted by the development of the new agent DA. A conversion from EB to DA was in the ratio 200:1 [99]. This poses another clinical question of whether patients can be adequately managed by a weekly regime. The converted values will be used in this analysis to allow direct comparison of the agents. The ratio of 200:1 is in accordance with the manufacturers recommendations within the European Union [99], since this is cost neutral

under European licensing. The issue of the conversion factor might be influential, specifically if this conversion is not precise, as the patients in the DA group would not be receiving the appropriate quantity of the agent, which means that patients would receive too little or too much of the agent, resulting in the consequences discussed previously. In fact, the first randomised trial which compared the DA and EB agents based on the 200:1 conversion [99] suggested that there is a reduction in dose for the DA group with an increase in Hb. If this is correct, then this is beneficial to the patient as they reap the benefits of receiving a lower dose together with their Hb being raised. Conversely, these patients may simply have been under prescribed. A conversion of 260:1 was also recommended by the Centers for Medicare and Medicaid Services [100], which means the patient would receive a higher dose based on this ratio. As the conversion of 200:1 is not a universal gold standard, it may be an inappropriate conversion, but that was the conversion rate adopted within our data.

3.3 Clinical Aims and Objectives

It is important to identify patients for whom Hb (Hb) is not well controlled, as they are more likely to be at greater risk of suffering adverse health effects than those whose Hb is stable and under control [98]. The nature of the Hb – epoetin relationship should also be explored, since this will contribute to the understanding of Hb control. For instance, if the incorrect relationship is modelled, then one would expect patients to exhibit poorly controlled Hb.

The initial aim of the trial for which the data were collected was to establish whether the two agents offer comparable control of Hb. Poor control may be identified by Hb cycling; a commonly occurring phenomenon in patients receiving epoetin therapy, which should be avoided due to its association with adverse patient outcomes. If Hb cycling is shown to occur, a strategy will be developed to improve Hb control in patients.

Summary of clinical aims:

- To identify if Hb cycling occurs in these patients.
- To identify potential sources of Hb cycling.
- To suggest corrections to the computerised decision support system to reduce Hb cycling.
- To compare the two epoetin agents with regard to control.
- To suggest corrections/ updates for the CDSS to provide a more tailored approach for patient–specific control.

3.4 Statistical challenges

There are a number of statistical challenges which are posed with this application. This section highlights these issues and discusses why they are problematic and how these issues may be resolved.

3.4.1 Autocorrelation and time lag

It is plausible that there is correlation between Hb measurements within the same patient due to the longevity of red blood cells and other biological mechanisms. In statistical terms this is autocorrelation. The process of Hb control with epoetin agents is dynamic and occurs over a time period which involves multiple doses and multiple Hb concentration assessments, highlighting that these are correlated repeated measurements of Hb and epoetin, over the 12 month time frame.

Due to the lifetime of the red blood cells being approximately 100 to 120 days [68], albeit reduced in patients with renal anaemia, it is plausible that Hb produced at month

(i) will still be present the following month (i+1) and possibly the next following month (i+2). Similarly with dose; epoetin administered at month (i) is likely to be present in the patient's blood at month (i+1) and also month (i+2). Epoetin is injected subcutaneously into the muscle and there is a time–lag before it reaches the bone marrow and red blood cell production is stimulated.

3.4.2 Haemoglobin – epoetin relationship

The relationship between Hb and epoetin is complex; it is anticipated that as dose increases Hb will increase. This relationship may not be linear. Although progression of renal disease and intercurrent complications are highly influential on the patients response to the epoetin dose, so Hb will not respond in the ideal way. Conversely, it could be that Hb influences the dose due to the process of the CDSS.

A standard linear regression would not be appropriate to model the Hb–epoetin relationship, as it cannot be assumed that Hb will continue to increase as epoetin dose is increased. A dose greater than 300 IU/kg is rarely required by patients, as stable Hb can be maintained with this dose or less [101]. It is therefore plausible that there is a plateauing of Hb levels for larger doses. Additionally, a dose greater than 300 IU/kg may elevate the patients Hb too high which would lead to complications associated with a large Hb level. The rate by which Hb responds to epoetin will vary from patient to patient: in particular some patients may be more sensitive, where a small dose adjustment achieves a large increase in Hb; or, a patient may be unresponsive, so a large dose adjustment achieves a small increase in Hb. The dose–response relationship is thus expected to be nonlinear, since the dose ladder is asymmetric; as dose steps increase so does the increment between them. It is not plausible to assume that Hb will continually increase as dose increases; it is therefore unlikely that the relationship is of a linear or even exponential nature. Cotter et al. suggested that the Hb-epoetin relationship is S-shaped [102]; this is also implausible as Hb would not decrease when epoetin is administered.

Since the CDSS bases current Hb on the dose administered one month previously, this will be the initial relationship which is investigated. Since clinical knowledge suggests that the response in Hb may be seen any period between 2 and 6 weeks, these are other relationships to investigate. The modelling structure should also consider that multiple measurements are collected for each patient. A modelling structure that accounts for both the variation between measurements and between patients is necessary.

3.4.3 Assessment of control

The aim is to model control using statistical methodology and provide a way of quantifying control, with a view to improving overall patient control.

3.4.4 Assessment of individuals

A key consideration is that control should be assessed on an individual basis. It might be possible to improve the CDSS by tailoring it to the individual. Statistical techniques which permit assessment of individuals would provide a useful insight, although using the full patient history of drug and blood measurements would yield a more powerful model. A number of authors [98] [71] suggest that it is necessary that treatment is individualised, as there is an increased risk of mortality and reduction in a patient's quality of life when recommendations are based on population guidelines.

Summary of statistical challenges:

- Autocorrelation in successive and subsequent repeated measurements.
- The correct / appropriate model form of the Hb – epoetin relationship.
- Time–lag between administration of epoetin to mean response in Hb.

- Evaluation of methods to provide suitable models for the assessment control of Hb.
- Individual assessment of patients whilst simultaneously determining the behaviour of the population.

Chapter 4

Functional Data Analysis

4.1 Background

Functional data analysis (FDA) is a relatively new statistical methodology, which has developed rapidly over the last 10 years. Limited material is available on this subject (particularly at the start of my PhD), although existing applications such as those presented by Ramsay and Silverman [103] emphasise its great potential. It is the exploitation of this potential within the biomedical systems under consideration that is the focus of this research. One of the few published applications of FDA in the medical domain is by Shi and colleagues [104], where the centre of mass of paraplegic patients (as they stand) is modelled using FDA.

FDA is the analysis of functions or curves, opposed to a large number of discrete data points. Inferences can then be drawn from a dataset of curves. The original format of the data is not usually in the form of a curve, although intuitively they may be represented as a function. Methods are therefore required to express the data in a suitable functional form, such as basis functions and smoothing techniques (these will be discussed in section 4.3).

Autocorrelation amongst repeated measures generates several problems for many

statistical analyses [1]. As the frequency of measurements increases, such that adjacent measurements become closer in time, their correlation also increases thereby exaggerating subsequent collinearity problems. Statistical methods, such as multiple regression, which rely on the independence or near independence of observations, are therefore not always applicable. When observations are considered as functions rather than a collection of correlated points, this issue is circumvented. Indeed, the more frequent the repeated measurements, the more justified the approach. A useful feature of FDA is that it is not necessary for the data points to be equally spaced. After the construction of the curve, the original data points are essentially discarded and analyses are carried out on the fitted function; though equally spaced measurements makes the fitting of the curves easier.

FDA will enable the analysis of each patient because individual curves will be fitted for and represent each patient. This still remains a large quantity of data to analyse. It will also be possible to analyse groups of patients as there are techniques available to combine the functions.

4.2 General examples

A number of examples are available to facilitate the understanding of the concept of FDA. One example is where the collection of curves can be averaged to estimate a mean curve. If the aim of the analysis is to compare two groups then a mean curve may be constructed for each group (with confidence limits) to determine whether the two group averages are the same. Ramsay and Silverman use mean functions in an example on weather records [105]. Temperatures are recorded over the period of one year from 35 weather stations across the globe. Fourier series basis functions are used to fit curves to the raw data as they exhibit sinusoidal properties (more details will be given in section 4.3.3). Mean functions are fitted essentially to summarise the curves into five regional groups, as well as an overall mean function (more details will be given in section 4.4). Climatologists can

then use these summaries to talk about typical weather patterns and about variability in these patterns over time.

The derivatives of a function are useful in investigating growth data; e.g. when examining height or weight of an individual in the lifecourse context, as one may be interested in the rate of change of weight or height (first derivative) or the intensity of growth spurts and their timing (second derivative). Data collection in this field is often complex and time consuming as subjects need to be followed up for long periods of their lives, though these tend to be good quality datasets and allow important questions to be addressed, such as; is birth weight associated with whether or not a person develops type 2 diabetes at age 30? More details will be given on these concepts in later sections in terms of greater statistical theory and how ideas have been transferred to the biomedical systems.

4.3 Basis Functions

It is important that the functions are smooth for certain analyses. In order to estimate the function, smoothing techniques are required. A wide range of basis functions are available to fit smooth functions: for example, polynomial bases, B-splines, P-splines, Fourier series or Wavelets. An informed choice should be made based on knowledge of the data. Smooth functions are defined mathematically as being continuous and differentiable, perhaps a number of times. Constraints can be imposed to ensure the fitted functions have continuous derivatives of a given order.

It is a key step to choose the most appropriate basis function for the data. It is important to represent the data so that key features are highlighted and hence can be efficiently and effectively analysed. For example, if the data are periodic and have sinusoidal features then a Fourier series basis would be most suitable. A wavelet basis copes well with discontinuous or rapid changes in behaviour. The data in the biomedical systems of interest do not exhibit such behaviour, therefore these bases are not applicable. The basis

functions to be considered are B-splines and polynomial regression functions in the form of loess curves; the reasoning behind these choices is discussed in the following sections.

4.3.1 Loess

Loess was introduced by Cleveland in 1979 [106] and later developed in 1988 [107]. It is an extension of an existing methodology where local polynomials have been used to smooth time series plots, with equally spaced data points [108]. Simple polynomials are fitted to local subsets of data, and this process is then repeated across the full range, to construct a smooth function known as a loess curve. The loess curve shows the dependent variable as a smooth function of the independent variable. An attractive feature of loess is its simplicity. Usually only lower-order local polynomials are fitted; the highest order polynomial often fitted is degree two (local quadratic polynomials). The theory of loess is that a function can be approximated by low order polynomials and that simple models will be sufficient to fit the data, which make it ideal as an exploratory technique. Further, no specification of a global function is required. Hence loess is a flexible technique.

A smoothing parameter (α) controls the flexibility of function. The larger the smoothing parameter, the smoother the function. Smaller values of α will fit close to the data. In loess analysis, too small a value is not desirable as this will capture too much random error in the data, though the aim of loess is to capture the underlying smooth relationship. The smoothing parameter is determined by the user and the choice is made depending on how smooth a function is desired. A choice of smoothing parameter is usually between $\frac{\lambda+1}{n}$ and 1, where λ is the degree of the local polynomial and n is the sample size of the data to which the function is being fit. Note that in R the default value is 0.75.

In order to fit a loess curve, the subsets (known as the neighbourhood) for which the local polynomials are fitted must be specified. They are determined by a nearest neighbours algorithm [106]. Loess also incorporates potential autocorrelation amongst

the explanatory variables, in terms of a weight function. This is based upon the theory that measurements close together are more related than those farther apart. Most weight is given to the points closest to the point of estimation (since these points are likely to be most correlated), with less weight being given to those points farther away. This suggests that loess would be a useful technique when data are unequally spaced, such as growth data.

An advantage of the loess methodology is that it provides an exploratory graphical tool, which will give insight to the data, and hence enable an informed decision to be made about which other statistical methodologies may be applied. Furthermore, as no specification of the function is required prior to the analysis, it means that the true features of the data will be revealed, as opposed to a method where the function is predefined (such as a sine wave or a straight line relationship). This means that a loess curve will reveal a nonlinear and linear relationship between the dependent and explanatory variable. The only features to consider prior to fitting the loess curve is the degree and smoothing parameter, this is an advantage as it requires the user to understand their data rather than fitting functions without prior knowledge of the data.

In terms of disadvantages of the methodology, loess is prone to the effects of outliers in the data, though this is the case with other least squares methodology. Loess also requires large densely sampled data to get the most of the methodology, therefore loess should be chosen as an analysis technique when this is a feature of the data, rather than with small samples of data where measurements are uncorrelated. Once fitted, it is not possible to represent the loess curve as a mathematical formula, although if loess is used simply to display the data graphically, then this drawback is minor.

4.3.2 B-splines

B-splines are commonly used in FDA due to their flexibility, ability to capture long-term trend, and ability to control the shape and smoothness of the function [109]. Fast computation of the basis function is an attractive feature that can be achieved with B-splines. B-splines can be implemented using the statistical software R, and these routines have been validated by many users. Derivatives of the smooth function can also be calculated, which facilitates further analysis.

A B-spline curve $C(t)$ is a linear combination of B-spline basis curves $N_{i,p}(t)$. B-spline basis functions are illustrated in figure 4.1 and defined as:

- $N_{i,0}(t) = \{1$ if $t_i \leq t < t_{i+1}$ and $t_i < t_{i+1}$
- $N_{i,0}(t) = \{0$ Otherwise
- $N_{i,p}(t) = \frac{t-t_i}{t_{i+p}-t_i} N_{i,p-1}(t) + \frac{t_{i+p+1}-t}{t_{i+p+1}-t_{i+1}} N_{i+1,p-1}(t)$

Basis functions are calculated recursively and are of degree p (order $= p + 1$). The B-spline curve $C(t)$ is defined as:

$$C(t) = \sum_{i=0}^{n} P_i N_{i,p}(t)$$

which is formed piecewise, joining smoothly over a vector of knots $T = \{t_0, t_1, ..., t_m\}$. The parameters $P_0, ..., P_n$ are known as the $n + 1$ control points. The degree is thus $p = m - n - 1$.

In general terms, for a B-spline of degree q:

- it consists of $q + 1$ (i.e m) polynomial segments
- each polynomial segment is of degree q

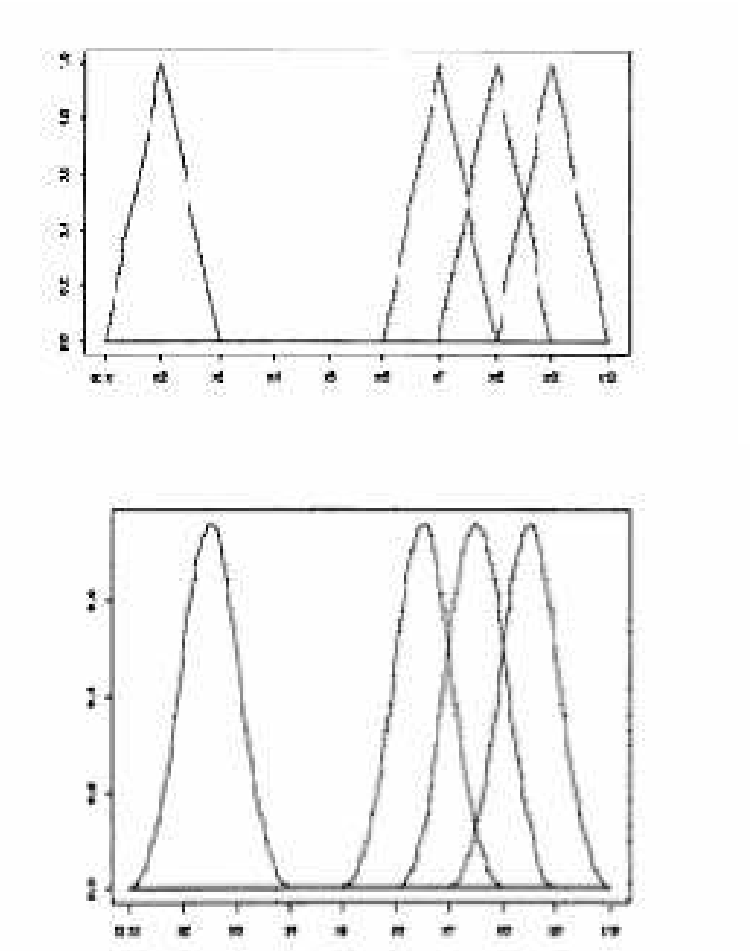

Figure 4.1: Examples of B-spline basis functions. Top: Uniform B-spline of degree 1. Bottom: Uniform B-spline of degree 2

- the polynomial segments join at q inner knots (discussed in the following section) and
- the total number of knots is $q + 2$.

Knots

The knots must be specified before the B-spline is calculated; from there calculation is done recursively. The number and position of knots is one of the important features to consider when fitting such curves. Note also that the functional basis is truncated: for B-splines this might be at order 4, for example, in order to fit piecewise cubic polynomials. The order of truncation is a further feature to be specified in functional fitting.

The spacing between the knots helps to define the shape of the basis function. Where knots are equidistant the B-spline is uniform, otherwise non–uniform. Equidistant knots, however, only allow limited control over smoothness (discussed in the following section) and the fit of the curve. If the knots are positioned close together, then the curve moves close to the corresponding control points, and if the knots are farther apart the curve moves away from the control points. For example, when investigating long-term or seasonal trend the placement of knots is important: when the knots are close together this allows examination of seasonal trend and knots farther apart to estimate long and medium term trend. A further issue is the choice of the number of knots; too few leads to underfitting and too many leads to overfitting. In a life-course example, an ideal scenario for the placement of knots would be at the ages where each measurement is recorded, as it is often the case that these data are unbalanced as more measurements are taken in the early stages of a persons life and less frequently in the later stages, with the first and last knots placed at the first and last ages. This would capture the rapid growth that occurs in this early period and the steadying off of growth that occurs after puberty. On the other hand, this may lead to overfitting, if, for example, measurements are recorded monthly for the

first two years and yearly thereafter.

Kooperberg and Stone [110] and Friedman and Silverman [111] have proposed schemes for the optimum number and position of the knots. They suggested that the knots should be placed either at or near selected data points, with the first and last knots at the first and last data points. In addition, the middle knots should be equally spaced. This seems a reasonable proposal. However, if the knots are placed at each data point then this is likely to overfit the data, which would ignore the general and overall trend. It is important that the knots are placed reasonably close to the data points, as otherwise features of the individual curves would not be captured. Knots should be chosen and placed in context of the data, which may not be achievable if relying solely on a computer algorithm to make the decision.

Friedman and Silverman [111] use an algorithm which uses the data to automatically select the number and position of the knots. Focus is placed on accurate estimation of the curve and not its derivatives. The strategy for knot placement is stepwise. The first knot is placed at the position which yields the best corresponding piecewise linear fit. Thereafter each additional knot is placed at the location which gives the best piecewise linear fit, which includes the previously placed knots. Knots are added in this manner until some maximum number is reached. At each eligible knot location a linear least squares fit must be performed to obtain the corresponding piecewise linear smooth curve $C(t)$.

Smoothness

Besides the order of the functional basis, smoothness is ensured by imposing a roughness penalty on either the fitted function or a derivative of the fitted function. This is most clearly expressed mathematically. Typically, the roughness penalty is defined with a squared norm, so that least-squares fitting is augmented as follows:

$$min \sum_{j=1}^{k} \{y_j - \sum_{i=1}^{m} P_i N_{i,p}(t)\}^2 + \alpha \int_{x_{min}}^{x_{max}} \{\frac{d^k}{dt} \sum_{i=1}^{m} P_i N_{i,p}(t)\}^2 dt$$

where $\alpha \geq 0$ is the roughness penalty parameter, k is the number of data points [112].

Numerical example

Suppose we have a knot vector $T = \{0, 0.5, 1, 1.5, 2\}$, thus $m = 4$ and $t_0 = 0$, $t_1 = 0.5$, $t_2 = 1$, $t_3 = 1.5$ and $t_4 = 2$. The basis functions of degree 0, 1 and 2 referring to these parameters have been calculated by hand to illustrate how the functions are fitted.

The basis functions of degree 0, $N_{i,0}(t)$ are quite simple (see table 4.1); they only exist on the range they are defined, and elsewhere they are equal to zero.

Basis Functions	Range	Equation
$N_{0,0}(t)$	[0, 0.5)	1
$N_{1,0}(t)$	[0.5, 1)	1
$N_{2,0}(t)$	[1, 1.5)	1
$N_{3,0}(t)$	[1.5, 2)	1

Table 4.1: Basis functions of degree 0

Table 4.2 shows the basis functions of degree 1 and table 4.3 shows the basis functions of degree 2. An illustration is given below of how $N_{0,1}(t)$ is fitted (basis function of degree 1). First note that:

$$N_{0,1}(t) = \frac{t - t_0}{t_1 - t_0} N_{0,0}(t) + \frac{t_2 - t}{t_2 - t_1} N_{1,0}(t)$$

with $N_{0,0}(t)$ defined over the range [0, 0.5) and $N_{0,1}(t)$defined over the range [0.5,1). Thus:

$$N_{0,1}(t) = \frac{t - 0}{0.5 - 0}(1) + \frac{1 - t}{1 - 0.5}(1)$$

which yields:

$$N_{0,1}(t) = 2t + 2(1 - t)$$

Basis Functions	Range	Equation
$N_{0,1}(t)$	[0, 0.5)	$2t$
$N_{0,1}(t)$	[0.5, 1)	$2(1-t)$
$N_{1,1}(t)$	[0.5, 1)	$2t-1$
$N_{1,1}(t)$	[1, 1.5)	$3-2t$
$N_{2,1}(t)$	[1, 1.5)	$2(t-1)$
$N_{2,1}(t)$	[1.5, 2)	$2(2-t)$

Table 4.2: Basis functions of degree 1

.

Basis Functions	Range	Equation
$N_{0,2}(t)$	[0, 0.5)	$2t^2$
$N_{0,2}(t)$	[0.5, 1)	$-1.5+6t-4t^2$
$N_{0,2}(t)$	[1, 1.5)	$4.5-6t+2t^2$
$N_{1,2}(t)$	[0.5, 1)	$0.5-2t+2t^2$
$N_{1,2}(t)$	[1, 1.5)	$-5.5+10t-4t^2$
$N_{1,2}(t)$	[1.5, 2)	$2(2-t)^2$

Table 4.3: Basis functions of degree 2

The basis functions that have been calculated can be fitted to a function, for example $sin(t^2)$. This is a non-periodic function so B-splines may be used as a basis. Figure 4.2, 4.3 and 4.4 show the basis functions of degree 1, 2 and 3 (not calculated in the above tables) fitted to the $sin(t^2)$ function. It was not possible to fit the basis of degree 0 as this was unstable. Figure 4.2 is an impractical representation of $sin(t^2)$ as the piecewise curves fitted at the knots are linear, additionally the function is not smooth and does not pass through all the data points. Figure 4.3 is a better representation of the function $sin(t^2)$ as it passes through all data points, except one where it almost passes through.

The basis function is smooth and captures the curvature of the true function. Figure 4.4 shows the basis function of degree 3, which is very similar to the basis function of degree 2 but now passes through all data points, which means it is a better representation of the true function $sin(t^2)$. Note that the B-spline basis function of degree 3 (order 4) seems to fit the function $sin(t^2)$ almost perfectly, although this has been fitted to a small range of data and outside of the range the B-spline may be erratic and not fit so well.

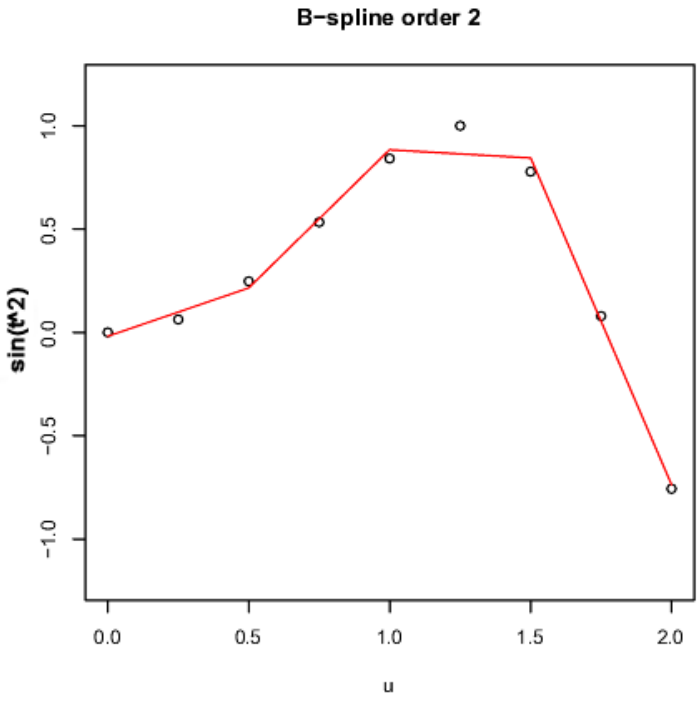

Figure 4.2: B-spline basis function order 2 fitted to $sin(t^2)$

4.3.3 Other basis functions

As mentioned previously, P-splines, Wavelets, and Fourier series are other frequent choices of basis functions, though will not be used in this thesis. Features of the B-splines and loess curves make them the most suitable choice for the data from the

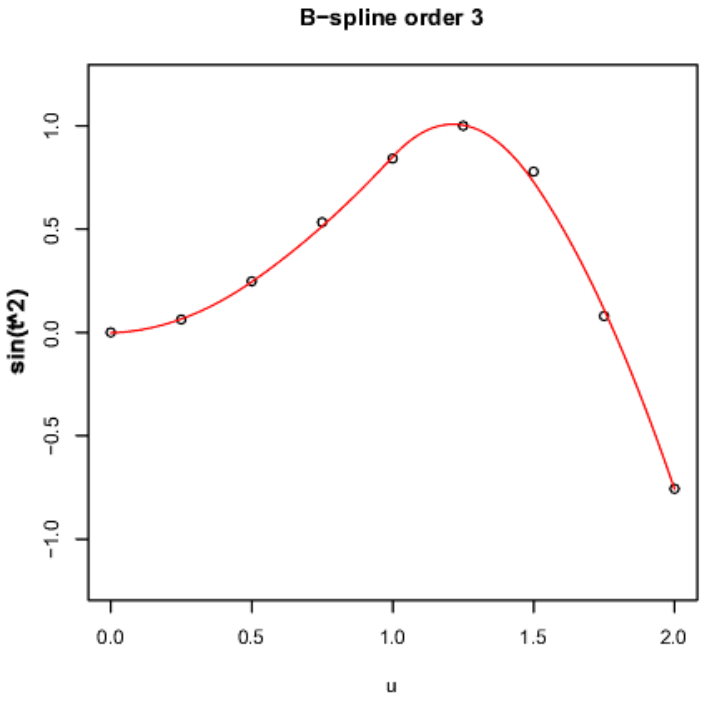

Figure 4.3: B-spline basis function order 3 fitted to $sin(t^2)$

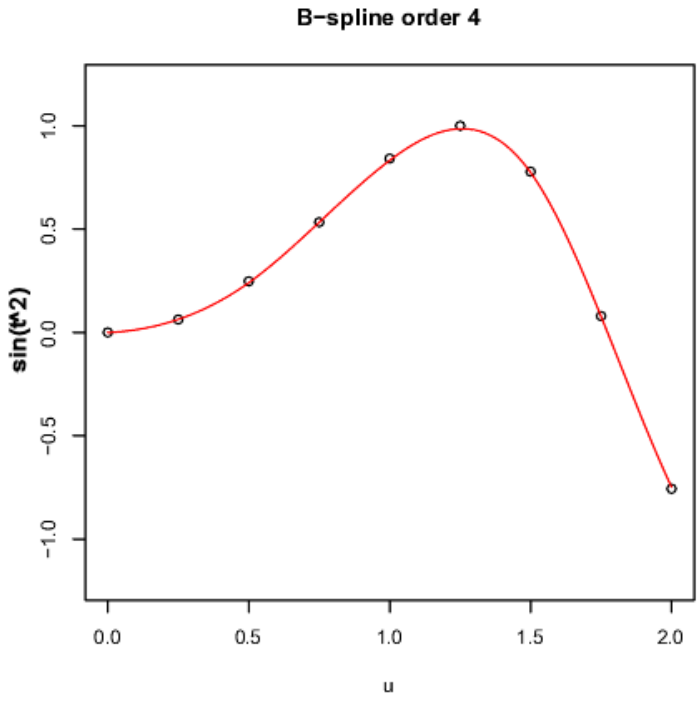

Figure 4.4: B-spline basis function order 4 fitted to $sin(t^2)$

biomedical systems; similarly features of Fourier series, Wavelets and P-splines make them an inappropriate choice. Fourier series are an excellent choice if the data are periodic and have sinusoidal features [113]. If it is not clear prior to the analysis whether the data exhibit such features, it may be implied that the sinusoidal nature of the data occurs due to the inherent nature of the function. Fourier series are the traditional basis of choice for long time series. Fourier series would be most useful, for example, to model temperature over a period of years, as it is clear that weather trends tend to be cyclical from year to year. Within the Fourier series basis function it is possible to specify the length of the period, thus with weather data this would be 365.25 days. This type of basis is most appropriately used with a stable function, where there are no strong local features. In particular Fourier series are inappropriate for functions that are discontinuous (or discontinuity in lower order derivatives). A wavelet basis is most suitable when the data is discontinuous or exhibits rapid changes in behaviour. If the data are on a bounded interval then wavelet bases deal well with this feature. In computational terms, Fourier series and wavelets are fast and efficient, though produce complex harmonics. The use of wavelet bases is relatively recent and experience of them is limited.

P-splines (polynomial splines) were developed from the bases that did not accommodate local features, hence P-splines have the capacity to capture changing local behaviour. P-splines are a combination of splines and penalties on the estimated spline coefficients [112]. B-splines are better used in practice than P-splines as the main difference between the two is that B-splines are zero everywhere except over a finite (specified) interval [109]. Another difference is that the knots used with a P-spline are predetermined; some users see this as advantage over B-splines where the choice of knots is more arbitrary. On the other hand, some may see this as meaning that the B-spline approach is more flexible since unneeded knots may be freely eliminated, and vice-versa.

4.3.4 Summary of basis functions

The success of basis function is reliant on the appropriate specification, congruent to the application, of the following parameters:

- the functional basis
- the order of truncation of the basis
- the number of knots
- the placement of knots
- the order of the roughness penalty
- the size of the roughness penalty parameter α.

4.4 Mean functions

The mean function is a simple analogue of the classical mean for univariate data. It can be calculated by averaging the functions pointwise across the replications. Similarly with the variance, and hence confidence intervals.

In some examples, smoothing of the mean function will be required, in addition to the already smooth functions. Whether the mean requires additional smoothing will depend on whether it is deemed that the function has high local variability. If it is necessary to smooth the mean function further, then the roughness penalty approach used to smooth the original curve will be applied. The best choice of smoothing parameter is often made subjectively, as this ensures that the data being analysed are understood.

4.5 Phase plots

Phase plots were discussed by Hirsch and Smale [114], who considered phase portraits as modelling the dynamics of a system. The path of a 'particle' over time is mapped in a 2–dimensional plane. The particle then creates a trajectory, which represents the dynamic system. Dynamic systems may also be constructed from the differential of the trajectory.

Phase plots have been adapted by Ramsay and Silverman [103], whereby they assessed the dynamics and energy in an econometrics system. They considered the control of supply and demand cycles within key periods in modern history, such as The Great Depression and different periods around World War II. Phase plots are defined as plots of the second derivative (acceleration of the function) versus the first derivative (velocity of the function). This means that the function needs to be of a relatively high order as the phase plots require the third derivative of the curve to be smooth, to ensure the second derivative is also smooth. Due to this complexity it will be investigated whether a plot of the first derivative against the value of the observation could be used to model the dynamics within a system. If this were possible then the function could be of a lower order.

Plotting pairs of derivatives is informative as the derivatives expose effects that are not seen in the original functions. The first derivative represents the rate of change of the process whilst the second derivative represents the instantaneous curvature of the process.

4.5.1 Application of phase plots for the renal anaemia system

Phase plots will be used within the renal anaemia system to determine whether it is possible to model control of Hb. The notion is that it will reveal dynamics of control of Hb in the patients. Phase plots will either be defined here as plots of the first derivative of the Hb function (units of g/dl/month) plotted against Hb values, or the second derivative

of the Hb function vs. the first derivative of the Hb function. The first derivative is key in assessing control, as this is the rate of change of trajectory and this considers the frequency of changes in Hb along with the nature of the change.

It is helpful to consider phase plots with an analogy. A pendulum swings with greatest velocity at the lowest point of its travel; when its kinetic energy is largest. At the point the pendulum is at its lowest position, its potential energy is lowest. Note that the position of the pendulum is analogous to the level of Hb and the velocity of the pendulum is analogous to the rate of change of the Hb trajectory. At either end of its swing, the pendulum velocity is instantaneously zero, and its kinetic energy therefore takes its lowest value at zero. At these points, at the top of the arc, the potential energy is highest since the pendulum is at its highest elevation. Thus, during its swing the pendulum exchanges energy from kinetic to potential, and back again. Where the pendulum point velocity and position are related graphically in a 'phase plot; such a sequence of energy exchange corresponds to something like a circle, with a centre of minimum potential energy [114]. At the start, these circles are large. As friction acts, the pendulum makes smaller swings and the total energy declines; the circle becomes a spiral and progresses inwards to the centre. A system displaying large circles in a phase plot can be considered to have a lot of overall 'energy'. In terms of the analogy of Hb trajectory, a large circle reflects little/ poor control (wasted 'energy'), while small phase circles correspond to less overall 'energy', reflecting close/ good Hb control. Such circles and spirals in the control of renal anaemia are centred on the optimum Hb value. The initial point will be noted from which patient progress is followed over time.

4.6 Implementation

Functional data analysis is available in R, S-plus and Matlab. As R [115] is available freely for download and usage, as are the FDA libraries, this was motivation for using this

software. A disadvantage of using the software R occurred when the FDA library was updated in 2008, whereby the code to fit the basis functions changed. This subsequently meant that the original code written for analysis of the data in this thesis required updating. This is a common feature of R; that updates often cause minor problems with the code already written.

Ramsay and Silverman wrote two books which are very useful sources for FDA. These were a major factor in the learning and implementation of FDA, as they also wrote the library for use in R. No training courses were able for FDA, although I attended a number of presentations on FDA at the International Biometric Conference in Montreal, 2006, where I was able to discuss FDA with experts in the field. De Boor [109] and Eilers [112] also provide useful sources for B-splines.

4.7 Aims and objectives

The aim of this section of the thesis is to determine whether FDA can contribute knowledge, insight and understanding to the chosen biomedical systems. FDA will be used as an exploratory tool with the CA system; by fitting loess curves to the data to gain an understanding of the relationship between cerebral blood flow and arterial blood pressure. For the renal system, smooth functions will be fitted to the data. The fitting of smooth functions will lead to derivatives of the curves being taken and the construction of phase plots. The results of this section will have a strong graphical aspect, which is advantageous as it provides insights into key aspects of both cerebral autoregulation and the control of renal anaemia. The more specific aims related to the systems are discussed in the following sections.

4.7.1 Cerebral Autoregulation

The aim within the CA system is to investigate if FDA, in the form of loess curves, will reveal the form of the relationship between blood flow and blood pressure. Insight may be gained about CA by considering CBF over a range of ABP, which is the usual approach for assessing CA. There is already a preconceived idea of the relationship, in the form of the CA curve [29], although assessment of CA is being made during a surgical procedure, for which the CBF–ABP relationship is unknown. This highlights the advantage of few assumptions being made about the underlying function of the data, in the construction of loess curves. The performance of loess will be tested by fitting a loess curve to a dataset [116] for which the ideal autoregulation curve is known to fit, to determine whether the ideal CA curve will be revealed if in fact present. Alternatively, the dependent variable (CBF) could be considered as a function over time, which would indicate the behaviour of CBF throughout the surgical procedure.

4.7.2 Renal Anaemia

It is noted that Hb levels for an individual may oscillate sometimes over a long time period. This at first might suggest the use of Fourier series as a functional basis. There is the possibility, however, that oscillations lack periodicity. More importantly it is necessary to demonstrate that the oscillatory behaviour arises from the patient response and is not induced by the choice of functional basis.

The aim within the renal anaemia system is to investigate if the fitting of smooth functions to patient Hb trajectories provides an insightful summary of patient response under the protocol supported by the CDSS for administration of epoetin. For these data it is reasonable to assume that Hb varies smoothly over time. In many stable patients, Hb has been shown to follow a smooth periodicity in association with epoetin dose changes [98]. Thus, fitting smooth B-spline curves to Hb measurements over time, for each patient, will

reveal how Hb varies throughout the course of treatment and provide understanding of patient response. The curves may also be useful in identifying if a patient's Hb is at the optimum level (11.8 g/dL) or within the target range (10.5–12.5 g/dL). It seems that this methodology will reveal many interesting features about the data, such as whether Hb is stable around the target, or exhibits the other extreme where Hb is oscillating off target. The knots will be carefully chosen so the curves are sufficiently flexible and fit close to the data points. The order and smoothing parameters will also be selected to ensure derivatives of the curves are smooth. Each function will be inspected to ensure they fit the data well.

The fitting of phase plots will allow for further extension of analysing control; something that has not been previously exploited in the literature.

A mean function with a 68% confidence interval (CI) will be calculated for Hb over time, for each of the two epoetin agents. This will enable comparison of the two agents and hence allow the question to be answered of whether the two agents are statistically significantly different in managing Hb. This is an extension of the idea used by [99] , though instead of calculating a mean and CI at the single monthly time points, the mean function and its CI will be continuous across the 12 month period, allowing for a more comprehensive comparison of the agents, as it will be seen how Hb varies throughout the year, as opposed to just at monthly intervals. The 68% confidence limits are the region of the graph where the CI trajectory is one standard error either side of the relevant mean curve. Where the confidence bands do not overlap, the two mean curves will be separated by 'at least two standard errors', which will indicate where the mean trajectories are significantly different from one another at the 5% – level.

4.7.3 Statistical Issues

FDA has previously been used where many measurements have been collected; for example, in the application by Shi [104] thousands of measurements were collected. This raises the question, therefore, as to whether 13 measurements per individual (as in the renal anaemia system) is enough to construct the function robustly and to apply the FDA techniques. If the data are particularly noisy then a smooth functional representation would be beneficial. If the function is correctly fitted, i.e with appropriately specified knots and smoothing parameters, this would enable the underlying relationship (clear signal) to be revealed (from the noise). This would highlight whether the patients are reasonably similar across the patient group or whether they all exhibit different behaviours (this will be useful in checking the assumption for multilevel modelling that individuals share characteristics but are not the same or totally different - see Chapter 5).

The statistical aims and objectives of this work are:

- To represent the data as smooth functions using appropriate basis functions.
- To organise the data so FDA techniques may be applied.
- To highlight important features of the data through graphical presentation.
- To investigate and model variability in the data.

4.8 Summary

The handling of functional data is similar to that of ‘usual’ data, though each datum is a continuous function. The aim of this work is to apply a variety of functional data techniques for analysis of the data from the two biomedical systems, to determine whether the methodology can contribute knowledge and insight to the medical domain. This

will be achieved by firstly constructing curves from the original repeated measurements using the appropriate basis functions. The data will then be analysed using FDA, such as constructing mean functions and taking derivatives of the curves.

The fitting of smooth curves to somewhat noisy data will allow the underlying relationship between two variables to be revealed. The expectation that the data in these biomedical systems are autocorrelated greatly justifies the FDA approach, as the smooth curve will be a better representation of the raw data than the data itself.

Chapter 5

Multilevel Modelling

5.1 Statistical aims and objectives

Multilevel modelling (MLM) is used in this thesis to illustrate that the statistical technique can provide distinct information about the two biomedical systems. MLM will be used to reveal different aspects about the applications compared to the FDA approach.

The applications require additional complexities other than fitting a simple multilevel model in order to represent the data suitably. General aspects of MLM will be discussed initially, with particular focus on the relevant aspects for the two applications. Discussion will then follow into how and why MLM will be applied in this thesis.

5.2 Background

Multilevel modelling is a powerful statistical technique, which is essentially an extension of ordinary linear regression modelling, though allows for more flexibility and complex data structures to be modelled. Multilevel analysis is used for data–sets with a clustered (hierarchical) structure differentiating it from the many 'traditional' statistical techniques

which make the assumption of independence. The hierarchical nature of the data is often viewed as a nuisance, however MLM embraces this complexity and allows additional features of the data to be exploited.

The term 'hierarchical linear model' was introduced by Lindley and Smith [117], a term which is used interchangeably with MLM. Early uses of MLM are found in the field of education [118] [119], where it continues to be used frequently. The structure of the education system lends itself to be analysed using MLM; for example, pupils (observation at lowest level i) are clustered within classes (level j), which in turn are clustered within schools (level k), which in turn are clustered within local education authorities (highest level l). More recently, MLM has been used within medical [120] and dental applications [121]. In dental research; sites around a tooth (level i), are clustered within teeth (level j), which are clustered within patients (mouths) (level k).

In medical research an obvious two level hierarchical structure arises, whereby patients are nested within hospitals. More complex data structures are also seen, as it is often the case that repeated measurements are collected on patients, also known as longitudinal data. For instance, measurements may be collected repeatedly over a relatively short time period; such as throughout a surgical procedure. Conversely, measurements may be collected over a number of weeks or months whilst patients are on a course of treatment and undergoing regular monitoring. Repeated measurements maybe considered as the lowest level of the hierarchy (level i), which are clustered within patients (level j).

Within the MLM framework, individuals are viewed as being similar apart from *random* variation and *fixed* measurable differences; note that MLM is also known as fixed and random effects modelling. This is a reasonable assumption to make, as it is likely that patients / individuals are sampled from a population who share similar attributes, i.e. patients undergoing a certain operation or patients with a specific disease. Within these groups it would not be plausible to assume that all patients would respond differently, or exactly the same; hence the assumption that patients are 'similar' is a balance between the

extreme views.

Random and fixed effects account for the differences which occur between patients. The fixed effects represent the average behaviour of the individuals in the sample, which are included in the model as mean value coefficients. The random effects allow for variation around the mean value to be modelled, and these are included in the model as either variance (e.g. in MLwiN) or standard error terms (e.g. in R). The advantage of fixed and random effects is that both population and individual characteristics can be modelled.

Due to the clustered nature of the data there is complexity in the structure of the variance. MLM allows for the total variance to be partitioned into a separate variance for each level. Considering the education example, one may assume that there is variation between the pupils in the class (i.e. within-class variation) and also between classes in the school (within-school variation). Accounting for the variation amongst individuals through random intercepts in the MLM frame work is much more parsimonious than fitting $n-1$ dummy variables in a standard regression model.

Within a generalised hierarchical linear model, the simplest structure for a MLM consists of only random intercepts, i.e. the relationship between the dependent and independent variable and all covariates remain fixed but the model intercepts may vary for each level in the model. For example, in the medical scenario this may represent a different intercept for each patient. Additional complexity is achieved to account for the covariate relationships, i.e. the relationship between the dependent and independent (or covariates) variables may differ. Thus:

- The random **intercept** accounts for the differences between baseline values.
- The random **slope** accounts for differences in the relationship between the dependent variables and covariates; and this may occur at any level of the hierarchy.

5.3 A two-level model

By means of understanding the concept, the methodological principle of MLM for a two-level model is illustrated. Denote the dependent variable to be y and the covariate with random slope x_1, and all other covariates $x_2, ..., x_N$ ($m = 2...N$)

$$y_{ij} = \beta_{0ij} + \beta_{1j}x_{ij} + \sum_{m=2}^{N} \beta_m x_{mij}$$

where: $\beta_{0ij} = \beta_0 + e_{0ij} + u_{0j}$ is the intercept term, $\beta_{1j} = \beta_1 + u_{1j}$ is the regression slope for x_1, and x_{mij} are additional covariates with related parameters β_m.

The intercept term can be divided such that: β_0 is the fixed (mean) intercept; e_{0ij} is the random intercept variation at level–1, where $e_{0ij} \sim N(0, \sigma^2_{e0})$ and σ^2_{e0} is the level 1 variance; u_{0j} is the random intercept variation at level–2, where $u_{0j} \sim N(0, \sigma^2_{u0})$ and σ^2_{u0} is the level 2 variance.

Similarly the regression slope for x_1 can be divided such that: β_1 is the fixed slope, and u_{1j} is the random slope variation at level 2, where $u_{1j} \sim N(0, \sigma^2_{u1})$ and σ^2_{u1} is the slope variance at level 2.

Note that σ_{uo1} as the covariance between the random intercept (u_{0j}) and the random slope (u_{1j}).

5.4 Considerations and assumptions of multilevel modelling

As with all statistical models, there are a number of key assumptions to consider with MLM. Certain assumptions for MLM are the same as for ordinary linear regression, although MLM has more flexibility as it allows some of the assumptions to be relaxed

and modelled explicitly. It is vital to check the assumptions: if they are not upheld there is likely to be bias in the estimated coefficients. Moreover, any conclusions drawn from the model will be erroneous. The assumptions and considerations for MLM are stated below:

- **Hierarchy :** By failing to recognize that there is a hierarchical structure in the data; i.e. assuming that the observations are independent, the standard errors of regression coefficients will be underestimated, leading to an overstatement of statistical significance, in particular the coefficients at the highest level will be most affected.

- **The Structure** of the model is of particular importance, as it is essential that the fixed and random parts of the model are correctly specified. This feature is not easy to 'test' as the model must make both clinical and statistical sense. The model should be constructed using prior knowledge of the data and specific clinical domain. The specification of the model is of importance to the two applications in this thesis and will therefore be discussed in detail in later sections.

- **Homoskedasticity** is where residual variance is constant at each level of the hierarchy, which is assumed. This assumption maybe relaxed and replaced by heteroskedasticity, where the variance will depend on an explanatory variable. If homoskedasticity is assumed where the variance is really heteroskedastic, the coefficients for the variance (and hence standard errors) will be underestimated. A plot of residuals vs. fitted values is often used to determine whether there is constant variance. Homoskedasticity and heteroskedasticity are illustrated in figures 5.1 and 5.2 respectively.

- **Independence** of the residuals at each level of the hierarchy is assumed. This assumption will be considered further in section 5.5, in particular the implications when this assumption cannot be upheld and the lowest level residuals are correlated.

Solutions will be suggested as to how the issue may be successfully resolved. Note that this assumption is different from assuming that within a MLM the measurements within an individual are not independent.

- **Normality** of random parameters and residuals at each level. I.e. the random intercepts are assumed to be normally distributed, with zero mean, independent and identically-distributed.

The assumption for independence of the lowest level residuals is often false when modelling longitudinal data, due to the nature of the data and measurement collection. Checking for correlated residuals is often overlooked as the procedure is more complex than producing simple graphical plots of the residuals, which enables homoskedasticity and normality to be assessed.

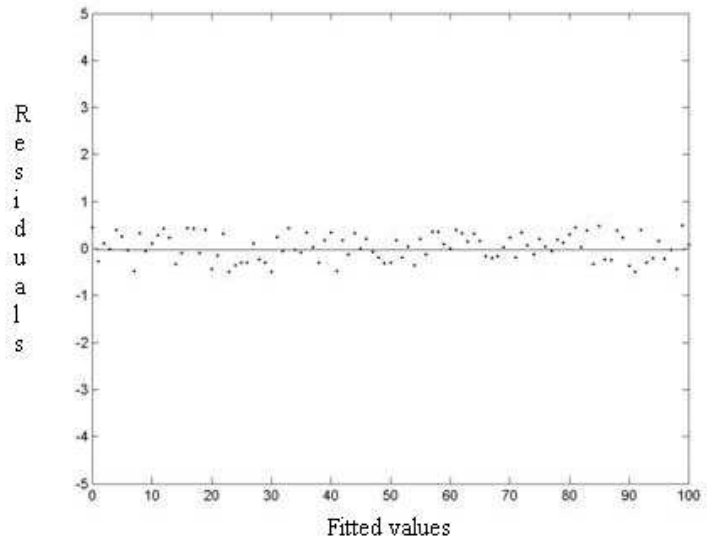

Figure 5.1: Figure showing plot of residuals vs. fitted values to illustrate homoskedasticity, where residual variation is constant for the range of data.

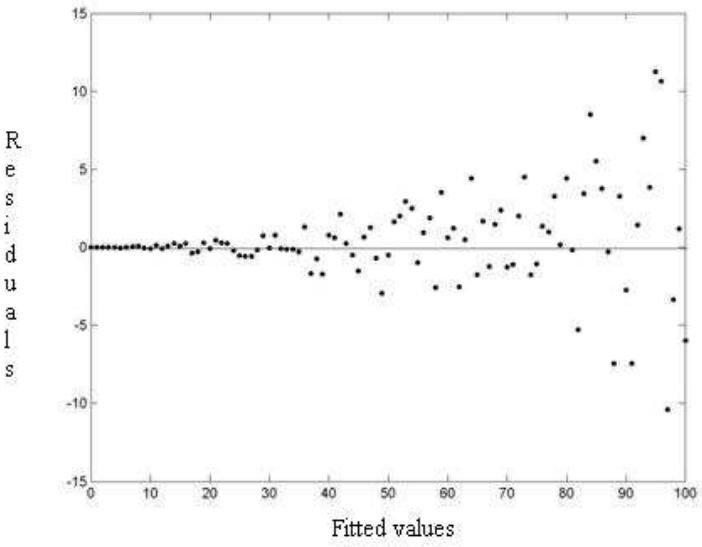

Figure 5.2: Figure showing plot of residuals vs fitted values to illustrate heteroskedasticity, where variation increases as the data values increase.

5.5 Autocorrelated residuals

In ordinary least squares regression (a basic single level model $y_t = \alpha + \beta x_t + \epsilon_t$) the within group residuals ϵ_t are assumed to be independent. This assumption can be relaxed in the MLM framework by incorporating a correlation structure for the residuals to account for dependency amongst observations. In particular, serial correlation structures are available to account for dependency in time series data, where data are observed sequentially over time.

As the residuals are a linear combination of the predictors, it is possible that the residuals will be serially correlated if one of the dependent or independent variables is serially correlated. Correlation in the error terms suggests that there is additional explanatory information in the data that has not been exploited, rather than a model that is specified incorrectly. When observations are collected on an individual close together in time (or a

natural sequential order), they will have similar departure from the underlying regression line which means that the residuals will be positively correlated; this correlation is referred to as autocorrelation.

Autoregressive models are particularly useful as the assumption can be made that the correlation between nearby measurements is stronger than measurements farther apart. In Autoregressive models, the distance between the residuals is known as the lag, and there will be stronger correlation between measurements at small valued lags compared with those at larger lags. Autoregressive models for the error structure express the current error ϵ_t, at current time t, as a linear combination of previous residuals plus a homoskedastic white noise term (defined here as: Z_t, where Z_t is normally distributed with mean 0).

$$\epsilon_t = \phi_1\epsilon_{t-1} + ... + \phi_p\epsilon_{t-p} + Z_t \tag{5.1}$$

Note that p refers to the number of past residuals to be included in the error structure model. The order of the autoregressive model is of order p, denoted $AR(p)$, which includes p correlation parameters: $\phi = (\phi_1, ..., \phi_p)$. Thus, a first order autoregressive model (AR(1)) of the errors is denoted: $\epsilon_t = \phi_1\epsilon_{t-1} + Z_t$, where $-1 \leq \phi_1 \leq 1$. AR(1) is a simple and effective autoregressive model. The single correlation coefficient ϕ_1 represents the correlation between observations one lag apart.

5.5.1 Diagnostics for autocorrelation

When assessing whether autocorrelation is present in the model residuals, it is useful to consider diagnostic plots of the normalised residuals; in particular the plot of the empirical autocorrelation function (ACF). Normalised residuals are defined as:

$$r_i = \hat{\sigma}^{-1}(\hat{\Lambda}_i^{-1/2})^T(y_i - \hat{y}_i).$$

Where $\hat{\sigma}^2\Lambda_i$ denotes the variance-covariance matrix for the i within group errors. If the within-group variance–covariance model is correct, the normalised residuals should be approximately distributed as independent random vectors. The Durbin-Watson statistic [122] tests for serial correlation amongst residuals, although from this test it is impossible to determine the nature of the correlation, for instance if stationarity or seasonality is present.

The ACF at lag l is defined as:

$$\hat{p}(l) = \frac{\sum_{j=1}^{M}\sum_{i=1}^{n_{j-l}} r_{ij}r_{i(j+l)}/N(l)}{\sum_{j=1}^{M}\sum_{i=1}^{n_{j-l}} r_{ij}^2/N(0)} \tag{5.2}$$

Where r_{ij} are the residuals from a fitted MLM, with $\sigma_{ij}^2 = Var(\epsilon_{ij})$, and $j = 1, \ldots, M$ observations at the uppermost level (i.e. patients) and $i = 1, \ldots, n_j$ repeated observations (for patient j).

The ACF is essentially a plot of the vector of correlation parameters $\phi = (\phi_1, ..., \phi_p)$, with a vertical bar representing each coefficient [3]. Critical bounds for the autocorrelations are usually plotted to denote correlations significantly different from zero. Approximate 95% bounds are $\pm\frac{2}{\sqrt{N}}$, where N is the number of observations [123]. The autocorrelations which extend beyond the limits are deemed statistically significant and signify that autocorrelation is present in the model at the time lag where a significant correlation coefficient occurs. A feature to be aware of is a correlation coefficient that is just significant or significant autocorrelation occurs at 'arbitrary' time lags then it may be that the correlation parameters are actually independent. Autocorrelation at larger lags is also less reliable as these are estimated with fewer residual pairs.

From the autocorrelation function it is possible to identify the specific correlation structure which is needed. For a first order process (AR(1)), the ACF decreases exponentially [123]. For higher order autoregressive processes the ACF may be a mixture of damped exponential or sinusoidal functions [3]. Figure 5.3 is used to illustrate the ACF for a

AR(1) process, i.e. when the residuals one lag apart are correlated.

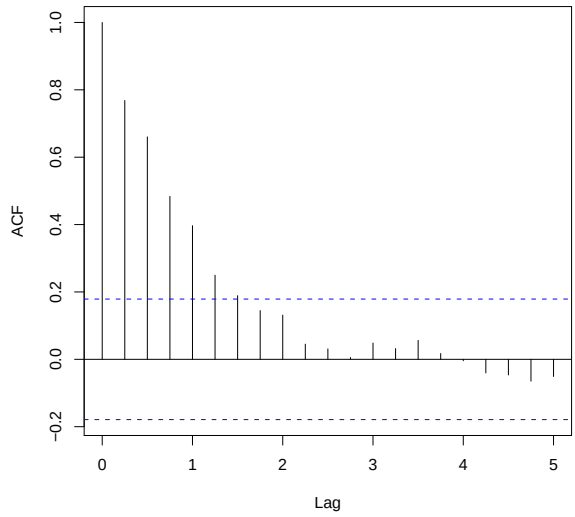

Figure 5.3: Autocorrelation function illustrating an AR(1) process. The dotted blue lines represent the 95 % confidence limits.

5.6 Assessment of model fit

Assessing the fitted model is important to ensure the correct and appropriate relationship between the outcome and explanatory variables is being modelled, as well as determining whether a correlation structure is needed for the residuals and if it is correct. There are many ways to do this, using numerous statistics constructed by numerous statisticians over the years. The models in this thesis will be assessed based on a number of factors, including Akaike's information criterion (AIC) and likelihood ratio tests (LRT). In addition to these numerical statistics, graphical summaries will also be used as they are a particularly effective way to assess model fit and should be examined routinely. The interpretability of the model is also of paramount importance, in particular prior clinical

knowledge should be considered and whether the model 'makes sense' for practical use.

Maximum likelihood estimation is a method for estimating model parameters satisfying the criteria of 'sufficiency' and 'efficiency' [124], though restricted maximum likelihood (REML) [125] is often preferred as maximum likelihood tends to underestimate the random parameters in a multilevel model [126] [127]. Furthermore, it is considered that REML is a more appropriate criteria for models with fixed and random effects [128], since REML produces less biased / more conservative estimates of the coefficients in the random part of the model [129]. Both maximum likelihood and restricted maximum likelihood are important to consider in the model fitting procedure, and will be discussed throughout this chapter.

5.6.1 Information criterion

Akaike's information criterion (AIC) [130] [131] is stated as yielding more plausible model solutions than solely relying on the likelihood value [132], because the likelihood value alone does not account for model complexity or parsimony. This makes the model more plausible as it takes into account a balance of bias and variance in the model construction, rather than a model based on a likelihood value which may include a number of nuisance parameters.

The AIC statistic is defined as:

$$AIC = 2p - 2ln(L)$$

where p is the number of parameters and L is the maximum likelihood for the estimated model. When 'REML' is used as the estimation method, L is replaced by the restricted maximum likelihood.

AIC takes into account the complexity of the model by considering both the statistical

goodness of fit and the number of parameters to be estimated which achieve this particular degree of fit, by imposing a penalty for increasing the number of parameters. Smaller values of the index indicate the models which are a relatively better fit, i.e. the one with the fewest parameters that still provides an adequate fit to the data. However, the AIC value is relative to the size of the dataset, such that larger datasets will yield larger AIC values.

Bayesian information criterion (BIC) [133] is similar to AIC, such that smaller values of this statistic indicates the 'better' model when comparing (two) models.

The BIC statistic is defined as:

$$BIC = (p)log(N) - 2ln(L)$$

where p is the number of parameters, L is the maximum likelihood for the estimated model and N is the sample size. When REML is used as the estimation method, L is replaced by the restricted maximum likelihood and $log(N)$ by $log(N - p)$.

Within this thesis AIC will be used as the preferred criterion between AIC and BIC. It is usually the case that the same inferences will be drawn from both criterion [134] and both are valid methods of model selection. Therefore AIC will be used because it tends to be a conservative criterion, whereas BIC more seriously penalises the introduction of additional parameters than AIC [135]. The models fitted in this thesis do not include many covariates, therefore strongly penalizing a model based on its parameters is not necessary.

5.6.2 Likelihood ratio test

Likelihood ratio tests (LRT's) are used to compare nested models, to determine whether the inclusion of additional parameters improves model fit. LRT's can be used to compare model fit by REML (usually maximum likelihood) if all models have been fit using REML and include the same fixed effects specification. The reason REML is used

over maximum likelihood is discussed later. LRT are particularly useful when deciding whether a correlation structure is required, as models with and without the correlation structure are nested. In this instance, the ACF should also be examined, to check whether inclusion of the correlation structure yields a model with uncorrelated residuals.

The likelihood test statistic is denoted :

$$LRT = 2log(L_1/L_2) = 2[log(L_2) - log(L_1)]$$

Where L_1 and L_2 are the likelihoods of models M_1 and M_2, where M_1 has k_1 parameters and is nested in M_2 which has k_2 parameters.

The null hypothesis H_0 is that M_1 is adequate, the alternative hypothesis H_1 is that the more complex model (M_2) is required. If the p-value is significant for the LRT compared to the $\chi^2_{k_2-k_1}$ distribution then it can be concluded that the alternative model is preferred as evidence to reject the null hypothesis.

It can be argued, however, that the p-value from the $\chi^2_{k_2-k_1}$ distribution is too large or inaccurate [136], if calculated from a small sample size. This suggests that the LRT should not be solely relied upon to decide whether one model is better than another and that a combination of criteria should be examined.

5.6.3 Graphical summaries

To check whether the within group errors are normally distributed a QQ-plot of the residuals should be used, or alternatively identifying whether the residuals are scattered randomly around zero by plotting the standardised residuals for each of the upper level units (i.e. patients).

A plot of standardised residuals vs. fitted values is used as a means of checking for homoskedasticity of the variance, as seen in figure 5.1 and figure 5.2.

Another assumption to check is whether the random effects are normally distributed with mean zero and no substantial correlation amongst the random effects. A plot of the random effects values for each of the upper level units is used for this, and this plot will also show if there is homogeneity in the variance structure. It is possible to determine whether random effects parameters are required by examining their respective confidence intervals. If the random effects standard deviations are significantly different from zero, this indicates that they are required in the model, as they add additional information to the model by indicating there is some variation in the random effect that should be explicitly modelled.

As discussed in the previous section, a plot of the ACF is an effective way of identifying correlation in the lowest level residuals and hence whether a correlation structure is required.

5.6.4 Interpretation and a priori knowledge of data generation

In addition to the factors for assessing model fit, which were discussed above, it is also important to assess the models based on prior knowledge of the clinical scenario and also knowledge of the data (such as; format and data collection). Firstly, the interpretation and specification of the model should make clinical sense. This can be achieved by understanding the clinical context of the data and also through collaboration with clinical colleagues. If a model has so many parameters, for example, this would be difficult to interpret in the clinical domain (and also for the statistician). Therefore, parsimonious models should be fit, whereby there is a balance between the complexity and the number of parameters.

Having knowledge of the data is of paramount importance. One should know where the measurements are collected, for example in controlled laboratory conditions or whilst the patient is under stress. This would give an indication as to whether a clear relationship will

be seen, or if the patient is under stress then the underlying relationship will be masked by noise. Furthermore, knowledge of the time between measurements is a key factor as a short time period, for example, may result in measurements that are autocorrelated. Using statistical methods which assume independence between measurements would then be incorrect. One should be aware that if a complex correlation structure is incorporated in the modelling, to account for the autocorrelated residuals, this correlation parameter may also have clinical meaning and relevance.

Within the MLM framework it is possible to analyse both balanced and unbalanced data. Unbalanced data will occur where a different number of measurements is collected for each individual. This may be due to missing values or where the time period of measurement collection varies between each patient. This is an advantage of MLM, as some statistical methods cannot be used with unbalanced data.

5.6.5 Summary

The criteria discussed here are the most effective way to assess the fit of the model, especially when fitting a large number of models and when working in the clinical domain, where ease of interpretation is necessary.

In summary, models will be assessed using the following:

- AIC
- LRT
- ACF
- Residual plots: e.g. standardised residual vs. fitted values
- Clinical insight / interpretation and knowledge of data generation

It should be noted that no individual assessment criterion should be solely relied upon; a better fitting model results from using a combination of carefully selected criteria and prior knowledge of the clinical context. It is often viewed that the graphical summaries are an 'easy' avoidance of calculating complex statistics to determine model fit. In fact this is not the case, as graphical summaries highlight complex and complicated features of the data, in addition to gaining understanding of the data, which could not be achieved purely from visually scrutinising numbers. The complexity should be in fitting the model and not in assessing model fit.

5.7 Implementation

Modelling with autoregressive correlation structures was introduced in 1994 [137]. An AR(1) correlation structure was then made available for implementation in MLwiN [138], though currently it is no longer possible to implement an extended correlation structure in this package. It is, however, possible to fit a MLM with a variety of correlation structures using the statistical software R [115]. A number of R libraries are available (nlme, lme4, mlmRev), though nlme was found to be the most effective.

A number of MLM text books usually mention that it is possible to implement correlation structures of the residuals, albeit briefly, suggesting that this particular aspect of MLM is seldom used. One exception is the text book by Pinherio and Bates [139], which is a useful guide to fitting MLM in R and includes a detailed section on correlation structures.

5.7.1 Presentation of models

The fitted models will be presented in the respective results chapters. The format of the models will be presented in two ways. To display the large number of models fitted, the

software notation will be used. For the 'final' models the more traditional notation will also be used, as described in section 5.3.

R notation

The R code below shows how a general MLM is specified in the software.

```
MODEL = lme(response ~ primary covariate(+ additional covariate),
random = primary covariate (+ additional covariates)
| grouping level upper (/ grouping level lower) ,
correlation=corARMA(p=a, q=b), method="ML" or "REML")
```

Where `lme` refers to the function from the `nlme` library in R used to fit the linear mixed-effects model. Initially the `response` is specified. The $\sim$ is read as 'to be modelled as'. The command following $\sim$ specifies the covariates in the fixed part of the model, there may be just the `primary covariate` or many `additional` covariates. The random part of the model is then specified, this may include the `primary covariate` and the same `additional covariates` as in the fixed part of the model, or different covariates may be included. The grouping structure is included in the random specification of the model, the upper most level is specified first followed by the lower levels, the lowest level does not need to be specified explicitly. If the model includes just the upper grouping level, this is a 2-level MLM. Alternative arguments may also be included: the `correlation=corARMA(p=a, q=b)` command incorporates a correlation structure for the residuals. This may be autoregressive (where $q = 0$ and $p > 0$), a moving average (where $p = 0$ and $q > 0$) and autoregressive moving–average (where $p > 0$ and $q > 0$) . The choice of fitting algorithm may also be specified as either maximum likelihood (`"ML"`) or restricted maximum likelihood (`"REML"`). The default is `"REML"` and will always be used in this thesis. Posterior variances will be larger and more realistic under REML than under ML. This will be especially true when the number

of upper level units is small [140]. It is also possible to use LRT's with REML, when all models are fitted with REML and the fixed specification of the model is the same between the models being compared. The RIGLS (restricted iterative generalised least squares) algorithm yields the REML estimates. In general the fixed effects estimates using ML and REML will be similar, though not identical. Inferences regarding the fixed effects in both estimation methods will be the same [139].

The `intervals` command provides approximate (95%) confidence intervals of all the model coefficients from the particular model specified. The intervals are calculated using a normal approximation to the distribution of REML estimators, where the estimators are assumed to be normally distributed (centred at the true parameter values and with covariance matrix equal to the negative inverse Hessian matrix of the restricted log-likelihood evaluated at the estimated parameters) [139].

5.8 Application to the data - Cerebral Autoregulation

The data from the CA biomedical system forms an inherent hierarchical structure: repeated measurements of CBF and ABP are clustered within 3 phases of CEA which in turn are clustered within multiple patients. MLM should be used in order to exploit the full power of the data, whereby all information is used to formulate the model. This would enable us to determine whether CEA has an effect on CA for the whole group of patients. By fitting this type of model, the assumption would be made that there is the same error and variance structure for all patients. It will therefore be necessary to determine if this is a reasonable assumption to make.

An important feature to consider is how the 3 phases of CEA are modelled. The possibilities are to include phase as its own level, as a fixed effect, or as a random effect. Before fitting the model it is clear that including phase as a fixed effect would yield too many parameters. This option would require an intercept for each patient in

each phase, hence 108 (3 x 36) parameters, and possibly another 108 parameters for a random slope. This would be an ineffective model and very difficult to interpret. A more parsimonious model would include phase as a random effect, whereby phase is represented as a standard deviation which would indicate the variation between patients in the intercept and slope. A total of 6 random parameters would be required: 3 representing variation in the intercepts for each phase and 3 for variation in the slope. Models will be investigated where phase is incorporated as the middle level or as a random effect, where random intercepts and slopes are investigated for each phase.

CA may be represented by the slope of the relationship between ABP and CBF. It will be possible to investigate whether a different slope is needed to represent each of the three phases. A significant slope in each phase would signify that CA changes across phases. Furthermore, it would be possible to determine whether CA is actually present (zero slope indicating perfect CA) and whether CA improves following CEA (the difference between slopes in phase 1 and phase 3).

It is highly likely that correlation will be present between adjacent CBF and ABP, due to the nature of CBF and ABP measurements and the time–frame in which they are collected, resulting in correlated residuals in the model. It may be necessary therefore to incorporate this correlation into the model by means of an autoregressive correlation structure.

When applying a correlation structure in this framework, it is assumed that the correlation structure and correlation coefficient is the same for all patients. This will be validated by fitting individual models (that allow for an additional correlation structure) for all patients. It will be possible to determine whether each patient requires a different correlation structure. If the same correlation structure is applicable, the range of values the coefficient takes will be investigated and compared with the global coefficient from the MLM. Whilst fitting a MLM to all data for all patients is the most efficient and effective modelling technique, there may be benefits from fitting individual models.

5.9 Application to the data - Renal Anaemia

Multilevel modelling has previously been introduced to the nephrology field [141], where the authors acknowledge the technique is a particularly useful tool for longitudinal data. Longitudinal data often arise in this area since it is necessary to collect frequently repeated measurements of haemoglobin, iron and ferritin concentrations on patients undergoing dialysis. However, long series of data would rarely occur, since patients would not undertake dialysis for a long enough period and the series would be terminated due to transplant or death. Even though patients may undergo a course of dialysis for a number of months or years, measurements will usually be taken once a month or less frequently, since low Hb levels would mean that frequent blood sampling is detrimental to the patient. Traditional methods of analysis for longitudinal data, such as time series analysis, would not be applicable, because this usually analyses one series, rather than many.

Multilevel modelling would be applicable to these data, since the data hierarchy would then be exploited, where the data form a two level structure: the first level pertains to the repeated haemoglobin and epoetin measurements and the second level pertains to the patient. Similar to the CA system, autoregressive correlation structures will be explored. The biological nature of Hb measurements means that it is unlikely that measurements will rapidly fluctuate in time. Furthermore, Hb is being controlled with the aim of maintaining a stable dose, thus large variation is not expected between Hb measurements from month to month. It is also unlikely that a patient's epoetin dose will fluctuate, as this incurs a larger cost than a stable dose, therefore autocorrelation would be expected amongst the epoetin measurements.

As discussed in Chapter 3 the relationship between Hb and epoetin is complex. In particular it will be necessary to investigate the time point between 2 and 6 weeks where the optimum response is seen in the patient's Hb as a result of the epoetin dose. The initial relationship that will be investigated is Hb(i) and dose(i-1). This will determine if

the relationship, upon which the current algorithm is based, yields plausible results. The wrong relationship to model is Hb(i) and dose(i), as the dose does not have an immediate effect on Hb, due to the erythropoiesis process. The most appropriate relationship will indicate the optimum time lag from epoetin administration to Hb response.

Further investigation will determine if it is necessary to limit the investigation of the ideal time lag to months or whether it is possible to deduce the optimum to weeks or days. For example, a linear combination of dose administered from several months could narrow the time–frame from months to weeks.

5.10 Summary

When the situation occurs where measurements are nested within individuals (or some other unit), this should be embraced in the statistical modelling (the standard errors of the regression coefficients would otherwise be underestimated, leading to wrong inferences being drawn). Furthermore, there is the flexibility with MLM to deal with unbalanced data structures. In particular, in the CA application, the length of the surgical procedure varies between patients, and hence the number of repeated measurements for each patient varies. Due to the nature of the repeated measurements, whereby measurements are collected closely in time and the smooth variation of the biological variable, autocorrelation is likely to be present amongst measurements. Autocorrelated measurements may lead to autocorrelated residuals, which can be accounted for in the MLM framework with additional autoregressive correlation structures. There are a number of features of the MLM modelling and also of the data within the biomedical systems, that justifies application to these data.

Chapter 6

Results - Cerebral Autoregulation

The results from the analysis of the data from the cerebral autoregulation biomedical system will be presented in this chapter, covering exploratory analysis with loess smoothing and detailed multilevel modelling incorporating an autoregressive error structure.

6.1 Exploratory Analysis

In this section, summary statistics for the data are presented, in the form of tables and boxplots. These figures and tables are useful (and necessary) to explore the data before undertaking more sophisticated statistical analyses. Table 6.1 summarises the number of patients and observations. Tables 6.2 and 6.3 present numerical summary statistics about ABP and CBF, respectively, between phases and within each of the three phases.

Table 6.2 shows that the standard deviations within each of the three phases are very similar, indicating that the variation of ABP values in the three phases is similar. The mean and median of ABP are also relatively similar across the three phases, albeit slightly higher in phase 2 than phases 1 and 3, although the increase is not clinically significant.

	Total number
Patients	36
Observations	4541
Observations (Phase 1)	1764
Observations (Phase 2)	1879
Observations (Phase 3)	898
	Summary
Age (median(range))	73(65-82)
Gender	Male 28 / Female 8
Anaethesia	Local 22 / General 14

Table 6.1: Summary statistics of the patients and observations for this study.

	Min	Q1	Median	Mean	Q3	Max	SD	Missing
Overall	38	88	103	103.9	119.0	188	23.02	2
Phase1	38	84	102	101.5	118.8	188	22.98	2
Phase2	48	92	105	106.6	125	183	22.60	0
Phase3	48	89	100	102.7	114	176	23.37	0

Table 6.2: Summary statistics of ABP. The first row shows the summary statistics between phases, rows 2, 3 and 4 show the summary statistics for each phase.

	Min	Q1	Median	Mean	Q3	Max	SD	Missing
Overall	9.84	38.01	48.32	51.70	61.61	138.90	19.92	57
Phase1	17.83	39.41	50.82	51.26	60.53	99.02	14.49	23
Phase2	9.84	33.78	43.29	46.09	54.40	134.20	19.31	28
Phase3	21.64	43.40	62.50	64.18	81.60	138.90	24.29	6

Table 6.3: Summary statistics of CBF. The first row shows the statistics between phases, rows 2, 3 and 4 show the statistics for each phase.

These features of the data suggest that CEA does not have a substantive impact on changing ABP, perhaps this is a result of careful controlling/ monitoring of ABP on behalf of the clinician. It should be noted that these statistics are calculated across all patients, and by combining the values it is appears that there is little change in the patients' ABP in the different phases of the operation. In further analysis, the patients should be inspected individually to determine whether this result is consistent for each patient or whether it is an artefact of combining the data.

Table 6.3 shows that the mean (and median) CBF decreases in phase 2 and increases in phase 3 compared with CBF in phase 1. Further investigation is required to determine whether these changes are clinically significant. The standard deviations increase as the phases progress, which suggests that patients begin surgery (phase 1) behaving similarly and this then diversifies in later phases. In phase 1 the standard deviation is relatively small. There is more variation in the patients response to their artery being clamped (phase 2) and much greater variation in CBF following surgery (phase 3). The SD in phase 3 is 70% larger than in phase 1, which indicates that patients' CBF varies much more following surgery than before. This suggests that the additional variation occurs as surgery progresses and that this may be due to differing responses to surgery.

Boxplots are essentially a graphical presentation of the information given in the tables. The boxplot format may be favoured over the tables as it enables the reader to directly

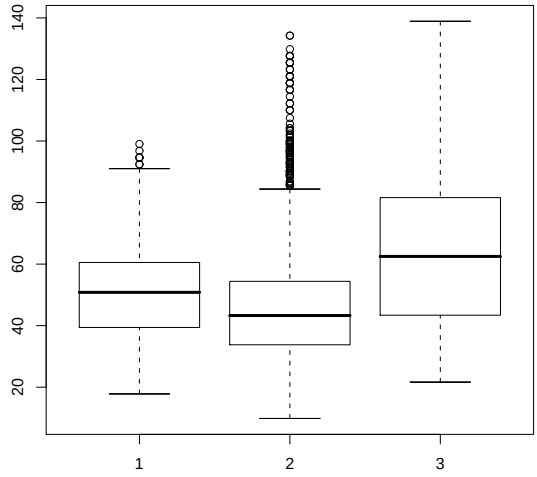

Figure 6.1: Boxplot of cerebral blood flow at each phase

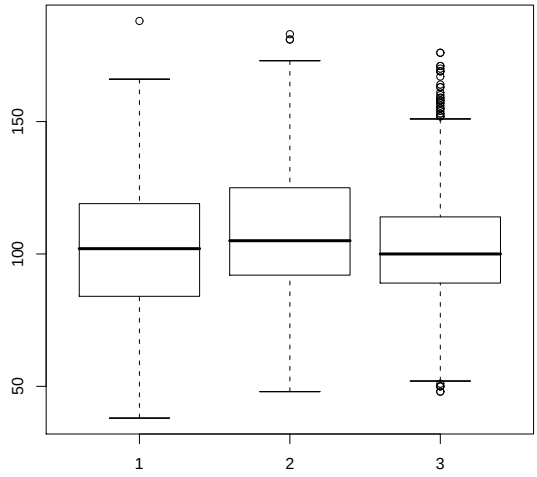

Figure 6.2: Boxplot of arterial blood pressure at each phase

compare the variables across the phases. A further attractive feature is that the differences are seen in relative context, thus allowing the reader to identify whether they are 'significant'. Figure 6.1 is particularly useful as it highlights that the variation of CBF increases throughout the phases, though unlike the table the boxplot demonstrates the nature of the variation. In phase 1 CBF appears to be normally distributed with small even spread either side of the mean. The boxplot for phase 2 shows that there are many outlying values at the upper end of the scale, though the majority of the data is normally distributed with a median lower than the median in phase 1. This suggests that the variation in CBF between phase 1 and 2 does not really increase but appears to due to the outliers. The outlying values in phase 2 suggest there are a number of instances where CBF exceeds the range that is experienced by the majority of cases and thus highlight that in phase 2 some patients may be complex and difficult to manage. The boxplot for phase 3 indicates that the variation between patients in this phase is larger, as the interquartile range and whole range of data is larger. Additionally the median value is largest in phase 3.

In regards to ABP, figure 6.2 shows that the central point and spread of the data is very similar at each phase of CEA. This information was gained from table 6.2. In addition to the information also given in the table, the boxplots show that the range of ABP is very large in each of the three phases, though further analysis is needed to determine whether each patient experience a large range of values or whether the range is large due to the combining of measurements. The boxplot for phase 3 also reveals that there are quite a large number of upper end outlying ABP's in this phase. This is important for the clinician to know that in some instances the ABP for the patients exceed the 'normal' range of values and hence require careful monitoring or control of their ABP.

Kernel density plots were also fitted to the data as another exploratory technique, to investigate the distribution of ABP and CBF. Inspection of the density plots did not provide further information or insight than that gained from the tables and boxplots, hence have not been presented.

Whilst the exploratory analysis reveals useful information about the group of patients as a whole, no insight is gained into the specific information about the individual. It is important to know about the data before further analysis can be implemented. The exploratory tables and boxplots highlighted that there is a lot of variation in the measurements, which may indicate that there is a lot of variation between the individuals or it could mean that there is a lot of variation amongst the individuals themselves. As this is unclear it is necessary to implement more complex analysis to gain better insight.

6.2 Functional Data Analysis

The aim within the CA biomedical system was to assess CA, which appeared to be a relatively ambiguous problem. It was not possible to apply existing methods that have been used for the assessment of CA for a number of reasons. First, other techniques for assessing CA have not been considered in the surgical scenario and thus were inappropriate and impractical. Second, repeated measurements are collected on these patients and the techniques previously used have not considered this type of data. Other methods of analysis were therefore sought.

Since the aim was to assess CA, the first avenue of exploration was to investigate the ABP–CBF relationship. The method of loess was deemed an ideal technique for investigating this relationship, following verification. A loess curve was fitted to 'ideal' data [116] where intact autoregulation occurs, see figure 6.3. This experiment was used to determine whether a loess curve would reveal intact CA in a situation where it is known to occur. Further details of the advantages and disadvantages of this curve will be discussed in Chapter 8.

Loess curves were subsequently fitted to CBF against ABP in each phase of carotid endarterectomy for all patients in the study, as it was predicted from the exploratory analysis that there would be variation between phases. These plots are useful for

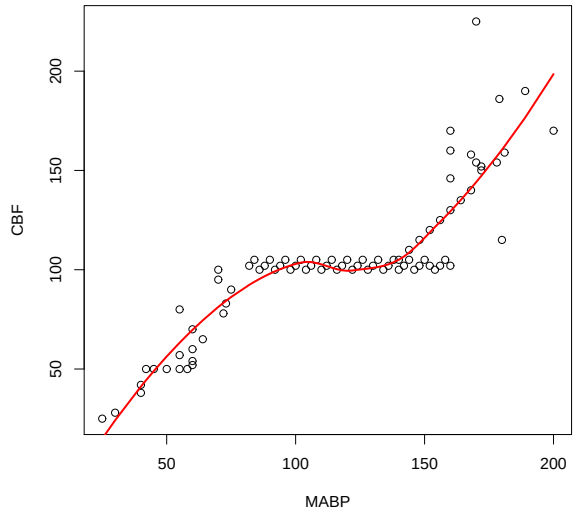

Figure 6.3: The Loess representation of the ideal autoregulation curve

determining graphically whether cerebral autoregulation is present in these patients in each phase. The black curve represents phase 1, the red curve phase 2 and the green curve phase 3.

Figure 6.4 to 6.12 show loess curves for a representative sample of patients. In the majority of patients the ordering of the curves is the same; from top to bottom the order is phase 3 (green), phase 1 (black) and phase 2 (red) (an exception is patient A (Figure 6.4)). This may be interpreted as cerebral blood flow being greatest in phase 3 and CBF lowest in phase 2, compared with the initial phase 1. Each of the figures have been plotted on the same axes to allow direct comparison of the patients ABP–CBF relationships.

Figure 6.4 represents a patient who experiences a large range of CBF and ABP values; which is consistent throughout the three phases. The range of CBF experienced in phase 2 is the largest. The range of ABP is wide in all of the three phases, which spans 50 – 175 mmHg. This is a particularly large range for one patient to experience, which is quite unexpected. If the theory of the CA curve is accepted then the data from this

Figure 6.4: Loess curve fitted to CBF against ABP in each phase, for patient A. Black curve - phase 1, red curve - phase 2 and green curve - phase 3.

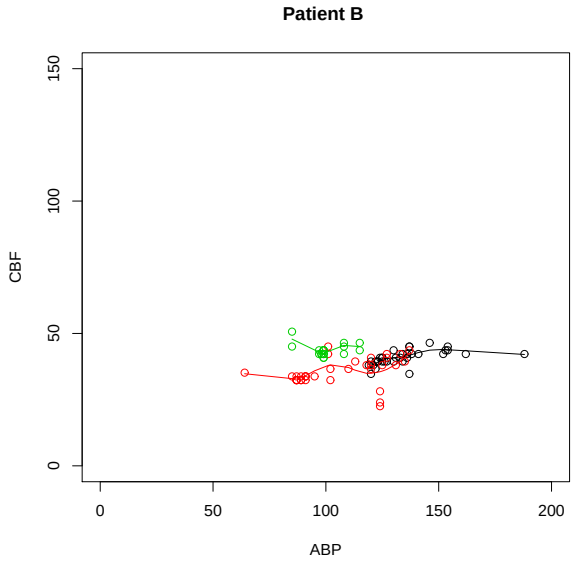

Figure 6.5: Loess curve fitted to CBF against ABP in each phase, for patient B.

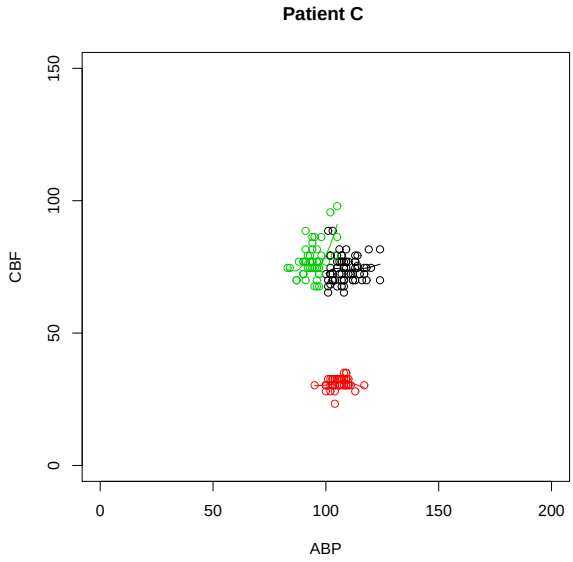

Figure 6.6: Loess curve fitted to CBF against ABP in each phase, for patient C.

Figure 6.7: Loess curve fitted to CBF against ABP in each phase, for patient D.

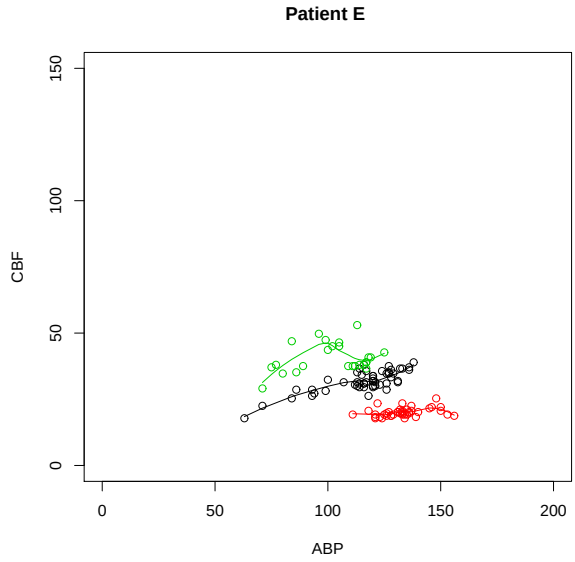

Figure 6.8: Loess curve fitted to CBF against ABP in each phase, for patient E.

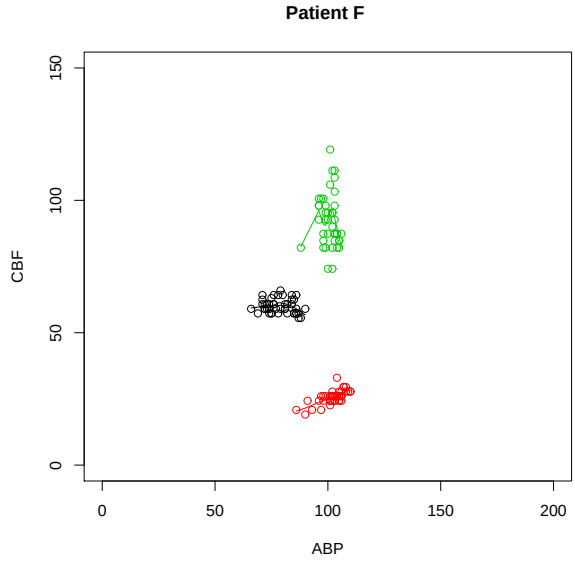

Figure 6.9: Loess curve fitted to CBF against ABP in each phase, for patient F.

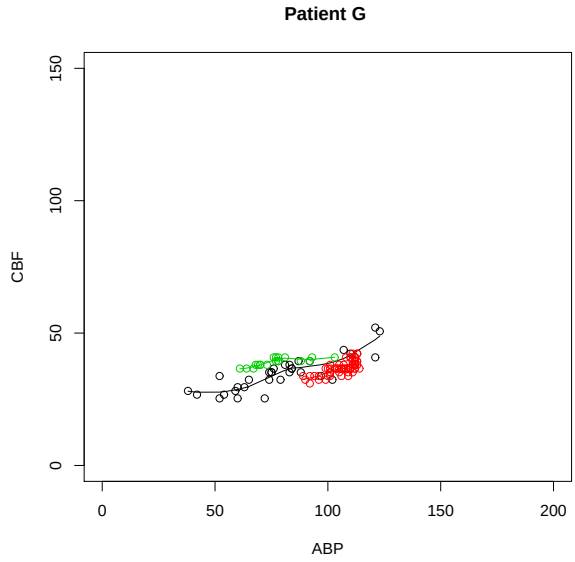

Figure 6.10: Loess curve fitted to CBF against ABP in each phase, for patient G.

Figure 6.11: Loess curve fitted to CBF against ABP in each phase, for patient H.

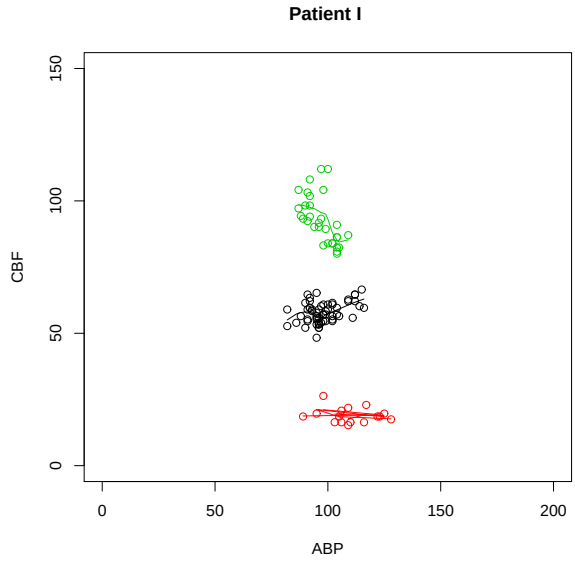

Figure 6.12: Loess curve fitted to CBF against ABP in each phase, for patient I.

patient could reproduce the plateau region of the curve as this is believed to exist between 50 – 150 mmHg. In this patient the slope of the ABP-CBF relationship is very steep, which would suggest that CA is not intact. There is little overlap of the different phase measurements, highlighting that the phases are distinct.

Figures 6.8 and 6.11 represent patients whose ABP and CBF are relatively similar. In both instances their range of ABP is narrower than patient A (although still wider than clinically desirable) and whose CBF is lower. There is distinct separation of phase 2 from phases 1 and 3, which themselves are close but with little actual overlap. Compared to phases 1 and 3, ABP is increased while CBF is decreased in phase 2.

Figures 6.9 and 6.12 display the curves for patients whose CBF and ABP are very erratic during CEA; all phases are distinct with no overlap. In both patients the relationship between CBF and ABP in phase 3 appears to have a negative gradient. CBF is greatly reduced in phase 2, which is also the case with the patient shown in figure 6.6. The ABP ranges are very narrow (patients C, F and I) in comparison to the other patients shown,

and Lassen [29].

Figure 6.5 and 6.10 show patients whose CBF changes very slightly between the three phases. The slopes of the CBF–ABP in each of the three phases, in both patients, are very shallow. The range of ABP differs with each phase, in particular for patient B there is very little overlap of ABP measurements for phases 1 and 2. In phase 1 the range is approximately 120–190 mmHg and phase 2 70–130 mmHg, which are large ranges in themselves though even more worrying when considering that throughout surgery the patients ABP ranges from 70–190 mmHg. ABP does seem to then come under control in phase 3 as the range is reduced to 90-110 mmHg and within stable ABP limits. Patient G (figure 6.10) also experiences a large range of ABP in phase 1 (40–175 mmHg), though this is reduced in phase 2 and 3. Figure 6.7 also shows a patient whose CBF changes very little between the phases, there is much overlap in the CBF measurements in this patient, albeit retaining the usual phase ordering (3,1,2). In this instance the range of ABP is relatively narrow (50–80 mmHg) in each of the three phases. The gradient of the relationship between CBF and ABP in this patient appears very similar in each of the three phases.

The important issue from fitting the loess curves was to determine whether (intact) CA was evident in these patients, i.e. do we see anything resembling the CA curve? The simple conclusion is that it is difficult to judge. The loess curves do, however, highlight that the majority of curves show a straight line relationship between CBF and ABP. The slope is relatively mild, though not a horizontal line representing intact CA. The range of ABP values are within the alleged region for intact CA (50–150 mmHg), thus suggesting that the patients' measurements lie within the middle section of the data and no turning point is expected.

6.3 Multilevel modelling

Subsequent to the exploratory analysis and the fitting of loess curves, multilevel modelling was applied. The reasons for using MLM are outlined below:

- Analysis of individual patients would be difficult due to some patients experiencing a limited range of ABP (for example). This will be overcome by MLM which shares information across patients.
- The important research question is to determine whether there is evidence, from measurements, that CEA influences CA. Not only is it important to address this question on an individual basis, such as with the loess curves, but also for the population. MLM provides a framework where the results from the patients can be combined.
- Measurements of ABP and CBF are equally spaced in time, which enables an autocorrelation structure to be incorporated.

6.3.1 Model fitting procedure

A number of models were fitted to investigate the relationship between CBF and ABP (see list below). For example model A may be described as: CBF is modelled by a fixed intercept (not specified explicitly), which represents mean CBF across all phases (as phase has not been specified in this model), and a random patient intercept (where the 1 in 1/Patient represents the intercept). The fixed slope of the model is specified by ABP and represents the mean slope of the CBF–ABP relationship. The slope is not random in this model (but is in models C, D and E) and so the same for all patients. The syntax of these models has been described in general in Chapter 5.

- Model A = lme(Flow $\sim$ ABP, random = 1/Patient)

- Model B = lme(Flow $\sim$ ABP+Phase, random = 1/Patient)
- Model C = lme(Flow $\sim$ ABP+Phase, random = ABP/Patient)
- Model D = lme(Flow $\sim$ ABP+Phase, random = ABP+Phase/Patient)
- Model E = lme(Flow $\sim$ ABP*Phase, random = ABP+Phase/Patient)
- Model F = lme(Flow $\sim$ ABP*Phase, random = ABP*Phase/Patient)

The information criterion of all models are presented in table 6.4.

Table 6.4: Information criterion of models A to E

Model	df	AIC	BIC	logLik
Model A	4	36116.64	36142.27	-18054.32
Model B	6	35108.04	35146.48	-17548.02
Model C	8	34942.95	34994.21	-17463.47
Model D	**15**	**30395.85**	**30491.95**	**-15182.92**
Model E	17	30409.37	30518.28	-15187.68

It is not possible to carry out likelihood ratio tests between all these models since the specification of the fixed effects is not the same in all models. The LRT requires strict nesting, which does not occur here. The complexity increases through models A to E; as model complexity increases the loglikelihood of the models should also increase. When phase is included as a random effect this greatly reduces AIC, BIC and increases the loglikelihood from the models without random effects, observed by the differences between Models C and D. Including an interaction between phase and ABP as in Model E (i.e. different slopes in each of the three phases) does not improve model fit, in terms of reducing AIC or BIC any further than from model D. Additionally, the interaction terms are not statistically significant. Model E is more complex than model D, where the loglikelihood should increase, though it decreases. Model estimates are obtained using

RIGLS, which is an iterative fitting procedure. This procedure aims to converge upon the global maximum of the loglikelihood profile. Thus, the estimated coefficients obtained in attempting to model the complexity of model E would suggest that the loglikelihood profile is not increasing strictly monotonically towards the global maxima (from either direction) and that the solution found is that of a local maxima. The model could be specified to find the global maxima, though this option was not implemented as the results of model E were not convincing.

Therefore the results for model E indicate that there may be numerical difficulties with this model, such that the model does not converge to the optimal solution. The solution obtained may be a solution of a local maximum rather than the optimal solution. The expected increasing trend is observed from models A to D. This leads us to question whether model E is overly complex and perhaps unstable. Thus, the preferred choice of model is model D. From models C to D there is an improvement in AIC, BIC and loglikelihood, thus believing in the numerical solution as this is the pattern that is expected to be observed. . There are a number of clinical justifications for the choice of model D, which will be discussed in Chapter 8. The full summary of Model E is shown in table 6.5. It was not possible to fit a model with a random interaction (Model F) as this was unstable and did not converge.

Model D = lme(Flow $\sim$ ABP+Phase, random = ABP+Phase/Patient)

Model D was deemed to be the most appropriate model from the models fitted above. The algebra for the general version Model D is presented in equation 6.1 and the specific model with coefficients is presented in equation 6.2. The model coefficients are presented in table 6.6 and further diagnostics in table 6.7. The standardised residuals vs. fitted values for this model is presented in figure 6.13 and the autocorrelation function of the normalised residuals is presented in figure 6.14.

Table 6.5: Coefficients of model E. Note that the degrees of freedom for this model is 4441.

Fixed effects:	Value	Std.Error	p-value		
(Intercept)	35.86	3.76	≤ 0.0001		
ABP	0.16	0.04	≤ 0.0001		
Phase2	-10.01	3.56	0.0049		
Phase3	7.58	3.70	0.0406		
ABP:Phase2	0.030	0.02	0.1551		
ABP:Phase3	0.004	0.02	0.8616		
Random effects:	Intercept	ABP	Phase 2	Phase 3	Residual
StdDev	20.65	0.22	16.59	16.37	6.66

$$Flow_{tj} = \beta_0 + u_{0j} + (\beta_1 + u_{1j})ABP + (\beta_2 + u_{2j})P2 + (\beta_3 + u_{3j})P3 + e_{tj} \quad (6.1)$$

$$Flow_{tj} = 34.81 + u_{0j} + (0.17 + u_{1j})ABP + (-6.83 + u_{2j})P2 + (8.01 + u_{3j})P3 + e_{tj} \quad (6.2)$$

Figure 6.13 shows that the range of standardised residuals is relatively large (-9.59 to 7.93), in relation to the ideal range of -2 to +2. The residuals also seem to display heteroskedasticity, whereby as the fitted values increase the range of residuals also increases (fans out). These features of this plot suggest that model 6.2 may not be the best fitting model.

Figure 6.14 highlights that the normalised residuals are strongly correlated (for the first five lags). This means that the assumption that the lowest level residuals are independent is not upheld. A solution for this would be to look at other specifications for the model and also incorporating an additional complexity for the residuals, to account for the dependency amongst the observations (which has resulted in correlated normalised residuals). After the issue of the correlated residuals is resolved by incorporating an

Table 6.6: Coefficients of model 6.2. Note that degrees of freedom for this model are 4443

Fixed effects:	Value	Std.Error	p-value		
(Intercept)	34.81	3.62	≤ 0.0001		
ABP	0.17	0.04	≤ 0.0001		
Phase2	-6.83	2.78	0.014		
Phase3	8.01	2.78	0.004		
Random effects:	StdDev		Corr		
(Intercept)	20.46		(Intr)	ABP	P2
ABP	0.22		-0.761		
Phase2	16.59		-0.154	-0.013	
Phase3	16.35		-0.191	0.262	0.086
Residual	6.66				

Table 6.7: Diagnostics of model 6.2

Information criterion:	AIC	BIC	logLik		
	30395.85	30491.95	-15182.92		
Residuals:	Min	Q1	Med	Q3	Max
	-9.59	-0.36	0.01	0.42	7.93

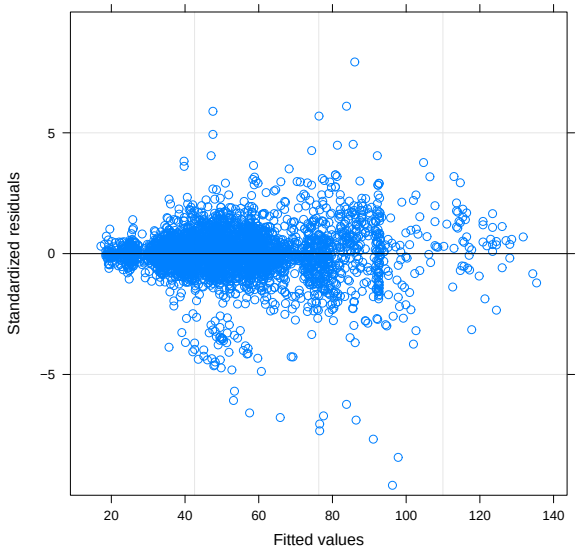

Figure 6.13: Standardised residuals vs fitted values for model 6.2

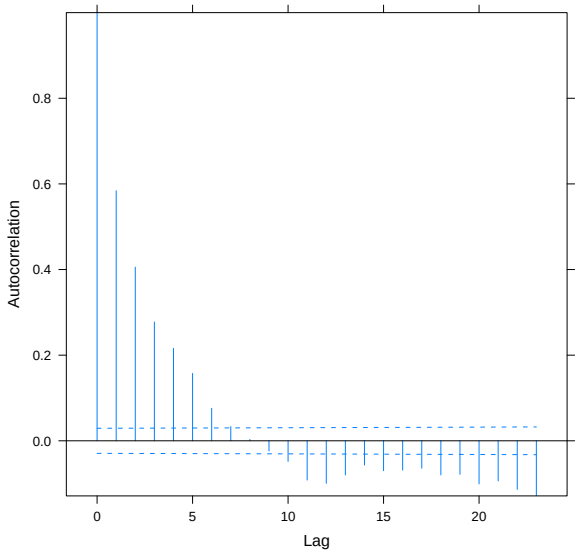

Figure 6.14: ACF of normalised residuals for model 6.2

autocorrelation structure, this may alleviate apparent heteroskedasticity seen in figure 6.13.

6.3.2 Model with autoregressive correlation

In order to resolve the issue of the correlated residuals models were fitted with a variety of correlation structures, including autoregressive, autoregressive moving average and moving average. Table 6.8 shows an example of two models with autoregressive correlation structures that were fitted. The additional correlation structures were incorporated with the best fitting model from the previous section (Model 6.2). An AR(2) correlation structure was shown to improve model fit the most (Model 6.4). See table 6.10 for the coefficients of this model, table 6.11 for model diagnostics and table 6.12 for 95% confidence intervals of the model coefficients. Various residuals plots are presented in figures 6.16 to 6.19. The general algebra for this model is presented in equation 6.3 and the specific equation in equation 6.4.

$$Flow_{tj} = \beta_0 + u_{0j} + (\beta_1 + u_{1j})ABP + (\beta_2 + u_{2j})P2 + (\beta_3 + u_{3j})P3 + \rho_1 e_{t-1,j} + \rho_2 e_{t-2,j} + Z_t \tag{6.3}$$

$$Flow_{tj} = 38.88 + u_{0j} + (0.14 + u_{1j})ABP + (-14.97 + u_{2j})P2 + (11.14 + u_{3j})P3 + 0.65 e_{t-1,j} + 0.31 e_{t-2,j} \tag{6.4}$$

Table 6.8: Diagnostics of models with autoregressive correlation structures

	df	AIC	BIC	logLik
Model 6.2	15	30395.85	30491.95	-15182.92
Model 6.2 + AR(1)	16	27913.12	28015.63	-13940.56
Model 6.2 + AR(2) (= Model 6.4)	**17**	**27509.62**	**27618.53**	**-13737.81**

Table 6.9: Likelihood ratio test of models with autoregressive correlation structures

Test	L.Ratio	p-value
Model 6.2 vs Model 6.2 + AR(1)	2484.73	≤ 0.0001
Model 6.2 + AR(1) vs Model 6.4	405.50	≤ 0.0001

Table 6.10: Coefficients of model 6.4. Note that the degrees of freedom in this model are 4443

Fixed effects:	Value	Std.Error	p-value		
(Intercept)	38.88	2.70	≤ 0.0001		
ABP	0.14	0.03	≤ 0.0001		
Phase2	-14.97	3.12	≤ 0.0001		
Phase3	11.14	4.04	0.0058		
Correlation Structure:	$\rho_1 = 0.65$	$\rho_2 = 0.31$			
Random effects:	StdDev		Corr		
(Intercept)	10.43		(Intr)	ABP	P2
ABP	0.14		-0.391		
Phase2	18.13		-0.095	-0.427	
Phase3	23.14		-0.832	0.548	-0.442
Residual	16.17				

Table 6.11: Diagnostics of model 6.4

Information criterion:	AIC	BIC	logLik		
	27509.62	27618.53	-13737.81		
Residuals:	Min	Q1	Med	Q3	Max
	-3.89	-0.45	-0.02	0.46	4.45

Table 6.12: 95 % confidence intervals for all coefficients in model 6.4

Fixed effects:	lower	est	upper
Phase1 intercept	33.59	38.88	44.18
ABP slope	0.09	0.14	0.19
Phase2 contrast	-21.09	-14.97	-8.85
Phase3 contrast	3.23	11.14	19.06
Random Effects:			
sd((P1 int))	5.96	10.43	18.24
sd(ABP)	0.10	0.14	0.19
sd(P2 con)	14.11	18.13	23.29
sd(P3 con)	17.75	23.14	30.15
cor((Int),ABP)	-0.75	-0.39	0.16
cor((Int),P2)	-0.70	-0.10	0.58
cor((Int),P3)	-0.99	-0.83	0.38
cor(ABP,P2)	-0.72	-0.43	-0.00
cor(ABP,P3)	0.21	0.55	0.77
cor(P2,P3)	-0.71	-0.44	-0.05
Correlation structure:			
ρ_1	0.64	0.65	0.67
ρ_2	0.28	0.31	0.34
Within-group standard error :	13.73	16.17	19.04

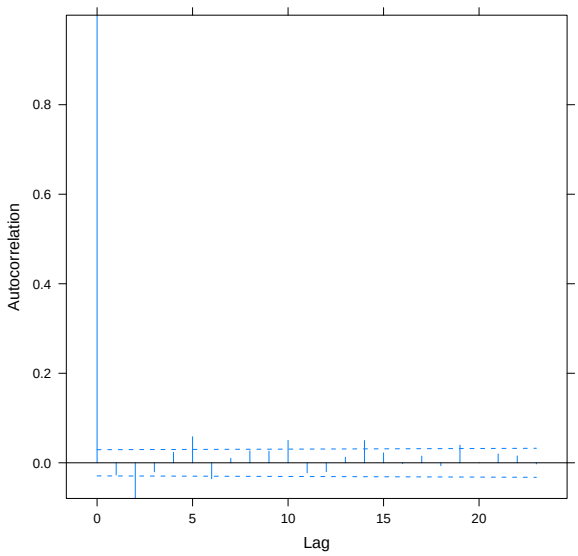

Figure 6.15: Autocorrelation function of normalised residuals for model 6.4

Figure 6.15 shows that the normalised residuals are no longer correlated with each other, as the autocorrelation is not (strongly) significant at the lags. At some lags the autocorrelation is significant, although this is very small and occurs at arbitrary and relatively large lags, hence is not considered important or problematic. It is therefore possible to conclude that the residuals are not correlated and the assumption of independence is upheld.

Figure 6.16 shows that the range of the standardised residuals has greatly reduced from (-9.59 to 7.93) to (-3.89 to 4.45) due to the inclusion of the correlation structure. Even though the range has reduced it still remains larger than the approximate range of -2 to 2. The residuals appear to be randomly scattered about zero, though perhaps for the largest fitted values (greater than 100) there are more negative residuals. When the residuals are split due to phase (figure 6.17) and patients (figure 6.18) it is possible to see where the outlying values occur (i.e. why there are more larger negative residuals for larger fitted values).

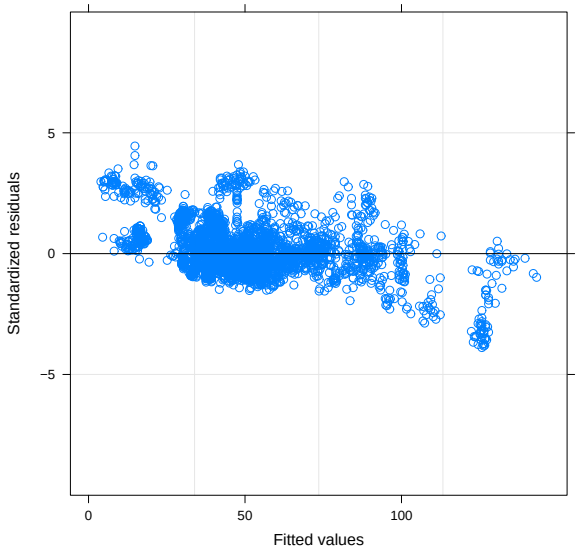

Figure 6.16: Standardized residuals vs fitted values for model 6.4

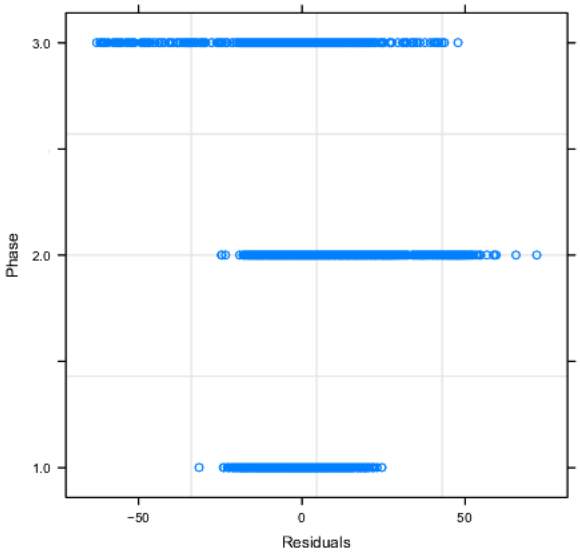

Figure 6.17: Phase residuals for model 6.4

Figure 6.17 shows graphically that as the surgical procedure progresses then the variation between patients also increase. This is evident from the range of residuals increasing as the phases increase. The range of residuals in phase 1 is relatively narrow in comparison to phases 2 and 3, where the range for phase 3 is only slightly larger than phase 2. The phase 1 residuals are centered on zero with an equal balance of positive and negative values, whereas the residuals in phase 2 are more heavily balanced with more positive residuals and in phase 3 more heavily balanced with more negative residuals.

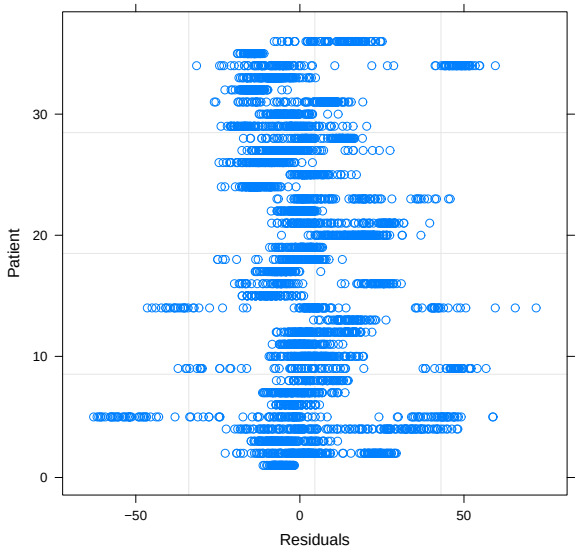

Figure 6.18: Patient residuals for model 6.4

Figure 6.18 shows the residuals for each patient, although not taking phase into account. The range of residuals varies greatly between patients, there does not appear to be a standard behaviour. This figure shows that some patients experience a huge range of residuals, in two cases there are patients whose residuals range -50 to 50. In these patients the three phases are clearly evident from the pattern of residuals. The other extreme is that a patient experiences a range -5 to 0. The majority of patients seem to experience a balance of positive and negative residuals, though there are a few who experience only

negative residuals. This is a particularly useful plot as it is possible to identify patients with interesting features. If there were a larger number of patients, however, it is likely that this plot would not be so effective.

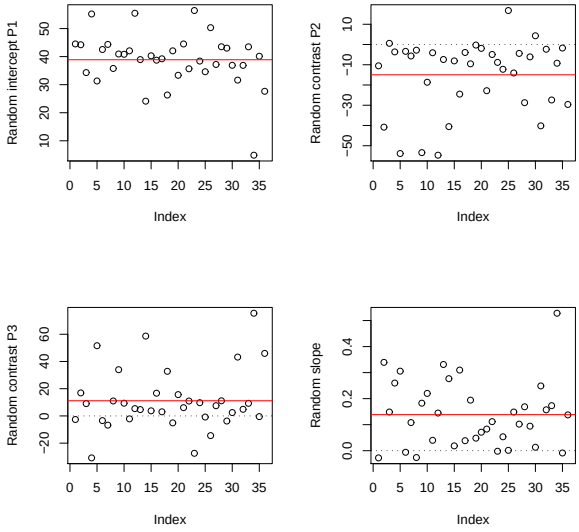

Figure 6.19: Random patient coefficients for model 6.4. Red line represents the fixed (mean) coefficient

A further assumption of MLM is that the random coefficients are normally distributed around the mean (fixed) coefficient. It is evident from figure 6.19 that this assumption is upheld. The figure shows four plots (3 for the random intercept and 1 for the random slope). Each plot shows that the random patient effects are scattered randomly and equally balanced around the mean.

6.3.3 Individual models

The assumption of multilevel models with an additional correlation structure is that the correlation parameters are the same for each patient. The aim of this section is to

determine whether this is a reasonable assumption to make. Models were fitted using generalised least squares with the `gls` command in R, which permitted the modelling of different correlation parameters due to patient (and also due to phase within patient). The fitting of these models will also allow different slopes to be fitted for each patient, which is similar to the random slopes philosophy in the MLM framework. The random patient element is incorporated by fitting single models for each patient. Similar to the previous section the model slope will represent static autoregulation, and the correlation parameters may represent dynamic autoregulation. The reasoning behind this theory will be discussed in detail in Chapter 8.

All models were fitted (described below) incorporating an AR(1) correlation structure for the residuals, as the AR(2) was not suitable for every patient, due to models being fitted with a reduced amount of data. The initial model to be fitted is shown in equation 6.5 shows the model fitted to each patient, where the static (slope) and dynamic (ρ) autoregulation varies between patients, though remains constant across phases.

$$Flow_{tj} = \beta_{0tj} + \beta_{1tj}MBP + \beta_{2tj}P2 + \beta_{3tj}P3 + \rho_1 e_{t-1,j} + Z_t \tag{6.5}$$

Equation 6.6 shows where the static parameter varies between phases and the dynamic remains constant between phases, though both vary between patients.

$$Flow_{tj} = \beta_{0tj}+\beta_{1tj}MBP+\beta_{2tj}P2+\beta_{3tj}P3+\beta_{4tj}MBP*P2+\beta_{5tj}MBP*P3+\rho_1 e_{t-1,j}+Z_t \tag{6.6}$$

Equations 6.7 6.8 and 6.9 show where both the static and dynamic parameters vary between phases, and between patients.

$$Flow_{1tj} = \beta_{0tj}P1 + \beta_{1tj}MBP + \rho_1 e_{t-1,j} + Z_{1t} \tag{6.7}$$

$$Flow_{2tj} = \beta_{2tj}P2 + \beta_{3tj}MBP + \rho_2 e_{t-1,j} + Z_{2t} \tag{6.8}$$

$$Flow_{3tj} = \beta_{4tj}P2 + \beta_{5tj}MBP + \rho_3 e_{t-1,j} + Z_{3t} \tag{6.9}$$

The plots of the dynamic and static parameters from these models are shown in figures 6.20 to 6.23

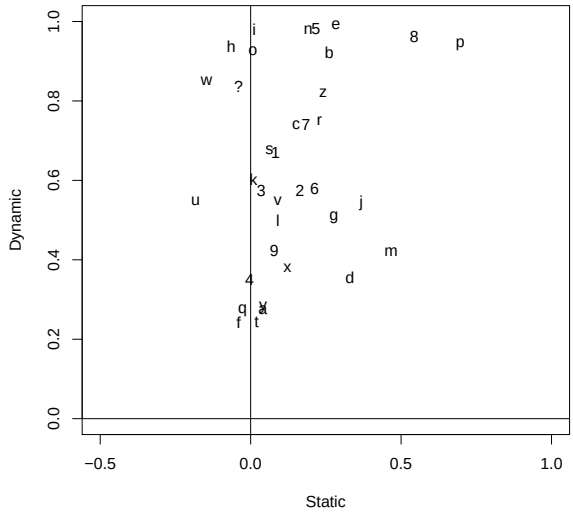

Figure 6.20: Dynamic vs static for equation 6.5

Figure 6.20 shows the dynamic parameter plotted against the static parameter for the model where each of the parameters varies due to patient, but not taking phase into account. There seems to be a slight increasing trend between the two parameters, such that static increases as dynamic increases. This suggests that the patients who experience a steeper slope (larger static parameter, which would indicate impaired autoregulation) will also experience a stronger correlation between adjacent measurements (larger dynamic parameter).

In this model there are a small number of patients whose static parameters are negative. It is not plausible that these values actually occur, as a negative slope would mean that CBF would decrease as ABP increases. The negative slopes, however, are relatively small. In figure 6.21 the negative static values have been constrained to zero, which is plausible as this phenomenon may occur if all the patients measurements are clustered together and thus the estimate of a negative slope is inaccurate.

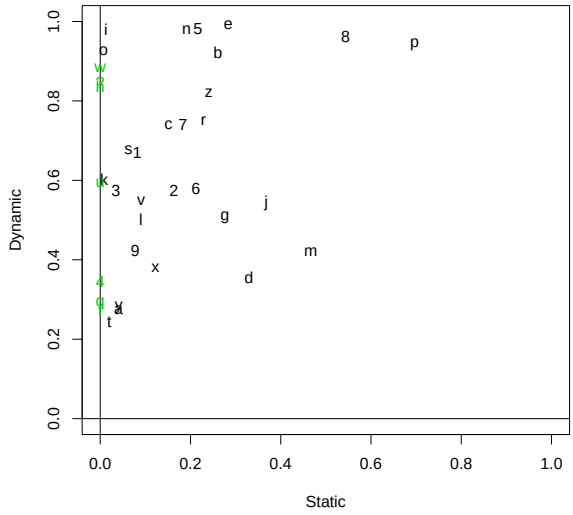

Figure 6.21: Dynamic vs static for equation 6.5 (restricting negative slope to zero)

Figures 6.22 and 6.23 show plots of dynamic vs static where dynamic is fixed and static varies between phase, and where dynamic and static vary between phases, respectively. These models seem a little impractical, however, as in figure 6.22 there are many patients who have one or more negative static parameters. In figure 6.23 there are many negative static parameters again and also negative dynamic parameters. Furthermore, there does not appear to be any significant relationship between the dynamic and static parameters for either of these models. It therefore seems that model 6.5 is the best representation of the data, when the few negative static parameters are restricted to zero. This finding also

supports the result that each patient requires different (random) parameters, though it is not necessary to vary these due to phase of CEA.

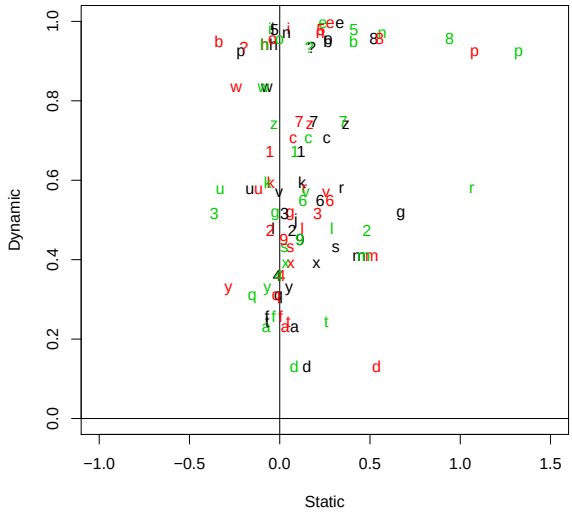

Figure 6.22: Dynamic vs static for equation 6.6

6.3.4 The effects of anaesthesia on cerebral autoregulation

A further aim of this work was to investigate whether CA was affected due to type of anaesthesia. This aim was addressed by using boxplots to compare the static and dynamic parameters, calculated from previous models, for local and general anaesthesia. Figure 6.24 shows the static parameter from the multilevel model with AR(2) correlation structure (model 6.4) by anaesthesia type. Figure 6.25 shows boxplots of the dynamic and static parameters from the individual models with AR(1) correlation structure (model 6.5) and the negative static parameters constrained to zero, by anaesthesia type.

Individual patient slopes were extracted from model 6.4 to permit comparison of their values between local and general anaesthesia. Figure 6.24 shows the gradient of the CBF-

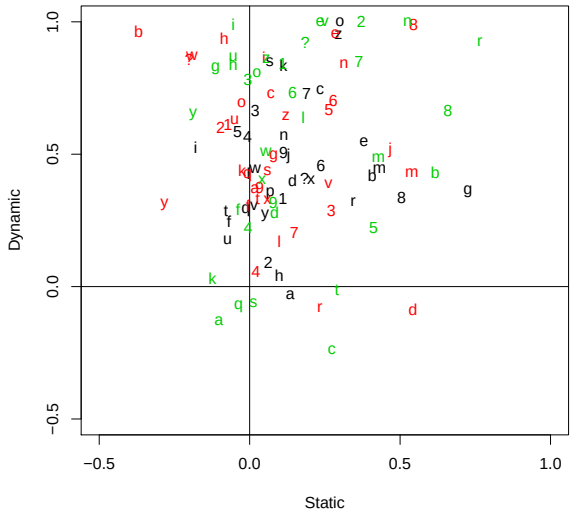

Figure 6.23: Dynamic vs static for equations 6.7 (phase 1 - black) 6.8 (phase 2 - red) and 6.9 (phase 3 - green)

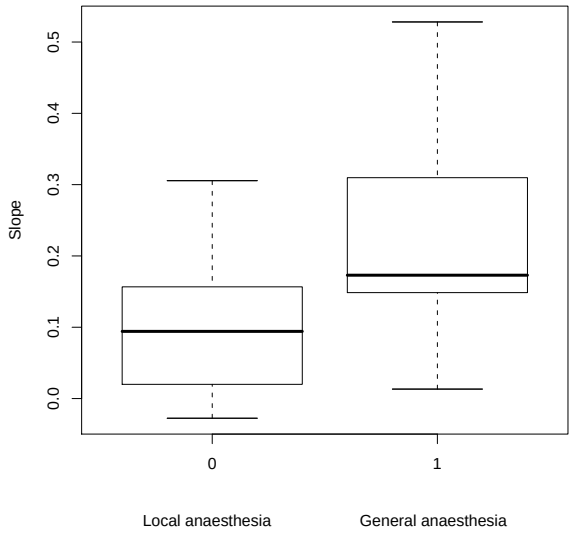

Figure 6.24: Boxplots showing static parameter (model slope) from equation 6.4 by anaesthesia type

ABP relationship for patients undergoing surgery under local and general anaesthesia. The median (range) gradient for local anaesthesia was 0.09 (-0.03 to 0.31) and for general anaesthesia 0.17 (0.01 to 0.53). A *t*-test showed that there is a statistically significant difference in the means of static autoregulation in local and general anaesthesia. Patients undergoing general anaesthesia generally had greater gradients, that is to say less efficient autoregulation.

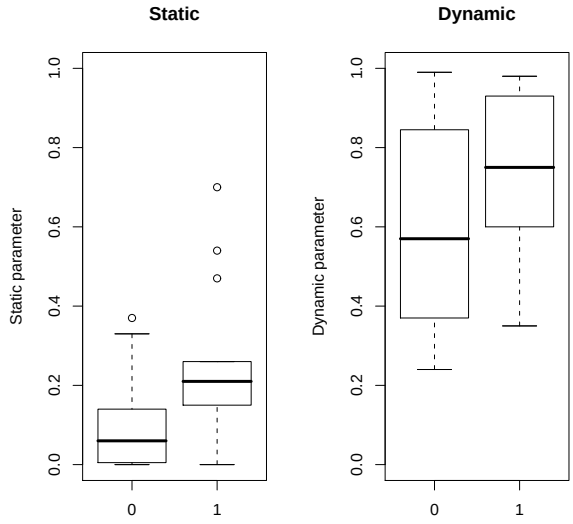

Figure 6.25: Boxplots showing static and dynamic parameters from individual models (equation 6.5) by anaesthesia type (0 = Local, 1 = General)

Figure 6.25 shows that the range and interquartile range of dynamic parameters is particularly large for both types of anaesthesia. In addition there is no significant difference in the dynamic parameters for the two types of anaesthesia. Similar to figure 6.24 the static parameters for patients undergoing general anaesthesia are generally greater than those undergoing CEA with local anaesthesia. The median (range) gradient for local anaesthesia was 0.05 (0.00 to 0.34 – excluding extreme values) and for general anaesthesia 0.22 (0.00 to 0.27 – excluding extreme values). A *t*-test suggests that there

is a statistically significant difference in the means of the static parameter, though not between the means of the dynamic parameter.

CEA is often associated with marked changes in blood pressure. Further, it is generally perceived that ABP generally falls after induction of general anaesthesia, which is true of these patients, as seen in figure 6.26. Figure 6.27 shows that this finding is consistent in each of the three phases. In phase 2 of CEA patient ABP is raised compared with ABP in phase 1. ABP then decreases in phase 3.

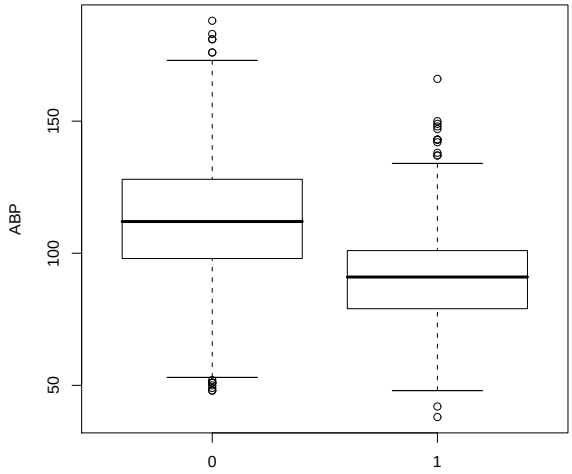

Figure 6.26: Boxplot of arterial blood pressure by anaesthesia type (0 = Local, 1 = General)

Figure 6.28 shows that across all phases CBF does not change due to type of anaesthesia, in fact the box plots appear almost identical. When this is broken down by phase (see figure 6.29 there is again no difference in CBF due to anaesthesia type. Furthermore, this figure again shows that the variability in CBF increases as CEA progresses.

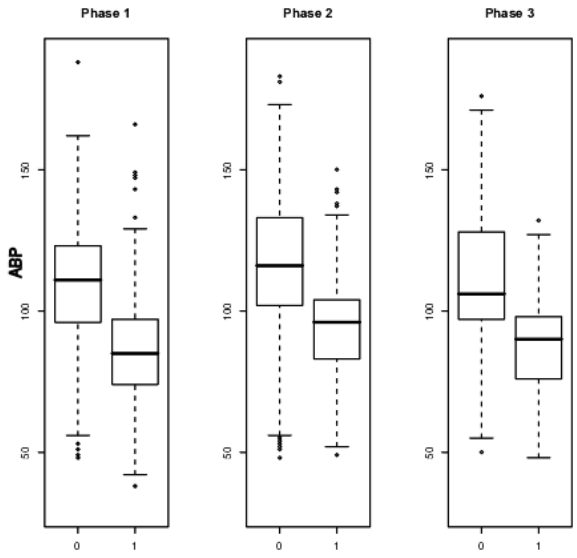

Figure 6.27: Boxplot of arterial blood pressure by anaesthesia type (0 = Local, 1 = General) in each phase

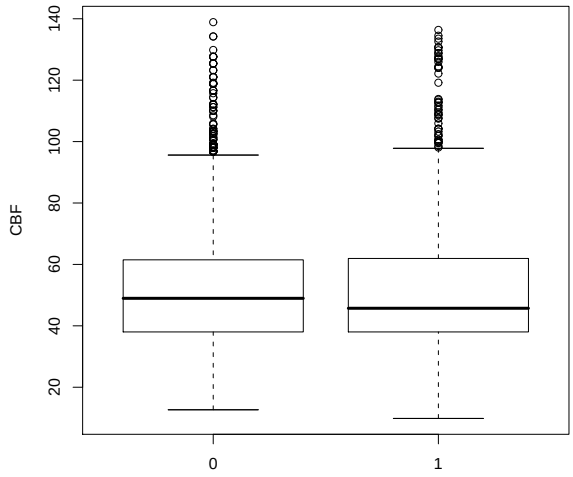

Figure 6.28: Boxplot of cerebral blood flow by anaesthesia type (0 = Local, 1 = General)

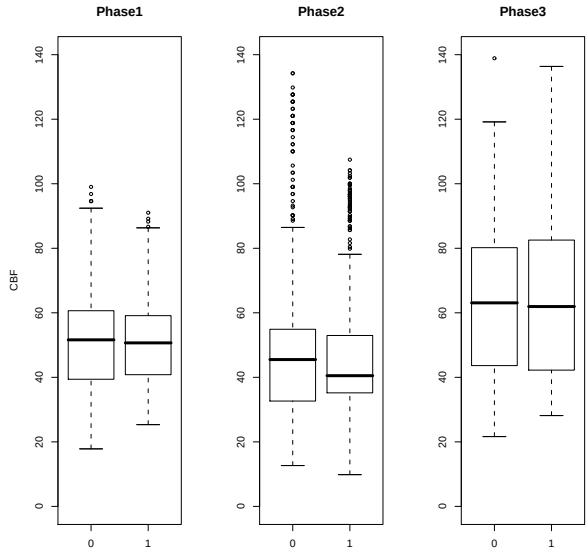

Figure 6.29: Boxplot of cerebral blood flow by anaesthesia type (0 = Local, 1 = General) in each phase

Chapter 7

Results - Renal Anaemia

The results from the analysis of the data from the renal anaemia biomedical system will be presented in this chapter. This includes exploratory tables and kernel density plots, the application of FDA (including B-spline curves and phase plots) and the application of MLM (which includes the fitting of complex models with autoregressive correlation structures and a more pragmatic clinical model).

7.1 Exploratory Analysis

In this section summary statistics for the data from the renal anaemia biomedical system are presented, in the form of tables and kernel density plots. These figures and tables are useful to give an initial insight to the data. Table 7.1 presents summary statistics about the patients in this study, for each agent. Table 7.2 presents numerical summary statistics about Hb and dose, for both agents combined and individually. Rows 2 and 5 of this table show the summary statistics of Hb and dose across all patients, rows 3,4,6,7 show the summary statistics of Hb and dose by epoetin agent. Kernel density plots are used to show the distribution of Hb and dose overall (figure 7.1) and also by agent (figure 7.2).

Agent	DA	EB
Patients	74	77
Observations	962	1001
Age (median(range))	64(51-73)	63(46-72)
Gender(Male:Female)	37:37	50:27

Table 7.1: Summary statistics of the patients and observations for this study by epoetin agent .

	No. patients	Min.	Q1	Median	Mean	Q3	Max.	SD
Hb (all data)	151	7.20	11.00	11.80	11.84	12.70	17.20	1.36
Hb DA	74	8.00	11.20	12.00	12.03	12.80	17.20	1.32
Hb EB	77	7.20	10.90	11.60	11.66	12.50	16.20	1.39
Dose (all data)	151	0.00	51.50	91.24	114.70	155.80	493.80	86.06
Dose DA	74	0.00	43.54	74.00	98.30	125.90	493.80	84.46
Dose EB	77	0.00	67.05	115.20	130.40	179.60	461.50	84.68

Table 7.2: Summary statistics for Hb (g/dL) and Dose (IU/kg)

The summary statistics (in table 7.2) show that across all patients the mean and median Hb is 11.8 g/dL, which is the same as the ideal level for patients with renal anaemia. When these values are calculated for individual agents, the mean and median Hb for patients receiving DA is 12.0 g/dL, which is slightly higher than the target. In the EB group the mean and median for Hb is 11.6 g/dL, which is slightly lower than target. The spread of the data appears to be very similar overall and by agent, evident from the standard deviations being approximately the same (≈ 1.36) and also the lower (Q1) and upper (Q3) quartiles of the data.

The mean and median values for the dose data are quite different from each other, whereby the median is less than the mean. This indicates that dose is positively skewed. In this instance the median should be used as a central measure rather than the mean, as the mean is highly influenced by outliers. The median dose for the DA group ($= 74$ IU/kg) is much lower than the median dose in the EB group ($= 115$ IU/kg), and the third quartile in the DA is also smaller then the EB group. These features of the data suggest that patients receive a higher dose in the EB than the DA group.

Figure 7.1 shows that Hb for both agents combined follows the normal distribution very closely, this is also the case when the data are separated by the agent (see figure 7.2). Figure 7.1 shows that overall dose is positively skewed, suggesting that patients tend to receive lower doses more than larger ones, and only a few patients on a few occasions receive doses greater than 300 IU/kg. The distribution for EB dose is much wider than for DA dose, seen in figure 7.2. There are a few patients on a few occasions that receive DA dose greater than 200 IU/kg; however in the EB group the distribution does not tail off until after 300 IU/kg. From these plots it seems that patients receiving EB are administered larger doses than patients receiving DA; however further and more sophisticated analysis will highlight whether this is truly the case.

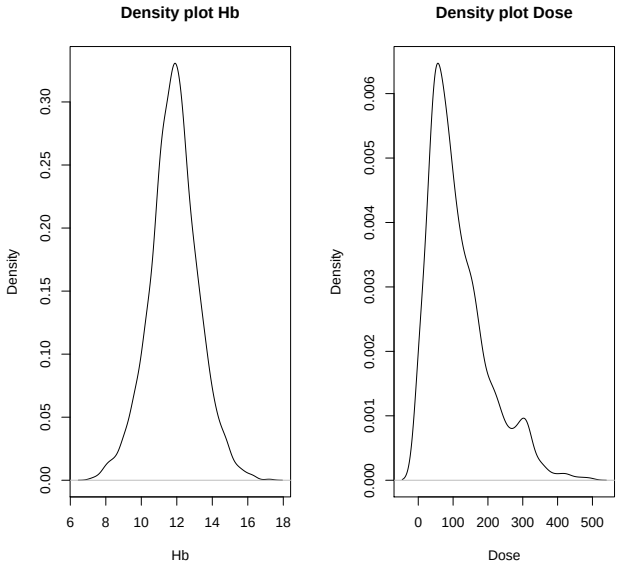

Figure 7.1: Kernel density plot of Hb (left) and Dose (right)

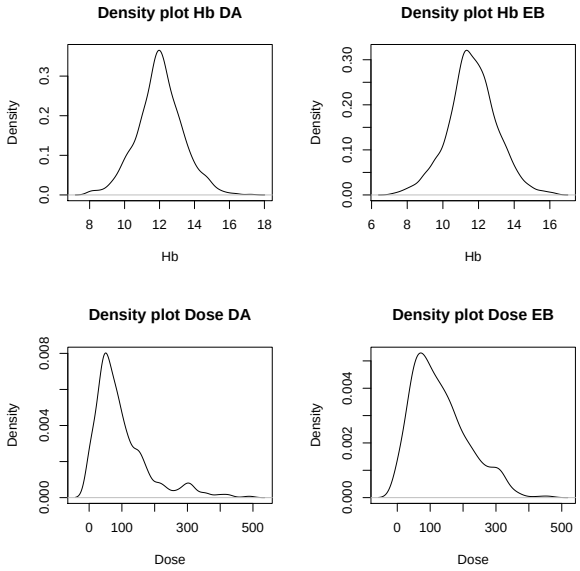

Figure 7.2: Kernel density plot of Hb (top) and Dose (bottom) by agent, DA (left) and EB (right)

7.2 Functional Data Analysis

The results from the application of FDA to the data from the renal anaemia biomedical system are presented in this section. FDA was used as features of the methodology and the data inferred that it would be an appropriate methodology to analyse the dynamics and control of renal anaemia in the patients, which are highlighted below.

- Hb (and dose) may be regarded as a smooth trajectory over time, which can be modelled using smoothed B-spline curves.
- It may be possible to model control of Hb using phase plots. In particular the first derivative of the Hb function would indicate the rate of change of the trajectory.
- Fitting individual curves will permit analysis of each individual patient.
- The mean functions may be a useful way to compare and contrast patients response to the two agents.

Four graphs are presented for a representative sample of patients representing a variety of scenarios that may occur with regards to control of Hb and Dose over time (figures 7.3 - 7.7). The top two graphs show the fitting of B splines to the Hb (left) and dose (right) data. The dotted pink line in the left plot indicates the optimum Hb of 11.8 g/dL. The bottom left graph shows the phase plot for Hb; this is the first derivative plotted against Hb measurement. A tight spiral in the center of the dotted lines would represent a patient who has good control of Hb. The red dot indicates the start of the trajectory in the phase plane and hence indicates the direction of the spiral. The plot in the bottom right is the first derivative of the Hb curve, in other words modelling the rate of change of Hb.

Ramsay and Silverman considered phases plots of the second derivative against the first derivative, in their examples [105]. This approach was implemented in this thesis with the renal data, but plots of the first derivative against Hb measurement produce very

similar results and identical conclusions are made. Therefore, in the interest of simplicity for the clinical domain, the original approach has not been used. This demonstrates a development of the FDA methodology.

A pragmatic solution is required when determining the appropriate position and number of knots for the B-splines, since the algorithmic rules that do exist are not readily implementable nor would they yield appropriate solutions for each curve. The aim was to choose a number of knots that represented the whole group of patients, opposed to having a different number of knots for each patient. Having a different number of knots for each individual would mean that the curves could not all be interpreted in the same way. Furthermore, the algorithm for the fitting of the curves would be more complex and the running time longer, if the knots varied between patients. Importantly, it was found necessary to place knots at the first and last data points (0 and 12 months) as the routine will not run unless the end points are knots. It was necessary to specify the knots to be evenly spaced, since these data are evenly spaced. In an example where the data are not balanced, it may be necessary to place more knots where there are more data points and fewer knots where there is less data.

This involved evaluating the relative merits of the permutations of knot number and position subject to the constraints outlined above. Placing a knot at each month was taken as a sensible starting point, as this option positions a knot at every time point the data were collected. This fully saturated possibility overfitted the curve as it passed through every measurement and the long term trend was not evident. As the B-splines needed more flexibility, the number of knots was reduced. Maintaining the equal balance in the position of knots, knots were considered at two month intervals (0,2,4,6,8,10,12), thus giving 7 knots points. While the B-splines became more representative of the data, further relaxation appeared possible, investigating 5 knot points at months 0,3,6,9 and 12. These B-splines displayed a good fit to the data. Reducing the number of knots further, i.e. to 4 (0,4,8,12) yielded some B-splines that needed further restriction (via

additional knots). Fewer than 5 knots resulted in curves that did not pass through many of the actual measurements. Therefore 5 knot points were deemed the most appropriate as statistically this produced flexible curves with enough freedom to represent the discrete measurements, while placing equidistant knots at sensible time positions.

Smoothness of the curves was ensured by fitting fifth degree B-splines polynomials and penalizing the fourth derivative to be smooth. This was done as it is reasonable to assume that Hb varies smoothly with time, in response to epoetin dose [98]. Furthermore, it permits phase plots to be constructed from smooth derivatives up to the third order, allowing assessment of Hb control.

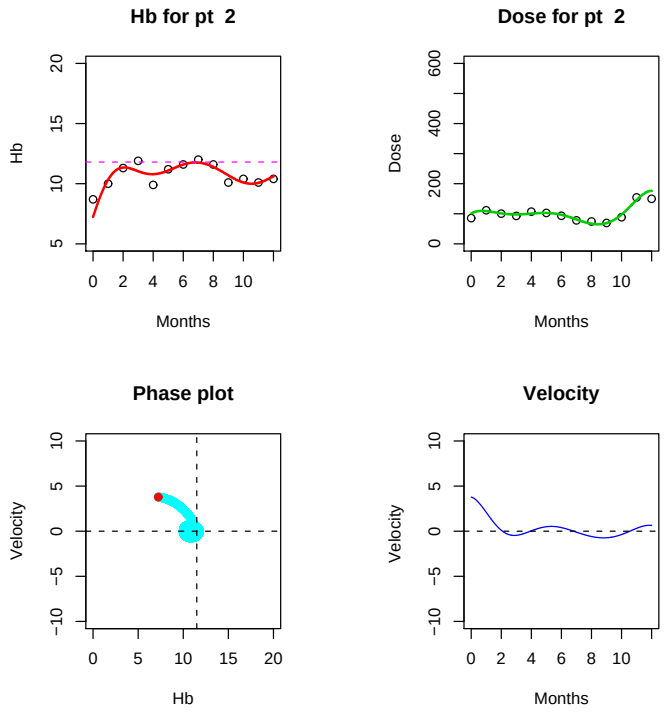

Figure 7.3: Graphs showing the fitting of B-splines to the Hb (left) and dose (right) data. The bottom graph shows the phase plot and velocity trajectory of the Hb curve for patient 2.

Figure 7.3 is used to show an example of a patient whose Hb begins off target then comes under control, albeit slightly below the optimum value. This is shown in the phase plot as the red dot indicates initially Hb is not under control or on target, the plot then spirals in to a very tight circle indicting good control. The dose seems to remain constant until month 10 around 100 IU/kg, this is then increased at the last two months when Hb is shown to be decreasing.

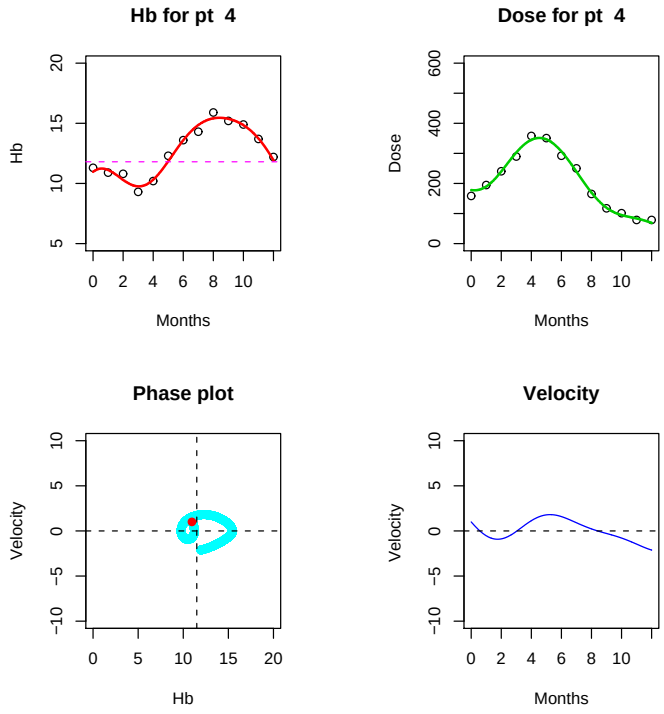

Figure 7.4: Graphs showing the fitting of B-splines to the Hb (left) and dose (right) data. The bottom graph shows the phase plot and velocity trajectory of the Hb curve for patient 4.

Figure 7.4 illustrates a patient whose Hb begins on target, then goes slightly below target, after month 5 Hb goes greatly above target but after month 8 is brought back to target. This is shown in the phase plot by an initial tight circle, which then spirals out to form a large circle that returns to the centre of the plot indicating the patient is coming back

under control. The dose for this patient reflects their Hb, where the patient is below target dose is increased in order to bring the patient to the optimum value, then from month 5 onwards the dose is decreased every month due to the patients Hb being out of control and well off target. The velocity curve is quite stable indicating that the change in dose is steady and there is no large dose change between adjacent months, although between month 4 and 12 the difference in dose is approximately 250 IU/kg.

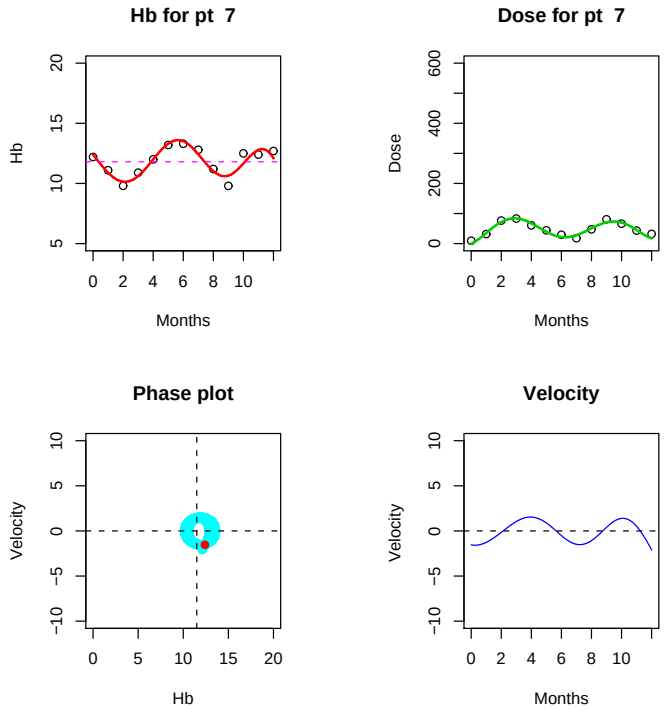

Figure 7.5: Graphs showing the fitting of B-splines to the Hb (left) and dose (right) data. The bottom graph shows the phase plot and velocity trajectory of the Hb curve for patient 7.

Figure 7.5 represents a patient whose Hb oscillates, albeit around the optimum value. There is a periodicity of approximately 5 months, whereby the patient's Hb will begin on target and will then increase or decrease for 2 months, then being brought back to the target. The dose for this patient shows the similar oscillating pattern, which mirrors the

respective patient's Hb curve. For example, where Hb is at its maximum, the dose is at its minimum. This is demonstrated by the phase plot by a large circle, representing poor control, around the centre of the plot. Similarly the velocity curve shows the oscillating behaviour.

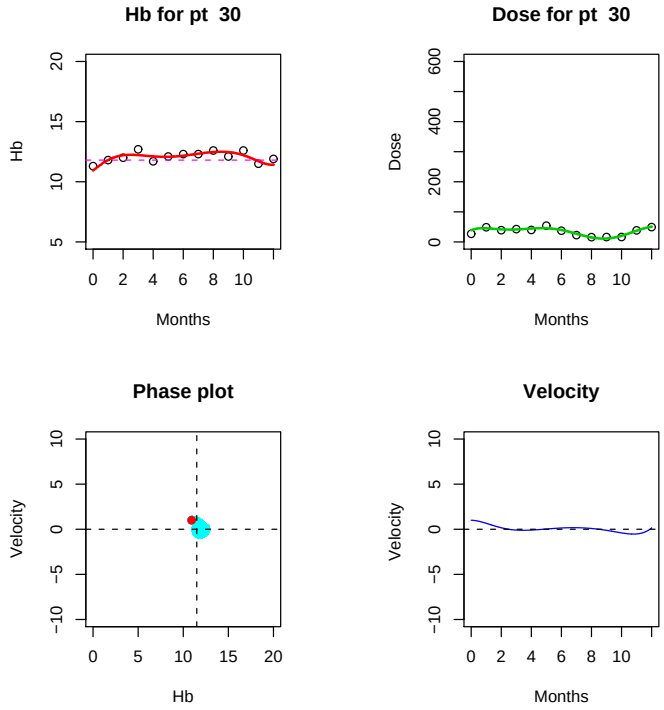

Figure 7.6: Graphs showing the fitting of B-splines to the Hb (left) and dose (right) data. The bottom graph shows the phase plot and velocity trajectory of the Hb curve for patient 30.

Figure 7.6 shows a patient whose Hb is extremely well controlled, such that it is almost constant at 11.8 throughout the whole study period. The dose is maintained at a low and constant level (approximately 50 IU/kg) throughout. The phase plot shows a small tight circle in the centre of the plot and the velocity curve is almost a horizontal line. This is highly likely to represent a patient who went on to receive a kidney transplant.

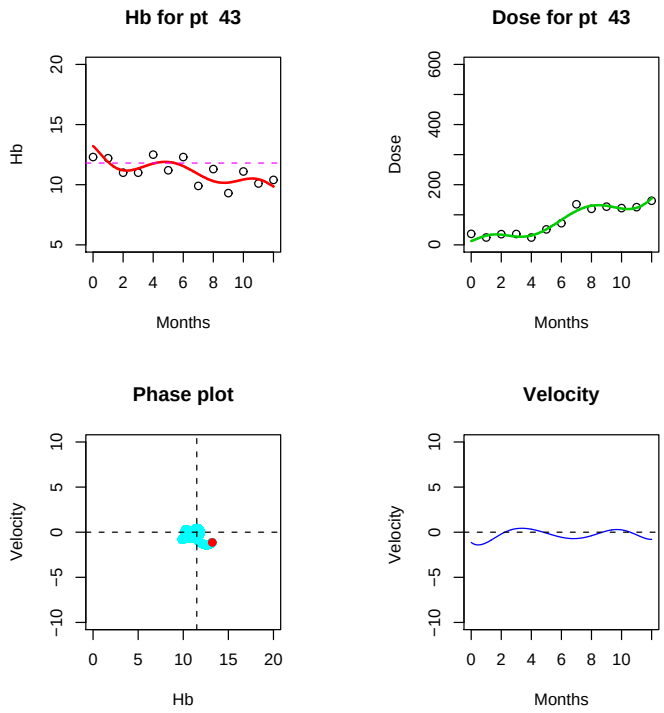

Figure 7.7: Graphs showing the fitting of B-splines to the Hb (left) and dose (right) data. The bottom graph shows the phase plot and velocity trajectory of the Hb curve for patient 43.

Behaviour	Number of patients	Percentage
Spiral in	33	22%
Spiral out	37	24%
Oscillate	48	32%
Stable	24	16%
Curve does not fit	9	6%

Table 7.3: Numerical summary of behaviour exhibited by renal anaemia patients phase plots

Figure 7.7 is used to represent an occasion where the B-spline curve for the Hb data does not pass through all of the individual data points; instead the curve passes through the middle of the data. The curve models the underlying behaviour of Hb rather than the slightly oscillating behaviour. The phase plot and velocity curve suggest that this patient's Hb is well controlled, although on this occasion less smoothing would be more appropriate.

Many patterns of Hb waveform variation were revealed by fitting the B-spline functions, some appeared to be sinusoidal and others irregular. The amplitude of the waves also varied widely. The decision was made not to categorise the wave patterns as the time frame and number of data points were limited. Table 7.3 is used to present the number (and percentage) of the 151 patients displaying the behaviour as shown in figures 7.3 to 7.7.

A particularly interesting group of patients that were apparent were those whose Hb wave shows evidence of Hb cycling. These patients are not adequatly controlled by the CDSS, which may occur due to external problems (such as experiencing an adverse event) or because they are receiving an inappropriate dose. This will be investigated further in section 7.3.

7.2.1 Functional Mean Curves

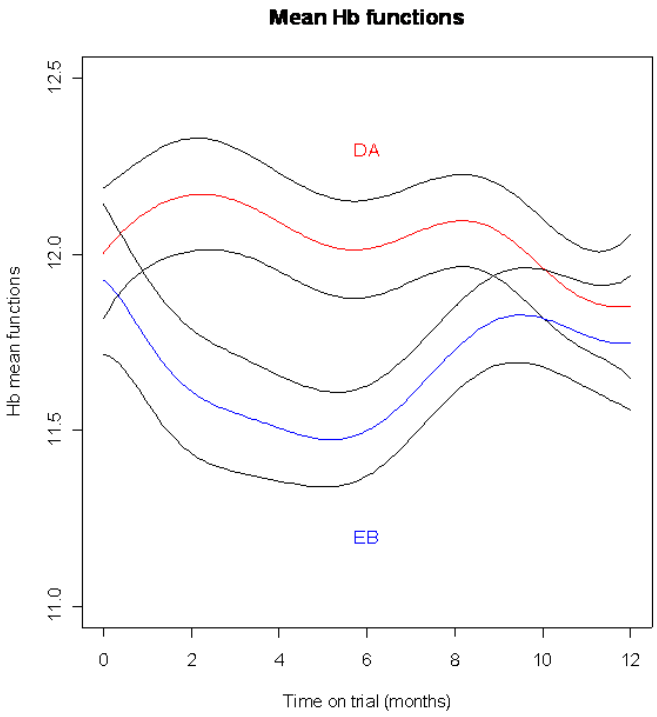

Figure 7.8: The two group mean curves for the two agents, with 68 % confidence limits. Where the confidence limits do not overlap the mean curves are separated by at least two standard errors and thus indicate where there is a significant difference between the two agents. The red curve represents the DA (Darbepoetin) group and the blue curve for the EB (Epoetin beta) group.

In order to compare the two agents in the FDA framework, mean functions were constructed from the Hb patient trajectories for the DA and EB groups, which are plotted with 68 % confidence limits. Figure 7.8 shows that there is a slight difference between the mean curves initially, with Hb for DA being slightly higher than Hb for EB. The patients were randomised at the beginning of the trial, therefore the slight difference in patients initially should be due to random sampling variation. After the first month the two mean

curves diverge, to the extent that the confidence limits no longer overlap, which indicates a statistically significant difference between the groups. From month 9 onward the mean functions converge until they reach similar levels at month 12.

7.3 Multilevel Modelling

The results from the application of multilevel modelling to the data from the renal anaemia biomedical system are presented in this section. The process of fitting the models is explained, together with the reasons behind the steps taken to find an appropriate model for the data.

7.3.1 Relationship between Haemoglobin and Epoetin

As discussed previously, epoetin doses are managed with the assistance of a computerised decision support system; which adjusts a patient's dose each month by examining their current Hb, which is believed to be the result of the previous months epoetin. Models were subsequently fitted to model this relationship. A further postulation is that the relationship between Hb and dose is nonlinear, since the dose ladder is asymmetric. It is not plausible to assume that Hb will increase linearly as dose increases; a clinical maximum dose is 300 IU/kg since beyond this no significant effect is seen in Hb levels. A quadratic effect was tested, since this would reflect that Hb increases as dose increases together with the plateauing affect that dose has when a certain level of Hb is reached.

The inclusion of random intercepts and slopes in the model was investigated. A random intercept would mean that patients Hb levels would vary (for a zero dose), this would give an assessment of renal health. A random slope would indicate variation in the relationship between Hb and dose for each patient. This has clinical meaning in terms of patient sensitivity. A steep slope would represent a patient who is more sensitive, whereby a small

Model	DF	AIC	BIC	loglik
1	4	5962.156	5984.160	-2977.078
2	**6**	**5949.692**	**5982.699**	**-2968.846**
3	5	5978.935	6006.438	-2984.468
4	7	5969.800	6008.304	-2977.900
5	10	5967.404	6022.410	-2973.702

Table 7.4: Information criterion for initial models investigating relationship between Hb(i) and dose(i-1)

dose adjustment achieves a large change in Hb. Conversely, a less sensitive patient maybe less responsive and a large adjustment in dose achieves a small increase in dose (i.e. a shallow slope). As the model coefficients have great clinical importance and meaning, this emphasises the practicality of these models in the clinical domain.

7.3.2 Haemoglobin(i) vs dose($i-1$)

The initial set of models fitted was to investigate the relationship between Hb(i) and Dose(i-1), these are shown in the list below. Information criterion from the models are presented in table 7.4

1. lme(hb(i) $\sim$ dose(i-1), random = 1 $\setminus$ patient)
2. lme(hb(i) $\sim$ dose(i-1), random = dose(i-1)$\setminus$ patient)
3. lme(hb(i) $\sim$ dose(i-1)+ dose^2(i-1), random = 1$\setminus$patient)
4. lme(hb(i) $\sim$ dose(i-1)+dose^2(i-1), random = dose(i-1)$\setminus$patient)
5. lme(hb(i) $\sim$ dose(i-1)+dose^2(i-1), random = dose(i-1)+dose^2(i-1)$\setminus$patient)

The best fitting model (according to AIC) between Hb(i) and dose(i-1), at this point, includes a random intercept and slope for each patient (model 2), a quadratic term was found not to improve the model fit. The residuals of this model were investigated and found to be autocorrelated (shown in figure 7.9), suggesting that a correlation structure for the residuals is required.

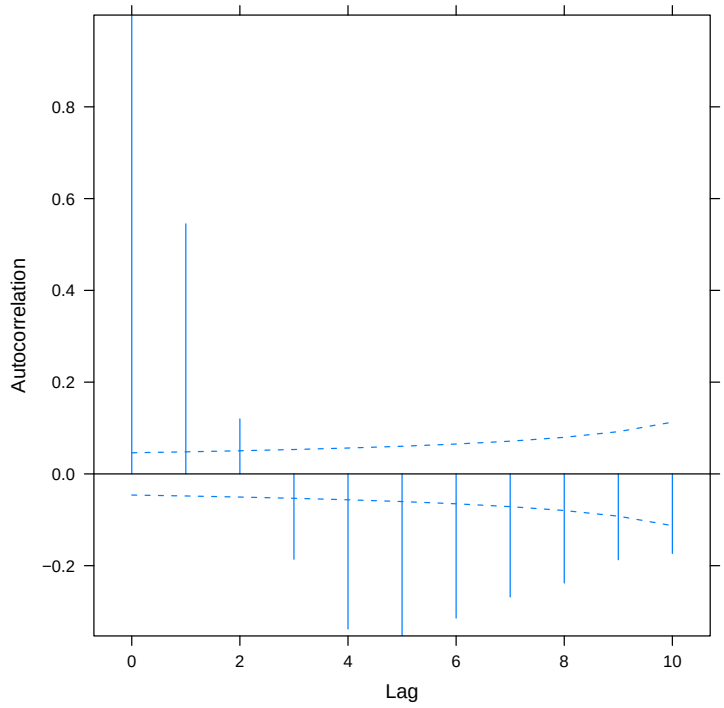

Figure 7.9: ACF of normalised residuals from model lme(hb(i) = dose(i-1), random = dose(i-1)\ patient)

When model 2 was fitted with AR(1) correlation structure this improved model fit and yielded uncorrelated normalised residuals, however this model yields a non positive definite variance-covariance matrix. The models with AR(2) and AR(3) also improved model fit, although in these cases the fixed effects dose term was not statistically significant, which further suggests that the models including dose(i-1) are unstable and do not represent the data.

7.3.3 Haemoglobin(i) vs. dose($i-\gamma$)

Further investigation for a suitable model was required, since the models between Hb(i) and dose(i-1) yielded erroneous results. Moreover, Hb cycling has been shown to occur in some of these patients, through the fitting of the B-spline curves. It is possible to speculate therefore that current Hb being based on dose(i-1) is sub-optimal. One suggestion is that a response in Hb, from epoetin dose, is seen within 2 to 6 weeks and there are a number of exogenous factors that may affect this and thus lengthen the time lag. A longer time lag than one month may occur from the administration of epoetin to response in Hb, which would mean that the 4 week lag currently implemented would be sub-optimal.

Models were fitted between Hb(i) and dose(i-2) and similarly Hb(i) and dose(i-3). It would not be plausible to include adjacent doses in the model as covariates, as there is very strong correlation (=0.95), hence collinearity, between both dose(i-1) and dose(i-2), which is clearly illustrated in figure 7.10. In general, including variables that are correlated would introduce problems associated with severe collinearity into the model, i.e. biased estimates of the coefficients and inflated standard errors. In this particular model, which shall be referred to as the collinearity model (see table 7.5), the fixed effect coefficient for dose(i-1) is negative and the variance-covariance matrix is non-positive definite. It is not plausible for the slope to be negative as this would suggest that Hb decreases as dose increases, which is not biologically feasible. A possible method of removing collinearity would be to include the mean and difference of dose(i-2) and dose(i-3), as both variables will not be included explicitly in the model [142]. In this case there would not be correlation between the variables and a combination of the effect from both doses can be modelled.

The difference and mean models are presented below, some include AR(1) correlation structures as thus far these have been necessary in yielding suitable models:

A1 lme(Hb(i)$\sim$ Mean+Diff, random= 1 \ pt)

Table 7.5: Coefficients of the collinearity model. This model includes dose(i-1) and dose(i-2) as covariates, which introduces collinearity into the model as the adjacent doses are correlated. There are 1508 degrees of freedom in this model.

Fixed effects	Coefficient	SE	p-value		
Intercept	11.18	0.10	≤ 0.0001		
Dose(i-1)	-0.015	0.0013	≤ 0.0001		
Dose(i-2)	0.023	0.0015	≤ 0.0001		
Random effects	SD	**Rand Corr**		**Fix Corr**	
Intercept	0.96	Int	Dose(i-1)	Int	Dose(i-1)
Dose(i-1)	0.01	-0.334		-0.304	
Dose(i-2)	0.014	-0.025	-0.824	-0.141	-0.794
Residual	0.88				
Diagnostics	AIC	BIC	logLik		
	4910.184	4964.318	-2445.092		
Residuals	Min	Q1	Med	Q3	Max
	-4.07	-0.53	0.05	0.57	2.82

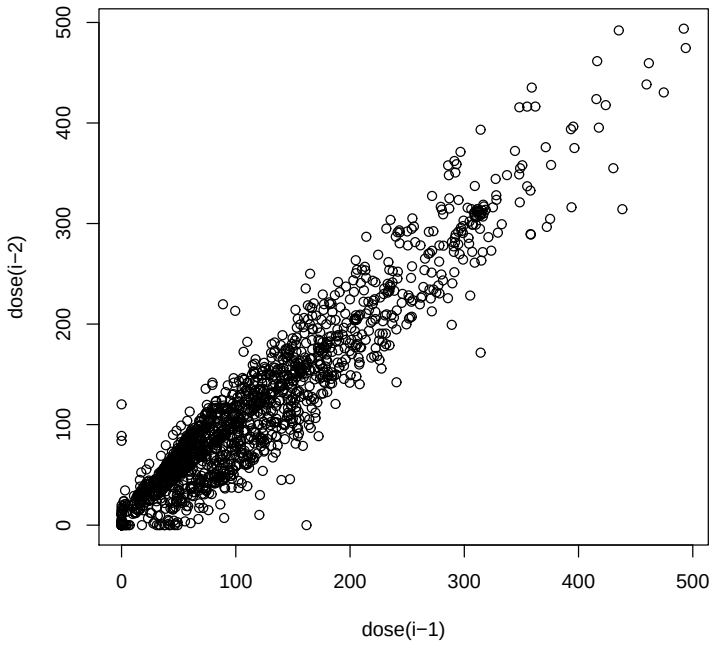

Figure 7.10: Scatter plot to show relationship between dose(i-1) and dose(i-2)

Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
A1	5	4769.775	4796.364	-2379.887			
A2	7	4584.280	4621.506	-2285.140	A1 vs A2	189.49438	≤ 0.0001
A3	10	4567.079	4620.258	-2273.540	A2 vs A3	23.20128	≤ 0.0001
A5	7	4596.839	4634.064	-2291.420			
A6	10	4532.341	4585.520	-2256.171	A5 vs A6	70.4982	≤ 0.0001
A7	**8**	**4191.565**	**4234.108**	**-2087.782**			
A8	11	4186.868	4245.365	-2082.434	A7 vs A8	10.6966	0.0135

Table 7.6: Information criterion for initial models investigating relationship between Hb(i) and the mean and difference of dose(i-2) and dose(i-3)

A2 lme(Hb(i)$\sim$Mean+Diff,random= Mean \ pt)

A3 lme(Hb(i)$\sim$Mean+Diff,random= Mean+Diff \ pt)

A4 lme(Hb(i)$\sim$Mean+Diff,random= Mean+Diff \ pt, corARMA(p=1))

A5 lme(Hb(i)$\sim$Mean+Mean^2, random= Mean \ pt)

A6 lme(Hb(i)$\sim$Mean+Mean^2, random= Mean+Mean^2 \ pt)

A7 lme(Hb(i)$\sim$Mean+Mean^2, random= Mean \ pt, corARMA(p=1))

A8 lme(Hb(i)$\sim$Mean+Mean^2, random= Mean+Mean^2) \pt, corARMA(p=1))

After fitting the model the difference is not statistically significant and does not improve model fit, which suggests difference is not required in the model. When the AR(1) correlation is included with the difference models, these did not converge, highlighting further numerical problems within these models. The best fitting model with Hb(i) as the response and $\frac{dose_{i-2}+dose_{i-3}}{2}$ as the covariate, includes a random intercept and slope and fixed quadratic effect together with an AR(1) correlation structure (Model A7), the

full summary is shown in table 7.7. The loglikelihood ratio test suggests that the random quadratic term is required in the model, but this model has a non positive definite variance-covariance matrix, hence is unstable.

Table 7.7: Coefficients of model A7. Note that there are 1357 degrees of freedom in this model.

Fixed effects	Coefficient	SE	p-value		
Intercept	10.53	0.15	≤ 0.0001		
Mean	0.017	0.00197	≤ 0.0001		
Mean^2	-0.000026	0.00000604	≤ 0.0001		
AR(1) coefficient (ρ)	0.66				
Fix Corr	Int	Mean			
Mean	-0.75				
Mean2	0.53	-0.85			
Random effects	SD	**Rand Corr**			
Intercept	0.96	Int			
Mean	0.01	-0.37			
Residual	1.15				
Diagnostics	AIC	BIC	logLik		
	4191.57	4234.11	-2087.783		
Residuals	Min	Q1	Med	Q3	Max
	-3.22	-0.48	0.04	0.52	3.41

7.3.4 Haemoglobin(i) vs dose($i-(2+\gamma)$)

The model shown in table 7.7 can be developed further; instead of simply using an arithmetic mean of dose(i-2) and dose(i-3), models will be fitted with a linear combination of dose(i-2) and dose(i-3) as the covariate. This approach means that it will be possible

to identify where between dose(i-2) and dose(i-3) the strongest relationship occurs with Hb(i). A model will be fitted with Hb(i) as the response and the dose covariate as equation 7.1, where $0 \leq \gamma \leq 1$. See equation 7.2 for the fitted model; this includes random intercept and slope parameters and fixed quadratic term, together with AR(1) correlation structure. Models will be fit with γ values between 0 and 1, at intervals of 0.1, the model which yields the lowest AIC will be concluded as being the best fitting model and $i-(2+\gamma)$ will be concluded as being the optimum time lag.

$$\text{dose}_{i-(2+\gamma)} = (1-\gamma)\text{dose}_{i-2} + (\gamma)\text{dose}_{i-3} \tag{7.1}$$

$$\text{Hb}_{ij} = (A+\alpha) + (B+\beta)\text{dose}_{i-(2+\gamma)} + (C)\text{dose}^2_{i-(2+\gamma)} + \rho e_{i-1,j} + \text{Z}_{ij} \tag{7.2}$$

Table 7.8: AIC for lag models 7.2, where $0 \leq \gamma \leq 1$

Lag (months)	AIC
2	4207.978
2.1	4200.183
2.2	4193.557
2.3	4189.261
2.4	**4188.408**
2.5	4191.565
2.6	4198.408
2.7	4207.834
2.8	4218.441
2.9	4229.02
3	4238.795

The lowest AIC occurs when the time lag is 2.4 months (γ=0.4), this suggests that the

optimum response in Hb is seen from the dose 2.4 months (10 weeks) previously. Note that, minimizing the AIC here is identical to maximizing the loglikelihood as the number of parameters is fixed. Figures 7.11 and 7.12 show the graphical representation of the model 7.3 and tables 7.9 and 7.10 showing model coefficients and confidence intervals.

$$Hb_{ij} = (10.52+\alpha)+(0.017+\beta)dose_{i-(2.4)}-0.000025dose^2_{i-(2.4)}+0.66e_{i-1,j}+Z_{ij} \quad (7.3)$$

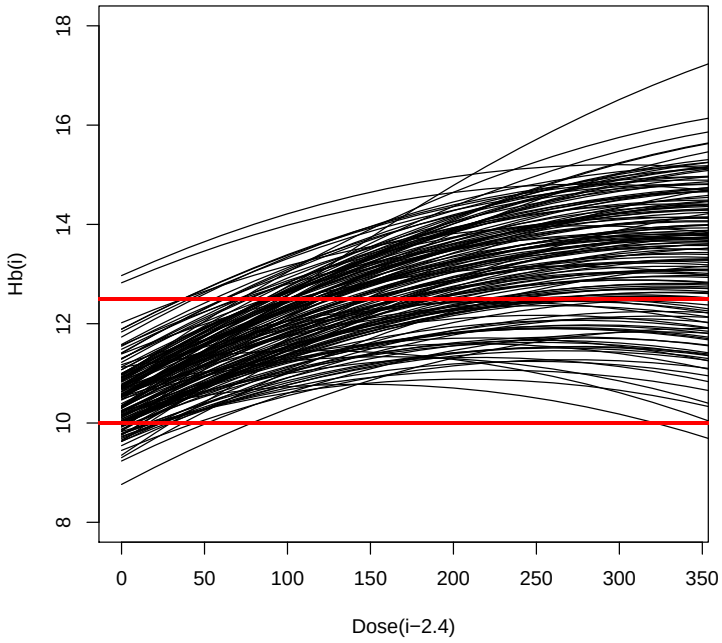

Figure 7.11: All patient curves from model 7.3. The red horizontal lines representing Hb control limits 10-12.5 g/dL

Figure 7.12 appears to be a reasonable representation of the patient curves shown in figure 7.11, as many patients follow the same trajectory although at different levels (due to the random intercepts). These patients tend to be in the middle of the so called spectrum

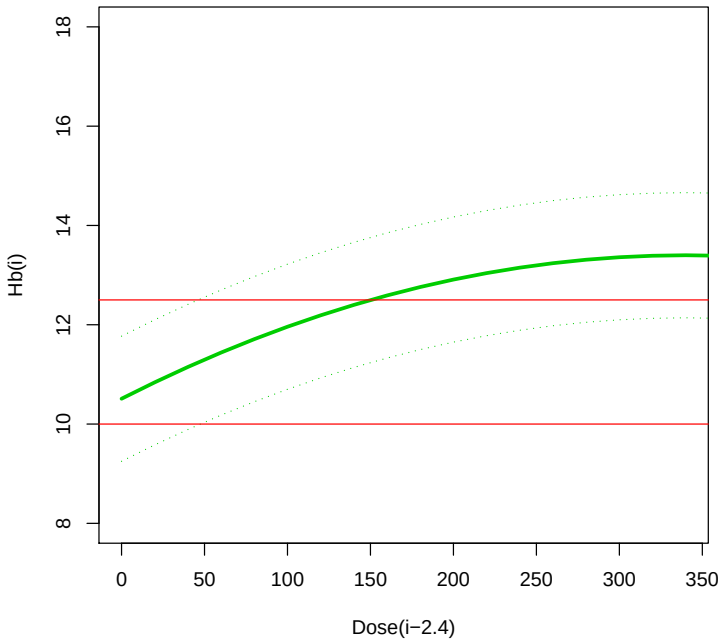

Figure 7.12: Mean curve from model 7.3, with 95 % confidence interval. The red horizontal lines representing Hb control limits 10-12.5 g/dL

of patients. However, the patients at the upper and lower end seem to have extreme behaviour, whereby there are patients with high Hb who experience a large slope. At the lower end there is a patient whose Hb seems to decrease with larger doses. The patients curves are summarised by a mean curve with 95% confidence limits, shown in figure 7.12. The 95 % confidence limits are calculated from the standard deviation of the random patient intercepts. Figure 7.12 shows that on average (from the mean curve) for reasonably low doses, between 0 and 150 IU/kg, the patients would attain a Hb within the target range (10 - 12.5 g/dL). The patients towards the lower confidence interval seem to require a larger dose. Below 50 IU/kg in these patients would result in Hb that is lower than 10g/dL, hence being dangerously low. A dose of around 175 IU/kg would put them in the middle of the target range. At the other end of the spectrum where patients Hb is relatively high, a dose of more than 50 IU/kg would result in a Hb exceeding the limits. These patients require a low dose to maintain their Hb. On the whole, in terms of the mean curve, it seems that a dose of around 100 IU/kg should maintain / help patients achieve Hb around 12 g/dL.

The intercept of model 7.3 is 10.52 g/dL, this is the mean Hb for a zero dose, across all patients and both agents. This is less than the optimum 11.8, though within the limits 10 to 12.5 g/dL. This suggests that patients do require some dose to maintain their Hb at a more acceptable level. The fixed slope is 0.017, meaning that for a 1 unit increase in dose then Hb will increase by 0.017 units. In more clinically meaningful terms a 100 unit increase in dose would achieve an increase of 1.7 units. The standard deviation of the random slope is 0.006, indicating that there is relatively small variation in the sensitivity between patients. The quadratic term should also be taken into account when considering this, although the coefficient is particularly small (= -0.000025) so will only have a substantial effect for large doses. As the coefficient is negative it accounts for the tailoring off of Hb at large doses, however at very large doses it would suggest that Hb decreases. This is implausible thus suggesting that the model is limited up to a maximum value which is too small to be considered as a true maximum.

Table 7.9: Coefficients of model 7.3. This model includes $\mathrm{Dose}_{(\mathrm{i}-2.4)}$ in the fixed and random specification and AR(1) correlation structure. Note that this model has 1357 degrees of freedom

Fixed effects	Coefficient	SE	p-value		
(Intercept)	10.52	0.155	≤ 0.0001		
$\mathrm{Dose}_{\mathrm{i}-2.4}$	0.017	0.0019	≤ 0.0001		
$\mathrm{Dose}^2_{1-2.4}$	-0.000025	0.00000603	≤ 0.0001		
AR(1) coefficient (ρ)	0.66				
Fix Corr	Int	$\mathrm{Dose}_{\mathrm{i}-2.4}$			
$\mathrm{Dose}_{\mathrm{i}-2.4}$	-0.75				
$\mathrm{Dose}^2_{1-2.4}$	0.53	-0.85			
Random effects	SD	**Rand Corr**			
Intercept	0.97	Int			
$\mathrm{Dose}_{\mathrm{i}-2.4}$	0.006	-0.38			
Residual	1.15				
Diagnostics	AIC	BIC	logLik		
	4188.408	4230.951	-2086.204		
Residuals	Min	Q1	Med	Q3	Max
	-3.19	-0.47	0.03	0.52	3.41

Table 7.10: 95 % Confidence intervals for all coefficients in model 7.3

Fixed effects:	lower	est.	upper
(Intercept)	10.21	10.52	10.82
$\mathrm{Dose}_{\mathrm{i}-2.4}$	0.013	0.017	0.021
$\mathrm{Dose}^2_{1-2.4}$	-3.69 e-05	-2.51 e-05	-1.32 e-05
AR(1) coefficient (ρ)	0.60	0.66	0.72
Random effects:			
sd(Intercept)	0.74	0.97	1.27
sd($\mathrm{Dose}_{\mathrm{i}-2.4}$)	0.004	0.006	0.010
cor(Intercept,$\mathrm{Dose}_{\mathrm{i}-2.4}$)	-0.695	-0.375	0.069
sd(Residual)	1.05	1.15	1.26

Table 7.10 shows that the standard deviation between patients is approximately equal to 1 (sd(Intercept) = 0.97), and so is the standard deviation within patients (sd(Residual) = 1.15). This suggests that there is as much variability between patients as there is within patients.

The model shown in figure 7.13, where the quadratic term varies between patients, is not an accurate representation of the data. Since the quadratic term varies between patients some patients experience curves with turning points at doses of around 100 IU/kg. In these patients doses greater than 100 would result in Hb being reduced, though from clinical knowledge it is known that this behaviour would not occur. In theory as the Hb increase so does the epoetin dose, up to a maximum where the effect in Hb tails of and a dose greater than this maximum is not influential. The maximum dose is stated as 300 IU/kg, however for the model in figure 7.13 model the maximum for the mean function is 250 IU/kg, which is less than the clinical maximum and hence unrepresentative. The maximum in model 7.3 is 328 IU/kg, which is a more realistic representation of the data albeit slightly higher than 300 IU/kg, yet still plausible as some patients do receive doses greater than 300 IU/kg when it is known that the effect may not be so influential.

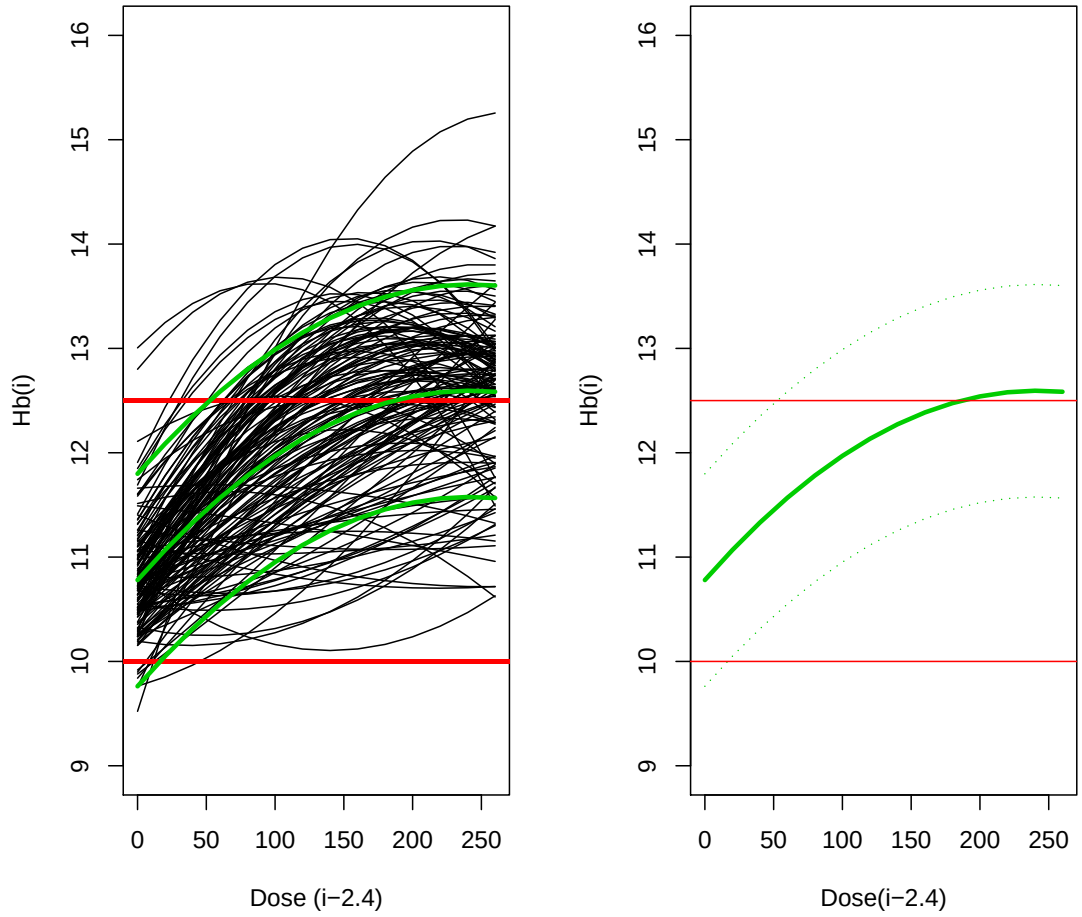

Figure 7.13: The figure on the left shows all patient curves from a model including different quadratic terms for each patient. The figure on the right shows a mean representation of the curves in the left figure. The green lines on both plots represent the mean curve with 95 % confidence intervals. Red horizontal lines representing Hb control limits 10-12.5 g/dL.

It is no coincidence that that maximum dose for the model is 300 IU/kg, nor is it that a dose of 100 IU/kg corresponds to a Hb response of 11.8 g/dL, as these are the values that have been found by trial and error in clinical practice [69] [94]. This highlights a successful outcome of the modelling, as clinically meaningful results have been yielded.

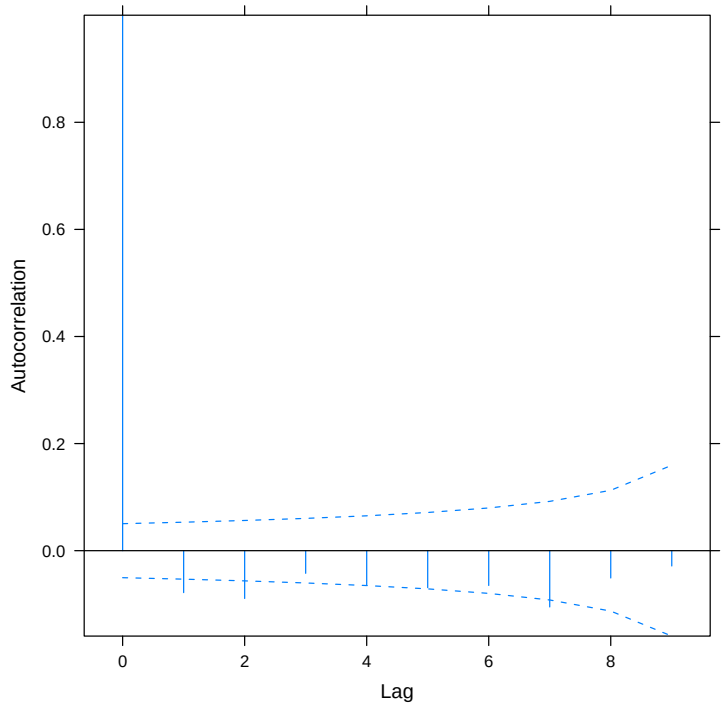

Figure 7.14: ACF of normalised residuals from model 7.3 (table 7.9)

Figure 7.14 shows that the normalised residuals from model 7.3 are (mostly) uncorrelated. There is significant autocorrelation at lags 1 and 2, but this is only very small. The same model was fitted with an AR(2) correlation structure, instead of AR(1), though did not improve the fit of the model significantly more than the AR(1), in contrast to the normalised residuals of the model without correlation structure (see figure 7.15), which are highly autocorrelated. Model 7.3 is therefore a substantial improvement.

Figure 7.16 shows that there is no relationship between the standardised residuals and

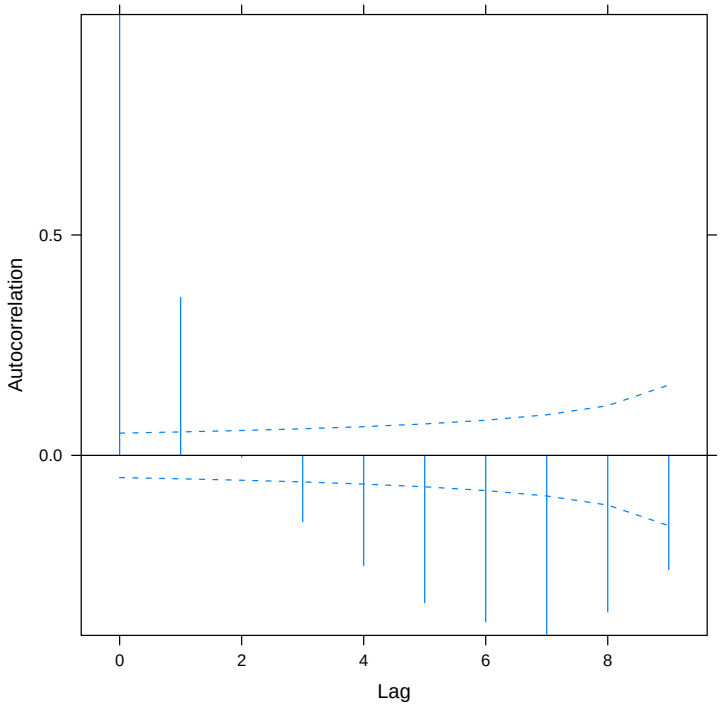

Figure 7.15: ACF of normalised residuals model 7.3 without the correlation structure

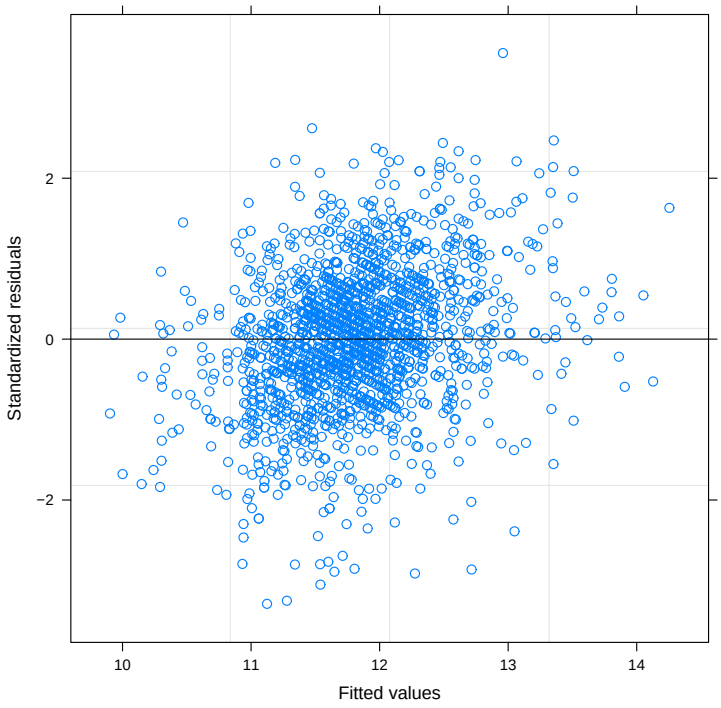

Figure 7.16: Standardised residuals vs fitted values for model 7.3

fitted values from model 7.3, i.e. that they are randomly scattered around zero, which is expected from a model that is appropriately specified. Furthermore, the majority of the residuals are within the range -2 to +2, thus approximately within the 95% confidence interval of the mean. Figure 7.17 shows that the random intercept and random slope from model 7.3 are normally distributed, with a mean zero. From each of the figures, the model assumptions have been checked and suitably adhered to. It therefore seems that model 7.3 is an appropriate representation of the data.

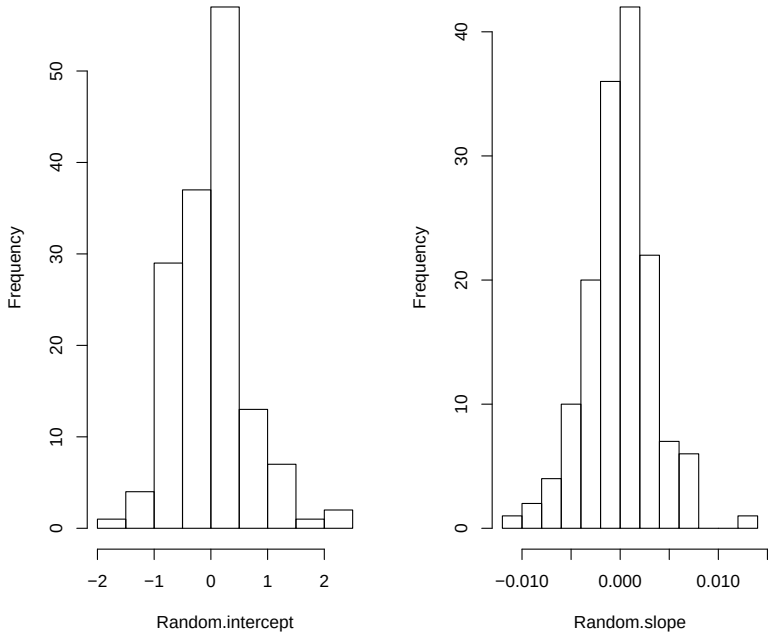

Figure 7.17: Histogram of random intercept and random slope for model 7.3

7.3.5 Haemoglobin(i) vs dose($i - \gamma$) for each agent

Models were then fitted for each agent using the same approach as in the previous section, in order to determine whether there is a difference in the lag time due to the agent. The

time lag was investigated between 1 and 3 months at intervals of 0.1, again the best fitting model was decided by the one which yielded the lowest AIC. AIC is used in favour of BIC as it is usually the case that the same conclusions will be made about the selection of model. In this case, the model yielding the lowest AIC will yield the lowest BIC, as constant adjustment for complexity is being made.

Epoetin Beta Model

Table 7.11 shows the EB dose lags between 1 and 3 months (at intervals of 0.1) together with the AIC from the respective model. The lowest AIC occurs in this instance at 2.3 months. The coefficients from the fitted model are shown in table 7.12.

Table 7.11: AIC for lag models for EB

Lag (months)	AIC	Lag (months)	AIC	Loglik
1.0	2388.861			
1.1	2384.438	2.1	2088.374	-1036.187
1.2	2378.267	2.2	2085.523	-1034.762
1.3	2370.091	**2.3**	**2084.686**	**-1034.343**
1.4	2360.144	2.4	2086.259	-1035.130
1.5	2349.086	2.5	2090.225	-1037.113
1.6	2337.923	2.6	2096.137	-1040.669
1.7	2327.645	2.7	2103.292	-1043.646
1.8	2318.903	2.8	2110.954	-1047.477
1.9	2311.923	2.9	2118.525	-1051.263
2.0	2306.609	3.0	2125.609	-1054.805

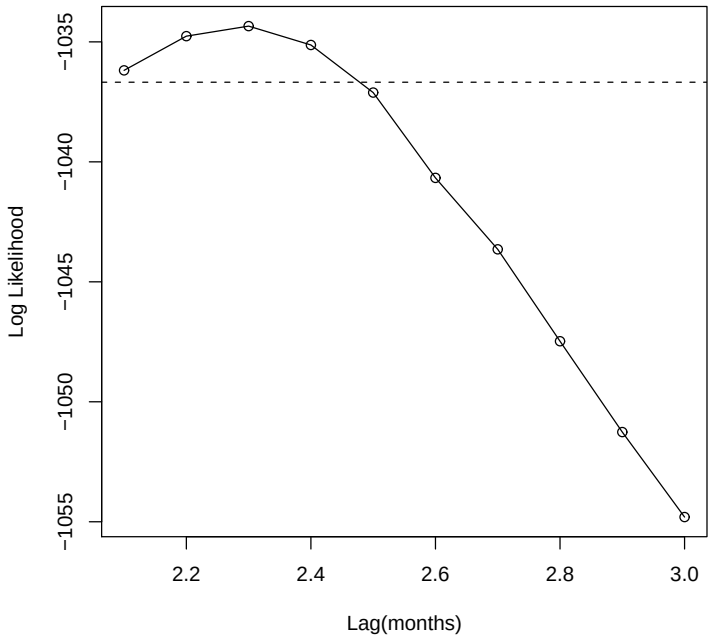

Figure 7.18: Profile log-likelihood for the optimal time lag for Epoetin Beta

Table 7.12: Model fitted to EB data, with $\mathrm{EB}_{(\mathrm{i}-2.3)}$ as the covariate and AR(1) correlation structure. Note that there are 691 degrees of freedom in this model.

Fixed effects:	Value	Std.Error	p-value		
(Intercept)	9.77	0.26	≤ 0.0001		
$\mathrm{EB}_{\mathrm{i}-2.3}$	0.020	0.0028	≤ 0.0001		
$\mathrm{EB}^2_{\mathrm{i}-2.3}$	-0.000030	0.00000834	3e-04		
AR(1) coefficient (ρ)	0.70				
Fix Corr	Int	$\mathrm{EB}_{\mathrm{i}-2.3}$			
$\mathrm{EB}_{\mathrm{i}-2.3}$	-0.76				
$\mathrm{EB}^2_{1-2.3}$	0.56	-0.89			
Random effects	SD	**Rand Corr**			
Intercept	1.26	Int			
$\mathrm{EB}_{\mathrm{i}-2.3}$	0.005	-0.476			
Residual	1.14				
Diagnostics	AIC	BIC	logLik		
	2084.686	2121.826	-1034.343		
Residuals	Min	Q1	Med	Q3	Max
	-3.19	-0.48	-0.001	0.50	3.01

Darbepoetin Alpha model

Table 7.13 shows the DA dose lags between 1 and 3 months (at intervals of 0.1) together with the AIC from the respective model. The lowest AIC occurs in this instance at 2.4 months. The coefficients from the fitted model are shown in table 7.14.

Table 7.13: AIC for lag models for DA

Lag (months)	AIC	Lag (months)	AIC	Loglik
1.0	2369.203			
1.1	2367.329	2.1	2123.379	-1053.690
1.2	2364.638	2.2	2120.614	-1052.307
1.3	2360.144	2.3	2118.636	-1051.318
1.4	2356.319	**2.4**	**2117.976**	**-1050.988**
1.5	2350.686	2.5	2118.991	-1051.494
1.6	2344.380	2.6	2121.612	-1052.806
1.7	2337.784	2.7	2125.337	-1054.669
1.8	2331.769	2.8	2129.510	-1056.755
1.9	2326.766	2.9	2133.596	-1058.798
2.0	2322.9	3.0	2137.287	-1060.644

Comparison

There are slight differences between the EB (table 7.12) and the DA (table 7.14) models, although in terms of the lag time from dose administration to response in Hb the conclusions are approximately the same; such that the optimal lag time is 2.3 months for EB and 2.4 months for DA. It does not seem therefore that there is a difference in the response time due to the agent. The 95% confidence interval for the EB optimal time lag, derived from the profile log-likelihood is (2.09, 2.49), this is shown in figure 7.18.

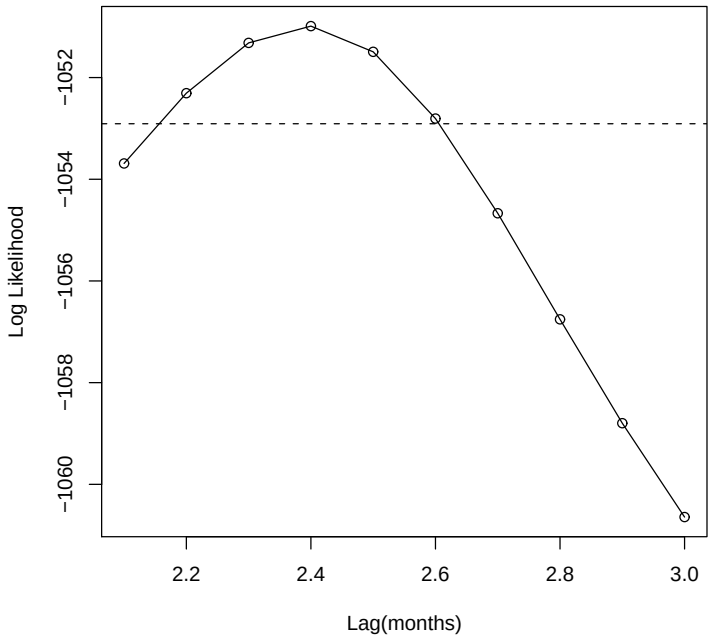

Figure 7.19: Profile log-likelihood for the optimal time lag for Darbepoetin Alpha

Table 7.14: Model fitted to DA data, with $\mathrm{DA}_{(\mathrm{i}-2.4)}$ as the covariate and AR(1) correlation structure. Note that there are 664 degrees of freedom in this model

Fixed effects:	Value	Std.Error	p-value		
(Intercept)	11.17	0.18	≤ 0.0001		
$\mathrm{DA}_{\mathrm{i}-2.4}$	0.014	0.0026	≤ 0.0001		
$\mathrm{DA}^2_{\mathrm{i}-2.4}$	-0.00002	0.00000873	0.02		
AR(1) coefficient (ρ)	0.64193				
Fix Corr	Int	$\mathrm{DA}_{\mathrm{i}-2.4}$			
$\mathrm{DA}_{\mathrm{i}-2.4}$	-0.78				
$\mathrm{DA}^2_{1-2.4}$	0.50	-0.79			
Random effects	SD	**Rand Corr**			
Intercept	0.61	Int			
$\mathrm{DA}_{\mathrm{i}-2.4}$	0.01	-0.64			
Residual	1.19				
Diagnostics	AIC	BIC	logLik		
	2117.976	2154.796	-1050.988		
Residuals	Min	Q1	Med	Q3	Max
	3.11	-0.51	0.01	0.55	3.92

Figure 7.19 shows the 95% confidence interval for DA is (2.15, 2.60). It seems that the point estimate and 95% confidence interval of the optimal time lags are relatively similar for each agent. Additionally the confidence intervals are quite narrow in both cases, and it would not be clinically important whether a patient receiving EB, for example, would have their dose adjusted at every 2.1 months or 2.5 months. Note that, the 95% confidence interval for the optimum lag time was calculated by subtracting 1.92 from the point estimate [143].

There are a few slight differences between the agents, shown from the fitting of the models to each agent. The first notable difference is the value of the fixed intercept, which is 9.77 g/dL for EB and 11.17 g/dL for DA. This suggests that on average for the patients under the DA regime, their Hb is within the target range even when they don't receive any of the agent, but for the EB patients they do require some agent before their Hb is within the target range (as they experience 9.77 g/dL with a zero EB dose). The difference in the fixed effects intercepts may be as a result of the conversion factor from EB dose to DA dose not being precise, thus resulting in patients receiving too much of the DA agent. It is evident that the DA conversion predicts a too high dose because for a zero dose the baseline Hb is relatively high (11.17 g/dL). A further possibility is that the initial randomisation process was not effective and patients with low Hb were selected for the EB group and those with high Hb to the DA group. This feature of the patients was seen from the functional mean curves, though with the MLM the time is not evident, and it is not possible to determine if the overdosing occurs through the whole 13 months or a shorter time period. The patients receiving DA may be receiving slightly larger doses than required, especially as it is predicted that they are within the target range without receiving any of the agent. The value of the fixed slope in the EB group (0.020) is slightly larger than that of the DA group (0.014), which suggests that the patients in the EB group may be more responsive to the agent than those in the DA group.

In the EB model the between patient variation (1.26) is similar to the within patient

variation (1.14), which is consistent with the overall model (model 7.3). This varies from the DA model where the between patient variation (0.61) is half the within patient variation (1.19). This suggests there is less variation in patients Hb and dose in the DA group than the EB group, which further emphasises that there was initially a difference in the baseline randomisation and in the conversion of EB to DA.

The model fitted to all the data (model 7.3) is essentially a balance of the features from two agents, such that the coefficients from model 7.3 are approximately the mid values of the coefficients from the separate agent models.

7.4 Clinical algorithm

Model 7.3 was found to be too complex to be applied in clinical practice, since it was difficult to rearrange the model in terms of dose and hence predict dose due to the large number of complex parameters (such as the correlation structure) - further details are given in Chapter 8. The model was made more pragmatic for a clinical approach. There were advantages of fitting the complex models, since this has advised the clinical approach. The clinical model has been fitted using the patients who received EB, since this was the standard drug used before the introduction of the newer DA and patients were already on the EB regime. Note that, in the previous section the models could have been fitted with dose as a fixed effect, which would have been useful in describing the statistical difference between the two groups. The choice was made to fit different models for the two agents as this was used to develop the clinical algorithm. Dose administered two months previously will be modelled as the covariate. The model (equation 7.4) includes a random patient intercept and fixed slope and quadratic terms, no correlation structure is included.

$$\mathrm{Hb_{ij}} = 9.74 + \mathrm{Patientoffset_{ij}} + 0.021\mathrm{EB_{i-2,j}} - 0.000032\mathrm{EB^2_{i-2,j}} + \mathrm{e_{ij}} \qquad (7.4)$$

The patient offset (PO) represents the random intercept term which varies between time points and patients, the mean value across all patients and time points is -0.045 g/dL. This was substituted in figure 7.20 which shows the mean behaviour of the patients together with 95 % confidence interval of the curve; the standard deviation for the patient offset was calculated as 1.28 g/dL. The blue curve shows that 'on average' to reach the target 11.8 g/dL, a dose of 120 IU/kg is required. For a zero dose, mean Hb is 9.74 g/dL, which is slightly less than the lower limit of the target range, which suggests that on average most patients require 'some' dose of the agent. Patients close to the lower confidence limit will not necessarily reach the 11.8 g/dL target, although a large dose ($>$ 250 IU/kg) would ensure the patients Hb is within the target limits. Patients at the upper confidence interval achieve greater than the target value without receiving any dose. A low dose ($<$ 40 IU/kg) would ensure they are within the target range. The parameter estimates in model 7.4 do not vary so much from those shown in the model presented in table 7.12.

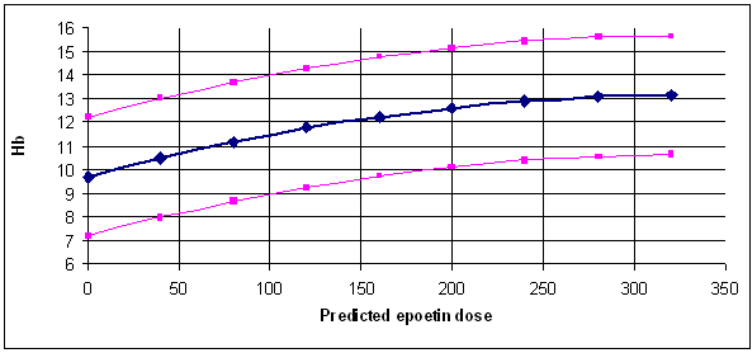

Figure 7.20: Hb vs predicted EB dose (as in model 7.4), the blue curve represents the mean and the pink curves represents the 95 % confidence interval

In order to predict dose, the following two equations (models 7.5 and 7.6) are solved simultaneously, and then rearranged for PO to yield model 7.7.

Table 7.15: Model for algorithm. Note that this model has 768 degrees of freedom.

Fixed effects:	Value	Std.Error	p-value		
(Intercept)	9.74	0.20	≤ 0.0001		
$\mathrm{EB}_{\mathrm{i}-2}$	0.021	0.0018	≤ 0.0001		
$\mathrm{Eb}^2_{\mathrm{i}-2}$	-0.000032	0.00000522	≤ 0.0001		
Fix Corr	Int	$\mathrm{EB}_{\mathrm{i}-2}$			
$\mathrm{EB}_{\mathrm{i}-2}$	-0.597				
EB^2_{1-2}	0.435	-0.908			
Random effects	SD				
Intercept	1.31				
Residual	0.97				
Diagnostics	AIC	BIC	logLik		
	2623.821	2647.512	-1306.911		
Residuals	Min	Q1	Med	Q3	Max
	-3.77	-0.52	0.02	0.59	3.07

$$\mathrm{Hb_{ij}} = 9.74 + \mathrm{PO_{ij}} + 0.021\mathrm{EB_{i-2,j}} - 0.000032\mathrm{EB^2_{i-2,j}} + \mathrm{e_{ij}} \quad (7.5)$$

$$\mathrm{Hb_{i+1,j}} = 9.74 + \mathrm{PO_{i+1,j}} + 0.021\mathrm{EB_{i-1,j}} - 0.000032\mathrm{EB^2_{i-1,j}} + \mathrm{e_{i+1,j}} \quad (7.6)$$

$$\mathrm{PO_{ij}} = (\mathrm{Hb_{i-1,j}}+\mathrm{Hb_{i-2,j}}-19.48-0.021(\mathrm{EB_{i-3,j}}+\mathrm{EB_{i-4,j}})+0.000032(\mathrm{EB^2_{i-3,j}}+\mathrm{EB^2_{i-4,j}}))/2 \quad (7.7)$$

The equation for predicting dose is found by using the quadratic equation to solve model 7.4 and substituting Hb as the target value (11.8 g/dL):

$$\mathrm{Predicteddose_{ij}} = -0.0021 + \sqrt{0.00017732 + 0.000128\mathrm{PO_{ij}}}/ - 0.000064 \quad (7.8)$$

Substituting equation 7.7 into 7.8 yields:

$$\begin{aligned}\mathrm{Predicteddose_{ij}} = -0.0021 + (0.00017732 + 0.000128((\mathrm{Hb_{i-1,j}} + \mathrm{Hb_{i-2,j}} - 19.48 \\ -0.021(\mathrm{EB_{i-3,j}} + \mathrm{EB_{i-4,j}}) + 0.000032(\mathrm{EB^2_{i-3,j}} + \mathrm{EB^2_{i-4,j}}))/2))^{1/2}/ - 0.000064\end{aligned} \quad (7.9)$$

7.4.1 Comparing predicted and true dose

The first predicted dose can be used from month 4 onwards; for the first four months the old algorithm should be used to calculate the dose. These months of data are required to validate the new algorithm. Table 7.16 shows summary statistics of the predicted doses at month 4. The mean of predicted dose (= 96.59 IU/kg) is less that the mean of the true dose (124 IU/kg) at month 4, which could suggest that patients receive too great a dose than is actually required. The NA's represent patients for whom it is not possible to predict a

Min	Q1	Med	Mean	Q3	Max	NA's
-73.42	28.92	92.64	96.59	173.30	274.20	13

Table 7.16: Summary of predicted doses at month 4

dose value, where the value under the $\surd$ is negative in equation 7.9 , this occurs when the offset value is less than -1.38 g/dL. These patients typically have low unstable Hb and seemingly require a large dose. The maximum value the equation predicts is 328 IU/kg, found by taking the derivative of equation 7.4 with respect to dose and setting this to zero to find the maximum (see equation 7.10). The maximum value (328 IU/kg) is just slightly more than the suggested clinical maximum dose, although patients do receive greater than 300 IU/kg, which suggested that the model is a good representation of the data. Figure 7.21 is used to show the relationship between predicted dose and patient offset, at month 4. The rules are given below of how to interpret a patient offset value:

- If Patient offset $<$ -1.385 $\rightarrow$ give maximum dose = 328 IU/kg/week (or clinical maximum)
- If Patient offset $>$ 2.06 $\rightarrow$ give dose = 0 IU/kg/week (or clinical minimum)
- If Patient offset = else $\rightarrow$ give dose according to figure 7.21

$$\begin{aligned} &\frac{\partial}{\partial \mathrm{EB}}(9.74 + \mathrm{Patientoffset_{ij}} + 0.021\mathrm{EB_{i-2,j}} - 0.000032\mathrm{EB^2_{i-2,j}} + \mathrm{e_{ij}}) \\ &\rightarrow 0 = 0.021 - 0.000064\mathrm{EB} \\ &\rightarrow \mathrm{EB} = \frac{-0.021}{-0.000064} \\ &\rightarrow \mathrm{EB} = 328 \end{aligned} \tag{7.10}$$

Figure 7.22 is used to compare predicted and true dose at month 4; the first possible occasion to predict dose using the new algorithm. There are five instances where predicted

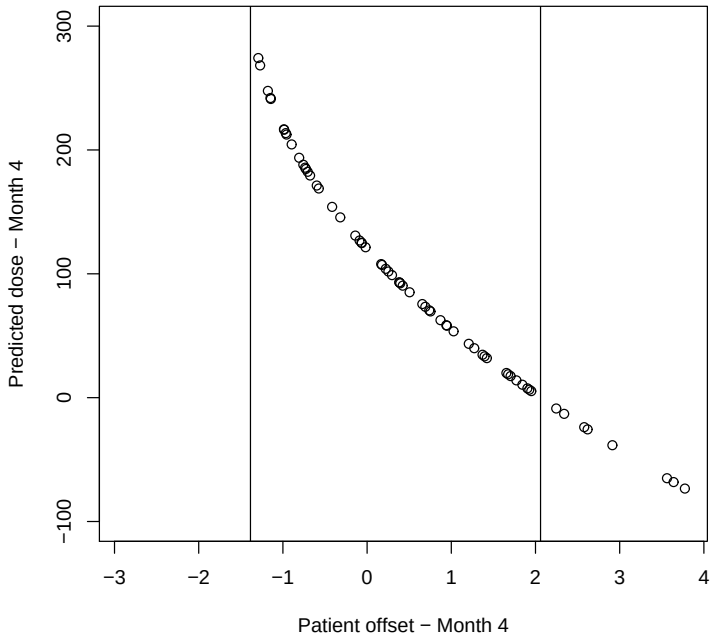

Figure 7.21: Predicted dose plotted against patient offset for month 4

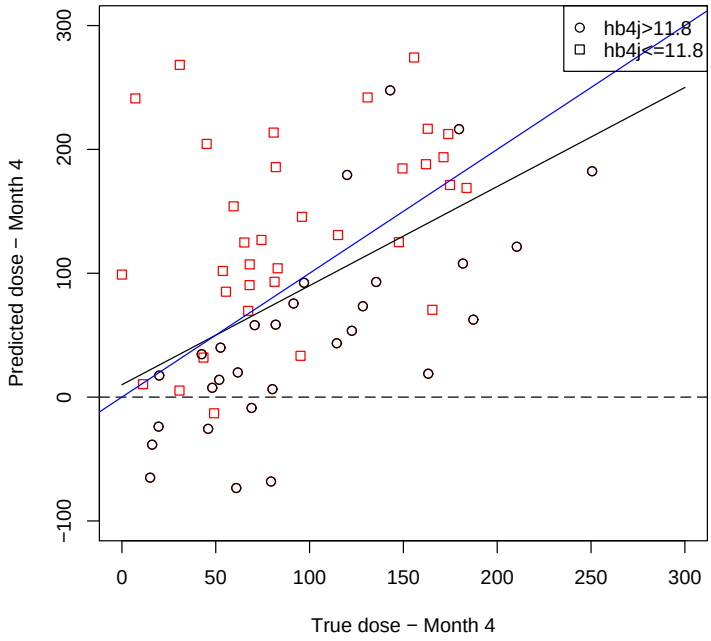

Figure 7.22: Predicted dose plotted against true dose at month 4. The red squares highlight where Hb is below or equal to 11.8 g/dL, the black circles highlight where Hb is greater than 11.8 g/dL. The blue line indicates where predicted dose and true dose are equal; the black line is an arbitrary sectioning of the data where the majority of points above the line are red squares and black circles below.

dose is equal to true dose, this is evident from the points which lie on the blue line. The black line is used to identify the separation between two groups of patients; patients in the upper sector (red squares) typically have lower than target Hb and their predicted dose is higher than the true dose; the patients in the lower sector (black circles) Hb is typically greater than the target value and their predicted dose is lower than the true dose.

In summary:

- If Hb $<$ 11.8 g/dL (Below target) - Predicted dose higher than true dose
- If Hb $>$ 11.8 g/dL (Above target) - Predicted dose lower than true dose

Figures 7.23, 7.24 and 7.25 are examples to show how predicted dose and true dose compare for all months, for selected patients; this is shown together with Hb.

Figure 7.23 is used to illustrate a scenario whereby predicted dose does not resemble true dose. The true dose is relatively stable, with a slight oscillation about 50 IU/kg, which is lower than the mean dose, thus suggesting that Hb is high but stable. In fact, Hb is high (around 13 g/dL) for the first three months, it then drops well below target for the next four months to around 9 g/dL, which is highlighted in the patients dose increasing though only slightly in this period. Hb then increases greatly to around 13/14 g/dL, which is reflected in the true dose slightly decreasing. The predicted dose is a more accentuated response to the patients Hb than the true dose; it is suggesting a very high dose when Hb is lower than average and then large decreasing steps when the patients Hb increases greatly. It would be of interest to know how the predicted dose regime would have affected the patient. For example, would their Hb be better controlled if the dose was more responsive to their fluctuating Hb levels rather than maintaining a relatively stable dose? It might be that the patient experienced an inter-current complication, and hence their Hb does not respond in the ideal way.

Figure 7.24 is another example where predicted dose does not resemble true dose. It can be seen from the plot of true dose over time that almost identical doses are given to the

patient each month, around 74 IU/kg. This would suggest that a stable Hb is maintained throughout, though possibly a little higher than target. The predicted dose suggests that the dose should be increased for the patient each month, which could imply that Hb is reducing, at a steady pace. In fact Hb begins quite low, then reaches target at month 3 where it remains for three months, it then oscillates around 9 g/dL and 10 g/dL. It seems that the predicted dose is trying to increase Hb to the 11.8 g/dL target value, but as Hb is not responding to the dose given, then more and more dose is required.

Figure 7.25 shows an example where the true and predicted doses are almost identical. In both cases there is slight oscillation of dose around 120 IU/kg, which was shown to be the required dose to achieve Hb = 11.8 g/dL, in figure 7.20. This is reflected in the Hb, which is almost on target at all time points. It has been suggested that for stable Hb patients a relatively low dose of epoetin should be administered, this would improve the efficiency of the treatment

In summary, the clinical algorithm is different from the full model due to:

- no correlation structure of the residuals is incorporated
- only a random intercept is modelled (no random slope)
- based on EB data only
- based on dose_{i-2} rather than $\text{dose}_{i-2.3}$ or $\text{dose}_{i-2.4}$

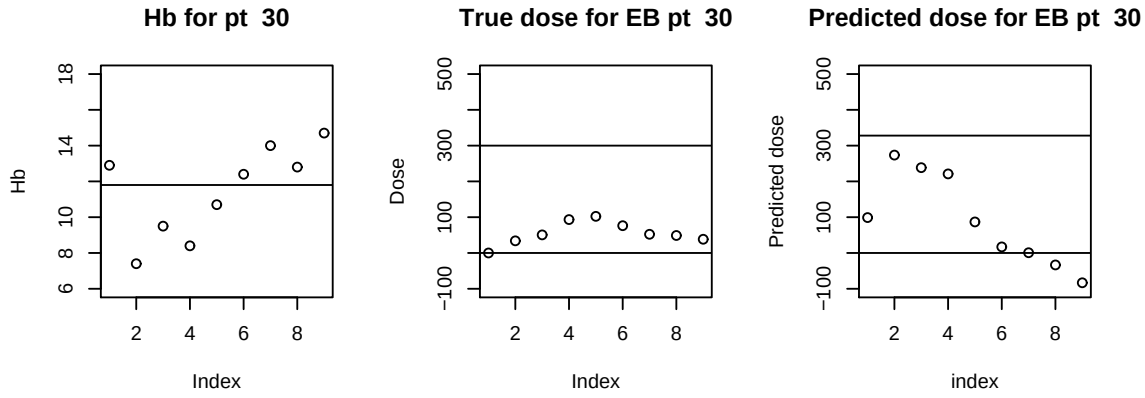

Figure 7.23: Plots of Hb, true dose and predicted dose for EB patient 30

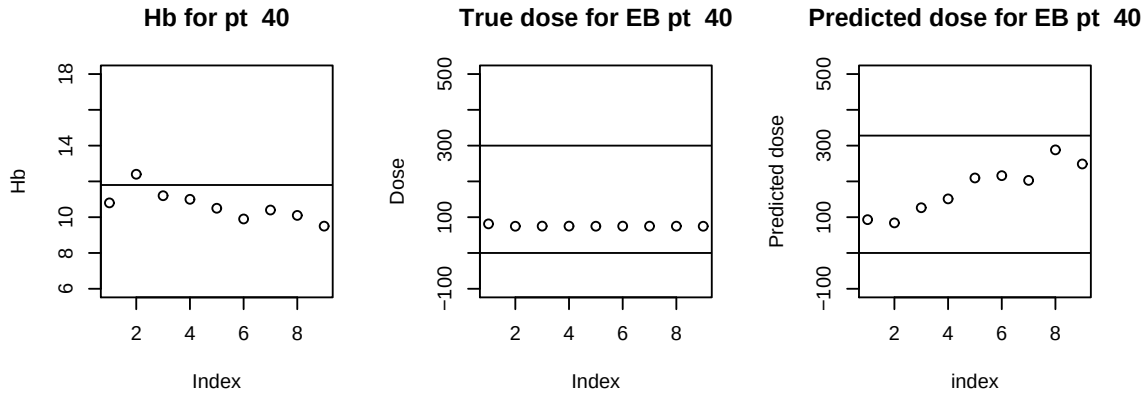

Figure 7.24: Plots of Hb, true dose and predicted dose for EB patient 40

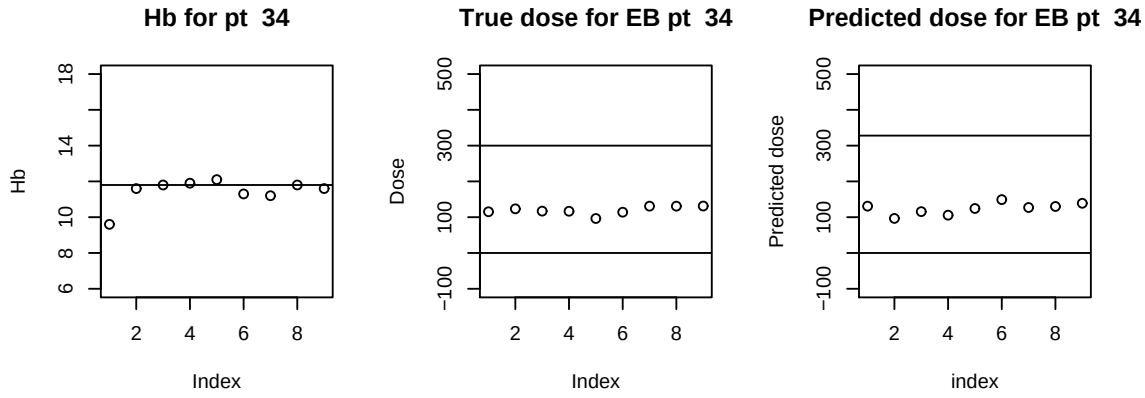

Figure 7.25: Plots of Hb, true dose and predicted dose for EB patient 34

Chapter 8

Discussion

8.1 Overview of chapter

The first part of this chapter will consider the aims of this thesis. The findings about the biomedical systems will then be discussed in detail; in terms of the important findings from the applications of functional data analysis and multilevel modelling. Following this the methodologies will then be discussed, in particular how the methods have developed from being applied to the two biomedical systems and also the limitations which arose. Discussion will then follow into the purely clinical aspects of the work. The chapter will conclude with a comparison of the two statistical methodologies and general discussion of the work.

8.2 Aim of research

The motivation for this research was to develop the unexploited statistical techniques of functional data analysis and multilevel modelling. There was particular interest in their novel application to diverse biomedical systems where repeated measurements were

collected. A further aim was to assess 'control' within the systems. Control was defined as the process of managing the patients by a particular means, such as surgery or course of treatment, and determining how to adjust this process so the patients achieve a 'good' quality of life. Examples of biomedical systems were from the fields of nephrology (renal anaemia management) where there was opportunity to refine the system to improve patient management, and stroke (cerebral autoregulation during carotid endarterectomy) where the aim was to find a superior method for assessing cerebral autoregulation. Even though there is the similarity of assessing control within the systems, the nature of the assessment is very different. Within the renal anaemia system the aim was to improve patient control, whereas assessment of control was the primary aim in the cerebral autoregulation system with the aim of improving patient safety.

A number of statistical challenges were raised by application to the biomedical systems, which included the clustering of repeated measurements within patients and the high degree of autocorrelation amongst repeated measures. These issues required resolution to enable the specific clinical questions posed about the systems to be answered successfully. This research demonstrates that through development of the statistical methodology it was possible to answer clinical questions and in addition to gain an improved understanding of the biomedical systems. The analysis undertaken also provided extra insight about the biomedical systems which provoked further inquiry, which could be addressed by the methods.

8.3 Cerebral Autoregulation

The biomedical system of cerebral autoregulation is self-regulatory, whereby mechanisms within the body ensure a constant blood flow to the brain is maintained over a range of arterial blood pressure. The aim of this work was to determine whether this natural mechanism is exhibited in patients undergoing carotid endarterectomy; surgery which

could alter blood flow (increase or decrease within the carotid artery) and so alter the relationship between perfusion of arterial blood pressure. This situation could potentially influence the entire biomedical system. In addition to actually determining whether any degree of CA is seen in these patients during surgery, it was investigated whether carotid endarterectomy directly influenced CA and hence caused it to vary.

The specific aim of the research within this system was to determine whether it was possible to assess autoregulation in an operating theatre setting, as this had not been reported in the literature. CA had previously been assessed in various groups (of patients) and scenarios. The techniques used, however, were deemed inappropriate for this patient group and the repeated measures data-set they generated. Statistical methodologies were sought to address this issue. It was necessary to determine if it was actually possible to assess and model CA; with particular interest in whether CEA resulted in an improvement in CA. This would allow us to establish if there is any immediate patient benefit from CEA.

The initial multilevel modelling allowed us to model static autoregulation. The modelling framework was extended to permit varying autocorrelation between patients, which meant that dynamic autoregulation could also be addressed. This idea and associated analyses arose as a consequence of considering the assumptions of the MLM, such that the autocorrelation structure is the same for all patients in a multilevel model. The assessment of static and dynamic autoregulation was progressed even further to answer whether the choice of anaesthesia had an impact on CA.

8.3.1 Application of functional data analysis

The fitting of loess curves WAS a useful exploratory tool for these data, as they allowed features of the data to be extracted with only limited assumptions. This methodology was ideal as the relationship between CBF and ABP in these patients was unknown.

The relationship was also unpredictable, since the patients were undergoing a surgical procedure and patients suffered with carotid artery stenosis, a condition which can severely impair CA.

Prior to fitting loess curves to the individual patient data, a loess curve was fitted to data where ideal CA had been shown to exist (see figure 6.3). Figure 6.3 was shown to approximate successfully the CA curve (figure 2.4). This suggests that if CA were present in the individual patient data then a loess curve would reveal the ideal relationship.

We anticipate that the majority of the patients ABP measurements will be within the range of the plateau region (50-150 mmHg), since the interquartile range is 88 to 119 and the whole range is 38 to 188 (see summary statistics in table 6.2). Before loess curves were fitted to the data it was predicted that a single (possibly horizontal) slope would be revealed, as the majority of patients measurements are within the theoretical range of autoregulation. It was not thought that it would be necessary to estimate a piecewise relationship or where the CA curve is estimated at the change points, although it would have been possible as demonstrated by figure 6.3 . Since the patients are elderly and possibly suffering with comorbidities, it is possible that they are already receiving medication for blood pressure control, hence the reason for the majority of measurements being within the ideal range.

The features of the individual curves are discussed in the respective results section (Chapter 6). In general, the range of ABP and CBF of the phase 2 curve is often quite different from the phase 1 and 3 curves. This is the clamped phase of the operation, so the patient was likely to be under stress when ABP is usually higher (see figure 6.2). In the majority of patients CBF is reduced during phase 2. This is contrary to the principle of the circle of Willis, which states that CBF is not impaired if flow down one artery ceases. A possible explanation of this inconsistency is that CBF measurements are taken in the region where the artery has been clamped. There is a minority, however, such as patient A (figure 6.4), for which CBF increases. Another finding is that CBF is increased for

some patients in phase 3 in comparison to phase 1. This is a key finding; if it is truly the general case it would indicate that CEA is a factor in increasing CBF, hence being a positive outcome of the operation.

Another key finding is that the loess curves fitted to the individual patient data do not exhibit evidence of fully intact CA, i.e. no evidence of a zero slope in any phase; not even phase 1, which is prior to surgery commencing. There might be a number of possible explanations for this. Firstly, the patients are elderly, frail and undergoing a major surgical procedure. The complex mechanism of CA may not therefore be fully intact in these patients. Secondly, it is possible there is measurement error in the CBF and ABP measurements, since collection throughout surgery is difficult and may not be accurate. Thus noise in the data obscures the true underlying relationship. In the operating theatre CA is not assessed as it would be in a laboratory, for example using tilt tables. This is a new scenario for the assessment of CA, hence it cannot be expected that CA would be evident, such as in other situations. Thirdly, in the surgical scenario it may not be possible to experience the full range of ABP to model the whole CA curve. It would not be ethical to induce blood pressure changes to acquire a larger range, as that would not be safe for the patients undergoing surgery. Thus fitting the loess curves over a small range of ABP, may be the reason that 'typical' behaviour is not exhibited. Perhaps not experiencing a wide range of ABP means it is not possible to see CA. Consequently, it is possible that a number of extraneous factors are masking the 'ideal' relationship between CBF and ABP. In particular, the measurement setting makes the relationship unclear.

Another possibility that the typical intact autoregulatory behaviour is not seen is that measurements are collected every 15 seconds. As discussed previously, it is not clear if 15 seconds signifies static CA or dynamic CA. Dynamic CA assessment occurs when there are rapid changes in CBF or ABP. Since there are no clear guidelines of the time frame that amounts to 'rapid' then it may be that 15 seconds is too long a time between measurements, or alternatively it may be an appropriate time to be assessed under the

dynamic framework. The assessment of static autoregulation is clearer, whereby 'static' amounts to a time period that allows for stabilisation between measurements, although it is unlikely that 15 seconds would result in stabilised CBF and ABP, as static CA is usually assessed with 20 minutes between measurements [144].

A disadvantage of the loess methodology is that outlying points can have high leverage, influencing the fit of the loess curve by drawing the curve towards that point; this is seen in figure 6.5 in phases 1 and 2, and 6.9 in phase 3. There are limitations also with figure 2.4, where the lower and upper limits (L_1 and L_2) are not well defined due to the curvature around the change points. Furthermore, the plateau region in figure 6.3 is not perfectly flat, which is in disagreement with the theory that perfect CA is represented by a zero slope. Loess is still useful in exploring the ABP–CBF relationship, however, as it makes no assumptions about the underlying function.

In summary, loess has shown:

- To be able to fit the ideal CA curve, albeit with some limitations/ caveats
- The relationship between CBF and ABP varies greatly between phases and patients
- The slope of the CBF – ABP relationship varies between patients
- The slope of the CBF – ABP relationship appears to approximately linear
- In many patients, there is an improvement in CBF in phase 3 compared to phase 1
- Fully intact CA is not detected amongst these patients
- Analysis is restricted as patients are explored individually.

8.3.2 Application of multilevel modelling

Following the loess analysis, MLM was subsequently applied. There were a number of reasons why MLM was deemed an appropriate technique to analyse these data:

- More sophisticated analysis is required as loess only provides an exploratory (albeit useful) approach.
- The data comprises a hierarchical structure: multiple repeated measurements are nested within patients. There will be greater power and insight from the modelling if all patients are considered concurrently, which will be possible using MLM.
- The relationship between CBF and ABP was approximately linear with a positive gradient. The assumption that the CBF–ABP relationship is linear originates from previous analysis undertaken where a correlation coefficient is calculated between the two variables.
- There were three distinct phases which may be modelled by three random intercepts, and 3 random slopes.
- To address the main clinical aim of the analysis, since within the MLM framework it is possible to test whether the slope differs between the three phases, providing a formal statistical test of the hypothesis that CA has changed (or not changed, for the null hypothesis).

A particular challenge in the construction of these multilevel models was in deciding whether phase of CEA should be included as a random effect or as an additional level of the model (between patient and repeated measurements), or even possibly both. It was not appropriate to include phase as a fixed effect for each patient, as this would yield too many parameters. Including phase as a random effect and additional level would also yield a model with too many parameters, as the phase effect would unnecessarily be included twice. The decision was made to incorporate phase into the model as a random effect, which would yield a more parsimonious model. The process of investigating whether random intercepts and slopes are actually required will be discussed later in this section.

The reason for choosing to include phase as a random effect as opposed to an another level is that phase as a third level resulted in a model that was more difficult to interpret.

A three level model would have yielded a phase level variance component. A statistically significant positive, non-zero, phase variance would then indicate variation in slope and formally test the main hypothesis: does CEA alter CA? The complexity of the third level, however, is greater than the inclusion of a random slope in the two level models. Hence the principle of parsimony dictates the two level random slope approach. Within the three level models there is a substantial variance – covariance matrix to determine. In particular, there were difficulties in the implementation of the three level models due to problems of convergence in R. A reason for this may be that a level with only three groups is too limited, as large samples are preferred to increase the precision of the parameter estimates.

The preferred MLM required a second order autoregressive correlation structure to account for autocorrelation amongst the residuals. A variety of autoregressive models were tried, including AR(1) and also higher orders, as well as ARMA models. The model with AR(2) yielded the best fit to the data. In this context, model fit was judged in terms of reducing AIC and also greatly reducing the range of the residuals. The model fitting procedure was repeated with the correlation structures and resulted in the same model being preferred with the correlation structure as well as without. The correlation structure was necessary as there was more correlation in the model than what could be accounted for by a traditional MLM, due to the autocorrelation within the CBF and ABP measurements. MLM with autoregressive correlation structures provides a novel assessment of static cerebral autoregulation, and also by extending the model this leads to assessment of dynamic autoregulation (which will be discussed in the following section).

Not only did this modelling approach yield novel research in the clinical domain, in statistical terms it is important to note that the correlation structure greatly improved model fit. In the model with AR(2) correlation structure the fixed effects coefficients were larger than in the model without. This shows that the effect size is decreased if the correlated residuals are not accounted for appropriately. The standard deviations for the random effects are also different in the two models. In the model with correlation

structure the standard deviations for the phase intercepts increase throughout the phases, highlighting that variation is more homogeneous at baseline and as time develops the extent of variation amongst the patients increases. This is expected as at the beginning of surgery patients are more similar and their CBF will be stable, as surgery progresses patients respond differently and hence the variation between the behaviour of their CBF is larger. In particular at the end of surgery the standard deviation is largest; as some patients may experience a dramatic increase in CBF after the occlusion in the carotid artery is removed, whilst other patients may experience a more modest increase in CBF.

It may be perceived as a problem that the variation/range of residuals increase as the phases increase, as it is assumed that the variances (in each phase) are homogeneous. This was accounted for as much as possible, by including phase explicitly as a random effect, and only became apparent after including the AR(2) correlation structure. In the model where no correlation structure was included, the phase variances were relatively similar with no obvious pattern. Once the modelling became more sophisticated, it was then revealed that the variation increased throughout the phases. The proportional increase in residual variance, although clear, is insufficient to cause major concern: that is less than a two fold increase. Additional terms could have been included for heteroskedasticity, though would have resulted in a model with more parameters, further complexity, and which would have been more difficult to interpret. It seems that the more sophisticated and complex the modelling procedure is, the more errors within the model are revealed. The preferred model may be presented with small caveats, although a major issue with the model was resolved by accounting for the dependency amongst the residuals with the autoregressive correlation structure.

The differences across phases were not constant within patients, which was evident from the loess curves, whereby the difference between phases 1 and 2 varies between patients. This is also evident from the MLM framework by allowing a random effect for phase intercept. The inclusion of random effects in the model was justified by

investigating a variety of models and showing model superiority (in terms of model fit) using likelihood ratio tests and by inspection of model residuals; the inclusion of random effects consistently provided a substantially better fit to the data.

Furthermore, the relationship between CBF and ABP is not predictable due to the large variability in the slope, which translates as much variation in CA between patients. There is also large variability in the range of ABP between patients and also between phases within patients. There are some patients who experience a small range of ABP values; in these cases the slope estimate will be less accurate than for ABP covering a larger range. The standard error of the slope will be much larger where a smaller range of values occurs. The gradient of the slope ($= 0.14$ cs.s^{-1} mmHg^{-1}) within patients was fixed across the three phases. The slope is, however, allowed to vary across patients. This was one of the main findings of the MLM: a non-statistically significant random phase slope suggests that there is no change in CA immediately following surgery. A slope of 0.14 suggests that, on the whole, CA is impaired in these patients, as it is different from zero. The value is statistically significantly different from zero, though it is unknown whether 0.14 is clinically significant. Views on this will be discussed later in this chapter.

8.3.3 Individual models

An assumption of the multilevel modelling is that the correlation parameters are the same for all patients. In the MLM framework it is not yet possible to readily allow this to vary between patients, i.e. by incorporating the correlation structure into the random specification of the model. Therefore in model 6.4 we assume that the correlation parameters 0.65 and 0.31 are a reasonable estimate for all patients.

Using the `gls` function in R, it is possible to fit individual patient models which incorporates complex correlation structures. This allows the correlation parameter to vary between patients. It is also not possible to fit individual patient models using `lme`, as

this requires a grouping structure to be specified. The `gls` function is within the `nlme` package, as is the `lme` command.

Models were initially fit with `gls` which included an AR(2) correlation structure. The AR(2) correlation structure was used as this yielded the most appropriate model in the MLM framework, so may have also been the 'best fitting' with `gls`. Not all models could be fit with AR(2), however, as some were numerically unstable. Models were subsequently fit with AR(1) correlation structure, which fit successfully for all patients. It seems that the AR(2) correlation structure was more appropriate in the MLM framework as those models were fit with a large amount of data, hence a more complex correlation structure could be fit. In the individual models there is less data to fit the models, hence a simpler structure was more appropriate and the AR(1) correlation coefficient is simple to interpret. This also suggests that an AR(1) correlation structure is more robust and is applicable in more circumstances than higher order autoregressive models.

Within these models the slope will represent static autoregulation which is analogous to the MLM. In the `gls` modelling scenario the AR(1) correlation parameter may be used as a proxy, and/or quantification, for dynamic autoregulation as this will indicate the strength of the relationship between measurements in a short time period. This is similar to dynamic autoregulation which incorporates a time element into the ABP-CBF relationship. Since measurements were collected at 15 second intervals the correlation parameter will reveal the strength of the relationship between adjacent measurements, thus representing CA in a short time frame. This varies from the static parameter which investigates the relationship between ABP and CBF across the whole time frame. The larger the dynamic parameter the stronger the relationship within the time frame. If the time frame were different from 15 seconds, then it may be possible to assume that the parameter would increase if the time frame was shortened and decreased if the time frame was longer. In contrast, 15 seconds may be too long between measurements to be considered dynamic CA and hence the correlation parameter would not provide an

accurate representation of dynamic CA.

Models were fit representing three different scenarios. Firstly, where the static and dynamic parameters varied between patients, but not between phases. These models included a random intercept for each phase. The primary interest, however, was in the slope and correlation structure, though the random intercepts are required to reliably estimate the slope and correlation coefficients. The second scenario allowed static autoregulation to vary between phases (i.e. random phase slopes), though the dynamic parameter remained constant for all phases. Note that only one model was required for these estimated parameters, whereas the third situation requires three different models to be fit for each patient. In this third instance the static and dynamic parameters varied between phases. These models were assessed by examining the plots of dynamic vs. static, as it would not be reliable to assess the models based on AIC as the sample size varies between phases, hence the AIC's are not directly comparable.

According to the plots of static CA against dynamic CA, it seemed that the first scenario best represented the data, which ascertains that CEA does not alter static autoregulation during surgery. It also shows that static autoregulation varies in degree between patients. This is consistent with the MLM analysis, such that the patient slope does not vary between phases, again emphasizing there is not sufficient evidence to suggest there is a difference in static autoregulation between phase 1 and phase 3 of CEA. The general relationship between dynamic and static, evident from the scatter plot (figure 6.20), is that the dynamic parameter increases as the static parameter increases. A static parameter larger than zero indicates that CA is not intact. A larger dynamic parameter indicates that there is stronger correlation between adjacent measurements. This means that high AR(1) correlation resulting in a large dynamic parameter, indicates less short timescale regulation of CBF: that is poorer dynamic autoregulation. Thus there is evidence to suggest that dynamic and static CA are impaired 'together'.

A limitation of this model is that for a small number of patients their static parameters are

estimated as negative. A negative slope for the ABP–CBF relationship would suggest that as ABP increases then CBF decreases, which seems clinically implausible. A possibility for this result is that the measurements are clustered close together and in fact there is no obvious slope within the data. It may be the case that a negative slope exists as a result of a patients CBF and ABP measurements being clustered together. This argument is further strengthened by the fact that the negative slopes are relatively small. Thus in the cases where patient's experience a negative slope, models were refitted, restricting the value of the slope to zero. This model was chosen above all others because the static parameter varying between phases is not consistent with the findings from the MLM, which takes into account information from all patients. Also in the models where the dynamic parameter varies between phases there does not appear to be a clear relationship between the static and dynamic parameters. This is indicative from the points being randomly scattered and there being no association between the two parameters, unlike the first scenario where there is a clearer increasing trend.

8.3.4 Effect of anaesthesia

Following the MLM and `gls` analysis and the subsequent investigation of the patients' static and dynamic parameters, the question was asked whether the static and dynamic parameters vary between patients undergoing CEA under local or general anaesthesia. Even though some of the patient data were collected from those involved in the GALA trial, it was not initially a primary goal of the analysis to investigate this hypothesis. A relatively simple approach was taken to investigate the hypothesis. Boxplots were constructed of the static parameter from the MLM (model 6.4) by anaesthesia type and also of the static and dynamic parameters (from model 6.5 with negative static parameters constrained to zero) by anaesthesia type. Formal t-tests were also carried out to investigate if there was a difference in the mean values between local and general anaesthesia.

In general (from model 6.4) the slope was found to be higher in those patients who

were under general anaesthetic during CEA, suggesting that static autoregulation is better preserved under local anaesthesia. The same conclusion can be made about the static parameter from the individual `gls` models. There is no statistical significant difference between local and general anaesthesia in the dynamic parameter. This suggests that it is not necessary to model different correlation structures for each patient and that the MLM assumption that the correlation structure is the same for all patients is reasonable.

It is a very important finding, in the clinical domain, that static CA varies due to anaesthesia type. Patients undergoing CEA under general anaesthesia were more likely to experience impaired CA than those under local anaesthesia. This is particularly important in terms of patient care as it may be the case that LA is safer. It may not be accurate, however, to infer this as the patients were not randomised to anaesthesia type. It may be more reasonable to suggest that patients undergoing CEA under general anaesthesia are those whose CA is more impaired prior to the surgical procedure. In order to draw unbiased conclusions about this hypothesis, analysis should be carried out on data where patients have truly been randomised to anaesthesia type.

In terms of ABP, this is higher in patients under local anaesthesia than general anaesthesia, which is consistent across all three phases. This is consistent with other research which has highlighted that carotid endarterectomy may be associated with marked changes in blood pressure in the perioperative period [9]. Blood pressure generally falls after induction of general anaesthesia. There is also an increase in ABP in phase 2 (compared with phase 1). Some patients are also thought to show signs of hypotension or hypertension in phase 3. In these patients, ABP appears more controlled in those undergoing general anaesthesia, particularly in phase 3. The range of ABP is much greater in the local anaesthesia patients particularly in phase 3, suggesting that ABP is less controlled. In terms of CBF there is no difference due to anaesthesia type, in any of the three phases.

8.3.5 Summary - Cerebral Autoregulation

The analysis of the data from the CA biomedical system has shown that:

- CEA is successful in immediately increasing the patients CBF from the level prior to the surgical procedure.
- CBF is reduced in phase 2, compared to phase 1. This emphasises the need to monitor CBF during surgery.
- Variation in CBF increases throughout CEA, suggesting that patients may be initially more homogeneous but tend to respond differently to surgery.
- There is no notable change in ABP due to CEA (albeit a slight increase in phase 2). This may be due to careful monitoring on behalf of the clinician or alternatively CEA does not affect ABP.
- A key finding of the MLM analysis is that there is no significant change in slope between phases of CEA, suggesting that CEA does not immediately effect (improve) CA.
- The slope does vary between patients, indicating that patients experience different degrees of CA.
- MLM with an AR(2) correlation structure has provided a novel way of assessing static CA. The models fitted using `gls` also allowed the frequent recording of measurement to be modelled explicitly for each patient, thus allowing modelling of dynamic CA.
- Patients undergoing CEA under general anaesthesia were found to experience larger slopes than patients under local anaesthesia. This could mean that CA is better preserved in local anaesthesia.

- In terms of implementation for clinical practice (of the above point): LA may be safer for patients undergoing CEA.
- MLM was well suited for the data and application in the intraoperative setting, where other methods would be difficult to apply.

8.4 Renal anaemia

The renal anaemia system is driven by the administration of epoetin agents to replace a naturally occurring hormone in the body (erythropoietin), which has declined due to failure of the kidneys. This differs from the CA system such that the renal anaemia patients are controlled externally and managed by the clinician, opposed to a natural biological mechanism. This shows that control may be viewed in different contexts; i.e. a patient may be controlled by external drugs and clinical management or alternatively a natural intrinsic mechanism of the body.

The aim within this system is to determine whether Hb is adequately controlled in patients with renal anaemia, who receive epoetin agents where the dose is determined by a computerised decision support system. In order to consider this aim it is necessary to determine whether Hb cycling occurs in these patients, since (large amplitude) Hb cycling indicates that a patients Hb is poorly controlled.

FDA revealed that Hb cycling did occur in some of the patients. The subsequent step was to identify possible sources of cycling and suggest ways for it to be reduced, which was undertaken in the MLM framework. There was also clinical interest in comparing the two agents in regards to control. It was important to adapt an approach to this analysis which considered patients as individuals, as the same mechanism for control may not be appropriate for all patients.

8.4.1 Application of functional data analysis

There are a number of features (listed below) of the FDA methodology which make it an appropriate tool for analysing the data within the renal anaemia system, such as:

- To visually represent the data.
- To investigate the patients Hb and dose trajectory over time.
- To assess patients on an individual level by fitting individual B-spline curves and phase plots.
- Phase plots may be a useful mechanism for modelling control.

There were a number of factors to overcome before the functions could be fitted to the data. In particular it was vitally important to choose the most appropriate basis function. Prior to the analysis it was unknown that some patients would exhibit evidence of Hb cycling, hence a basis function that assumes the data are periodic (such as Fourier series) would have been inappropriate. Furthermore, it is more credible that the Hb cycling was not imposed by the modelling, but rather arose from the data. Wavelets would not have been suitable as the data are not discontinuous or rapidly changing. Polynomial bases were not used either, as these are relatively simple and a more complex function was required. Furthermore, the derivatives (of the polynomial function) are not satisfactory estimates of the true derivative, due to rapid localisation which is typical of high order polynomials.

Figures 7.3 to 7.7 show a representative sample of the patients in this study and highlight examples of different characteristics exhibited by the patients. A patient whose Hb level is oscillating is an indication that Hb is not under control, even if the oscillations are around the target value (i.e figure 7.3). The reason that a patient's Hb comes under control after a long time period might be due to an adjustment to the treatment regime. Since the

patients Hb becomes stable then the regime is successful for this patient. The inward spiral of the phase plot illustrates that the patients' Hb is coming under control. After initially spiralling in, the phase circle is relatively large but centred around the target value. This implies there is still scope for improvement before the patient is considered to be optimally under control. A patient who is under control (figure 7.6) will have a Hb trajectory that shows small oscillations around the target Hb level. Their phase circle will be relatively small and tight around the target value. The dose for a stable patient will also be usually stable, as in this example. This patient represents someone who is under control for the period of investigation. Figure 7.4 represents a patient who begins on target, and then their Hb is out of control, but is brought under control to the target at the last month of the trial. This patients dose varies greatly, presumably to bring Hb back under control. The phase plot in this instance shows a tight spiral at first and then a large spiral outwards, before coming back to target and under control. The B-spline fits the Hb measurements very well, implying that the spiralling out of the phase plot is not an artefact of a poor fitting curve. Hence, this is a real representation that the patient is not under control. This patient was then investigated further and found to be suffering with adverse health, consequently affecting Hb, which in turn is increasing the risk of mortality due to oscillating Hb. Figure 7.5 is used to represent a patient whose Hb experiences large cycles around the target value. This patient is at risk of mortality due to oscillating Hb even though it is around the target.

Figure 7.7 is used as an example to show both a limitation and advantage of fitting B-spline curves to data. The limitation is that the B-spline curve does not pass through all of the individual measurements, but passes through the middle of the data. One possible view for this is that the curve is actually modelling the underlying behaviour of Hb over time and that it is this curve that is more reflective of the true relationship rather than the measurement itself. This is because the individual measurements are taken at 'snap shots' in time and these 'snap shots' may not be true due to biological or measurement error.

It is important to note that phase plots are not only useful in mathematical terms for representing control but also in clinical terms as they may be used as a measure of patient stability and control. They may also be a predictive tool to determine if the patient is at either end of the spectrum; either at increased risk of death or stable and therefore suitable for transplant.

Mean functions

The purpose of the original data collection [99] was to compare patients' response to the two epoetin agents (EB and DA). It was not an original aim of the thesis to compare the agents, however, but rather to demonstrate that the same aim could be considered in the FDA framework. This was achieved by comparing the mean functions of the two groups, by plotting them with their 68% confidence intervals on the same axes. The 68 % confidence intervals indicate the region of the graph where the trajectory is within 1.0 standard error of the respective group mean trajectory. Consequently, where the limits fail to overlap, the two mean trajectories are separated by at least 2 standard errors, 1 standard error from each group. Such a plot therefore indicates where the mean trajectories differed significantly from one another at the 5% level from. This is not a formal test to compare the two mean functions, the 68 % confidence limits are merely a visual guide. In general, further theory needs to be developed for hypothesis testing with FDA.

Figure 7.8 shows that there is a slight difference between the curves initially, with mean Hb for the DA group being slightly higher than mean Hb in the EB group. Patients were initially randomised to the agent, so the slight difference between the groups is due to random measurement variation. The mean functions then diverge to the extent that the confidence intervals no longer overlap, indicating a statistically significant difference between the two groups. From month nine onwards the mean functions converge, until they reach similar levels at month twelve. The mean Hb in the DA group is higher than the ideal target value of 11.8 g/dL, although remains within the optimum clinical limits

of (10.0 - 12.5 g/dL), though consistently above 12 g/dL until month 10. The mean in the EB group starts approximately at the optimum target, though is subsequently is below 11.8 g/dL, yet remains within the target range.

In terms of managing renal anaemia, the CDSS seems to perform slightly better for the EB group than the DA group. A possible explanation for this is that the conversion ratio 200:1 (EB:DA) used to convert the standard EB dose to DA, yields a too high dose for the DA group. Similarly, shifting from thrice weekly to once weekly doses of epoetin, is not sufficient or as simple as merely trebling the dose per injection. Another possibly is that using the same algorithm for the DA agent as the EB agent may be wrong and slightly different dosing strategies may be required for each agent. Note that there is no statistical difference between the two groups initially, as the patients were randomised to two groups, or finally once the CDSS protocol had time to bring patients under control. The mean functions do not show that the agents are equivalent, rather where they are different. The mean function was a useful comparative mechanism; yielding the same conclusion as the findings by Tolman et al [99]. In conclusion, EB and DA have comparable control of HB levels for patients with renal anaemia.

Functional principal components analysis was applied to these data to determine the sources of variation around the mean functions. The idea for this application was that it could have revealed similarities between patients (in their Hb trajectories) and identify groups of patients. This analysis did not prove useful, however, as it did not yield any additional insight into the patients or epoetin management regime.

Summary - FDA

In summary FDA has shown:

- B-splines were effective in fitting curves to Hb measurements over time. This allowed for Hb to be represented as the continuous trajectory that it is, opposed

to single monthly measurements.

- Phase plots were valuable in identifying the stability of the patients Hb measurements. They also provided insight of whether the patient is under control or not. This provides information on the success of the CDSS in managing Hb levels, through the administration of epoetin on a weekly basis.
- The mean functions show that DA and EB are comparable in managing Hb in patients with renal anaemia.
- This is an effective and useful technique in modelling these data, which has a temporal pattern. This suggests that FDA is a particularly effective tool in modelling data which are autocorrelated.
- The graphical representations are clear and allow for an efficient portrayal of control.

8.4.2 Application of multilevel modelling

Functional data analysis revealed that Hb oscillations occur in a number of these patients. The current algorithm considers that epoetin dose administered at month i is predicted from Hb at month i, which is believed to be the result of epoetin at month $i-1$. Since Hb cycling occurs in some of these patients, this suggests that Hb at month i may be based on the wrong past epoetin dose and in fact a longer time lag is required to elapse for epoetin to have its optimum effect on Hb. The FDA approach essentially investigated Hb and dose as individual entities in time. It was possible to compare the Hb and dose functions to determine how they were related. The MLM approach allows the complex relationship between Hb and dose to be investigated by exploring the variables together rather than how they vary in time.

The initial approach investigated Hb(i) and dose(i-1), as this is the relationship that the CDSS is based on. If this was a well fitting model, then it would have suggested that the CDSS is using an appropriate algorithm. A possibility then would have been that the Hb cycling is the result of other factors and not the wrong dosing strategy.

An important feature of the epoetin dosing algorithm is that it should be tailored to the individual, as each patient's Hb (and their past Hb measurements) will vary, and also in their response to epoetin. This can be incorporated into the MLM with random effects. A random intercept will reflect that each patient's Hb varies (for a zero dose) and a random slope accounts for varying Hb-dose relationships between patients. It was predicted that a random slope and intercept would be required. Models were fitted with this in mind but also considering the approach of choosing the 'best' model which was outlined in the MLM chapter.

Another possibility was that a nonlinear effect would be required: firstly, due to the asymmetry of the dose ladder; secondly, we anticipated that as dose increased Hb would also increase, although it is unlikely that this is purely linear, as the increase would not continue indefinitely. Using prior clinical knowledge it is known that a dose higher than 300 IU/kg/week is the maximum, and above this would not have an additional significant effect on Hb. It seems therefore that for larger doses the effect on Hb will level off, though increase slightly. The initial approach though was to model this effect with a quadratic term. Furthermore, it was suspected that there would be correlation between measurements within the same patient (resulting in correlated residuals within the model) due to longevity of red blood cells and other biological mechanisms. To overcome this autocorrelation, a correlation structure can be incorporated, specifically autoregressive.

A selection of models were fitted, shown in table 7.4, which included a variety of random effects, quadratic terms and correlation structures. The best fitting model (decided by the outlined scheme) with dose(i-1) as the covariate was: `lme(hb(i) ~ dose(i-1), random = dose(i-1)\ patient)`. The model, however, did not include the

expected quadratic term. The residuals of this model were also found to be autocorrelated, thus was extended to include a correlation structure. This model also had problems as it yielded a non-positive-definite covariance matrix, which means that the model is unstable and possibly incorrectly specified.

The combination of these features suggest that models with Hb(i) regressed on dose(i-1) are not an accurate representation of the data. Work undertaken by Gupta and Beserab [145] found that the most appropriate relationship was between current Hb and dose administered one month previously. This finding was based on large intravenous administered epoetin, with the one month being defined as first response in Hb. The most relevant lag should rather be between subcutaneous administration and peak Hb response whilst under steady dosing, which may be different from administering patients large impulse doses. Other work [86] [69] suggests that the optimum response time is 2 to 6 weeks. The reason the CDSS was based on 1 month is that this lies mid way through this time interval, though it could be anytime within this interval (or longer). It has been suggested that the CDSS is not updated immediately, in terms of the next dose to be administered calculated from the blood sample. It may be up to 8 weeks later before the dose is adjusted (personal communication - Dr. E.J. Will). This further suggests the reason the Hb(i) – dose(i-1) relationship is erroneous.

Two approaches were taken to investigate Hb(i) and its relationship to other dose lags. The first approach was to include adjacent doses (i.e dose(i-1) and dose(i-2)). The reasoning for this was that the effect on Hb may be from a combination of doses. This approach, however, was deemed inappropriate due to the high correlation between adjacent doses. If included together in the model this would introduce collinearity, which is associated with biased coefficients and inflated standard errors. The second approach was to include a linear combination of adjacent doses. This would remove collinearity as both variables are not included individually but are replaced by a latent variable representing dose. Initial models were fitted with the mean of dose(i-2) and dose(i-3). The decision was made not to

include dose(i-1) as this resulted in unstable models. Models were also fitted that included the mean and difference between dose(i-2) and dose(i-3). The difference did not improve model fit, nor were the coefficients statistically significant. The best fitting model included Hb(i) as the response and dose(i-2.4) as the covariate, see equation 7.3. This suggests that the optimum response in current Hb is from dose administered 10 weeks previously. The model also required an additional correlation structure as the models without yielded correlated normalised residuals. An autoregressive correlation structure of order 1 best improved the model fit (see table 7.9).

Model 7.3 also includes a random intercept and slope, a random quadratic term did not improve model fit and the term was not statistically significant. Figure 7.13 of the random quadratic model shows this is implausible as a larger number of the patients maximum values are for low doses (approximately 100 IU/kg), thus suggesting after they receive their maximum Hb will decrease. This is extremely unlikely and hence the model is not an accurate representation of the data. In model 7.3 the standard deviation of the random intercept is approximately 1 unit, thus we would expect 95 % of patients' Hb to be within the range 8.6 to 12.4 g/dL for a zero dose. The random slope suggests that the relationship between Hb and the dose lag at 2.4 months ($\equiv$ 10 weeks) varies significantly between patients. This means that some patients will have a steeper slope and their Hb will be more responsive to the dose, hence smaller doses may be required. Conversely there are those patients with a shallower slope, whereby these patients are less responsive, so may require a larger dose for a small change in Hb. It is of course possible that the 'best' lag varies between patients, although within the MLM framework it was not possible to account for this. It is reasonable to use the same lag for all patients, as this appears to be the best procedure at present. It would be difficult to implement clinically if each patient required their dose adjusting at different time periods, in particular there would be large possibility for error if the lag times became disorganised.

The correlation parameter is 0.66 suggesting that there is reasonably strong correlation

between adjacent residuals. We also saw previously that there was strong correlation coefficient between adjacent doses (figure 7.10). There is also strong correlation (= 0.71) between Hb measurements. Due to the nature of the system, Hb does not fluctuate greatly and the clinician aims to maintain stable doses as a fluctuating dose is more expensive.

8.4.3 Clinical algorithm

The aim of fitting the MLM was so that improvements to the CDSS could be suggested to reduce Hb cycling and hence improve patient care. The modelling resulted in a different lag time being found from that used currently, thus it was necessary to find ways to present the model to a clinical audience to highlight why the new time lag should be implemented. A more pragmatic approach was sought to express this complex model in an appropriate way.

The first step was to rearrange the model to make dose the outcome variable, however a number of difficulties arose. First, as current Hb is based on the dose administered 2.4 months previously, it would not make sense to predict this dose for the patient (as they will have already received this). It was essential therefore to rearrange the formula so current dose was made the outcome variable which is based on past Hb (and past doses also). This process was made more difficult due to the inclusion of the random effects and also the AR(2) correlation structure.

The decision was then made to refit the model without the correlation structure. Even though the residuals of this model would be correlated the model coefficients are very similar in the model without the correlation structure compared with the model with AR(1) correlation structure. The clinical model also only includes a random intercept and not a random slope as this parameter was particularly small and most of the variation between patients can be explained by the random intercept rather than in the relationship between Hb and dose. Furthermore, the clinical model is only based on the EB data as

this is the drug the patients were already being administered prior to the RCT. The dose lag in the clinical algorithm was 2 months (rather than 2.4 months). The reason for this is that less data were required to formulate the simultaneous equations.

The predicted dose is based on the patients' Hb from the previous two months and the EB dose from the months 3 and 4 months previously. This algorithm can only be used from month 4 onwards, which is a limitation to this new algorithm. This means that another method for choosing the patient's dose should be used in the patients' first 4 months of treatment. This is a vital period for the patient as they are being introduced to epoetin therapy, and since they are not currently being administered epoetin it may be that their Hb levels are relatively low. Therefore the algorithm currently employed should be used during this period as this will change the patients' epoetin levels monthly, which may be necessary during the initial period to monitor the patients more regularly. This may be seen, therefore, as a benefit to the new algorithm as after this close monitoring period the patient will be accustomed to the epoetin therapy which could enable the longer time lag to be more effective for the patient. The patient is likely to spend longer than 4 months receiving epoetin therapy, possibly a year or more, so the patient will experience the benefit of a number of dose changes based on the new algorithm.

The algorithm only using 4 months of the patients data history could be an advantage or disadvantage. In terms of disadvantages, if a long series of data is collected on each patient it seems ineffective to use such a small selection of this data and that a better prediction could be based on more data. On the other hand it could be seen as more effective to base the prediction on less data, in particular computationally when fitting the algorithm. Also if a patient was very unstable at the beginning of their treatment but then came under control, then their past history may adversely affect their current prediction rather than if it was based on just the past 4 months of data.

A positive feature of the algorithm is that there is a limit to the maximum dose predicted (= 328 IU/kg/week). The patients where the maximum was predicted were investigated

and typically experienced low and unstable Hb. This value is only slightly higher than the stated clinical maximum, thus could be adjusted to 300 IU/kg/week. Even though the maximum in the current algorithm is 300 IU/kg/week, patients have been given doses much larger than this. Therefore, with the new algorithm, the maximum dose would be constrained. In other instances, negative doses are predicted, however these are where a patient's Hb is too high and ideally need some epoetin removing from their bodies. As this is clearly not possible, patients should receive a zero dose, though as discussed earlier, patients do not tend to receive a zero dose, but a clinically accepted minimum since a complete cessation of epoetin may cause a sudden breakdown of the new cells and hence would be harmful for the patient.

The instances described above are when the patients either require the minimum or maximum dose, for all other patients a specific dose is predicted. In Chapter 7 examples are shown which compare the true dose given to the patient and the predicted dose from the algorithm. These will not give a true reflection as the predicted value is based on the values from the old algorithm and does not account for what would have happened if the patients received the predicted dose from the new algorithm the month previously. An example is shown, however, where the patients true dose and predicted dose are very similar, though the reason for this is probably because the patients' Hb is relatively stable around the 11.8 g/dL target. The only way to determine how the new algorithm performs is to implement this in clinical practice, this will be discussed further in Chapter 9 (Future work).

Albeit with limitations, this model and the whole process of MLM has demonstrated that the methodology can be used in clinical practice, yields meaningful results and contributes knowledge to the medical domain.

8.4.4 Summary - Renal anaemia

The analysis of the data from the renal anaemia biomedical system has shown that:

- The B-spline curves indicated that Hb cycling occurs in some patients, indicating poorly controlled Hb.
- Phase plots were useful to identify the extent by which patients were controlled, identifiable from the spiralling nature of the plots.
- The two agents EB and DA are comparable in managing renal anaemia, evident from the mean functions and their confidence limits.
- FDA is an effective technique, in particular its graphical nature, though to an extent the methodology is only exploratory. For example, it was not possible to quantify the phase plot in respect of assessing control.
- MLM revealed how dose administered 10 weeks previously yields the optimum response in Hb (opposed to the 4 weeks previously anticipated).
- The complexity of the MLM advised a more pragmatic approach, which developed an update to the current clinical algorithm. The clinical algorithm was gained from a fuller model with the whole cohort of patients, in terms of knowledge of behaviour (lag) and understanding. In practice simplicity was favoured above complexity (in the model fit).

8.5 Methodological discussion

To be able to assess control within these systems, a number of statistical challenges needed to be addressed and overcome, before the analysis could be undertaken. There were some similarities between the systems which meant that there was overlap in how the issues

were tackled. This means that some aspects may be applied to assessment of control in general. Through using both FDA and MLM to analyse both biomedical systems this meant that information learnt from the application to one system fed into the knowledge and development for the second system. The systems also posed their own challenges which were addressed individually. In this instance these features are only relevant to the particular area.

In this section the methodological results will be discussed. In particular: how the methodologies were developed through their application to the two biomedical systems; how well the methodologies performed, in terms of their success and also their limitations. It will be considered what the applications brought to the methodologies. New approaches have been taken by using FDA and MLM to answer questions from the respective biomedical systems since standard analytical techniques were not appropriate (such as; ordinary least squares modelling or time series analysis) and more sophisticated methodology was required. The presentation of the methods required careful consideration for clinical implementation and presentation to non-specialist audiences.

8.5.1 Functional Data Analysis

The data within the biomedical systems under consideration were not suitable to apply traditional time–series analysis. Time–series analysis usually consists of just one series, where the aim is to investigate a variable over a long period. In the biomedical systems of interest, multiple time series would need to be considered. Furthermore, the length of time considered within the systems is relatively short, compared with a period of years where time series is more often used. Other methods were sought which would allow investigation of repeated measurements over time.

A number of features of FDA made it suitable for analysing the data from both systems.

The question arose of whether 13 observations per individual (from the renal anaemia system) would be valid for robust analysis; the answer to this question was that 13 data points per individual with 151 subjects were suitable to apply functional data analysis techniques. Once the curve is fitted to the data, the individual points are essentially discarded. As long as the curve is a reasonable representation of the data, then it is irrelevant whether the curve was fit from 13 data points or 1300 data points.

The fit of the curve was judged by screening the curves to determine whether they accurately fitted the data. As there were only a small number of patients in the biomedical systems this procedure was feasible, which may not have been the case with a larger group of patients. With a larger number of patients a more realistic approach would be to select a sample of curves to determine the suitability of the fit.

The fitting of individual curves to the data raised a number of issues, in particular the fit of the curves at the extremes was difficult to implement. This problem occurs with most data–sets before the optimum parameters are found, although it was intensified in the renal anaemia application due to the short series. This issue was resolved by placing knots at the first and last data points and increasing the smoothing parameter. This highlights a potential limitation of using FDA with relatively short series, as the smoothing parameter would be greater than hitherto considered.

These issues with the fitting of the curves were successfully overcome, which demonstrates development of the methodology. The application of FDA to the renal anaemia data shows an illustration of the method that uses relatively few values compared with 100's or even 1000's of measurements. A possibility for the successful implementation of the method to the data is that the data were complete and did not contain any missing values. Furthermore, the data were balanced (i.e. the same length for each patient and with equally spaced time points) which made the fitting of the B-splines more efficient.

The phase plots were a useful way of modelling control of Hb in renal anaemia patients,

albeit with limitations. First, it was not clear how to quantify the phase plots, hence there was no numerical summary of control. It was therefore necessary to examine each phase plot individually to determine whether the patient's Hb was well or poorly controlled, with a large number of patients this is not so practical. This may not be as problematic as first considered; as in reality patients would experience one-to-one communication with a clinician, for example, at the patients dialysis session. In this scenario the clinician would be able to examine the patient's individual phase plot and determine at that moment if the patient Hb is well controlled. This suggests that the phase plots may have a practical use. A clinical trial could be implemented to investigate whether using the phase plots, to determine if the patients Hb is well or poorly controlled, would lead to an improvement in a patients condition by enabling the clinician to clearly identify control and monitor the patient more closely.

The FDA methodology has also been developed through its application to the renal anaemia system. The phase plots used by Ramsay and Silverman were of acceleration against velocity (i.e. the second derivative of the curve against the first derivative). The phase plots used to model control of Hb in the patients with renal anaemia were plots of velocity against Hb measurement (i.e. the first derivative of the curve against displacement). The plots of acceleration against velocity were initially plotted for the renal anaemia data which yielded identical conclusions about the patients as the velocity vs. Hb plot. There are a number of benefits to using the velocity vs. Hb plot, firstly the added complexity of the accelerations vs. velocity was not justified, particularly as the phase plots were for use in the clinical domain. Secondly, the B-spline curves could be of a lower order as one less derivative needed to be penalized. Finally, the added complexity of the smoothing parameter was essential, as this yielded smooth functions and derivatives were smooth. This gave clarity to the phase plots and allowed control to be clearly expressed.

Limitations of Functional Data Analysis

Phase-plots were not applied to the data from the CA biomedical system (as they were in the renal anaemia system), because the definition of CA depends on the relationship between CBF and ABP, not CBF alone. The phase plots therefore, would not have been a useful mechanism for assessing and modelling CA. Prior to the analysis the relationship between CBF and ABP in the patients was unclear, thus exploratory FDA techniques were sought, which would permit the exploration of CBF against ABP. This was possible with the loess curves. In the renal anaemia biomedical system, the relationship between Hb and epoetin dose was not clear (before the MLM analysis). The phase plots therefore, were a more appropriate tool for this biomedical system as it was necessary to determine the control of Hb alone. The relationship between Hb and dose was explored using more complex methodology in the MLM framework.

A further difficultly was encountered in the FDA analysis in the fitting of B-splines to the CA data. The onerous difficultly arose due to the different lengths of data for each patient and phase, as B-splines require a full matrix of measurements. Furthermore, it was not possible to simply fit the B-splines to each patient as single entities. One option to overcome this was to interpolate each patients measurements, which would have resulted in each phase and patient having the same length of data. This option was rejected, however, as this would have removed the important information that the length of the whole procedure and the length of each phase varies from patient to patient. It seems therefore that B-splines are better suited to balanced data sets where there are no missing values. B-splines should be used where the primary interest is in investigating the trajectory of a variable over time, whereas loess curves are an appropriate tool for exploring (graphically) the relationship between a response and explanatory variable, where the explanatory variable may be time or some other quantity. Furthermore, loess may be used for varying lengths of data.

Mean function were constructed by taking a smooth version of the smoothed (Hb)

trajectories. This could be regarded as potentially an over-smoothed function, if there is a large amount of smoothing in the functions and then including further smoothing in the mean curve. In the renal anaemia application a smoothing parameter was chosen so that the mean function was not over-smoothed and thus provided an accurate representation of the data. The smoothed trajectory, however, can be regarded as a better representation than the individual points. This is because measurements are taken at 'snap-shots' in time, whereas in reality the entities (such as Hb, CBF and ABP) are continuous. The curves therefore represent the variable across the whole time frame, rather than individual measurements. Hence, the FDA approach increases accuracy and confidence in the conclusions.

8.5.2 Multilevel Modelling

The important issues arising from the MLM analysis are discussed under the respective headings below.

Correlation Structures

Modelling with autoregressive correlation structures is an existing technique in the statistical domain, though remains relatively unexploited, to the extent that it is currently not possible to implement correlation structures for the residuals in a major multilevel modelling statistical package. Thus the aim of this work was to demonstrate that the technique is essential and to advocate that it should be more widely used, whilst indicating the pitfalls that occur if it is not used in certain instances.

Due to the nature of the measurements (that there is strong correlation autocorrelation present) the residuals of the MLM are highly correlated. This also applies in the renal anaemia system where measurements are taken a month apart. A possible reason for this occurring is that the variation in dose and Hb from month to month is not so high, which

means that there is not frequent changes between measurements. The fitting of models with autoregressive correlation structures is a major aspect of this work and where data are potentially correlated this methodology should be considered.

Traditional MLM does account for the fact that measurements within patients are more similar than measurements between patients. The difference in the measurements in the two biomedical systems considered here is that measurements within individuals are actually autocorrelated, where traditional MLM would assume they were just ‘similar’. Checking for autocorrelation in the residuals is often overlooked in the model validation procedure. It is usually the assumption of normality of residuals and random effects that are more often examined.

The autoregressive correlation structure greatly improved model fit; in particular in reducing the range of model residuals. Most importantly, however, the autoregressive correlation structure yields uncorrelated normalised residuals. This is assessed from the autocorrelation functions of the normalised residuals. The majority of autocorrelations are not significant at the 5% level. There are some, however, that are. Note that multiple testing is undertaken here which increases the chances of a false positive. Simple corrections such as Bonferroni are not appropriate since the autocorrelation estimates have a complex correlation structure themselves. This situation is usually considered to be one where judgment based on experience is the best guide. Given the few significant values scattered at different lags, and the relatively small excursion beyond the 5% limits, the ACF is taken to be satisfactory: thus evidence of lack of autocorrelation.

Complexity of models

The models fitted are rather complex due to the inclusion of the correlation structure; the question should be asked whether the use of complex models is justified, particularly when they are applied in the clinical domain. In the renal anaemia models, the models

were too complex when rearranging the expressions so the dose could be predicted. The models were subsequently tailored (and simplified) so they could be used to suggest updates to the CDSS. Without the complex modelling, and the results gained from this approach, it would not have been possible to formulate the simpler strategy. Therefore one can conclude that the complexity of the models is justified, though perhaps less directly applicable, rather it is useful to advise a more pragmatic course and yield simpler, yet more robust models.

In the CA system the complex models were necessary for important clinical findings to be revealed from the MLM models. The correlation structure was particularly important, such that it was deemed a proxy for dynamic autoregulation. In this application in particular, the modelling had complexity that would have been useful to investigate further, though it was not possible to vary the correlation structure for each patient. It was assumed therefore that the same correlation structure was applicable for all patients. A different correlation structure for each patient can be investigated where individual patient models were fitted (using the `gls` command within the `nlme` library). In these models, however, an AR(2) correlation structure was too complex for many patients, which resulted in a non-positive-definite variance-covariance matrix, though AR(1) could be fitted to all patients. This demonstrates that additional necessary complexity provides more knowledge about the patients that can be achieved within the model when all patient data is used. An advantage of the MLM is that the different lengths of data in the CA system did not result in problems when fitting the models, such as those encountered in the FDA methodology which requires balanced data.

Another finding is that the sample size of the data seems to affect which correlation structure is most appropriate. In the renal anaemia application, where the series of data for each patient is relatively short, an AR(1) correlation structure yielded the best model. In the CA application, where the series of data for each patient is longer, an AR(2) structure yielded a better fit in the MLM setting. Although, when individual models were fitted to

these data, and hence fitted to less data overall, the AR(1) provided a better fit.

In the renal anaemia application, the Hb–dose relationship was nonlinear, as was suspected before the analysis was undertaken; hence was taken into consideration when fitting the models. The best fitting model for these data included a quadratic term and this was incorporated into the `lme` model by simply specifying that dose was linear and quadratic. This could have been achieved in the nonlinear framework within the `nlme` library in R, though in the interest of interpreting the model in the clinical domain this approach was not developed. Furthermore, the quadratic term within the model yields a suitable representation of the Hb – dose relationship.

Choosing the model

A number of factors were involved when choosing the most suitable model, such that there is no specific rule that can be used to judge the fit of the models. First, clinical knowledge was used to formulate the models. Without using prior clinical knowledge, the conclusions drawn from the model may be inaccurate or medically implausible. It may be the case, however, that the results of the modelling draw new findings about the specific clinical area. If prior clinical knowledge is used to formulate model development there will generally be more confidence in the results.

In terms of judging the most appropriate model when determining which parameters should be fixed, random or included at all, the model which yields the lowest AIC will be investigated further. Initially the models were investigated to determine if the parameters were statistically significant, and likelihood ratio tests were carried out to determine whether the additional parameters resulted in an improvement in model fit.

The next step of the modelling was to investigate a variety of plots of the model residuals; in particular the autocorrelation function to check whether the normalised residuals were autocorrelated. The plot of standardised residuals vs. fitted values was useful for checking

a number of other assumptions. From this plot it was possible to determine whether the residuals were homogeneous and normally distributed, and the range of the residuals was revealed.

It is the combination of statistical and clinical viewpoints that allows a medical statistician to make reliable judgments about the models. It is important that a balance is maintained between the complexity of the model and ease of interpretation.

Limitations of Multilevel Modelling

The inclusion of the autoregressive correlation structures greatly improved model fit,; e.g. in reducing AIC and the range of the standardised residuals from (-10, 10) to (-4, 4). The range of the standardised residuals in the preferred model (-4, 4) is relatively large, however, in comparison to the ideal range (-2, 2). There may be a number of reasons for this. First, one may speculate that the model is indeed a poor fit, although it should be noted that this is 'real' (noisy) and thus one cannot expect the model to fit perfectly to every individual. Models were fitted using a limited number of covariates, as other covariates were not available, i.e. not collected. The inclusion of other covariates (such as: age of patient, additional comorbidities) may have improved the model fit and hence reduced the range of the residuals. It may be argued that the missing variable causes bias in the model, however, the decision of which variables to collect was made by the clinician. The variables collected therefore, would be the most important and clinically relevant. A possibility for the large range of residuals in the CA model is that the model fitted poorly at the extremes of the series in the three phases. One way to resolve this may have been to truncate the data in the phases, however this would reduce the number of data points in the model. This would have required discussion with the clinician to determine appropriate truncation. Alternatively, observing the surgical procedure at first hand would enable a more informed outlook on the analysis. For example, it would be possible to see how the patients react under different type of anaesthesia and also determine how distinct

the three phases really are within CEA.

An alternative way of improving model fit may have been achieved by accounting for heteroskedasticity, such that we could have accounted for the random variables having different variances. Heteroskedasticity will result in the variance estimates of the model being biased. To account for heteroskedasticity, variance functions are used to model the variance structure of the within group errors using covariates. The decision was made not to account for heteroskedasticity in the models in this thesis as this would have yielded models with additional complex parameters, thus making the model too difficult for interpretation in the clinical domain. There was no worrying pattern evident when examining the residuals of the model. Furthermore, the main complexities of the data were resolved through random and fixed effects modelling and the addition of the autoregressive correlation structure.

On numerous occasions, some models could not be fit since the covariance matrix was non-positive definite. When this issue occurred the conclusion was made that this model was not suitable and alternatives were sought. The models fitted are not perfect. In particular, the range of residuals is still larger than desired and that the phase variation increases throughout the phases for the CA models. The models were greatly improved with the correlation structures, other adjustments are unlikely to improve the model to this extent.

8.5.3 Software

The statistical software package R was used for all analysis presented within this thesis. The idea was that the statistical methods would be easier to implement if they could be applied using one software package. The methods would be more generalisable if they could be implemented in more than one package, but there are a number of advantages however for using R as the chosen package, for practical and statistical reasons. First, it

is better to be more proficient with one package than more superficial with a number of packages.

Initially the MLM analysis was carried out in MLwiN, particularly the CA analysis, which is a specialist package for MLM. When it became apparent that a correlation structure was needed for the autocorrelated residuals, other software was sought, as it is not straightforward to implement correlation structures in MLwiN. In the first edition of MLwiN, macros were available for only an AR(1) correlation structure. In the recent versions of MLwiN these macros are not yet available or failed to work This further highlights that MLM with correlation structures is an unexploited methodology. Therefore the `nlme` library was used for all mlm analysis, in particular the `lme` command.

It would be possible to undertake the analysis in other packages, though a number of packages would be required, whilst R is freely available thus making it more attractive than using costly packages such as M-plus (for MLM) or Matlab (for FDA).

There were no particular problems with the implementation of the MLM within the `nlme` library, other than when fitting a three level model. This was possible, though with a 3-level model there is a substantial variance-covariance matrix to determine. Thus, the 3-level model proved difficult to fit due to numerical fitting reasons and also in the efficiency of the program.

There were a number of issues with the `FDA` library, which was used for the functional data analysis. At the start of this PhD the `FDA` library was relatively new and contained a small number of bugs; the library was updated at the end of the first year of this PhD. In the updated version of the library a number of commands changed, which resulted in the existing code no longer working. Consequently the code was revised so the B-spline curves and phase plots could be fitted to the renal anaemia data in the new `FDA` library. The loess curves were implemented using the `loess` function, for which there was no problems.

8.6 Clinical Discussion

In this section the clinical implications will be discussed, i.e. how the statistical methodologies have yielded results that are meaningful and clinically interpretable. In this section there will be discussion of the results of applying FDA and MLM to each of the clinical applications and what these methodologies brought. Subsequently, there will be discussion of the more general clinical issues which arose.

8.6.1 Cerebral Autoregulation

The data collected for the CA biomedical system was for monitoring of patients in an intra operative setting. There were 62 patients undergoing CEA who were approached for participation in this study, however there were 26 patients who were not eligible for participation, resulting in measurements collected on 36 patients. There is the possibility of bias here due to the number of non–responders for the study. The exclusion criteria included absence of a temporal window for TCD monitoring and the presence of atrial fibrillation and other arrhythmias. It may be the case that the data included in the study are of relatively stable and routine patients, and those with complications are not included, which means that the estimation of CA was conducted in the healthy (eligible) patients.

MLM was deemed an appropriate analysis technique as other methods of assessing CA would not be suitable because of the difficultly to apply during the intra operative setting. Techniques used to examine CA such as the thigh cuff release test, the transient hyperaemic response test, drugs to manipulate ABP, and transfer function analysis are not ideal for use during carotid surgery. The transient hyperaemic response test requires manual compression of the carotid artery. It would have to be performed on the side opposite to surgery and would not be safe in patients with bilateral carotid artery disease, as it could not be guaranteed that CBF would be maintained if the flow down both carotid

arteries is obstructed. Measuring static autoregulation by giving drugs would require the administration of an agent that causes a rise in blood pressure to many patients who already had a markedly elevated blood pressure. Transfer function analysis makes the assumption that the underlying form of mean arterial pressure and cerebral blood flow (over time) is a sinusoidal wave, suggesting that ABP and CBF vary periodically over time. This assumption is suitable for CBF and ABP measured at a steady state, for example on an intensive care unit, but not for CBF and ABP during CEA. These techniques require data to be collected for a period of minutes whilst the patient is left undisturbed. This would cause significant delays during surgery which would not be acceptable whist the carotid artery is cross clamped. Furthermore, the sinusoidal decomposition used in transfer function analysis would not sufficiently represent the step changes in ABP and CBF between the phases of CEA. In this research, CA is investigated in a original way by applying multilevel modelling, with the additional complexity of a correlation structure, to cerebral blood flow and blood pressure data recorded during and after surgery.

CA was quantified as the gradient of the slope of the CBF–ABP relationship. In the preferred MLM the mean (fixed) slope was estimated as 0.14. In theory, a slope different to zero would indicate impaired CA. If this is to be accepted, then the conclusion could be made that on the whole this patient group does not exhibit intact CA (and hence is impaired). In the literature, there is no guidance which states a cut-off between intact and impaired CA; it is therefore not possible to clarify for certain whether 0.14 is impaired or intact. Furthermore, the ideal CA (zero slope) is based on a laboratory setting, there is again no guidance of ideal CA within the operative setting. This raises questions of validity in this new setting. It therefore seems that the MLM with correlation structure is a useful tool for assessing CA in the operative setting.

One must recognise that 0.14 is a slight slope and in the opinion of the collaborating clinician this value would imply intact autoregulation. The wide variation in the slope

was deemed clinically important as this would suggest that in the patients where ABP was manipulated through CEA their CBF would change. Hence, for the most susceptible patients this would result in a slope greater than 0.14, developing into impaired CA. An additional finding is that the slopes were higher in the patients undergoing CEA under general anaesthesia, suggesting that CA is better preserved in local anaesthesia. This is informative for clinical practice. If patients initially have poor CA the more appropriate choice of anaesthesia for surgery would be local anaesthesia.

Furthermore, it was anticipated that CA would improve following CEA. Previous research has shown that CA improves in the weeks and months following surgery **??**. It has not been reported, however, if there is an immediate improvement/change in CA. The application of MLM on these data has revealed that there is no significant evidence that there is a difference in CA from the beginning of CEA to the end of CEA. A possibility for this is that there has not been sufficient time of the complex mechanism for CA to adjust, especially as the patients have just undergone a surgical procedure. Ideally, to investigate this hypothesis further, follow–up data should have been collected on the patients in the hours or days following CEA, which may have allowed CA to return to the patients.

MLM has not previously been used within this field or to assess CA. This finding is for the group of patients as a whole, thus representing the behaviour in the majority of cases. This conclusion cannot be made for all patients, as it may be that there are a minority who do experience a change in CA. Within clinical practice, however, it is often the case that treatments/operations are administered to benefit the majority of patients and that there is often anomalous results which differ from the norm. The novel application of this research is evident from the novelty of the statistical methodology used and also the important clinical findings yielded from the analysis.

In the individual models, the correlation parameter was used as a proxy for dynamic autoregulation. The question should be asked whether 15 seconds between measurements amounts to immediate changes in ABP and CBF required for assessment of dynamic

CA. Again, there was no clarification in the literature of the exact time between measurements suitable for dynamic CA. Since there was a large correlation between adjacent measurements this suggests that 15 seconds is applicable. Within 15 seconds there is no evidence to suggest that measurements have had time to stabilize, which would mean static autoregulation. It has been possible to assess static CA with these data, though the measurements were taken every 15 seconds, since the relationship between CBF and ABP was explored, which does not account for the time when the measurement was taken (other than acknowledging phase of CEA).

8.6.2 Renal Anaemia

The data provided for analysis within this thesis were originally collected as part of a randomised controlled trial to compare the two epoetin agents EB and DA. The data were of excellent quality, such that for all 151 patients there were complete Hb and dose measurements for 12 months (plus an initial value). These patients were specifically chosen as they yielded complete data, though may introduce bias. The data essentially represent the patients in the middle of the spectrum, i.e. the data will not contain those who have died or those who have transplant (within the trial period), it is the patients who are regularly attending dialysis. Since there are so many patients who provide complete measurements it is likely to be the case that these patients are representative of patients with end stage renal disease. In fact the clinician was only interested in those patients with complete data, as this was the focus of the CDSS.

Another factor to consider is whether the dose conversion from EB to DA was accurate in the ratio of 200:1. It was highlighted, in the mean curves from the FDA, that initially the patients were similar but in the intermediate period of the trial there was a statistically significant difference in the mean Hb levels of the two agent groups. Following the intermediate separation, the patients mean Hb came together at the last month of the trial. The differences in Hb for the two agents were marginal for a period, albeit a detectable

difference. It seems that the DA group were slightly overtreated, the reason for which being the wrong conversion from EB to DA. In contrast, the EB group were seemingly undertreated, which may be a result of the conversion of once weekly EB from thrice weekly EB. Consequently, the doses were corrected over time by the CDSS, seen from the coalescing of the doses in the final month.

Even though there was a marginal difference in the two agent groups (for a short period) this is comparing the patient curves to the optimum Hb of 11.8 g/dL. It should be noted that the patients mean curves and confidence limits are clearly within the range 11.0 - 12.5 g/dL, which is stated as an acceptable range of Hb for renal anaemia patients. This therefore raises the question as to whether 11.8 g/dL is a realistic target value? This is the subject of much debate and there is still no clear guidelines. Table 3.2 showed that there are a number of different organizations that publish different optimum Hb guidance. It seems therefore that a target range (approximately 10 - 12.5 g/dL) is the most appropriate means of attaining an appropriate Hb for the patients, rather than a specific value.

Following the FDA analysis, it became apparent that the levels of control in the patient group varied. There were the patients who were under control for the whole 12 months, who are likely to be the patients that would be considered for kidney transplant soon after the end of the trial. The second group is the patients who are not on target initially but then come under control, which is a possibly a result of the patients adjusting to the new dosing regimen before settling. Another group is those that are under control but then go out of control, which may be due to an adverse event. The final group is those that experience fluctuating (cycling) Hb throughout the whole trial period. It is these patients who are of most concern, as there are a number of possibilities for Hb cycling; such as the patient being administered an inappropriate epoetin dose. Subsequent to the discovery that a number of patients experienced Hb cycling, analysis was under taken to investigate how current Hb relates to different epoetin lags.

The CDSS used within this trial based current Hb on the epoetin administered 4 weeks

previously, which suggests that there is 4 week lag from administration to response in Hb. The 4 week suggestion is the midpoint of the 2 to 6 week guidance from Epogen. Following MLM analysis, the relationship between current Hb and dose administered 1 month previously was found to be erroneous and a more appropriate relationship was deemed to be 10 weeks. This is much longer than the 4 week time lag, though still plausible. It has been suggested that even though the patients dose should be adjusted monthly, it may actually take a further two weeks before it is actually updated, which means that the wrong dose is being administered to the patient before it is changed. Additionally a patient may require a longer period that 4 weeks to receive a stable dose, which may allow the epoetin to work more efficiently and allow the patients to see a better response in Hb.

It would be necessary to implement a trial to investigate the 10 week lag vs. 4 week lag, although another possibility would be to incorporate an additional parameter into the model which allows the time lag to vary between patients. The second option would require complex programming to enable a model to be fitted in the MLM framework that varies the lag. Hence, there are two suggestions for further work: first to maintain all patients on the same regime though investigate the 10 week lag vs. 4 week lag; or second, devise a scheme where patients dose adjustment is based on an individual time lag.

8.7 Comparison of methods

FDA is a relatively new technique that is not yet fully established and has a more limited network of users, whereas MLM is more established with a wide range of users and applications. In particular there are a number of software packages that are primarily for MLM, that have been developed over the years. Since FDA is relatively new there is no specialised software, though is implementable in a few packages.

When fitting a MLM there are a number of ways to check the validity of the model (i.e.

AIC, LRT, plots of residuals), though with FDA there are no formal checks to assess model fit. This is because the choice is relatively arbitrary for the smoothing parameter, order of function and choice and position of knots, however as one becomes more familiar with the method, it will aid in the decision making of the 'arbitrary' choices.

In regards to the formality of the methods, the fitting of MLM yields an equation to represent the relationship between the outcome and explanatory variables. In the FDA framework there is no equation to represent the function.

The functions fitted using FDA enable the (MLM) assumption of patients being similar to be checked. Simple examination of the curves can determine whether there is an overall similarity or difference amongst the patients. FDA curves would highlight whether there is a linear relationship between the response and explanatory variables. This would be useful in graphically representing random intercepts and curves.

The similarities and differences between the two statistical methods are summarised below.

MLM

- Models can be expressed by a formal equation
- Formal tests and checks available to test model validity, i.e. residual plots, LRT's and ACF's
- Autoregressive correlation functions to account for correlated residuals (due to measurements being autocorrelated)
- Powerful technique which allows data from all individuals to contribute to model fit
- Complex modelling technique (particularly with autoregressive correlation structures)

FDA

- Functions are not specified as formal equations
- No formal test of model validity, i.e. arbitrary choice of smoothing parameters and knots
- The smooth functional representation of the data benefits from the temporally correlated measurements
- Models are fit to individuals, with limited population analysis
- Mostly an exploratory data analysis tool

8.8 General Discussion

This thesis was driven by clinical application. In relation to both biomedical systems the clinicians had a number of clinical questions that needed answering, using statistical methodology. In both situations the clinicians were aware that the methodology used previously in the respective fields was not applicable for the biomedical system and that consultation with a statistician would allow for new avenues of analysis to be explored. Initial consultations indicated that there was potential to advance the statistical modelling employed and through better modelling be able to answer fully and more appropriately clinical questions. For example, in the cerebral application the clinician believed that simply calculating a correlation coefficient between ABP and CBF is too naive to model such a complex entity as CA. Consultation with a statistician would be useful to suggest more sophisticated methodology to be used for analysis from a so–called ‘mathematical tool box’. This thesis is a novel application of multilevel modelling and functional data analysis to the biomedical systems; it was their use in other application

areas and particular features of the methodology, which have shown their potential and appositeness.

New clinical questions arose from the analysis of the initial questions. In the cerebral system the multilevel modelling revealed that it was possible to model parameters that represent static and dynamic autoregulation. In the renal anaemia system, the initial aim of the analysis was to find an appropriate methodology to graphically express control. This question was successfully addressed in the FDA domain through phase plots being used to model control. When the relationship between epoetin and Hb was modelled using MLM it was determined that the CDSS was basing current Hb on the wrong previous dose. The MLM methodology required further development to determine the time lag from administration of epoetin to response in Hb. These additional questions were clinically relevant and also allowed for further development of the statistical methodology.

In the biostatistics field there is a delicate interface between the medical domain and the statistical domain. From a pure theoretical statistical point of view, answering questions such as those posed in this thesis would be difficult for a number a reasons. First, the data used in theoretical work is not as ‘real’ as the data from a group of patients undergoing an operation for example. This would mean, therefore, that the models fitted to the data would not be ‘perfect’. In the clinical domain, it is not plausible for one to expect a perfect model and that a somewhat approximate solution is necessarily acceptable.

A further task for the biostatistician is to express complex statistics to clinicians and other non–statisticians. It is necessary for the biostatistician to learn about the context of the data they are analysing; this may be from the clinician themselves or other more traditional sources (books, Internet, etc). For someone from the statistical domain learning the technical medical details and understanding the clinician is in itself challenging. It is important therefore that both the statistician and clinician adapt their language and communication techniques to portray their work and ideas to non-specialist audiences. It was possible to use ‘pictures’ to present the two methodologies used in this thesis,

which was appreciated by clinical (and statistical) audiences. It was possible to educate the audience about the methodologies and the applications through the graphical representation of the statistics. There would be no purpose in using and developing complex statistical methods if they are not accessible. A reasonable question that may then be asked is: could we have managed with simpler methodology? In short the answer is no. The statistician should explore greater complexity, but then be able to simplify and make the methodology and results accessible to a variety of audiences.

In order to publish material in clinical journals, multidisciplinary collaboration is required between the clinician and statistician. Outputs that focus on clinical issues are best received, though ideally this requires the statistician to grasp the clinical details as well as having statistical expertise. It is also important that the clinical domain is open and accommodating to novel approaches.

Chapter 9

Conclusions and Future work

9.1 Conclusion

The application of FDA and MLM were novel approaches for assessing control within the biomedical systems. The assessment of cerebral autoregulation in patients with carotid artery stenosis and the management of renal anaemia with epoetin agents are existing issues that have been addressed previously, though with no ‘gold standard’ method of assessment. It was necessary therefore to find a suitable technique that may become this standard. Furthermore, there were a number of features of the data collection and data itself, which made the previously used methods for the respective systems unsuitable.

Within the CA system the aim was to assess CA during CEA, which had not previously been attempted. Assessment of CA was made under conditions that did not reflect a controlled laboratory setting, nor was assessment being made with fit healthy individuals. In fact patients suffered with a condition (carotid artery stenosis), which adversely affects CA, and patients were most likely ‘under stress’ due to the invasive surgical procedure. The result was that the ‘ideal’ autoregulatory behaviour was not seen in these patients. CA (static) was quantified as the slope of the relationship between CBF and ABP.

Furthermore, it was found that there was no change in CA immediately following CEA (compared with CA before CEA). This is an additional novelty of the work, such that assessment of CA is being made immediately following surgery, whereas previously this has only been done in the weeks and months after CEA. It was also revealed that the CEA was successful in increasing CBF immediately following surgery, which in the longer term (i.e. hours and days) may improve the patients' CA.

The loess curves were an ideal exploratory tool for the CA data as they revealed the underlying relationship between CBF and ABP (since the data were relatively noisy due to collection during surgery and patients not experiencing ideal CA). This exploratory analysis revealed features of the data and CBF–ABP relationship, which highlighted that MLM (with an autoregressive correlation structure), was an appropriate analysis technique for these data. The analysis of the data in the MLM framework highlighted that there were advantages to analysing the group of patients, such that this yielded a more powerful model and also the conclusions drawn could be about the patients in general. Thus, clinical practice could be advised about how the surgical procedure affected (and benefited) the majority of patients. MLM with autocorrelation structure directed further analysis, which meant that dynamic CA could also be modelled, as well as static CA. This analysis was used to determine whether there was a significant difference in dynamic and static CA due to anaesthesia type. It was found that static CA was better preserved under local anaesthesia compared with general anaesthesia, though there was no statistically significant difference in dynamic CA for the anaesthesia types. This suggests that it may be safer for patients to undergo CEA under local anaesthesia.

In the renal anaemia biomedical system, the interest was in assessing control of Hb with epoetin agents, in patients with renal anaemia. These patients were managed with a CDSS which adjusted their epoetin dose each month, by basing their current Hb on the epoetin administered one month previously. FDA of these data revealed that some patients were adequately managed by this regime, although there were some patients who experienced

Hb cycling. In the patients where Hb cycling was shown to occur, this was represented in the phase–plots by large spirals. The patients who experience Hb cycling may be experiencing an adverse event or not receiving the correct dose of epoetin. In these patients the CDSS may not have predicted the correct dose that they should receive. A reason for this may be that current Hb is being based on the sub-optimum past dose.

MLM was used to investigate this theory further. It was discovered that the relationship between Hb(i) and dose(i-1) yielded erroneous statistical models, which highlighted that this relationship was in fact not ideal. The MLM with autoregressive correlation structure did in fact reveal that the most appropriate relationship was between current Hb and dose administered 10 weeks previously. This is a vital finding for the nephrology field as it means that the current CDSS may not be letting the epoetin agent have enough time to have its optimum effect on Hb and that the current CDSS may require updating.

The statistical methodologies have also been developed, in addition to the novel findings that were revealed about the biomedical systems. MLM with an autocorrelation structure was an existing, though unexploited and relatively unused technique. The research in this thesis has demonstrated that this is a particularly useful technique for repeated measures data, where measurements are temporally correlated. Checking for autocorrelated residuals is an often overlooked assessment of model fit, though if autocorrelation is present this could induce bias in the model (underestimated standard errors) and result in incorrect inferences being drawn. The approach of MLM with an autoregressive correlation structure is a relatively complex modelling procedure, though has been successfully presented and accepted by clinical audiences.

FDA is a relatively new statistical analysis technique, with relatively few applications. This thesis presents two additional successful applications of FDA. The FDA methodology has also been developed, in particular through its application to the renal anaemia biomedical system. This application has demonstrated that FDA may be successfully applied to relatively short time series (i.e. 13 measurements per individual).

It was key in this application that the measurements (within individuals) were temporally correlated and that the data could be reasonably represented by a smooth trajectory over time. The application of phase–plots have also been developed, such that the construction of the phase plots has been simplified from plots of the second derivative against the first derivative, to plots of first derivative against the actual measurement. The simplified approach resulted in identical conclusions being made about the patients. This is beneficial for interpretation in clinical practice and also the initial function fitted has less constraints (i.e. being able to penalize one less derivative).

The use of B-splines and loess curves within this thesis has demonstrated that there is a variety of ways to fit curves to data. Using both techniques has highlighted the advantages and limitations of each method. The application of each technique revealed some of these advantages and limitations, though other features were revealed in the comparison of techniques.

A key feature of the research in this thesis has being to communicate somewhat complex statistics to clinical and other non–specialist audiences. Furthermore, it was important not to compromise the statistical integrity of the work in doing so. This work has demonstrated that it is possible to use and apply existing interesting statistical technique to achieve novel findings in the clinical domain, and also develop and validate the methodology.

9.2 Future Work

There are a number of suggestions for future work that have arisen from this thesis, which will be discussed in this section. In addition to future applications within the same biomedical areas, the methodologies may be used to assess control in other biomedical areas; such as managing glucose level with insulin therapy in patients with diabetes. It would be possible to collect repeated measures data in this area as patients with diabetes

are required to measure their blood sugar daily (or more frequently), which would yield a wealth of measurements. A device known as a GlucoWatch is available, which does not measure blood glucose directly, but measures slight chemical reactions on the surface of the skin. It potentially has the ability to help insulin users to fine tune their overnight minimum insulin rates. The GlucoWatch gives readings every 20 minutes, thus yielding a wealth of measurements each day. It would be possible to assess control in these patients to monitor how the administration of insulin manages the patients' glucose level.

In both applications, if further studies were to taken place, it would be beneficial for additional variables to be collected or provided to the statistician, so confounding factors could be accounted for. Explanatory variables such as age, sex, comorbidities, deprivation, and risk factors for acquiring the disease may be shown to be important factors. Specific important confounding factors for the renal anaemia application may be iron status, serum ferritin level and time on dialysis. For the cerebral autoregulation application confounding factors may include whether the patient has bilateral or unilateral stenosis, length of CEA and degree of atherosclerosis.

9.2.1 Cerebral Autoregulation

In order to determine appropriately whether anaesthetic type has an effect on CA then patients should be randomised to anaesthetic type to enable unbiased comparison of general and local anaesthesia. It may not be the case that this is possible for all patients as in certain situations the patient or clinician may favour a particular anaesthetic choice on medical grounds. Although in the cases where it is possible to randomise, the data from these patients should be available for statistical analysis. This would be the ideal scenario, however, it is unlikely that another RCT would be justified, particularly as the GALA trial [51], which was carried out on such a large scale, was inconclusive. A more reasonable possibility would be to analyse the data where there is no medical preference for anaesthetic type, just simply the choice of the patient. The patient could be matched

and analysis undertaken with the reasonable assumption that choice of anaesthesia was made randomly.

The collection and recording of CBF and ABP measurements throughout surgery was a difficult task to undertake, thus further data collection would require much effort on behalf of the surgical team. Following the difficult data collection, it would be time consuming to extract (and clean) the data from the surgical equipment. This suggests that there would be a high cost in a future investigation of this kind.

Since there was doubt whether measurements collected every 15 seconds amounted to 'rapid' changes in CBF and ABP, in any further investigation, measurements could be collected at intervals shorter than every 15 seconds (which is possible with TCD and ABP monitoring). This would further establish the method for assessing dynamic CA.

In this thesis CA was being assessed immediately following CEA, which was a novel setting for assessing CA. In the same patients, it would have been useful to collect longitudinal measurements on CBF and ABP, every 15 seconds for a 5 minute period, in a period following surgery. For example, after patients have left recovery, or even 24 hours after. This would be a particularly interesting area of investigation, as it would demonstrate whether there is a time lag in CA changing due to CEA.

There was difficultly when assessing CA in the MLM framework in determining whether a slope not equal to zero meant that patients were not autoregulating. This work has revealed the need for guidance or a grading system of autoregulation, as it was established that it is not as straightforward as declaring that a zero slope indicates perfect CA and a non-zero (positive) slope indicates impaired CA. This would require collaboration between the clinician and statistician.

Making the correlation structure a random parameter in the MLM framework was not implementable, therefore in regards to prospective future work it may be possible to write the code for this to be possible in R. This would be particularly relevant for the

cerebral autoregulation application as it would make it possible to model dynamic and static autoregulation in the MLM framework. This would also yield development for the MLM methodology.

9.2.2 Renal Anaemia

As suggested previously, the CDSS may be basing current Hb on the suboptimal past dose. In order to determine whether this is truly the case a RCT should be carried out, whereby one group is randomly allocated to having their epoetin adjusted each month based on the previous months dose (i.e. the CDSS already in place). The second group's epoetin dose will be adjusted, basing current Hb on epoetin administered 10 weeks previously, though still updating the dose each month. An RCT is likely to be justified in this area as patients are already undertaking haemodialysis, thus it would simply be a matter of enrolling patients and randomising to the two groups. There are a large number of patients who require haemodialysis, thus a large sample is likely.

The success of this trial could be judged using FDA, whereby the Hb trajectories and phase plots could be used to determine whether there is significantly more Hb variability in one group than the other. Furthermore, the R code to create the phase–plots could be made user friendly. This would enable clinicians to input patient data, with the output being the phase–plot for that patient. This would allow clinicians (and patients) to see the data in a way that 'control' is graphically represented and meaningful to the patient.

Bibliography

[1] De Stavola BL, Nitsch D, dos Santos Silva I. Statistical issues in lifecourse epidemiology. Am J Epi. 2006;163:84–96.

[2] Will EJ, West RM, Harris K. Assessment of control in biomedical systems with specific application to the control of renal anemia in haemodialysis patients. The Yorkshire Kidney Research Fund; 2006.

[3] Chatfield C. The Analysis of time series - An Introduction. London: Chapman and Hall; 2004.

[4] Association TS. Stroke incidence and risk factors in a population based cohort study. London: The Stroke Association; 2001.

[5] Nice. Early assessment and treatment of people who have had a stroke or transient ischaemic attack (TIA). Information about NICE clinical guideline 68; 2008.

[6] Organization WH. Cerebrovascular Disorders. Geneva: Offset Publications; 1978.

[7] Wolfe C, Rudd T, Beech R. The Burden of Stroke. London: The Stroke Association; 1996.

[8] Adamson J, Beswick A, Ebrahim S. Is stroke the most common cause of disability? Journal of Stroke and Cerebrovascular Diseases. 2004;13(4):171–177.

[9] Howell SJ. Carotid Endarterectomy. British Journal of Anethesia. 2007;99(1):119–131.

[10] Chandratheva A, Mehta Z, Geraghty OC. Population-based study of risk and predictors of stroke in the first few hours after a TIA. Neurology. 2009;72:1941–1947.

[11] Swain S, Turner C, Tyrrell P, Rudd J. Diagnosis and initial management of acute stroke and transient ischaemic attack: summary of NICE guidance. BMJ. 2008;67:1–37.

[12] Findlay JM. Guidelines for the use of carotid endarterectomy: current recommendations from the Canadian Neurosurgical Society. Can Med Mssoc. 1997;157:653–659.

[13] Anatomy G. The Anatomical Basis of Clinical Practice, 40th edition. Edinburgh : Churchill Livingstone; 2008.

[14] Reinhard M. Effect of Carotid Endarterectomy or Stenting on Impairment of Dynamic Cerebral Autoregulation. Stroke. 2004;35:1381–1387.

[15] Paulson OB, Strandgaard S, Edvinsson L. Cerebral Autoregulation. Cerebrovascular and Brain Metabolism Reviews. 1990;2(2):161–192.

[16] Uston C. Dr. Thomas Willis' famous eponym: the circle of Willis. Journal of the history of the neurosciences. 2005;14(1):16–21.

[17] de Boorder MJ, van der Grond J, van Dongen AJ, Klijn CJ, Jaap Kappelle L, Van Rijk PP, et al. Spect measurements of regional cerebral perfusion and carbondioxide reactivity: correlation with cerebral collaterals in internal carotid artery occlusive disease. J Neurol. 2006;253(10):1285–1291.

[18] Halliday A. Prevention of disabling and fatal strokes by successful carotid endarterectomy in patients without recent neurological symptoms: randomised controlled trial. Lancet. 2004;363(9420):1491–1502.

[19] Schroeder T, Sillesen H, Engell HC. Hemodynamic effect of carotid endarterectomy. Stroke. 1987;18:204–209.

[20] Schroeder T, Sillesen H, Sorensen O, Engell HC. Cerebral hyperperfusion following carotid endarterectomy. Journal of Neurosurgery. 1987;66(6):824–829.

[21] van Mook W, Rennenberg R, Schurink GW, van Oostenbrugge RJ, Mess WH, Hofman PAM, et al. Cerebral hyperperfusion syndrome. The Lancet. 2005;4(12):877–888.

[22] Reinhard M. Transfer function analysis for clinical evaluation of dynamic cerebral autoregulation - a comparison between spontaneous and respiratory-induced oscillations. PhysiolMeas. 2003;24:27–43.

[23] Tiecks FP. Comparison of Static and Dynamic Cerebral Autoregulation Measurements. Stroke. 1995;26:1014–1019.

[24] Eames PJ, Blake MJ, Dawson SL, Panerai RB, Potter JF. Dynamic cerebral autoregulation and beat to beat blood pressure control are impaired in acute ischaemic stroke. The Journal of Neurology, Neurosurgery, and Psychiatry. 2002;72(4):467–472.

[25] Naylor AR, Wildsmith JA, McClure J, Jenkins AM, Ruckley CV. Transcranial Doppler Monitoring during carotid endarterectomy. British Journal of Surgery. 1991;78(10):1264–1268.

[26] Panerai RB. Assessment of cerebral pressure autoregulation in humans - a review of measurement methods. PhysiolMeas. 1998;19:305–338.

[27] Telman G, Kouperberg E, Nitecki S, Karram T, Schwarz HA, Sprecher E. Cerebral hemodynamics in symptomatic and asymptomatic patients with severe unilateral carotid stenosis before and after carotid endarterectomy. European Journal of Vascular and Endovascular Surgery. 2006;32(4):375–378.

[28] White RP, Markus HS. Impaired dynamic cerebral autoregulation in carotid artery stenosis. Stroke. 1997;28:1340–1344.

[29] Lassen NA. Cerebral Blood Flow and Oxygen Consumption in Man. The American physiological society. 1959;39:183–238.

[30] Bondar RL. Simultaneous transcranial Doppler and arterial blood pressure response to lower body negative pressure. J Clin Pharmacol. 1994;34:584–589.

[31] Rosenblum WI. Autoregulatory plateau: does it exist? J Cereb Blood Flow Metab. 1995;15(1):174–177.

[32] Brian JE. Recent insights into the regulation of cerebral circulation. ClinExpPharmacolPhysiol. 1996;23:449–457.

[33] Jorch G, Jorch N. Failure of autoregulation of cerebral blood flow in neonates studied by pulsed doppler ultrasound of the internal carotid artery. Eur J pediatr. 1987;146:468–472.

[34] Panerai RB, Kelsall AWR, Rennie JM, Evans DH. Cerebral autoregulation dynamics in premature newborns. Stroke. 1997;26:74–80.

[35] Czosnyka M, Smielewski P, Kirkpatrick P. Monitoring of cerebral autoregulation in head injured patients. Stroke. 1996;27:1829–1834.

[36] Czosnyka M, Smielewski P, Kirkpatrick P, Guazzo E. Testing of cerebral autoregulation in head injury by waveform analysis of blood flow velocity and cerebral perfusion pressure. Acta Neurochir. 1994;60:468–471.

[37] Lam JMK, Hsiang JNK, Poon WS. Monitoring of autoregulation using laser Doppler flowmetry in patients with head injury. J Neurosurg. 1997;86:438–445.

[38] van Beek AH, Claassen JA, Rikkert MG, Jansen RW. Cerebral autoregulation: an overview of current concepts and methodology with special focus on the elderly. J Cereb Blood Flow Metab. 2008;28(6):1071–1085.

[39] Zhang R, Zuckerman JH, Giller CA, Levine BD. Transfer function analysis of dynamic cerebral autoregulation in humans. Am J Physiol Heart Circ Physiol. 1998;274(1):233–241.

[40] Reinhard M, Roth M, Muller T, Czosnyka M, Timmer J, Hetzel A. Cerebral autoregulation in carotid artery occlusive disease assessed from spontaneous blood pressure fluctuations by correlation coefficient index. Stroke. 2003;34(9):2138–2144.

[41] Aaslid R, Lindegaard KF, Sorteberg W, Nornes H. Cerebral autoregulation dynamics in humans. Stroke. 1989;20:45–52.

[42] Levine BD, Giller CA, Lane LD, Buckey JC, Blomqvist CG. Cerebral versus systemic hemodynamics during graded orthostatic stress in humans. Circulation. 1994;90:298–306.

[43] Claydon VE, Hainsworth R. Cerebral autoregulation during orthostatic stress in healthy controls and in patients with posturally related syncope. Clin Auton Res. 2003;13:321–329.

[44] Harris K, West RM, Dellagrammaticas D, Gough MJ, Gilthorpe MS, Chapman GA, et al. Cerebral Autoregulation during Carotid Endarterectomy Under Regional and General Anesthesia. European Journal of Vascular and Endovascular Surgery; 2010 In press.

[45] Tang SC, Huang YW, Shieh JS, Huang SJ, Yip PK, Jeng JS. Dynamic cerebral autoregulation in carotid stenosis before and after carotid stenting. J Vasc Surg. 2008;48(1):88–92.

[46] Mitsis GD, Zhang R, Levine BD, Marmarelis VZ. Modeling of nonlinear physiological systems with fast and slow dynamics. II. Application to cerebral autoregulation. Ann Biomed Eng. 2002;30(4):555–565.

[47] Giller CA, Mueller M. Linearity and non-linearity in cerebral hemodynamics. Med Eng Phys. 2003;25(8):633–646.

[48] Rerkasem K, Bond R, Rothwell PM. Local versus general anesthesia for carotid endarterectomy. Cochrane database of systematic reviews. 2004;40:584–585.

[49] Wong JH, Findlay JM, Suarez-Almazor ME. Hemodynamic instability after carotid endarterectomy: risk factors and associations with operative complications. Neurosurgery. 1997;41:35–41.

[50] McCleary AJ, Gough MJ, Dearden NM, Dickson DH, Watson A. The differing effects of regional and general anesthesia on cerebral metabolism during carotid endarterectomy. European Journal of Vascular and Endovascular Surgery. 1996;93:171–176.

[51] Lewis SC, Warlow CP, Bodenham AR. General anesthesia versus local anesthesia for carotid surgery (GALA): a multicentre, randomised controlled trial. Lancet. 2008;372:2132–2142.

[52] Gough MJ. The GALA trial – a summary of the findings. European Journal of Vascular and Endovascular Surgery. 2008;36(5):505–506.

[53] Bouma GJ, Muizelaar JP, Bandoh K, Marmarou A. Blood pressure and intercranial pressure-volume dynamics in severe head injury: relationship with cerebral blood flow. J Neurosurg. 1992;9:15–19.

[54] Goosekens I, Schmidt EA, Czosnyka M. Pressure-autoregulation, C02 reactivity and asymmetry of haemodynamic parameters in patients with carotis artery stenotic disease: A clinical appraisal. Acta Neurochir. 2003;145:527–532.

[55] Chaturvedi S, Bruno A, Feasby T, Holloway R, Benavente O, Cohen SN, et al. Carotid endarterectomy–an evidence-based review: report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology. Neurology. 2005;65(6):794–801.

[56] Heckmann JG, Brown CM, Cheregi M, Hilz MJ, Neundrfer B. Delayed cerebrovascular autoregulatory response to ergometer exercise in normotensive elderly humans. Cerebrovasc Dis. 2003;16(4):423–429.

[57] Franke WD, Allbee KA, Spencer SE. Cerebral blood flow responses to severe orthostatic stress in fit and unfit young and older adults. Gerontology. 2006;52:282–289.

[58] Carey BJ, Eames PJ, Blake MJ, Panerai RB, Potter JF. Dynamic cerebral autoregulation is unaffected by aging. Stroke. 2000;31:2895–2900.

[59] Lipsitz LA, Mukai S, Hamner J, Gagnon M, Babikian V. Dynamic regulation of middle cerebral artery blood flow velocity in aging and hypertension. Stroke. 2000;31(8):1897–1903.

[60] Ma JZ, Ebben J, Xia H, Collins AJ. Hematocrit level and associated mortality in hemodialysis patients. J Am Soc Nephrol. 1999;10:610–619.

[61] McMahon LP, Mason K, Skinner SL, Burge CM, Grigg LE, Becker GJ. Effects of haemoglobin normalization on quality of life and cardiovascular parameters in end-stage renal failure. Nephrol Dial Transplant. 2000;15:1425–1430.

[62] Schmidt RJ, Dalton CL. Treating anemia of chronic kidney disease in the primary care setting: cardiovascular outcomes and management recommendations. Osteopathic Medicine and Primary Care. 2007;1:1–14.

[63] Adamson JW, Eschbach JW. Treatment of the anemia of chronic renal failure with recombinant human erythropoietin. Annu Rev Med. 1990;41:349–360.

[64] Nurko S. Anaemia in chronic kidney disease: Causes, diagnosis, treatment. Cleveland clinic journal or medicine. 2006;73(3):289–297.

[65] Roger SD. Managing the anaemia of chronic kidney disease. Aust Prescr. 2009;32:129–131.

[66] Klutstein MW, Tzivoni D. Anaemia and heart failure: aetiology and treatment. Nephrology Dialysis Transplantation. 2005;20(7):7–10.

[67] Sakiewicz P, Paganini E. The use of iron in patients on chronic dialysis: mistakes and misconceptions. J Nephrol. 1998;11(1):5–15.

[68] Marieb EN. Human Anatomy and Physiology. London: Pearson Education International; 2004.

[69] Eschbach JW, Egrie JC, Downing MR, Browne JK, Adamson JW. Correction of the anemia of end-stage renal disease with recombinant human erythropoietin. New England Journal of Medicine. 1987;316(2):73–78.

[70] Ly J, Marticorena R, Donnelly S. Red blood cell survival in chronic renal failure. Am J Kidney Dis. 2004;44(4):715–719.

[71] Fishbane S, Nissenson AR. The New FDA label For Erythropoietin Treatment: How Does It Affect Hemoglobin Target? Kidney International. 2007;72(7):806–813.

[72] Rossert J, Fouqueray B, Boffa JJ. Anemia Management and the Delay of Chronic Renal Failure Progression. J Am Soc Nephrol. 2003;14:173–177.

[73] Will EJ, Richardson D, Tolman C, Bartlett C. Development and exploitation of a clinical decision support system for the management of renal anemia. Nephrol Dial Transplant. 2007;22(4):31–36.

[74] Payne TH. Computer decision support systems. Chest. 2000;118:47–52.

[75] Organization WH. WHO Expert committee on biological standardization technical report series. WHO Technical Report Series; 2000.

[76] Kruse A, Uehlinger D, Gotch F, Kotanko P, Levin N. Red Blood Cell Lifespan, Erythropoiesis and Hemoglobin Control. Contrib Nephrol. 2008;161:247–254.

[77] Kaufman JS, Reda DJ, Fye CL, Goldfarb DS, Henderson WG, Kleinman JG, et al. Subcutaneous Compared with Intravenous Epoetin in Patients Receiving Hemodialysis. The New England Journal of Medicine. 1998;339(9):578–583.

[78] Aggarwal HK, Nand N, Singh S, Singh M, Kaushik G. Comparative efficacy of Subcutaneous versus Intravenous dose of erythropoietin in pre-dialysis patients of chronic renal failure. Journal Indian academy of clinical medicine. 2002;3(1):46–50.

[79] McMahon FG, Vargas R, Ryan M. Pharmacokinetics and effects of recombinant human erythropoietin after intravenous and subcutaneous injections to healthy volunteers. Blood. 1990;76:1718–1722.

[80] Horl WH. Optimal route of administration of erythropoietin in chronic renal failure patients: Intravenous versus subcutaneous. Acta Haematologica. 1992;87(1):16–19.

[81] Muirheadm N, Churchill DN, Goldstein M. Comparison of intravenous and subcutaneous recombinant human erythropoietin for anemia in hemodialysis patients with significant comorbid disease. Am J Nephrol. 1992;12:303–310.

[82] Bright R. Cases and observations: illustrative of renal disease accompanied by the secretion of albuminous urine. Guys Hosp Rep. 1836;1:338–338.

[83] Bonsdorff E, Jalavisto E. A humoral mechanism in anoxic erythrocytosis. Acta Physiol Scand. 1948;16:150–170.

[84] Miyake T, Kung CK, Goldwasser E. Purification of human erythropoietin. J Biol Chem. 1977;252:5558–5564.

[85] Lin FK, Suggs S, Lin CH. Cloning and expression of the human erythropoietin gene. Proc Natl Acad Sci USA. 1985;82:7580–7584.

[86] Eschbach JW, Egrie JC, Downing MR, Browne JK, Abdulhadi MH, Delano BG, et al. Recombinant human erythropoietin in anemic patients with end-stage renal disease. Results of a phase III multicenter trial. Ann Intern Med. 1989;111(12):992–1000.

[87] Strippolo GFM, Navaneethan SD, Craig JC. Haemoglobin and haematocrit targets for the anemia of chronic kidney disease (Review). The Cochrane Collaboration. 2009;3.

[88] de Benoist B, McLean E, Egli I, Cogswell M. Worldwide prevalence of anaemia 1993 to 2005. WHO Global Database on Anemia; 2008.

[89] Moreno F, Sanz-Guajardo D, Lopez-Gomez JM, Jofre R, Valderrabano F. Increasing the hematocrit has a beneficial effect on quality of life and is safe in selected hemodialysis patients. J Am Soc Nephrol. 2000;11:335–342.

[90] Besarab A, Bolton K, Browne JK, Egrie JC, Nissenson AR, Okamoto DM, et al. The effects of normal as compared with low hematocrit values in patients with cardiac disease who are receiving hemodialysis and epoetin. N Engl J Med. 1998;339:584–590.

[91] Parfrey PS, Foley RN, Wittreich BH, Sullivan DJ, Zagari MJ, Frei D. Double-blind comparison of full and partial anemia correction in incident hemodialysis patients without symptomatic heart disease. J Am Soc Nephrol. 2005;16:2180–2189.

[92] Singh AK, Szczech L, Tang KL. Correction of Anemia with epoetin alfa in chronic kidney disease. N Engl J Med. 2006;355:2085–2098.

[93] Drueke TB, Locatelli F, Clyne N. Normalization of hemoglobin levels in chronic kidney disease and anemia. N Engl J Med. 2006;355:2071–2084.

[94] West RM, Harris K, Glithorpe MS, Tolman C, Will EJ. Functional Data Analysis Applied to a Randomized Controlled Clinical Trial in Hemodialysis Patients Describes the Variability of Patient Responses in the Control of Renal Anemia. J Am Soc Nephrol. 2007;18(8):2371–2376.

[95] Hayat A, Haria D, Salifu MO. Erythropoietin stimulating agents in the management of anemia of chronic kidney disease. Patient Preference and Adherence. 2008;2:195–200.

[96] Ebben JP, Gilbertson DT, Foley RN, Collins AJ. Hemoglobin level variability: Associations with co-morbidity, intercurrent events, and hospitalizations. Clin J Am Soc Nephrol. 2006;1:1205–1210.

[97] Richardson D, Lindley EJ, Bartlett C, Will EJ. A randomized controlled study of the consequences of hemodialysis membrane composition on erythropoietic response. Am J Kidney Diseases. 2003;42(3):551–560.

[98] Fishbaine S, Berns JS. Hemoglobin cycling in hemodialysis patients treated with recombinant human erythropoietin. Kidney Int. 2005;68:1337–1343.

[99] Tolman C, Richardson D, Bartlett C, Will EJ. Structured conversion from thrice weekly to weekly erythropoietic regimens using a computerized decision-support system: A randomized clinical study. J Am Soc Nephrol. 2005;16:1463–1470.

[100] Barnett AI, Crmieux PY. Dose Conversion from Epoetin alfa to Darbepoetin alfa for Patients with Chronic Kidney Disease Receiving Hemodialysis. Pharmacotherapy. 2003;23(5):1–4.

[101] Eschbach JW, Egrie JC, Downing MR. The use of recombinant human erythropoietin: Effect in end-stage renal disease. Prevention of Chronic Uremia. 1989;1:148–155.

[102] Cotter D, Zhang Y, Thamer M, Kaufman J, Hernan MA. The effect of epoetin dose on hematocrit. Kidney International. 2008;73(3):347–353.

[103] Ramsay JO, Silverman BW. Functional Data Analysis. New York: Springer-Verlag; 1997.

[104] Shi JQ, Wang B, Murray-Smith R, Titterington DM. Gaussian process functional regression modeling for batch data. Biometrics. 2007;63:714–723.

[105] Ramsay JO, Silverman BW. Applied Functional Data Analysis - Methods and Case Studies. New York: Springer-Verlag; 2002.

[106] Cleveland WS. Robust Locally Weighted Regression and Smoothing Scatterplots. J Am Soc Stat Ass. 1979;74(368):829–836.

[107] Cleveland WS, Devlin SJ. Locally Weighted Regression: An approach to Regression analysis by local fitting. J Am Soc Stat Ass. 1988;83(403):596–610.

[108] Macauley FR. The Smoothing of Time Series. New York: National Bureau of Economic Research; 1931.

[109] de Boor C. A practical guide to splines. New York: Springer; 1978.

[110] Kooperberg C, Stone CJ. Logspline Density Estimation. Computational Statistics and Data Analysis. 1991;12:327–347.

[111] Friedman JH, Silverman BW. Flexible Parsimonious Smoothing and Additive Modeling. Technometrics. 1989;31(1):3–21.

[112] Eilers PHC, Marx BD. Flexible smoothing with B-splines and penalties. StatSci. 1996;11(2):89–121.

[113] Jackson RS, Griffiths PR. Comparison of Fourier self-deconvolution and maximum likelihood restoration for curve fitting. Analytical Chemistry. 1991;63(22):2557–2563.

[114] Hirsch MW, Smale S. Differential equations, dynamical systems and linear algebra; 1974.

[115] Team RDC. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria; 2005.

[116] Strandgaard S. Antihypertensive drugs and cerebral circulation. EurJClinInvest. 1996;26:625–630.

[117] Lindley DV, Smith AFM. Bayes estimates for the linear model. Journal of the Royal Statistical Society B. 1972;34:1–41.

[118] Goldstein H. Statistical information and the measurement of educational outcomes (editorial). Journal of the Royal Statistical Society A. 1992;155:175–177.

[119] Goldstein H. Multilevel statistical models. London: Edward Arnold; 1995.

[120] Leyland AH, Goldstein H. Multilevel modelling of health statistics. Chichester: Wiley; 2001.

[121] Gilthorpe MS. The application of multilevel modelling to periodontal research data. Community Dental Health. 2000;17:227–235.

[122] Durbin J, Watson GS. Testing for Serial Correlation in Least Squares Regression, I. Biometrika. 1950;37:409–428.

[123] Box GEP, Jenkins GM, Reinsel GC. Time Series Analysis, Forecasting and Control. New Jersey: Prentice-Hall; 1994.

[124] Fisher RA. The goodness of fit of regression formulae and the distribution of regression coefficients. J Roy Statist Soc. 1922;85:597–612.

[125] Bartlett MS. Properties of sufficiency and statistical tests. Royal Statistical Society, Series A. 1937;160:268–282.

[126] Patterson HD, Thompson R. Recovery of interblock information when block sizes are unequal. Biometrika. 1971;58:545–554.

[127] Harville DA. Maximum likelihood approaches to variance component estimation and to related problems. Journal of the American Statistical Association. 1977;72:320–340.

[128] Laird NM, Ware JH. Random effects models for longitudinal data. Biometrics. 1982;38:963–974.

[129] Snijders TAB, Bosker RL. Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling. London: Sage Publications Ltd; 1999.

[130] Akaike H. Information theory and an extension of the maximum likelihood principle. International Symposium on Information Theory. 1973;p. 267–281.

[131] Akaike H. A new look at the statistical model identification. IEEE Transactions on Automatic Control. 1974;19(6):716–723.

[132] Lindsey JK. Models for Repeated Measurements. New York: Oxford University Press; 1999.

[133] Schwarz G. Estimating the Dimension of a Model. The Annals of Statistics. 1978;6(2):461–464.

[134] Wang Y, Liu Q. Comparison of Akaike information criteria (AIC) and Bayesian information criteria (BIC) in selection of stock recruitment relationships. Fish Res. 2006;77:220–225.

[135] Ojo JF, Olatayo TO. On the Estimation and Performance of Subset Autoregressive Integrated Moving Average Models. European Journal of Scientific Research. 2009;28(2):287–293.

[136] Stram DO, Lee JW. Variance components testing in longitudinal mixed effects model. Biometrics. 1994;50:1171–1177.

[137] Goldstein H, Healy MJR, Rasbash J. Multilevel time series models with applications to Repeated Measures data. Statistics in Medicine. 1994;13:1643–1655.

[138] Rasbash J, Browne W, Healy M, Cameron B, Charlton C. MLwiN version 2.02. Multilevel Models Project, Institute of Education, London; 2005.

[139] Pinheiro JC, Bates DM. Mixed Effects Models in S and S-Plus. New York: Springer-Verlag; 2002.

[140] Bryk AS, Raudenbush SW. Hierarchical linear models : applications and data analysis methods. Thousand Oaks, CA ; London : Sage Publications; 2002.

[141] Holden JE, Kelley K, Agarwal R. Analyzing Change: A Primer on Multilevel Models with Applications to Nephrology. Am J Nephrol. 2008;28(5):792–801.

[142] Blance A, Gilthorpe MS, Tu YK. A multilevel modelling solution to mathematical coupling. Statistical Methods in Medical Research. 2005;14:553–565.

[143] Stryhn H, Christensen J. Confidence intervals by the profile likelihood method, with applications in veterinary epidemiology. Contributed paper at ISVEE X. 2003;.

[144] Czosnyka M, Smielewski P, Lavinio A, Pickard JD, Panerai R. An Assessment of Dynamic Autoregulation from Spontaneous Fluctuations of Cerebral Blood Flow Velocity: A Comparison of Two Models, Index of Autoregulation and Mean Flow Index. Anesthesia and Analgesia. 2008;106:234–239.

[145] Gupta AK, Besarab A. A distributed lag model with first order auto-regressive error term for hemoglobin response to erythropoietin in dialysis patients. AMIA Annu Symp Proc. 2007;11:967.